

Titre: Développement et validation d'un système automatique de
Title: classification de la dégénérescence maculaire liée à l'âge

Auteur: Thanh Vân Phan
Author:

Date: 2015

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Phan, T. V. (2015). Développement et validation d'un système automatique de
classification de la dégénérescence maculaire liée à l'âge [Mémoire de maîtrise,
Citation: École Polytechnique de Montréal]. PolyPublie.
<https://publications.polymtl.ca/1790/>

Document en libre accès dans PolyPublie

Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/1790/>
PolyPublie URL:

Directeurs de recherche: Farida Cheriet, & Lama Séoud
Advisors:

Programme: Génie biomédical
Program:

UNIVERSITÉ DE MONTRÉAL

DÉVELOPPEMENT ET VALIDATION D'UN SYSTÈME AUTOMATIQUE DE
CLASSIFICATION DE LA DÉGÉNÉRÉSCENCE MACULAIRE LIÉE À L'ÂGE

THANH VÂN PHAN

INSTITUT DE GÉNIE BIOMÉDICAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE BIOMÉDICAL)

JUILLET 2015

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

DÉVELOPPEMENT ET VALIDATION D'UN SYSTÈME AUTOMATIQUE DE
CLASSIFICATION DE LA DÉGÉNÉRÉSCENCE MACULAIRE LIÉE À L'ÂGE

présenté par : PHAN Thanh Vân

en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de :

M. SAWAN Mohamad, Ph. D., président

Mme CHERIET Farida, Ph. D., membre et directrice de recherche

Mme SÉOUD Lama, Ph. D., membre et codirectrice de recherche

M. HURTUT Thomas, Ph. D., membre

DÉDICACE

À ma famille, mon amour et mes amis, ...

REMERCIEMENTS

Tout d'abord, je tiens à remercier chaleureusement ma directrice de recherche, Pr. Farida Cheriet, professeure à l'École Polytechnique, pour la confiance et le soutien qu'elle m'a porté tout au long de ce projet. Ses commentaires éclairés m'ont toujours permis de progresser dans ma recherche en apportant de nouvelles perspectives et d'aboutir à tout ce travail accompli. Je lui suis très reconnaissante d'avoir proposé et supervisé mon projet.

Je tiens également à remercier ma co-directrice, Lama Seoud, pour son aide précieuse et ses conseils avisés. En particulier, ses encouragements et ses corrections pour la rédaction m'ont été d'une grande aide, et j'ai toujours apprécié son écoute et sa disponibilité, ainsi que les idées apportées pour suivre différentes pistes de recherche.

Je voudrais également remercier Philippe Debanné pour sa disponibilité et son aide lorsqu'il y avait des problèmes techniques au laboratoire.

Un grand merci à toute l'équipe de Diagnos Inc., qui m'a fourni les données et les connaissances nécessaires pour effectuer mon projet. En particulier, Dr. Hadi pour ses connaissances médicales qui m'a appris beaucoup de choses intéressantes dans ce domaine et Houssem Ben Tahar pour le support matériel sans lequel le projet n'aurait pas pu être réalisé.

Mes remerciements aux membres du Laboratoire d'imagerie et de vision 4D de l'École Polytechnique pour leur soutien moral, leur écoute, leurs disponibilités, leurs conseils. J'ai beaucoup apprécié discuter avec tout le monde de tout et de rien durant les lunchs, et j'aimerais remercier tout particulièrement Mathieu Tournier, Mahnaz Fasih, Anaïs Dume, Florian Berton, Fantin Girard, Argyrios Christodoulidis, Pierre-André Brousseau qui étaient d'une agréable compagnie au laboratoire.

Merci à toute ma famille. À mes parents qui m'ont permis de vivre cette expérience et qui m'ont toujours soutenue. À ma famille au Canada qui m'ont très bien accueilli et qui m'ont apporté tout ce dont j'avais besoin. À mes cousins en Europe, qui n'hésitaient pas à m'aider quand je leur demandais et qui m'ont toujours bien fait rire.

Je tiens enfin à remercier Vincent Putterie, pour son amour et sa patience, ne me laissant jamais tomber et me supportant toujours dans mes choix et mes projets.

RÉSUMÉ

La dégénérescence maculaire liée à l'âge (DMLA) est une des principales causes de déficience visuelle menant à une cécité irréversible chez les personnes âgées dans les pays industrialisés. Cette maladie regroupe une variété d'anomalies touchant la macula, se présentant sous diverses formes. Un des moyens les plus couramment utilisés pour rapidement examiner la rétine est la photographie de fond d'œil. À partir de ces images, il est déjà possible de détecter et de poser un diagnostic sur l'avancée de la maladie. Une classification recommandée pour évaluer la DMLA est la classification simplifiée de l'AREDS qui consiste à diviser la maladie en quatre catégories : non-DMLA, précoce, modérée, et avancée. Cette classification aide à déterminer le traitement spécifique le plus optimal. Elle se base sur des critères quantitatifs mais également qualitatifs, ce qui peut entraîner des variabilités inter- et intra-expert. Avec le vieillissement de la population et le dépistage systématique, le nombre de cas de DMLA à être examinés et le nombre d'images à être analysées est en augmentation rendant ainsi le travail long et laborieux pour les cliniciens. C'est pour cela, que des méthodes automatiques de détection et de classification de la DMLA ont été proposées, afin de rendre le processus rapide et reproductible. Cependant, il n'existe aucune méthode permettant une classification du degré de sévérité de la DMLA qui soit robuste à la qualité de l'image. Ce dernier point est important lorsqu'on travaille dans un contexte de télémédecine. Dans ce projet, nous proposons de développer et valider un système automatique de classification de la DMLA qui soit robuste à la qualité de l'image. Pour ce faire, nous avons d'abord établi une base de données constituée de 159 images, représentant les quatre catégories de l'AREDS et divers niveaux de qualité d'images. L'étiquetage de ces images a été réalisé par un expert en ophtalmologie et nous a servi de référence. Ensuite, une étude sur l'extraction de caractéristiques nous a permis de relever celles qui étaient pertinentes et de configurer les paramètres pour notre application. Nous en avons conclu que les caractéristiques de texture, de couleur et de contexte visuel semblaient les plus intéressantes. Nous avons effectué par après une étape de sélection afin de réduire la dimensionnalité de l'espace des caractéristiques. Cette étape nous a également permis d'évaluer l'importance des différentes caractéristiques lorsqu'elles étaient combinées ensemble. Il en ressortait que les caractéristiques les plus discriminantes pour classifier la DMLA étaient les patrons binaires locaux calculés sur le canal vert. Finalement, nous avons modélisé et testé différents systèmes de classification afin d'étudier comment la méthode proposée classifiait les images de fond d'œil selon les différentes catégories de la DMLA. Les résultats obtenus ont

démontré la robustesse du système à la qualité de l'image ainsi qu'une performance de classification meilleure que celle réalisée dans la littérature. Les erreurs commises correspondent à des images de rétinopathie diabétique, des images avec des vaisseaux de choroïde visibles ou des images trop dégradées par des artefacts. Nous avons donc, lors de ce projet, développé et validé le premier système de classification du degré de sévérité de la DMLA qui soit robuste à la qualité de l'image. Avec une validation plus approfondie et une base de données plus complète, ce système permettrait d'établir un diagnostic fiable sur la DMLA, ce qui servirait dans le suivi clinique de patients atteints de DMLA et dans l'évaluation de l'effet des traitements proposés pour guérir la maladie lors d'études cliniques.

ABSTRACT

Age-related macular degeneration (AMD) is the leading cause of visual deficiency and legal blindness in the elderly population in industrialized countries. This disease is a group of heterogeneous disorders affecting the macula. For eye examination, a common used modality is the fundus photography because it is fast and non-invasive procedure which may establish a diagnostic on the stage of the disease. A recommended classification for AMD is the simplified classification of AREDS which divides the disease into four categories: non-AMD, early, moderate and advanced. This classification is helpful to determine the optimal and specific treatment. It is based on quantitative criteria but also on qualitative ones, introducing inter- and intra-expert variability. Moreover, with the aging population and systematic screening, more cases of AMD must be examined and more images must be analyzed, rendering this task long and laborious for clinicians. To address this problem, automatic methods for AMD classification were then proposed for a fast and reproducible process. However, there is no method performing AMD severity classification which is robust to image quality. This last part is especially important in a context of telemedicine where the acquisition conditions are various. The aim of this project is to develop and validate an automatic system for AMD classification which is robust to image quality. To do so, we worked with a database of 159 images, representing the different categories at various levels of image quality. The labelling of these images is realized by one expert and served as a reference. A study on feature extraction is carried out to determine relevant features and to set the parameters for this application. We conclude that features based on texture, color and visual context are the most interesting. After, a selection is applied to reduce the dimensionality of features space. This step allows us to evaluate the feature relevance when all the features are combined. It is shown that the local binary patterns applied on the green channel are the most the discriminant features for AMD classification. Finally, different systems for AMD classification were modeled and tested to assess how the proposed method classifies the fundus images into the different categories. The results demonstrated robustness to image quality and also that our method outperforms the methods proposed in the literature. Errors were noted on images presenting diabetic retinopathy, visible choroidal vessels or too much degradation caused by artefacts. In this project, we propose the first AMD severities classification robust to image quality. With further validations on a more complete database, an automatic clinical classification of AMD will allow reliable diagnostic, helpful for the follow-up of the patient and for clinical trials where the aim is to develop a cure to the disease.

TABLE DES MATIÈRES

DÉDICACE.....	III
REMERCIEMENTS	IV
RÉSUMÉ.....	V
ABSTRACT	VII
TABLE DES MATIÈRES	VIII
LISTE DES TABLEAUX.....	X
LISTE DES FIGURES	XI
LISTE DES SIGLES ET ABRÉVIATIONS	XV
LISTE DES ANNEXES	XVII
CHAPITRE 1 INTRODUCTION.....	1
1.1 Photographie de fond d’œil	1
1.1.1 Principe de la modalité.....	1
1.1.2 Artefacts dans les photographies de fond d’œil	3
1.2 Évaluation clinique de la DMLA	5
1.2.1 Description de la maladie.....	5
1.2.2 La maculopathie liée l’âge	5
1.2.3 La DMLA atrophique (forme sèche avancée)	8
1.2.4 La DMLA exsudative	10
1.2.5 Méthode clinique de diagnostic.....	11
1.3 Classification de la DMLA à partir des images de fond d’œil	14
1.3.1 Limitations des méthodes de diagnostic cliniques actuelles	15
1.4 Organisation du mémoire	18

CHAPITRE 2 REVUE DE LITTÉRATURE	19
2.1 Concept de reconnaissance de forme statistique	19
2.1.1 Extraction de caractéristiques.....	19
2.1.2 Sélection de caractéristiques	32
2.1.3 Modélisation de classifieur.....	37
2.2 Système automatique pour l'évaluation de la DMLA.....	47
2.2.1 Méthodes basées sur la segmentation des drusen.....	47
2.2.2 Méthodes basées sur le calcul direct de caractéristiques.....	50
2.3 Objectifs et hypothèses de recherche	54
CHAPITRE 3 ARTICLE 1 : UNE CLASSIFICATION AUTOMATIQUE POUR LA DÉGÉNÉRESCENCE MACULAIRE LIÉE À L'ÂGE BASÉE SUR LES CARACTÉRISTIQUES D'IMAGES DE FOND D'ŒIL.....	55
3.1 Présentation de l'article	55
3.2 An Automatic Classification for Age-related Macular Degeneration Based on Fundus Image Features	55
3.2.1 Introduction	55
3.2.2 Materials and methods	58
3.2.3 Experimental setup	63
3.2.4 Results	66
3.2.5 Discussion	75
3.2.6 Conclusion.....	78
CHAPITRE 4 DISCUSSION GÉNÉRALE	79
CONCLUSION	84
BIBLIOGRAPHIE	87

LISTE DES TABLEAUX

Table 3-1: Number of images in each category with corresponding image quality	64
Table 3-2: Binary classification assessment with RF.....	67
Table 3-3: Trinary classification assessment with RF	67
Table 3-4: Binary classification assessment with RF trained with good quality images and tested on poor quality images	68
Table 3-5: Trinary classification assessment with RF trained with good quality images and tested on poor quality images	69
Table 3-6: Binary classification with threshold assessment with RF.....	70
Table 3-7: Binary classification assessment with SVM and features selected with Fisher criterion	73
Table 3-8: Trinary classification assessment with SVM and features selected with Fisher criterion	73
Table 3-9: Binary classification assessment with SVM and features selected with SFFS	74
Table 3-10: Trinary classification assessment with SVM and features selected with SFFS	75

LISTE DES FIGURES

Figure 1-1: (a) Appareil d'acquisition (Diagnos, 2015), (b) Photographie de fond d'œil : (1) macula, (2) disque optique, (3) vaisseaux sanguins rétiniens.....	2
Figure 1-2: Exemples d'artefacts dans les photographies de fond d'œil : (a) acquisition de cils, (b) cils et clignement des yeux entraînant une réflexion intense, (c) réflexion de bord, (d) mauvaise illumination, (e) réflexions spéculaires à cause de dépôts sur la lentille, (f) arcs et dépôts sur la lentille, (g) halo brumeux, (h) flou, (i) artefact causé par une petite pupille.	4
Figure 1-3: Couche de fond d'œil pour (a) un œil sain et (b) un œil avec maculopathie liée à l'âge (The Angiogenesis Foundation, 2013).	5
Figure 1-4: Types de drusen : (a) drusen miliaires, (b) drusen séreux et (c) drusen cuticulaire (Jonisch & Shah, 2011).	7
Figure 1-5: Progression de la GA : (a) Deux régions distinctes de GA, (b) progression de GA qui a changé de forme et s'est élargi, huit ans plus tard (Laud, Mukkamala, Brue, & Slakter, 2011).	9
Figure 1-6: Progression de la DMLA sur 11 ans : (a) ensemble de drusen séreux dans la macula, (b) formation de drusen additionnels et déposition de pigment, (c) résorption de drusen au centre formant le début de l'atrophie après 8 ans, (d) formation de l'atrophie géographique bien définie après 11 ans (Laud, Mukkamala, Brue, & Slakter, 2011).	9
Figure 1-7: Formation de la néovascularisation : (a) réaction inflammatoire produisant les VEGF et (b) traversée des néovaissaux à travers la membrane de Bruch (The Angiogenesis Foundation, 2013)	10
Figure 1-8: DMLA exsudative avec hémorragie (Jonisch & Shah, 2011).	11
Figure 1-9: Grille d'Amsler vue par (A) un patient normal et (B) un patient qui présente des signes de DMLA (Université Médicale Virtuelle Francophone, 2013)	12
Figure 1-10: Image OCT d'une rétine avec DMLA exsudative présentant une néovascularisation choroïdienne (CNVM), drusen miliaires (HE) et une infiltration de fluide sous-rétinien (SFR) (Chiang, Witkin, Regillo, & Ho, 2011).	13

Figure 1-11: Angiographie de la rétine : (a) avec fluorescéine où les vaisseaux sanguins rétiniens et les drusen sont mis en évidence et (b) avec ICG où les néovaisseaux choroïdiens sont plus démarqués (Chiang, Witkin, Regillo, & Ho, 2011).	14
Figure 1-12: Photographie de fond des différentes structures caractérisant la rétinopathie diabétique.	16
Figure 1-13: Photographie de fond d'œil d'un cas où les exsudats de la rétinopathie diabétique se trouvent au niveau de la macula.	17
Figure 2-1: Espace de couleur RGB : (a) représentation du coin inférieur, (b) représentation du coin supérieur et (c) exemple d'une couleur en RGB (Horst Frank, 2004). Tiré de http://commons.wikimedia.org/wiki/File:RGB_farbwuerfel.jpg	20
Figure 2-2: Espace de couleur L*a*b à (a) luminance 25% et (b) luminance 75% (JacobVoss, 2004). Tiré de http://commons.wikimedia.org/wiki/File:Lab_color_at_luminance_25_percent.png et http://commons.wikimedia.org/wiki/File:Lab_color_at_luminance_75%25.png	21
Figure 2-3: Espace de couleur HSV (SharkD, 2010). Tiré de http://commons.wikimedia.org/wiki/File:HSV_color_solid_cylinder_alpha_lowgamma.png	22
Figure 2-4: Exemples de calcul de la GLRLM : (a) matrice d'une image et (b) GLRLM correspondant dans la direction 0°	23
Figure 2-5: Exemple de calcul du signe et de l'amplitude du LBP pour un pixel central donné dans un voisinage à $P = 4$ et de rayon R	26
Figure 2-6: Illustration de la construction locale des histogrammes de gradient orienté (Levi, 2013). Tiré de https://gilscvblog.wordpress.com/2013/08/18/a-short-introduction-to-descriptors/	28
Figure 2-7: Espace des échelles gaussiennes pour 3 octaves et 5 sous-niveaux (Indif, 2010). Tiré de http://commons.wikimedia.org/wiki/File:Pyramid_of_gradients.png	29
Figure 2-8: Espace des échelles de différence de Gaussienne pour 3 octaves et 5 sous-niveaux (Indif, 2010). Tiré de http://commons.wikimedia.org/wiki/File:Differences-of-Grandients.png	30

Figure 2-9: Recherche des extrema avec les DoG (Indif, 2010). Tiré de http://fr.wikipedia.org/wiki/Scale-invariant_feature_transform#/media/File:D%C3%A9tection-d%27extr%C3%A9mums-dans-l%27espace-des-%C3%A9chelles.png	31
Figure 2-10: Projection des échantillons sur l'axe des caractéristique C1 et C2 pour calculer le critère de Fisher : (a) sans chevauchement et (b) avec chevauchement.....	33
Figure 2-11: Schéma résumant les méthodes SFFS et SFBS.....	36
Figure 2-12: Illustration des deux familles de classifieurs : (a) à modèle génératif et (b) à modèle discriminant.....	38
Figure 2-13: LDA pour deux classes dans l'espace de caractéristiques C1-C2 avec f1 un exemple de bonne fonction discriminante car les deux classes sont bien séparées et f2 un exemple de mauvaise fonction discriminante car les deux classes sont superposées.	39
Figure 2-14: Schéma de l'hyperplan d'une machine vecteur de support séparant deux classes....	40
Figure 2-15: Diagramme des frontières binaires pour la méthode one-versus-all pour un problème à trois classes.....	43
Figure 2-16: Schéma d'un arbre de décisions avec Tf les nœuds de l'arbre, xi les prédicteurs, $tile$ seuil sur le prédicteur et yk l'étiquette des classes.	45
Figure 2-17: Vote des différents arbres décisionnels pour déterminer la classe y d'un élément x	46
Figure 2-18: Méthode de calcul des caractéristiques LBP en multi-résolution proposée par Garnier et al (2014).	51
Figure 2-19: Méthode de calcul des points SURF proposée par Kankanhalli et al (2013).	52
Figure 3-1: Images of macula area for different AMD categories: (a) healthy and (b) diabetic retinopathy in category {1}, (c) hard drusen in category {2}, (d) soft drusen in category {3}, (e) hemorrhages and (f) geographic atrophy in category {4}.	56
Figure 3-2: Preprocessing method: ROI corresponding to the square inscribed in the circle formed by the retina and the result of preprocessing with illumination normalization and contrast enhancement in green channel.	59

Figure 3-3: Examples of poor quality images: (a) shadows and intense reflection, (b) haze, (c) arc and specular reflections and (d) blur	64
Figure 3-4: TPR and 1-FPR for classification {1} vs {2&3} for different thresholds using LBP+Color+HOG features.....	70
Figure 3-5: The ten most relevant features for {1 & 2} vs. {3 & 4} classification as measured by mean decrease in accuracy.	71
Figure 3-6: The ten most relevant features for {1} vs. {2& 3} vs. {4} classification as measured by mean decrease in Gini index.	72
Figure 3-7: Examples of misclassified images with our method: (a) early case with dark macula, (b) moderate case with choroidal vessels in background, (c) non-AMD cases with diabetic retinopathy, (d) advanced case with shadow on the macula	77
Figure 4-1: Exemples d'images mal classifiées : (a) rétinopathie diabétique détectée comme DMLA, (b) cas de non-DMLA avec du flou et des vaisseaux de la choroïde détecté comme DMLA, (c) cas précoce et (d) cas modéré avec des vaisseaux de choroïde visibles détectés comme non-DMLA, (e) cas avancé avec une ombre et des réflexions lumineuses et (f) cas avancé où l'image est centrée sur le disque optique détectés comme non-DMLA.....	83

LISTE DES SIGLES ET ABRÉVIATIONS

AREDS	Age-related Eye Disease Study
DMLA	Dégénérescence maculaire liée à l'âge
DoG	Difference of Gradient ou Différence de gradient
GA	Geographic atrophy ou atrophie géographique
GLCM	Grey level cooccurrence matrix ou Matrice de cooccurrence en nivrau de gris
GLRLM	Grey-level run-length matrix ou matrice de longueur de passe en niveau de gris
HOG	Histogram of Oriented Gradients ou Histogrammes de gradient orientée
HSV	Hue, Saturation, Value ou Teinte, saturation, valeur
IA	Instantaneous Amplitude ou Amplitude instantanée
ICG	Indocyanine green ou Vert indocyanine
IF	Instantaneous Frequency ou Amplitude instantanée
LBP	Local Binary Pattern ou Patron binaire local
LDA	Linear Discriminant Analysis ou Analyse linéaire discriminante
MLA	Maculopathie liée à l'âge
OCT	Optic coherence tomography ou Tomographie à cohérence optique
RGB	Red, Green, Blue ou Rouge, Vert, Bleu
ROI	Region of Interest ou Région d'intérêt
SBS	Sequential Backward Selection ou Sélection séquentielle arrière
SFS	Sequential Forward Selection ou Sélection séquentielle croissant
SFFS	Sequential Floating Forward Selection
SFBS	Sequential Floating Backward Selection
SIFT	Scale Invariant features transform
SURF	Speeded Up Robust features

VEGF

Vascular Endothelial Growth Factor ou Facteur de croissance vasculaire

LISTE DES ANNEXES

Annexe A – Anatomie de l’œil	95
Annexe B – Caractéristique de la Matrice de longueur de passe en niveau de gris.....	97
Annexe C – Article de conférence : Towards an Automatic Clinical Classification of Age-related Macular Degeneration	99

CHAPITRE 1 INTRODUCTION

La dégénérescence maculaire liée à l'âge (DMLA) est une maladie qui détériore progressivement la macula. Cette partie centrale de la rétine est le siège de la plus grande concentration de cellules photoréceptrices dans l'œil et permet ainsi la vision centrale et celle des détails. Cette maladie est une des principales causes de déficience visuelle et de cécité irréversible chez les personnes âgées dans les pays industrialisés (Organisation Mondiale de la Santé , 2014). Aux États-Unis, le nombre de personnes atteintes de DMLA est estimé à plus de 10 millions de personnes, dont 1,75 millions qui présentent une forme avancée de la maladie. En 2020, on prévoit que près de 3 millions de personnes souffriront d'une forme avancée de DMLA (Kasuga, Chen, & Zhang, 2011).

Une des méthodes couramment utilisées pour l'examen de la rétine est la photographie de fond d'œil. En effet, cette modalité permet de visualiser de manière rapide et non-invasive les structures importantes de la rétine, comme le disque optique, les vaisseaux sanguins et la macula. À partir de ces images, il est déjà possible de détecter la DMLA et d'en déterminer le stade. En fonction de ce dernier, un traitement spécifique est alors proposé.

Dans la suite de ce chapitre, nous introduisons la technique de photographie de fond d'œil ainsi que la DMLA, des concepts essentiels à la compréhension de ce projet. Puis, nous présentons l'organisation générale de ce mémoire.

1.1 Photographie de fond d'œil

Dans cette section, nous aborderons l'imagerie de fond d'œil (Abràmoff, Garvin, & Sonka, 2010), et plus particulièrement la photographie de fond d'œil en couleur.

1.1.1 Principe de la modalité

L'imagerie de fond d'œil est définie comme le mécanisme de représentation 2-D de la rétine issue de la projection de la rétine en 3-D par l'utilisation de lumière réfléchie. Il existe différents types d'imagerie de fond d'œil, comme par exemple la photographie de fond d'œil en stéréo ou l'angiographie à fluorescéine. Dans notre cas, nous nous intéressons plus particulièrement à la photographie de fond d'œil en couleur, dont les intensités représentent la quantité de longueurs d'onde de rouge, de vert et de bleu qui sont réfléchies par la rétine. Cette méthode est un moyen rapide et non-invasif d'obtenir une image planaire de la rétine. Notamment, cette imagerie très

informative est très largement utilisée et est recommandée pour une détection précoce des maladies de la rétine.

L'acquisition se fait par une caméra digitale qui capture l'image de l'intérieur de l'œil à travers la pupille. Les structures importantes de la rétine peuvent alors être visualisées comme le disque optique, les vaisseaux sanguins rétiniens et la macula (Figure 1-1). Les structures anatomiques de l'œil sont expliquées plus en détail dans l'Annexe A.

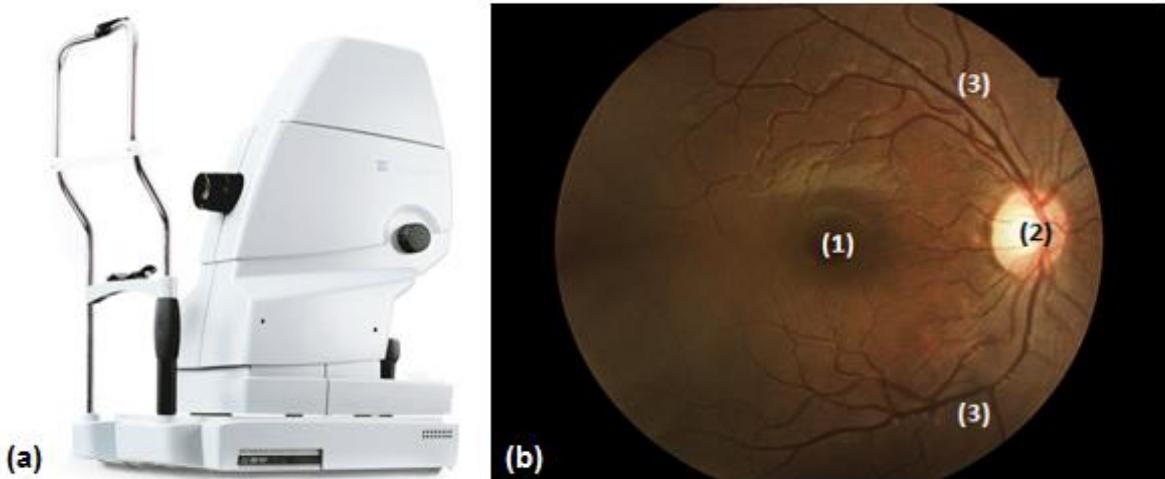


Figure 1-1: (a) Appareil d'acquisition (Diagnos, 2015), (b) Photographie de fond d'œil : (1) macula, (2) disque optique, (3) vaisseaux sanguins rétiniens

La photographie de fond d'œil présente quelques défis techniques. Un des problèmes majeurs est que la rétine est peu réfléchissante. De ce fait, une illumination externe via un flash lumineux est donc nécessaire pour imager la rétine. Cependant, l'illumination de la rétine et la réflexion des rayons lumineux vers le capteur s'effectuent sur un même chemin optique à travers la pupille. Ces deux faisceaux ne peuvent pas se superposer sinon cela résulte en des réflexions de la cornée et du cristallin qui vont diminuer ou éliminer le contraste de l'image. Ceci est corrigé par un système qui sépare les faisceaux pour combiner l'illumination et l'acquisition (miroir semi-réfléchissant ou miroir avec un trou central) et un système de correction qui élimine les réflexions de la cornée (obscurateur central dans le système d'illumination). La conséquence d'utiliser des chemins séparés dans le plan de la pupille est qu'au niveau de l'objectif, les ouvertures optiques sont de

l'ordre de seulement quelques millimètres. La formation des images s'effectue par un système constitué d'un objectif et de lentilles de zoom. Enfin, un système d'accommodation est positionné en fin de parcours de manière à corriger les erreurs de réflexion propres au sujet (DeHoog & Schwiegerling, 2009). Les artefacts de réflexions et d'illumination ne sont pas rares sur les images de fond d'œil et constituent un des problèmes majeurs lors de leur analyse automatique.

1.1.2 Artefacts dans les photographies de fond d'œil

Dus aux difficultés techniques et aux conditions d'acquisition, différents artefacts peuvent se présenter au niveau des photographies de fond d'œil. Ces artefacts peuvent perturber l'analyse des images, notamment pour poser un diagnostic fiable. On en vient donc à introduire le concept de qualité de l'image. Elle peut être définie par deux propriétés de l'image : la finesse et l'illumination. La finesse de l'image correspond à la capacité d'apercevoir clairement les structures. Elle dépend principalement du focus de la caméra au moment de l'exposition, mais aussi de facteurs externes qui pourraient générer du flou comme par exemple le mouvement. L'illumination représente à quel point l'objet est exposé à la lumière, que ce soit sous- ou surexposé. Ces deux facteurs dépendent donc généralement de l'appareil utilisé et du niveau d'expérience de la personne qui fait l'acquisition (Bartling, Wanger, & Martin, 2009).

Les artefacts qui viennent influencer la qualité de l'image sont causés par différents facteurs. Le clignement des yeux peut amener à une acquisition où les cils et les paupières sont imagés (Figure 1-2.a et Figure 1-2.b) et introduire également des réflexions très intenses (Figure 1-2.c) ou des ombres (Figure 1-2.d). Des défauts d'éclairage dus au flash peuvent également introduire des zones sous- ou surexposées à la lumière. Des lentilles salies par la poussière ou d'autres dépôts peuvent engendrer des images floues et des points blancs de différentes tailles peuvent également apparaître, cachant ainsi certaines structures de la rétine (Figure 1-2.e et Figure 1-2.f). Des artefacts d'arc qui obscurcissent l'image se présentent lorsque la distance entre le patient et la caméra n'est pas adéquate (Figure 1-2.f). Un mauvais alignement de la caméra peut entraîner un halo « brumeux » dans le pourtour de la rétine, diminuant ainsi le contraste de l'image (Figure 1-2.g). Le flou peut également être causé par le mouvement des yeux ou de la tête du patient, ainsi que par un mauvais focus de la caméra (Figure 1-2.h). Enfin, selon l'appareil utilisé, des images peuvent être acquises avec une basse résolution et ne contiennent alors pas suffisamment de détails sur les petites structures, comme les drusen. De plus, la compression d'image dans un contexte de

télémédecine peut également engendrer des pertes d'information de l'image originale (Fasih, Langlois, & Cheriet, 2014). Dans ce projet, nous considérons qu'une image est de mauvaise qualité si elle contient au moins un de ces artefacts.

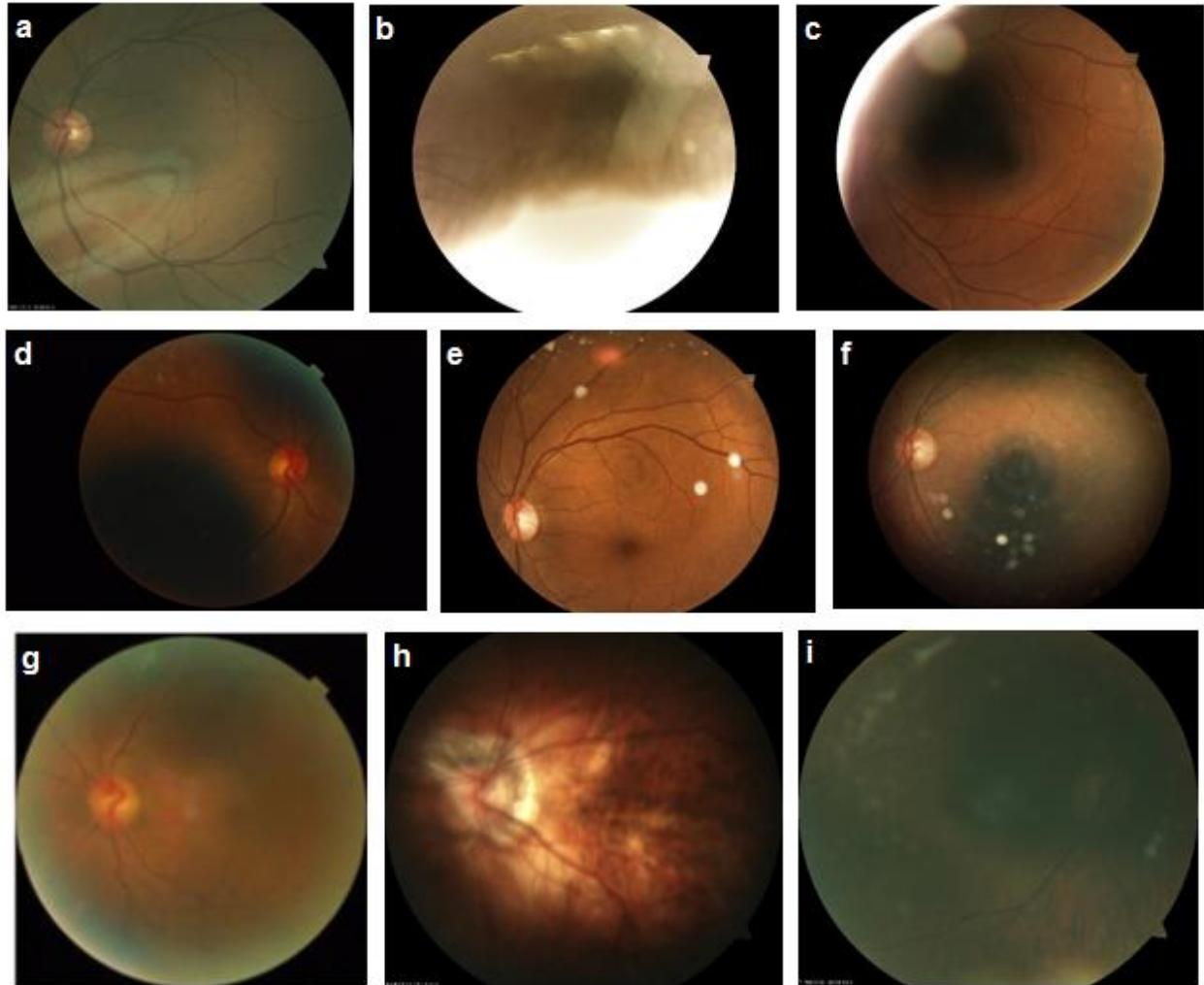


Figure 1-2: Exemples d'artefacts dans les photographies de fond d'œil : (a) acquisition de cils, (b) cils et clignement des yeux entraînant une réflexion intense, (c) réflexion de bord, (d) mauvaise illumination, (e) réflexions spéculaires à cause de dépôts sur la lentille, (f) arcs et dépôts sur la lentille, (g) halo brumeux, (h) flou, (i) artefact causé par une petite pupille.

1.2 Évaluation clinique de la DMLA

1.2.1 Description de la maladie

La dégénérescence maculaire liée à l'âge (DMLA) est une des principales causes de déficience visuelle et de cécité irréversible chez les personnes âgées dans les pays industrialisés (Kasuga, Chen, & Zhang, 2011). Les facteurs de risque sont nombreux, d'origines génétiques ou acquises par le tabagisme, l'alimentation, l'alcool ou suite à une maladie cardiovasculaire (Coleman, 2011). Cette maladie consiste en un groupe hétérogène d'anomalies conduisant à une détérioration progressive de la macula. À un stade précoce appelé maculopathie liée à l'âge (MLA), le patient ne présente aucun symptôme apparent et aucune altération importante de la fonction visuelle. Des signes sont cependant visibles sur les photographies de fond. La maladie peut rester stable ou évoluer vers une des deux formes avancées : la forme sèche, aussi appelée la DMLA atrophique, et la forme humide, appelée la DMLA exsudative. Dans les deux cas, le patient perd sa vision centrale de manière irréversible (Haute Autorité de la Santé, 2012). Les différentes formes de la maladie sont explicitées plus en détail dans cette section.

1.2.2 La maculopathie liée l'âge

La MLA est principalement caractérisée par la présence de drusen. Ils correspondent à des accumulations de résidus protéino-lipidiques issus de la phagocytose des photorécepteurs par les cellules de l'épithélium pigmentaire. Ils se déposent au niveau de la membrane de Bruch qui sépare l'épithélium pigmentaire et la choroïde (Figure 1-3).

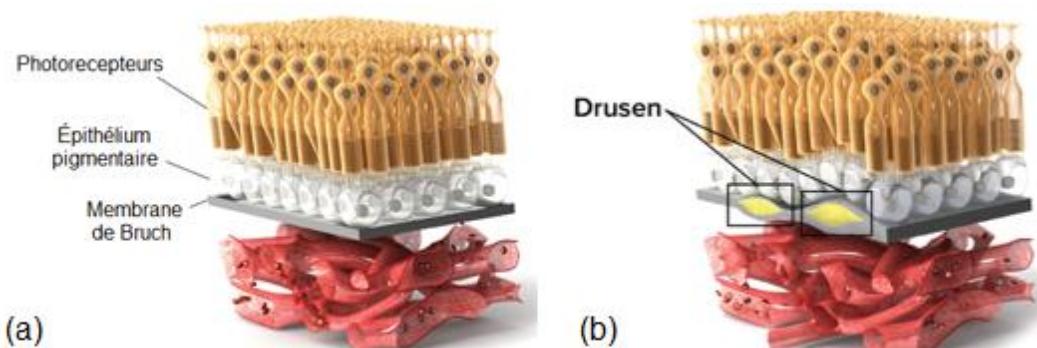


Figure 1-3: Couche de fond d'œil pour (a) un œil sain et (b) un œil avec maculopathie liée à l'âge (The Angiogenesis Foundation, 2013).

Sur les images de fond d'œil, les drusen apparaissent dans la zone maculaire comme de petites taches, blanchâtre ou jaunâtre, de forme et de taille variables. Il existe différents types de drusen (Jonisch & Shah, 2011) :

- Drusen miliaires ou *hard drusen* (Figure 1-4.a) : ce sont de petites taches jaunes rondes et bien définies, de diamètre inférieur à 50 μm . Ils sont asymptomatiques et peuvent se présenter chez de jeunes personnes (moins de 50 ans), ne conduisant pas forcément vers une forme avancée.
- Drusen séreux ou *soft drusen* (Figure 1-4.b) : ce sont des drusen de forme plus diffuse, dont la taille est supérieure à 125 μm . Ils sont souvent asymptomatiques mais peuvent s'accompagner d'une diminution des capacités visuelles à l'obscurité. Ces drusen conduisent à un épaississement de la membrane de Bruch et par conséquent, le risque de progresser vers un stade avancé est grand. Ce phénomène abime les photorécepteurs qui se dégradent et cela prédispose à la néovascularisation.
- Drusen cuticulaire (Figure 1-4.c) : ce sont des drusen qui sont identiques aux drusen séreux du point de vue histologique et anatomique, sauf qu'on en trouve chez les jeunes personnes et qu'ils ne prédisposent pas à la DMLA.

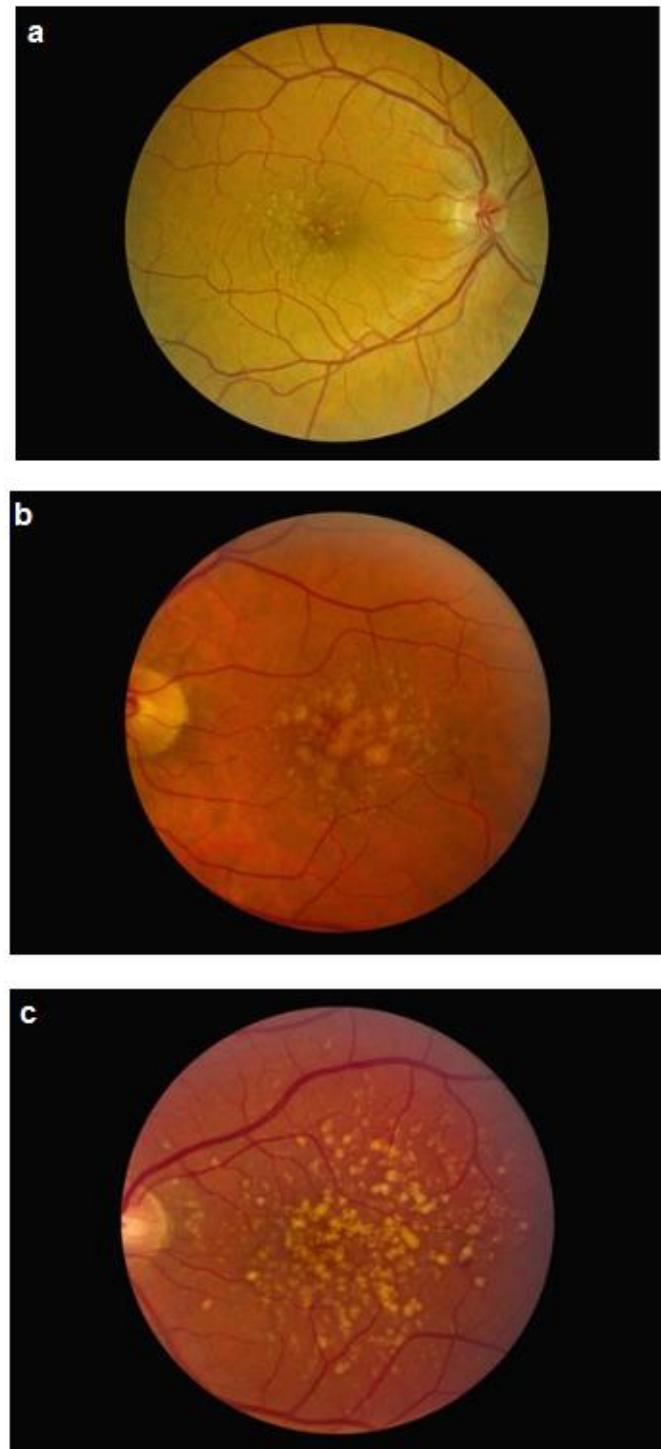


Figure 1-4: Types de drusen : (a) drusen miliaires, (b) drusen séreux et (c) drusen cuticulaire
(Jonisch & Shah, 2011).

Autres que les drusen, des altérations pigmentaires de type hypopigmentation ou hyperpigmentation de l'épithélium pigmentaire peuvent être également observées dans les cas de MLA. Ces altérations correspondent à un début de mort cellulaire de l'épithélium pigmentaire (Haute Autorité de la Santé, 2012).

À cause des hautes activités métaboliques et aux expositions régulières à la lumière, la rétine est très sensible aux dommages oxydatifs, surtout en présence de drusen. Avec l'âge, la prévalence des drusen augmente, étant donné les déchets métaboliques produits par des cellules vieillissantes. La dégénérescence de la macula s'intensifie alors à cause des disfonctionnements des tissus rétiniens et de l'accumulation des dommages oxydatifs. L'utilisation d'antioxydants, comme les bêta-carotènes ou certains acides gras, et de suppléments de vitamines, comme les vitamines C, E et B, est alors recommandée pour atténuer ces dommages. Ceux-ci sont donnés à titre préventif, afin de réduire la progression de la maladie, notamment en empêchant la génération de néovaisseaux suite à des inflammations (Meleth , Raiji, Krishnadev, & Chew, 2011). Un suivi régulier du patient est également important pour détecter le plus rapidement les premiers signes de perte d'acuité visuelle qui pourraient mener vers un stade avancé.

1.2.3 La DMLA atrophique (forme sèche avancée)

La DMLA atrophique ou sèche est la forme avancée d'une DMLA non-exsudative, caractérisée par l'amincissement progressif de l'épithélium pigmentaire, appelé atrophie géographique (GA). Cette atrophie est associée à une forte présence de drusen et à des anomalies de l'épithélium. La forte pression due aux néovaisseaux sur l'épithélium pigmentaire affaibli conduit à la mort progressive des photorécepteurs, causant lentement la perte d'acuité visuelle (Haute Autorité de la Santé, 2012).

Sur les images de fond d'œil, les GA apparaissent sous des formes très variées. Cela peut être une zone jaune singulière ou plusieurs zones d'atrophie (Figure 1-5), où les vaisseaux sanguins de la choroïde deviennent anormalement visibles. Ces atrophies sont issues de l'élargissement et l'union de drusen séreux venant de lieux d'atrophie distincts (Figure 1-6). Dans une forme précoce de la maladie, l'atrophie se trouve autour de la fovéa. Parmi les formes diverses, certaines atrophies existantes peuvent fusionner avec de nouvelles zones d'atrophies résultant en des configurations en fer à cheval. Lorsque l'atrophie géographique touche la zone fovéolaire, on se trouve alors dans un cas avancé avec perte de vision (Laud, Mukkamala, Brue, & Slakter, 2011).

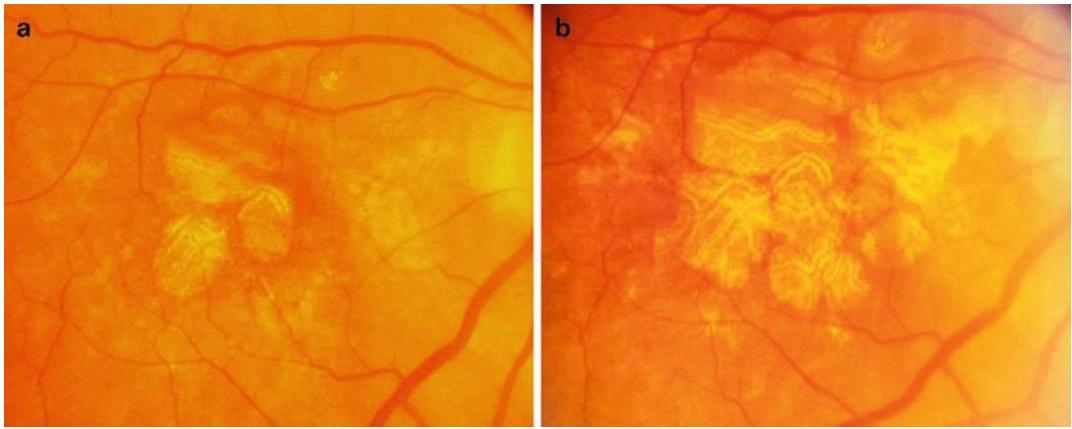


Figure 1-5: Progression de la GA : (a) Deux régions distinctes de GA, (b) progression de GA qui a changé de forme et s'est élargi, huit ans plus tard (Laud, Mukkamala, Brue, & Slakter, 2011).

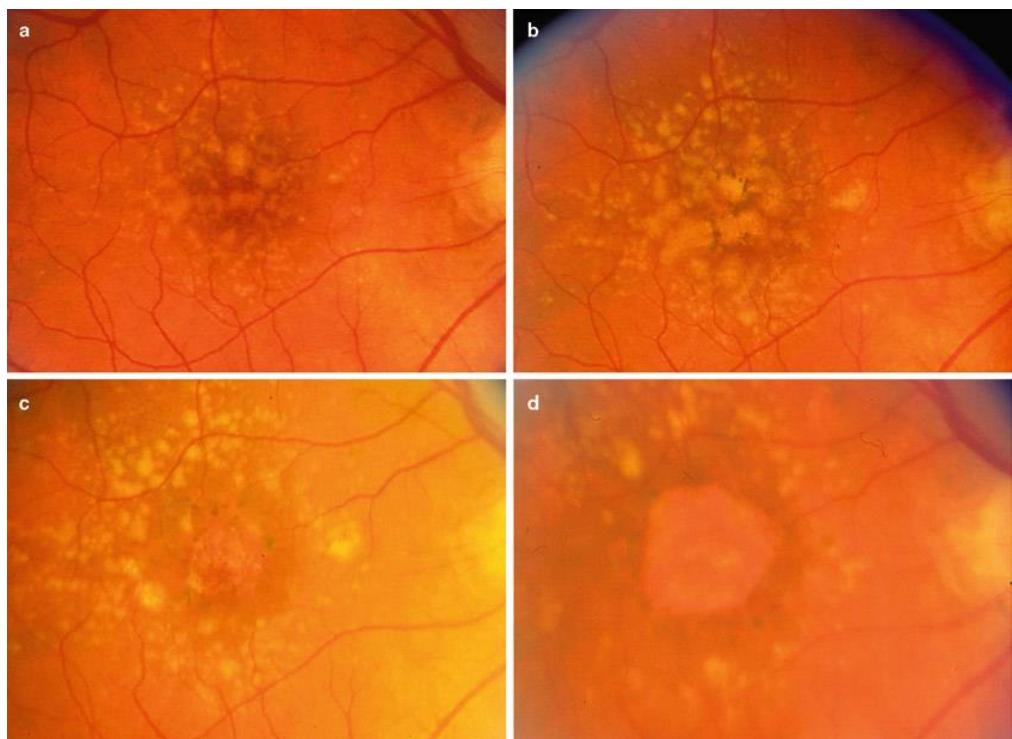


Figure 1-6: Progression de la DMLA sur 11 ans : (a) ensemble de drusen séreux dans la macula, (b) formation de drusen additionnels et déposition de pigment, (c) résorption de drusen au centre formant le début de l'atrophie après 8 ans, (d) formation de l'atrophie géographique bien définie après 11 ans (Laud, Mukkamala, Brue, & Slakter, 2011).

À ce jour, il n'existe aucun traitement qui a pu démontrer son efficacité pour traiter la DMLA atrophique. Cependant, la progression est lente et se fait sur plusieurs années. Accompagnées de compléments pour ralentir la maladie, une rééducation orthoptique et des aides visuelles peuvent alors être proposées au patient. Des études cliniques sont menées pour trouver des moyens de modérer et potentiellement restaurer la perte de vision, avec l'introduction de nouveaux médicaments, de la thérapie génique ou l'application de prothèses rétiniennes (Laud, Mukkamala, Brue, & Slakter, 2011).

1.2.4 La DMLA exsudative

La DMLA exsudative ou humide est la forme avancée caractérisée par la prolifération de néovaissaux anormaux d'origine choroïdienne. Cette néovascularisation (Figure 1-7) est causée par des produits dérivés de réactions inflammatoires, tel que des facteurs de croissance vasculaire (VEGF). La DMLA exsudative se présente suite à une traversée de ces néovaissaux à travers la membrane de Bruch, qui entraîne alors un décollement maculaire exsudatif, une accumulation de fluide dans l'espace sous-rétinien et des hémorragies à l'intérieur de l'œil. Cette forme est responsable d'une perte rapide de la vision de manière définitive (Penha & Rosenfeld, 2011).

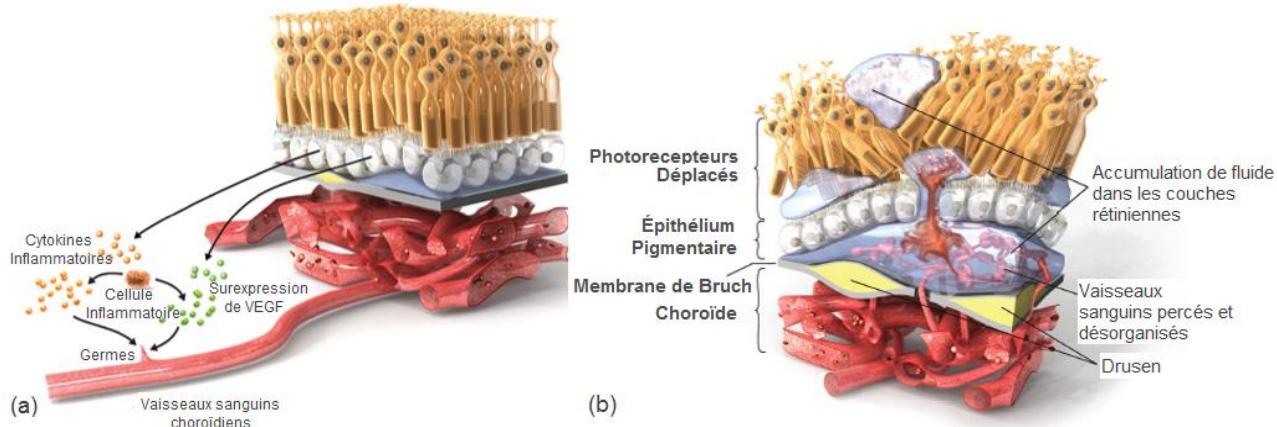


Figure 1-7: Formation de la néovascularisation : (a) réaction inflammatoire produisant les VEGF et (b) traversée des néovaissaux à travers la membrane de Bruch (The Angiogenesis Foundation, 2013)

Sur les images de fond d'œil, les hémorragies apparaissent comme des taches rouges sombres (Figure 1-8). Après un épisode hémorragique, il peut également y avoir des zones nécrosées qui apparaîtront comme des zones jaunes plus claires.



Figure 1-8: DMLA exsudative avec hémorragie (Jonisch & Shah, 2011).

Les traitements existants pour la forme exsudative visent principalement les néovaissieux choroïdiens. Des traitements par injection intraoculaire anti-VEGF, comme le ranibizumab ou le pegabtanib, sont généralement recommandés afin d'inhiber la croissance vasculaire. D'autres alternatives sont également proposées comme la photocoagulation au laser, la thérapie photodynamique par visudyne ou encore le traitement chirurgical. Ces derniers sont en particulier utilisés pour la destruction des néovaissieux visibles se trouvant au niveau de la fovéa (Haute Autorité de la Santé, 2012).

1.2.5 Méthode clinique de diagnostic

Dans cette section, sont décrites les méthodes cliniques utilisées par les médecins pour diagnostiquer la DMLA. Tout d'abord, l'apparition de symptômes de perte d'acuité visuelle doit être observée. Les patients atteints d'une forme précoce peuvent se plaindre d'une gêne en vision nocturne ou d'une diminution de la perception de contraste. Des pertes de vision se présentent chez les personnes atteintes d'une DMLA avancée. Ceci peut être vérifié par des tests de vue habituelle, et également via la grille d'Amsler (Figure 1-9). Un patient sain voit cette grille comme des lignes

horizontales et verticales parfaitement droite. Un patient qui voit des déformations dans le centre de la grille, appelées des métamorphopsies, est alors suspecté d'être atteint de DMLA (Haute Autorité de la Santé, 2012).

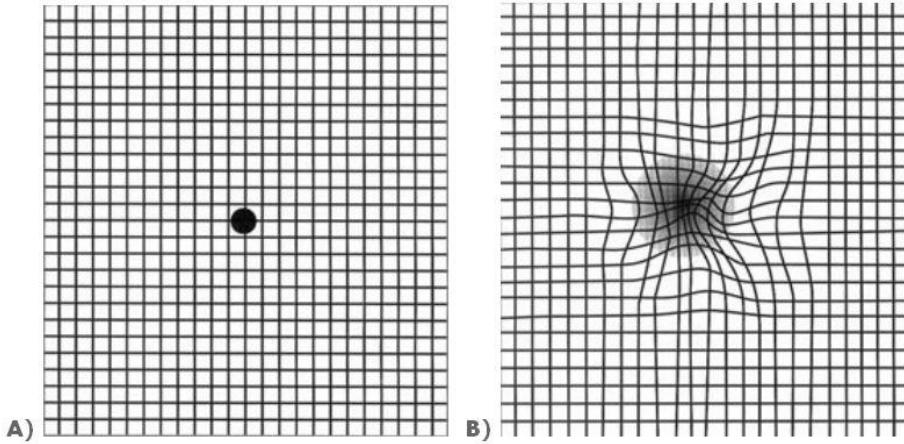


Figure 1-9: Grille d'Amsler vue par (A) un patient normal et (B) un patient qui présente des signes de DMLA (Université Médicale Virtuelle Francophone, 2013)

Ensuite, on a recourt à un examen par photographie de fond d'œil (voir Chapitre 1.1), qui permet un diagnostic rapide, et préciser l'étendue des altérations et leur topographie par rapport à la fovéa. En les archivant et les documentant, les images de fond d'œil permettent également de suivre la progression ou détecter des changements au fil du temps. Cet examen permet de détecter et d'évaluer le degré de sévérité de la DMLA. Pour confirmer le diagnostic, on a recourt à d'autres modalités comme la tomographie à cohérence optique (OCT) ou l'angiographie à fluorescéine.

Les images OCT de rétine (Figure 1-10) permettent de visualiser les différentes couches de la rétine et sont utilisées pour confirmer s'il y a bien un amincissement des couches de la rétine dans le cas d'une DMLA atrophique ou pour quantifier et localiser l'xsudation dans le cas de la DMLA exsudative. Ces informations servent également à établir le traitement adéquat (Haute Autorité de la Santé, 2012).

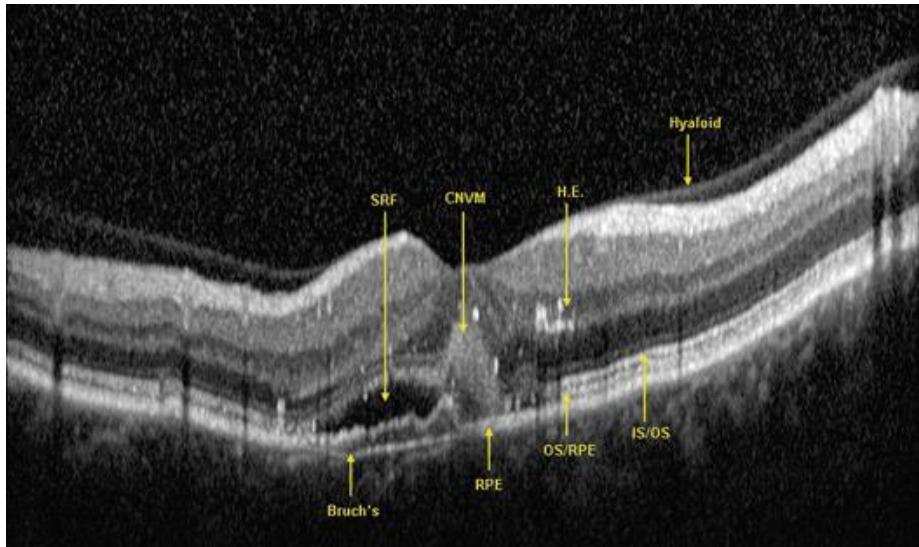


Figure 1-10: Image OCT d'une rétine avec DMLA exsudative présentant une néovascularisation choroïdienne (CNVM), drusen miliaires (HE) et une infiltration de fluide sous-rétinien (SRF) (Chiang, Witkin, Regillo, & Ho, 2011).

L'angiographie à fluorescéine (Figure 1-11) permet une évaluation détaillée de la vascularisation rétinienne. Le système d'acquisition est le même que la photographie de fond d'œil avec, en plus, l'ajout de filtres à bande réduite. La technique utilise comme colorant de la fluorescéine qui est injectée dans le système sanguin. Lors d'une exposition à la lumière bleu-vert, le colorant va absorber la lumière et réémettre une couleur de longueur d'onde plus grande (jaune-vert). Les vaisseaux sanguins de la rétine sont alors bien mis en évidence. Un autre type de colorant, le vert indocyanine (ICG), peut être également utilisé pour visualiser plus en détail les néovaisseaux (Chiang, Witkin, Regillo, & Ho, 2011). Cette modalité est utilisée pour confirmer le diagnostic en affirmant la présence de néovaisseaux choroïdiens. Elle permet de connaître leur topographie, ainsi que l'évaluation des lésions.

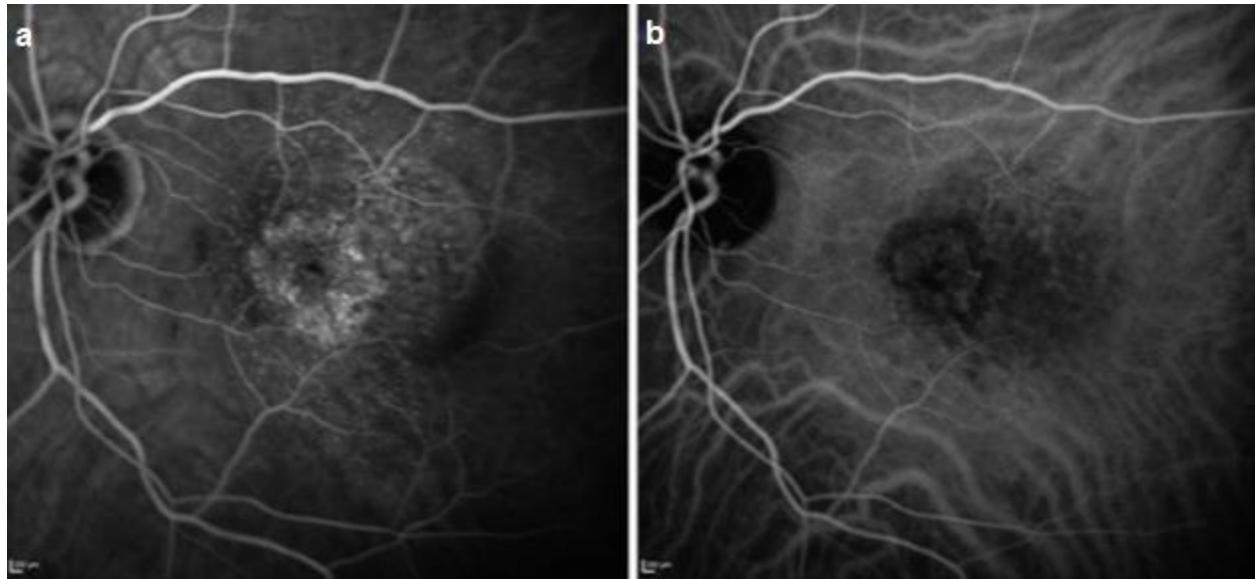


Figure 1-11: Angiographie de la rétine : (a) avec fluorescéine où les vaisseaux sanguins rétiniens et les drusen sont mis en évidence et (b) avec ICG où les néovaisseaux choroïdiens sont plus démarqués (Chiang, Witkin, Regillo, & Ho, 2011).

1.3 Classification de la DMLA à partir des images de fond d'œil

À partir des méthodes de diagnostic, le degré de sévérité de la DMLA est alors déterminé, notamment pour indiquer le traitement adéquat. Actuellement, il n'existe pas de classification clinique standard pour la DMLA mais celle qui est recommandée et généralement utilisée est la classification simplifiée proposée par l'AREDS (Age-Related Eye Disease Study Research Group (AREDS), 2005). Cette classification se divise en quatre catégories et se base sur les caractéristiques observées sur les photographies de fond d'œil, notamment le nombre et la taille des drusen, la localisation des GA et la présence de néovaisseaux ou d'hémorragies (Haute Autorité de la Santé, 2012):

- **Catégorie 1** : correspond à l'absence de DMLA. Cette catégorie est caractérisée par l'absence de drusen ou la présence en faible quantité de drusen dont le diamètre est inférieur à 63 µm.
- **Catégorie 2** : correspond au stade précoce de DMLA (MLA). Cette catégorie est caractérisée par la présence, soit de multiples petits drusen, soit quelques drusen de

diamètre entre 63 et 125 μm , soit une GA excluant la fovéa. Si au moins un de ces éléments est présent, on peut alors classer l'image de fond d'œil dans cette catégorie.

- **Catégorie 3** : correspond au stade modéré de la DMLA. Lorsqu'il y a présence de multiples drusen de diamètre entre 63 et 125 μm , ou des drusen de diamètre supérieur à 125 μm , ou une GA excluant la fovéa, on peut classer l'image dans cette catégorie.
- **Catégorie 4** : correspond au stade avancé de la DMLA. Cela correspond à des images de fond d'œil qui présentent une GA touchant la fovéa et/ou présence de DMLA exsudative.

Les cas de DMLA en catégorie 3 sont souvent importants à distinguer, de par le fait que les personnes ne présentent pas encore de cécité mais qu'ils sont à haut risque de développer une forme avancée de la maladie. Une fois ciblés, les patients seraient surveillés régulièrement et suivraient un traitement pour ralentir la progression.

1.3.1 Limitations des méthodes de diagnostic cliniques actuelles

Il existe différentes limitations aux méthodes cliniques réalisées par des évaluateurs humains. Avec le vieillissement de la population dans les pays industrialisés, le nombre de patients atteints de DMLA est en augmentation (Kasuga, Chen, & Zhang, 2011), par rapport au nombre de médecins disponibles pour une consultation. La télémédecine permet de centraliser les images de fond d'œil pour les analyser, mais le nombre d'images devient de plus en plus important rendant le travail long et laborieux pour les cliniciens. De plus, l'évaluation de la DMLA se base sur quelques critères subjectifs comme le nombre de drusen et la localisation par rapport à la fovéa. Les seuils sur ces deux critères ne sont pas vraiment précisés et le diagnostic se base fortement sur le jugement du clinicien. Ceci peut donc entraîner des variabilités inter- et intra-experts dans la classification de la DMLA.

Une des difficultés majeures est de pouvoir distinguer la DMLA parmi les autres cas qui se présentent dans les photographies de fond d'œil. Il existe notamment des maladies qui partagent des caractéristiques similaires à la DMLA, comme la rétinopathie diabétique (Figure 1-12). Les éléments confondant de la rétinopathie diabétique sont les exsudats et les cotton wool spots. Les exsudats sont des lésions qui ont des structures ressemblant à celle des drusen miliaires. Ils se différencient cependant par une plus grande luminance et par leur localisation qui est plus aléatoire. Les drusen miliaires ont une forme ronde bien définie avec une luminance peu marquée et ont plus

tendance à se regrouper dans la zone maculaire. Les cotton wool spots sont des lésions importantes et plus diffuses qui pourraient être confondues avec des drusen séreux. Ils correspondent en fait à une ischémie qui entraîne la mort cellulaire de la rétine. Les drusen séreux sont plus diffus et se regroupent au niveau de la macula. Dans la rétinopathie diabétique avancée, il peut également y avoir des hémorragies mais celles-ci ne sont pas spécifiquement localisées au niveau de la macula (Niemeijer, van Ginneken, Russel, Suttorp-Schulzen, & Abràmoff, 2007; van Grinsven, et al., 2013). Lorsque les symptômes de la rétinopathie diabétique se présentent au niveau de la macula (Figure 1-13), les médecins ont souvent besoin de connaître le dossier médical du patient, ses antécédents et ses habitudes de vie, pour établir le diagnostic.

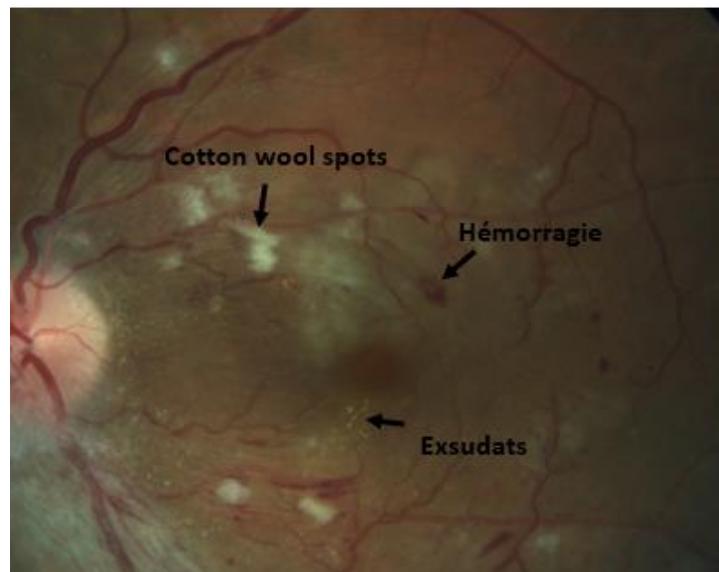


Figure 1-12: Photographie de fond des différentes structures caractérisant la rétinopathie diabétique.

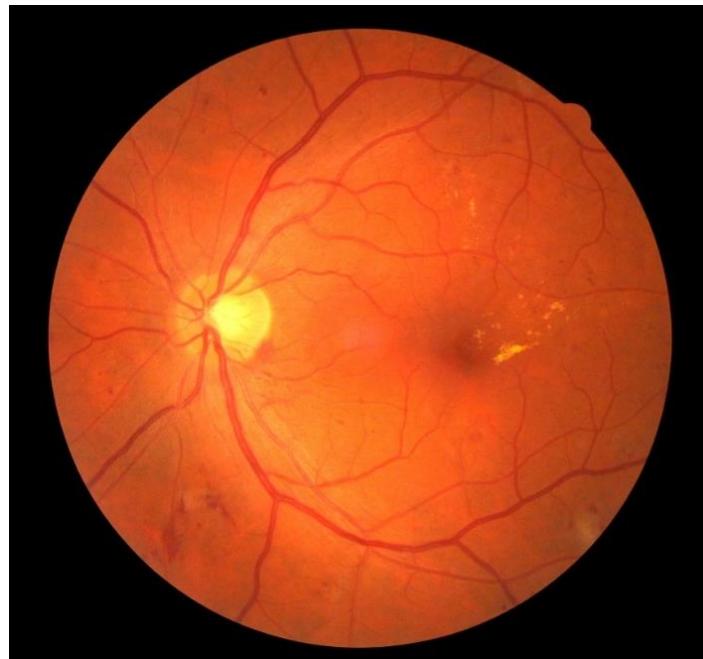


Figure 1-13: Photographie de fond d'œil d'un cas où les exsudats de la rétinopathie diabétique se trouvent au niveau de la macula.

De ce fait, différents travaux ont proposés des systèmes automatiques pour effectuer une classification de la DMLA, afin de rendre son évaluation plus rapide, et de permettre la reproductibilité des résultats sur le diagnostic. L'objectif de ce projet de recherche est donc de développer et valider un système automatique qui permet l'évaluation de la DMLA à partir d'images de fond d'œil en couleur.

1.4 Organisation du mémoire

Le présent mémoire se subdivise en quatre chapitres. Suite à ce chapitre introductif, une revue de littérature est présentée afin de comprendre les concepts clés pour la construction d'un système de classification à partir d'images médicales, et également de détailler les méthodes automatiques existantes sur l'évaluation de la DMLA. Le chapitre suivant est constitué d'un article scientifique, soumis en juin 2015, à *IEEE Transactions on Biomedical Engineering*. Celui-ci explique la méthodologie utilisée pour la classification des images de fond d'œil selon le degré sévérité de la DMLA. Cet article présente les études menées sur les performances de différentes caractéristiques, sur le choix d'un système de classification et sur la robustesse du système à la qualité de l'image. Un chapitre de discussion va permettre de revenir sur les résultats importants et d'émettre les avantages et les limites du système proposé. Enfin, le mémoire se termine avec le chapitre de conclusion qui propose entre autres les recommandations pour des travaux ultérieurs. Des annexes permettent en plus d'ajouter des informations complémentaires au présent projet, notamment l'Annexe C, qui présente des résultats préliminaires au développement du système de classification.

CHAPITRE 2 REVUE DE LITTÉRATURE

Ce chapitre présente une revue des connaissances relatives au sujet de recherche. La première partie porte sur les concepts de base pour l'élaboration d'un système de classification. Ensuite, la deuxième partie présente les différents travaux qui ont été proposés sur la détection et l'évaluation automatique de la DMLA. Finalement, la dernière partie du chapitre expose les objectifs et hypothèses de recherche du présent projet.

2.1 Concept de reconnaissance de forme statistique

Cette section permet d'expliquer les outils qui sont utilisés pour développer le système de classification pour la DMLA. Tout d'abord, des caractéristiques de l'image sont extraites afin de représenter l'image sous forme de vecteur. Si le nombre de caractéristiques est trop grand par rapport au nombre d'exemples d'apprentissage, une étape de sélection de caractéristiques permet de réduire la dimensionnalité du problème. Enfin, les valeurs de caractéristiques sont utilisées en entrée d'un système de classification. Celui-ci est alors modélisé à partir d'exemples lors de la phase d'apprentissage et sa performance est évaluée avec de nouvelles images lors de la phase de test.

2.1.1 Extraction de caractéristiques

Afin d'analyser plus facilement les images, celles-ci sont généralement représentées dans une forme plus simple qui garde une certaine information de l'image. L'extraction de caractéristiques consiste à extraire cette information, qui est encodée ensuite sous forme d'un vecteur de caractéristiques. Par exemple, on peut extraire des histogrammes d'intensité, la moyenne des intensités de l'image, leur médiane ou encore leur écart-type. Ensuite, le vecteur de caractéristiques est calculé pour chaque échantillon, ce qui génère la matrice de caractéristiques servant d'entrée au système de classification. Dans ce chapitre, les caractéristiques communément utilisées dans l'analyse d'images biomédicales sont brièvement abordées, et les caractéristiques utilisées spécifiquement dans le présent projet sont explicitées plus en détail.

2.1.1.1 Caractéristiques de couleurs

L'information de couleur est une caractéristique qu'on a tendance à intuitivement choisir. Par exemple, dans le cadre d'analyse d'images de fond d'œil, les différentes structures de la rétine

présentent des couleurs différentes : la macula est orange sombre, le disque optique est jaune très clair et les vaisseaux sanguins sont rouge sombre. Cette caractéristique est notamment utilisée dans ce projet, car les différents éléments de la DMLA sont eux-mêmes caractérisés par des couleurs spécifiques : les drusen sont jaunâtres-blanchâtres, les hémorragies sont rouges, les GA sont jaunes et les rétines saines sont oranges. Il existe plusieurs espaces de couleurs qui sont présentés dans cette section.

L'espace de couleur le plus communément utilisé est l'espace RGB (Figure 2-1), qui recompose les couleurs comme une combinaison d'intensité de rouge (R), vert (G) et bleu (B) qui correspondent chacune à un canal de l'image. Ce mode de représentation est couramment utilisé pour afficher les images, mais à cause de la très forte corrélation entre les trois canaux, il est très peu pratique pour le traitement d'image. Par exemple, l'évaluation de la similarité des couleurs est impossible dans cet espace (Cheng, Jiang, Sun, & Wang, 2001). Dans le cadre de l'évaluation de la DMLA, les différents travaux utilisent uniquement le canal vert pour le traitement de l'image et l'extraction de caractéristiques. En effet, c'est dans le canal vert que les lésions et les vaisseaux sanguins offrent le plus de contraste. De plus, le canal rouge est souvent considéré comme saturé, offrant peu de contraste, et le canal bleu présente beaucoup de bruit dans les images de fond d'œil (Walter, et al., 2007). Cependant, ces deux canaux ne sont pas à exclure, contenant potentiellement des informations complémentaires au vert, notamment pour la discrimination des lésions claires comme les drusen et les exsudats (van Grinsven, et al., 2013).

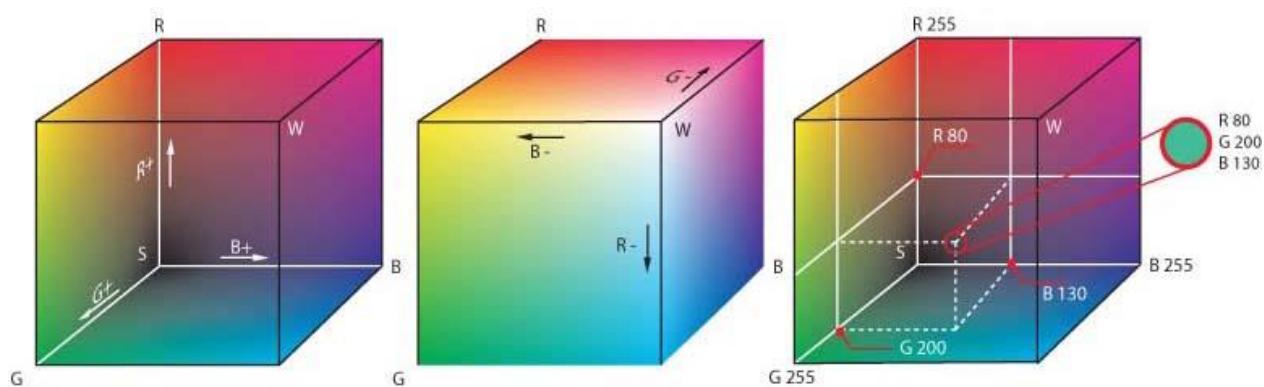


Figure 2-1: Espace de couleur RGB : (a) représentation du coin inférieur, (b) représentation du coin supérieur et (c) exemple d'une couleur en RGB (Horst Frank, 2004). Tiré de

http://commons.wikimedia.org/wiki/File:RGB_farbwuerfel.jpg

Un espace de couleur également utilisé dans le cadre de l'évaluation de la DMLA est l'espace $L^*a^*b^*$ (Kankanhalli, Burlina, Freund, & Bressler, 2013). Cet espace, illustré sur la Figure 2-2, représente la couleur par un paramètre d'intensité, la luminance (L), et deux paramètres de chrominance, un qui représente les teintes du vert au magenta (a), et l'autre du bleu au jaune (b). L'avantage de cet espace de couleur est que la similarité des couleurs peut être mesurée du fait que les distances entre les couleurs sont euclidiennes. De plus, les tons contenus dans le canal L et les couleurs contenues dans les canaux a et b sont indépendants, et donc un changement de l'un n'affecte pas l'autre (Cheng, Jiang, Sun, & Wang, 2001). Un espace similaire est l'espace Y^*Cb^*Cr , qui représente également la couleur en une composante de luminance (Y) et deux de chrominance (Cb et Cr qui correspondent respectivement à la différence de bleu et rouge par rapport à Y). Cet espace a notamment été utilisé comme caractéristique pour la discrimination entre drusen et exsudats (van Grinsven, et al., 2013).

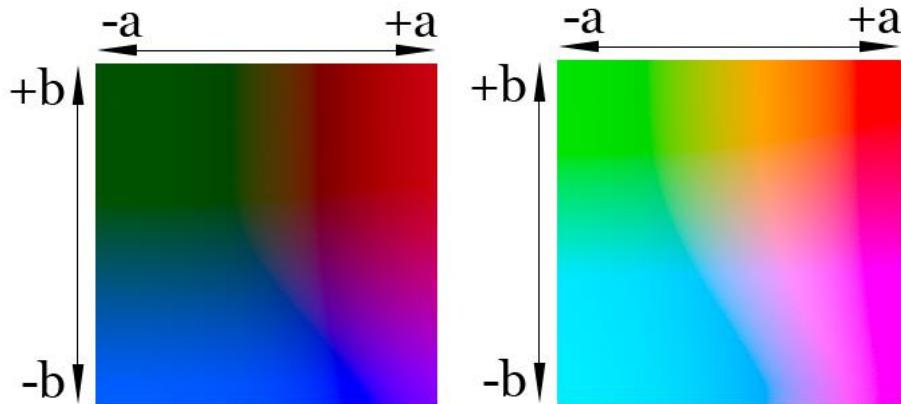


Figure 2-2: Espace de couleur $L^*a^*b^*$ à (a) luminance 25% et (b) luminance 75% (JacobVoss, 2004). Tiré de http://commons.wikimedia.org/wiki/File:Lab_color_at_luminance_25_percent.png et http://commons.wikimedia.org/wiki/File:Lab_color_at_luminance_75%25.png

Enfin, la représentation en HSV (Figure 2-3) est un système plus intuitif à la vision humaine et est très utilisée en traitement d'image. Elle consiste à représenter la couleur par la teinte (H), la saturation (S) et la valeur (V). La teinte représente les couleurs basiques déterminées par la longueur d'onde dominante dans le spectre de la lumière. La saturation correspond à la mesure de

la pureté de la couleur et de la quantité de lumière blanche mélangée à la teinte. La valeur correspond tout simplement à l'intensité lumineuse de la couleur et fait référence à l'ajout d'intensité de blanc ou de noir aux teintes (Cheng, Jiang, Sun, & Wang, 2001).

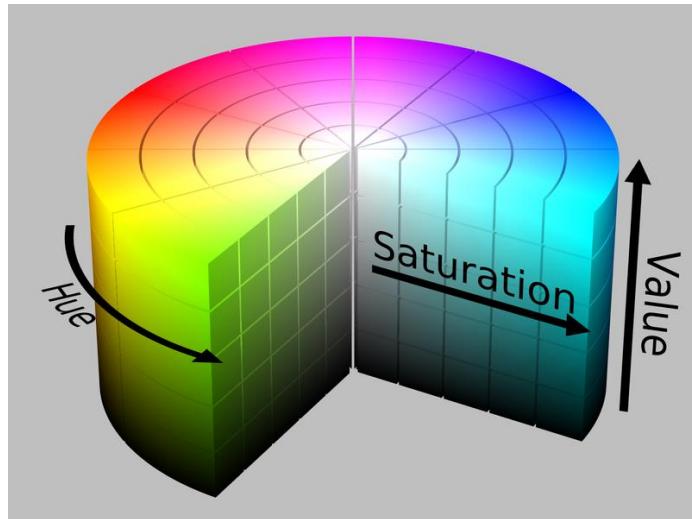


Figure 2-3: Espace de couleur HSV (SharkD, 2010). Tiré de http://commons.wikimedia.org/wiki/File:HSV_color_solid_cylinder_alpha_lowgamma.png

Chaque canal de couleur peut être utilisé en tant que caractéristique, via par exemple les histogrammes de couleurs, ou bien peut servir comme image d'entrée à d'autres méthodes d'extraction de caractéristiques.

2.1.1.2 Caractéristiques de texture

La texture d'une image est une information sur l'arrangement spatial de ses intensités. C'est une caractéristique très importante souvent utilisée en analyse d'images. La rétine présente notamment des textures bien ordonnées qui varieraient selon son état de santé. Cette section présente quelques caractéristiques de texture couramment utilisées pour la classification.

Une grande famille de caractéristiques de texture est celle des caractéristiques basées sur un opérateur. Ces méthodes consistent généralement à convoluer l'image avec un masque qui correspond à un opérateur, pour obtenir les caractéristiques de texture (Reed & Hans du Buf, 1993).

Les filtres de Gabor sont des fonctions de noyau gaussien modulé avec une sinusoïde, souvent

utilisés pour la détection de contour. En analyse de texture, la méthode consiste à filtrer l'image avec des filtres de Gabor à différentes orientations et d'extraire l'information sur la variation dans la réponse aux filtres sur des surfaces différentes (Porter & Canagarajah, 1997).

La transformée en ondelettes permet de décomposer l'image en ses composantes fréquentielles et a pour avantages, par rapport à la transformée de Fourier, de permettre une localisation spatiale et une flexibilité des fonctions de base qui sont appelés ondelettes. Le résultat de la transformée en ondelettes dépend de la fonction de base choisie, dont des exemples sont les ondelettes de Haar, de Daubechies, ou de Lemarié. Le résultat obtenu à chaque décomposition est une image de résolution plus basse, appelé coefficient d'approximation, et 3 images contenant les informations de haute résolution en horizontal, vertical et diagonale, appelés coefficients de détails. De plus, le fait que l'image soit décomposée répétitivement dans des gammes de fréquences plus basses permet une analyse en multi-résolution (Porter & Canagarajah, 1997).

Une autre grande famille de caractéristiques est celle des caractéristiques basées sur les statistiques.

La matrice de longueur de passe en niveau de gris (GLRLM, pour *Grey Level Run-Length Matrix*) $p(i, j)$ est définie comme étant le nombre de lignes dont les pixels sont d'intensité de gris i et de longueur j , pour une certaine orientation donnée (Tang, 1998). Cette matrice est souvent calculées pour les quatre directions principales : 0° , 45° , 90° et 135° . Un exemple de calcul de la GLRLM est présenté sur la Figure 2-4. Des caractéristiques de la GLRLM peuvent ensuite être extraites pour former le vecteur de caractéristiques (voir Annexe B). L'extraction de la texture par les GLRLM est une méthode simple mais les applications sont limitées comparés à d'autres méthodes d'extraction de texture qui ont surpassés cette méthode, comme par exemple les matrices de cooccurrence.

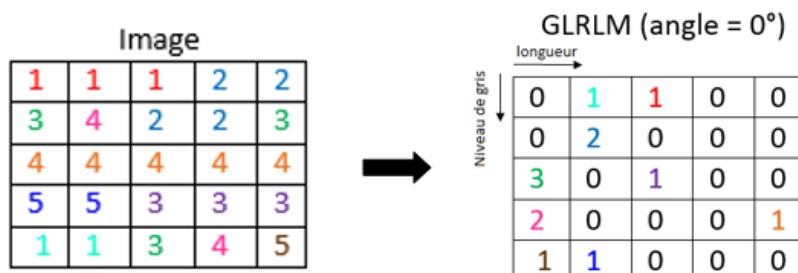


Figure 2-4: Exemples de calcul de la GLRLM : (a) matrice d'une image et (b) GLRLM correspondant dans la direction 0° .

Les matrices de cooccurrence (GLCM, pour *Grey-Level Cooccurrence Matrix*) définissent également des matrices de niveaux de gris spatialement dépendant (Haralick, Shanmugam, & Dinstein, 1973). À partir d'une image $I(x, y)$ de taille $L_r \times L_c$, la matrice de cooccurrence $P(i, j)$ se construit sur la mesure de la probabilité d'occurrence d'une paire d'intensités (i, j) séparées par une distance d dans la direction d'angle θ :

$$P(i, j, d, \theta) = \# \{ ((x_1, y_1), (x_2, y_2)) \in L_r \times L_c \mid I(x_1, y_1) = i, I(x_2, y_2) = j \} \quad (2-1)$$

Avec $\#$ qui symbolise le nombre d'éléments, et les positions des pixels dans l'image (x_1, y_1) et $(x_2, y_2) = (x_1, y_1) + (d\cos(\theta), d\sin(\theta))$. La taille de la matrice dépend du nombre de niveaux de gris contenus dans l'image. Afin de réduire le temps de calcul, la GLCM peut être calculée pour un nombre réduit de niveaux de gris en sous-échantillonnant les valeurs d'intensité de l'image. Tout comme la GLRLM, on calcule généralement la GLCM pour les quatre directions principales et plusieurs caractéristiques, comme par exemple l'entropie ou le second moment angulaire, sont calculées pour servir à la classification (Reed & Hans du Buf, 1993).

La modulation d'amplitude et modulation de fréquence (AM-FM, pour *amplitude modulation-frequency modulation*) représente une image comme étant la somme de ses composantes AM-FM, c'est-à-dire que l'image est décomposée en un ensemble de sinusoïdes basiques et leur amplitude (Havlicek, 1996). La décomposition AM-FM d'une image $I(x, y)$ de type bande restreinte est modélisée par :

$$I(x, y) = \sum_{n=1}^M a_n(x, y) \cos[\Phi_n(x, y)] \quad (2-2)$$

Où M est le nombre de composants AM-FM, $a_n(x, y)$ l'estimée de l'amplitude instantanée et $\Phi_n(x, y)$ la phase instantanée. Afin de couvrir la large gamme de fréquences contenues dans une image, un banc de filtres passe-bande est appliqué sur l'image. Ensuite, les valeurs d'amplitudes instantanées (IA, pour *instantaneous amplitude*) et de fréquences instantanées (IF, pour *instantaneous frequency*) correspondant à la dérivée spatiale de la phase instantanée sont calculées pour chaque pixel. Une des méthodes de calcul est d'appliquer l'algorithme de séparation discrète d'énergie (Maragos & Bovik, 1995). D'autres méthodes se basent sur l'approximation de fonction quasi-propre ou par des méthodes quasi-locales (Murray, Rodriguez, & Pattichis, 2010). Les

valeurs d'IA et d'IF donnent des informations sur le contraste local et permettent de caractériser des structures à différentes fréquences.

Le patron binaire local (LBP, pour *local binary pattern*) est l'une des caractéristiques les plus puissantes en classification de texture. La méthode consiste à détecter des primitives locales de textures, appelées patrons, qui sont alors encodées en binaire, et à mesurer leur occurrence dans l'image. Les LBP sont calculés pour chaque pixel de l'image d'intensité de gris g_c sur un voisinage de rayon R , constitué de P voisins d'intensité de gris g_p . Les coordonnées des P voisins sont données par $\left(-R \sin\left(\frac{2\pi p}{P}\right), R \cos\left(\frac{2\pi p}{P}\right) \right)$ autour d'un pixel central. Si les coordonnées des voisins ne tombent pas au milieu d'un pixel, les intensités de gris g_p sont alors estimées par interpolation. La texture est alors considérée comme étant un ensemble de différences d'intensité de gris des voisins par rapport au pixel central, et non plus comme la valeur des intensités. Le vecteur de différences locales caractérise la structure locale de l'image. Comme la reconnaissance de texture ne peut se faire directement à partir des valeurs de différence $g_p - g_c$ à cause de leur grande sensibilité au bruit, à la translation et la rotation, on utilise plutôt le signe de cette différence pour caractériser la texture (Ojala, Pietikainen, & Maenpaa, 2002). La composante de signe du LBP est alors un opérateur invariant en échelle de gris et est définie par :

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \quad (2-3)$$

$$\text{où } s(x) = \begin{cases} 1, & \text{si } x \geq 0 \\ 0, & \text{sinon} \end{cases}$$

Plus le nombre de voisins P est grand, plus les combinaisons pour former des patrons augmentent. Un histogramme mesurant le nombre d'occurrences des différents types de LBP est alors construit pour caractériser la texture de l'image de taille $I \times J$:

$$H(k) = \sum_{i=1}^I \sum_{j=1}^J f(LBP_{P,R}(i,j), k), \quad k \in [0, K] \quad (2-4)$$

$$\text{Avec } K \text{ la valeur maximale des patrons et } f(x, y) = \begin{cases} 1, & x = y \\ 0, & \text{sinon} \end{cases}$$

Afin de réduire le nombre de combinaisons de LBP, on utilise les patrons uniformes fondamentaux. Il s'agit d'un ensemble de patrons, qui ont une structure circulaire uniforme avec très peu de transitions spatiales. En binaire, les transitions spatiales sont des transitions de bit 0/1. Ce sont des patrons qui représentent des structures pertinentes comme des points, des coins, ou des courbures. L'uniformité U d'un patron se mesure via la formule suivante :

$$U(LBP_{P,R}) = |s(g_{P-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^P |s(g_p - g_c) - s(g_{p-1} - g_c)| \quad (2-5)$$

En fait, U correspond au nombre de transitions spatiales. Un patron, dont la valeur U ne dépasse pas 2, est alors considéré comme « uniforme ».

L'invariance en rotation peut être également introduite avec l'opérateur $LBP_{P,R}^{riu2}$ où « riu2 » signifie « uniforme et invariant en rotation » :

$$LBP_{P,R}^{riu2} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c), & \text{si } U(LBP_{P,R}) \leq 2 \\ P + 1, & \text{sinon} \end{cases} \quad (2-6)$$

En complémentarité, la composante de magnitude du LBP vient s'ajouter au signe afin de généraliser le concept de LBP. Celle-ci se calcule de la manière suivante :

$$LBP_{P,R} = \sum_{p=0}^{P-1} t(m_p, c) 2^p \quad (2-7)$$

$$\text{avec } t(x, c) = \begin{cases} 1, & x \geq c \\ 0, & \text{sinon} \end{cases}$$

Le paramètre c est un seuil déterminé selon l'application, mais souvent mis à la valeur moyenne d'intensité de l'image. Similairement au $LBP_{P,R}^{riu2}$, on peut construire un opérateur invariant en rotation $LBP_{P,R}^{riu2}$ pour la magnitude. La combinaison des histogrammes de signe du LBP et de la magnitude du LBP, qu'on appelle patron binaire local complet (Guo & Zhang, 2010), donne ainsi des meilleures classifications de texture, par rapport à l'information de signe seule. Un exemple de calcul de LBP complet pour un pixel central est donné par la Figure 2-5.

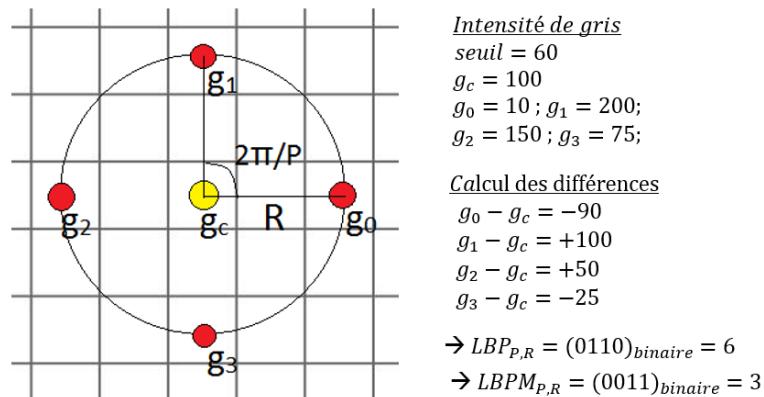


Figure 2-5: Exemple de calcul du signe et de l'amplitude du LBP pour un pixel central donné dans un voisinage à $P = 4$ et de rayon R .

2.1.1.3 Caractéristiques de contexte visuel

Les caractéristiques de contexte visuel sont des caractéristiques qui ont pour but de faire ressortir les objets par rapport au fond de l'image. Elles sont généralement utilisées pour la détection d'objet, mais également pour extraire des caractéristiques visuelles de l'image.

Les histogrammes de gradient orientés (HOG, pour *histogram of oriented gradients*) se basent sur le principe que l'apparence locale et la forme de l'objet peuvent être décrites par la distribution d'intensité des gradients ou de la direction des contours. Le gradient d'une image $\nabla I(x, y)$ est une grandeur vectorielle qui indique de quelle façon l'intensité du pixel varie dans l'espace. Le calcul du gradient se fait par convolution de l'image avec un masque de dérivée première. Les masques les plus simples, comme les masques 1-D de dérivée centrée en un point, semblent donner les meilleures performances et sont donc généralement utilisés pour calculer les HOG (Dalal & Triggs, 2005). Les gradients horizontal et vertical peuvent alors être calculés de la manière suivante :

$$G_x = [1 \quad 0 \quad -1] * I(x, y) \quad \text{et} \quad G_y = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} * I(x, y) \quad (2-8)$$

L'amplitude G et la direction θ du gradient sont alors obtenus à partir de G_x et G_y par :

$$G = \sqrt{G_x^2 + G_y^2} \quad (2-9)$$

$$\theta = \arctan\left(\frac{G_y}{G_x}\right) \quad (2-10)$$

Les descripteurs sont ensuite extraits par construction locale d'histogrammes de gradient (Figure 2-6). Pour ce faire, l'image est divisée en petites cellules connectées. Pour chaque cellule, un histogramme est construit par mesure de l'occurrence des pixels dans la cellule qui sont dans une certaine orientation, pondérée par l'amplitude du gradient. Un pixel dont l'amplitude de gradient est plus grande dans une direction donnée donne plus de poids à cette direction dans l'histogramme. Le nombre de cellules et le nombre d'intervalles pour discréteriser les directions sont des paramètres qui influencent la taille du vecteur de caractéristiques.

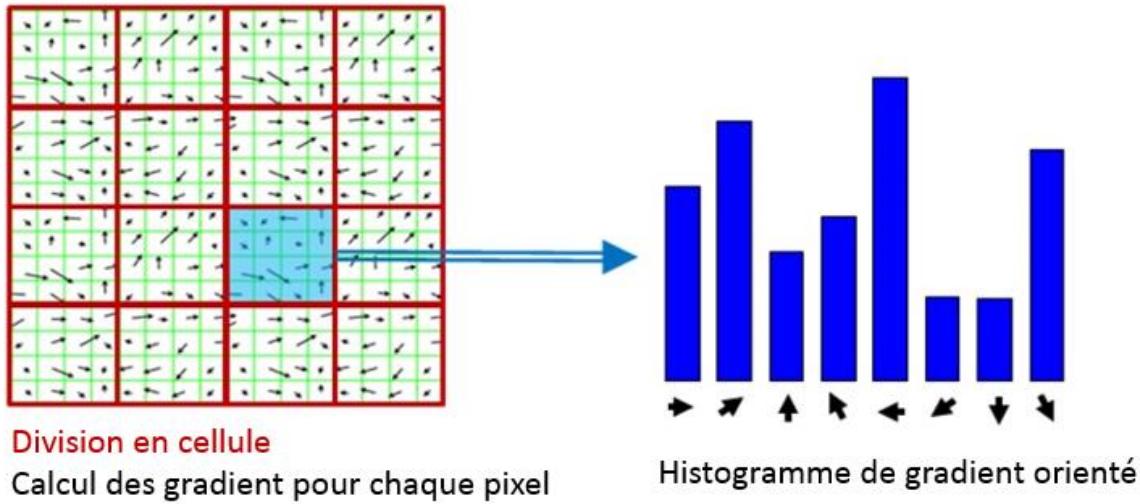


Figure 2-6: Illustration de la construction locale des histogrammes de gradient orienté (Levi, 2013). Tiré de <https://gilscvblog.wordpress.com/2013/08/18/a-short-introduction-to-descriptors/>

Afin d'éviter les disparités dues aux variations d'illuminations, on effectue une normalisation des descripteurs. Pour ce faire, plusieurs cellules sont regroupées en un bloc sur lequel l'opération de normalisation est appliquée. Chaque cellule contribue donc plusieurs fois dans le descripteur final. Le regroupement des cellules pour former les blocs s'effectue par une géométrie soit rectangulaire (R-HOG), soit circulaire (C-HOG). La méthode R-HOG est similaire à la méthode SIFT réalisée sur une seule échelle. Les opérations de normalisations sont :

$$L2 - norme : f = \frac{v}{\sqrt{\|v\|_2^2 + \varepsilon^2}} \quad (2-11)$$

$$L1 - norme : f = \frac{v}{\|v\|_1 + \varepsilon} \quad (2-12)$$

$$L1 - racine : f = \sqrt{\frac{v}{\|v\|_1 + \varepsilon}} \quad (2-13)$$

Avec v le vecteur non-normalisé des histogrammes contenus dans un seul bloc, $\|v\|_k$ la k-norme du vecteur, et ε une constante de faible valeur (Dalal & Triggs, 2005).

La Scale-Invariant feature transform (SIFT) est une méthode similaire à la méthode R-HOG, qui est ici appliquée à différentes échelles. Cette caractéristique a comme propriétés d'être invariante au changement d'échelle, en plus d'être invariante aux changements d'illumination et

invariante en translation et rotation et pour les autres transformations affines (Lowe, 1999). Le calcul de cette caractéristique se fait d'abord par la construction d'un espace à différentes échelles. Pour une image $I(x, y)$, cette étape est réalisée avec des fonctions gaussiennes :

$$G(x, y, \sigma) = (g_\sigma * I)(x, y) \quad (2-14)$$

Avec (x, y) les coordonnées spatiales, σ le facteur d'échelle et

$$g_\sigma = \frac{1}{\sqrt{2\pi}\sigma} \exp{-(x^2 + y^2)/2\sigma^2} \quad (2-15)$$

L'espace des échelles gaussiennes (Figure 2-7) est alors décomposé en octave et en sous-niveaux. Un changement d'octave implique une taille d'image réduite de moitié et les sous-niveaux correspondent à des variations de la valeur de l'écart-type. Le facteur d'échelle peut alors être défini selon l'indice des octaves o et l'indice de sous-niveaux s :

$$\sigma(o, s) = \sigma_0 2^{(o+s/s)} \quad (2-16)$$

Avec σ_0 le niveau d'échelle de base. Par exemple, on peut fixer sa valeur à $\sigma_0 = 1,6 \cdot 2^{1/S}$, avec S , le nombre maximum de sous-niveaux (Lowe, 1999).

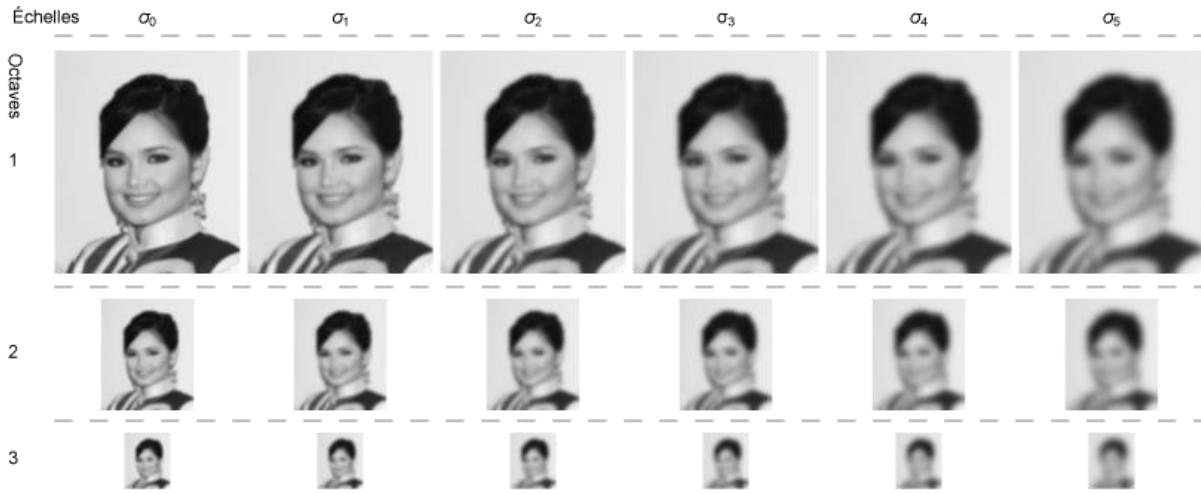


Figure 2-7: Espace des échelles gaussiennes pour 3 octaves et 5 sous-niveaux (Indif,2010). Tiré de http://commons.wikimedia.org/wiki/File:Pyramid_of_gradients.png

Ensuite, des points-clés sont identifiés et localisés par la recherche de maxima ou de minima dans l'espace des échelles (Figure 2-8 et Figure 2-9) par l'utilisation de fonctions de différence de Gaussienne (DoG) :

$$D(x, y, \sigma(o, s)) = G(x, y, \sigma(o, s + 1)) - G(x, y, \sigma(o, s)) \quad (2-17)$$

La construction d'HOG dans une région autour de ces point-clés permet de leur assigner une orientation, qui est alors intégrée au vecteur de caractéristiques des points-clés. Afin de rendre l'algorithme invariant en rotation, on normalise les vecteurs de caractéristiques. Cette caractéristique est souvent utilisée pour la mise en correspondance de points entre deux images (Vedaldi, 2006).

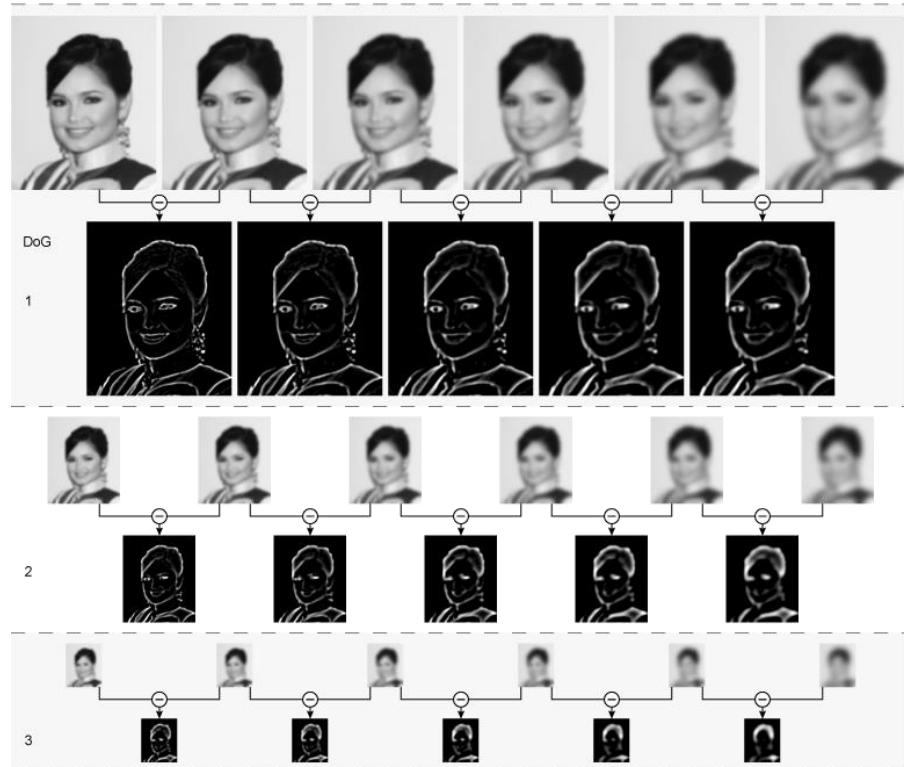


Figure 2-8: Espace des échelles de différence de Gaussienne pour 3 octaves et 5 sous-niveaux (Indif, 2010). Tiré de <http://commons.wikimedia.org/wiki/File:Differences-of-Grandients.png>

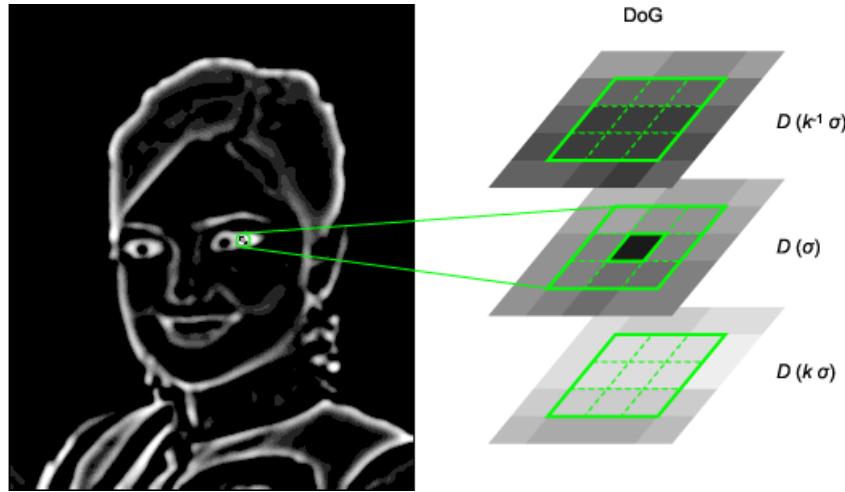


Figure 2-9: Recherche des extrema avec les DoG (Indif, 2010). Tiré de http://fr.wikipedia.org/wiki/Scale-invariant_feature_transform#/media/File:D%C3%A9tection-d%C2%A7ext%C3%A9rums-dans-l%C2%A7espace-des-%C3%A9chelles.png

Les Speeded Up Robust Features (SURF) sont des caractéristiques qui sont aussi souvent utilisés pour la mise en correspondance de points et se basent également sur le principe d'espace d'échelles. Le calcul des SURF (Bay, Ess, Tuytelaars, & Van Gool, 2008) se fait différemment et plus rapidement que celui des SIFT. Afin d'accélérer les calculs, des filtres de forme rectangulaire sont utilisés pour approximer les fonctions gaussiennes. La convolution avec ce type de filtres s'effectue plus aisément sur l'intégrale de l'image $I_\Sigma(x, y)$ qui est donnée par :

$$I_\Sigma(x, y) = \sum_{i=0}^x \sum_{j=0}^y I(i, j) \quad (2-18)$$

Une fois $I_\Sigma(x, y)$ calculée, le calcul des intensités dans une région rectangulaire s'effectue par trois opérations d'addition. La construction de l'espace des échelles se fait alors par convolution de l'image intégrale avec des filtres Gaussiens rectangulaires de différentes tailles. Les images pour chaque octave et chaque sous-niveau sont alors indépendantes. Les points d'intérêt sont alors trouvés par le calcul de la matrice Hessienne, qui consiste à calculer le Laplacien de Gaussienne:

$$H(x, y, \sigma) = \begin{bmatrix} \frac{\partial^2 g(\sigma)}{\partial x^2} & \frac{\partial^2 g(\sigma)}{\partial x \partial y} \\ \frac{\partial^2 g(\sigma)}{\partial x \partial y} & \frac{\partial^2 g(\sigma)}{\partial y^2} \end{bmatrix} \quad (2-19)$$

Avec (x, y) les coordonnées spatiales, σ le facteur d'échelle et $g(\sigma)$ la convolution de l'image avec un filtre gaussien. Les points-clés sont alors les points pour lesquels le déterminant de la matrice H est positif et donc qui correspondent à des extrema locaux. La position des points est affinée à l'aide d'une interpolation (Brown & Lowe, 2002). Pour extraire les caractéristiques SURF, on analyse la réponse d'un ensemble de points autour des points-clés aux ondelettes de Haar verticale et horizontale. L'orientation relative à la réponse la plus importante est alors intégrée au vecteur de caractéristiques des points.

2.1.2 Sélection de caractéristiques

Après le calcul des caractéristiques, il n'est pas rare d'obtenir un vecteur de caractéristiques dont la taille est beaucoup plus grande que le nombre d'échantillons d'apprentissage. C'est souvent le cas dans les problématiques biomédicales où l'accès à des données est souvent limité. La bonne pratique veut que le nombre d'échantillons soit au moins dix fois plus grande que le nombre de caractéristiques, sinon on tombe dans des problèmes de malédiction de la dimensionnalité ou le sur-apprentissage. Bien que cette étape ne soit pas toujours nécessaire, une sélection de caractéristiques permet de réduire la dimensionnalité du système pour la classification. Il existe différents algorithmes de sélection de caractéristiques qui se différencient par leur complexité et leur optimalité à trouver le meilleur sous-ensemble de caractéristiques. Cette section présente les différentes méthodes de sélection de caractéristiques.

2.1.2.1 Sélection selon le mérite individuel

La sélection par filtrage selon le mérite individuel utilise un critère d'évaluation qui se calcule sur chaque caractéristique indépendamment les unes des autres. Ce critère évalue la pertinence d'une caractéristique selon les propriétés mesurées sur les données d'apprentissage. Le but est alors de calculer un score et de prendre les caractéristiques qui ont montré le meilleur score.

Le critère de Fisher (Duda, Hart, & Stork, Non-parametric techniques, 2000) évalue la divergence entre deux classes i et j en se basant sur la moyenne μ et l'écart-type σ estimés de la caractéristique F pour ces deux classes :

$$D(F) = \frac{(\mu_i - \mu_j)^2}{\sigma_i^2 + \sigma_j^2} \quad (2-20)$$

Ce critère doit être maximisé, puisqu'on veut que la séparation entre les classes soit la plus grande possible. On veut donc le moins de chevauchement possible entre les deux classes lorsque les échantillons sont projetés sur l'axe de la caractéristique pour laquelle on calcule le critère de Fisher (Figure 2-10). Dans un cas multi-classe, le critère devient la divergence moyenne globale des critères de Fisher pour des classes deux à deux. Une autre méthode (Gu, Li, & Han, 2012) consiste à évaluer la divergence de la moyenne de la classe par rapport à la moyenne des échantillons μ^F pour la caractéristique F :

$$D(F) = \frac{\sum_{c=1}^C n_c (\mu_c^F - \mu^F)^2}{\sum_{c=1}^C n_c (\sigma_c^F)^2} \quad (2-21)$$

Avec n_c , μ_c^F et σ_c^F qui sont respectivement le nombre, la moyenne et l'écart-type des échantillons pour la classe c et C le nombre de classe.

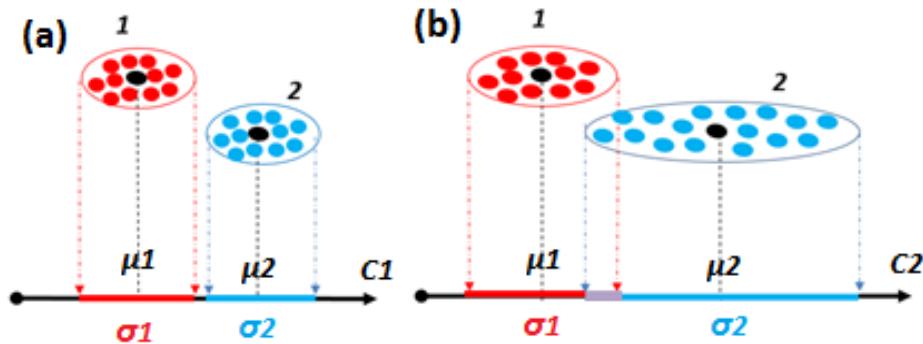


Figure 2-10: Projection des échantillons sur l'axe des caractéristiques C_1 et C_2 pour calculer le critère de Fisher : (a) sans chevauchement et (b) avec chevauchement.

Le critère de corrélation (Guyon & Elisseeff, 2003) se calcule généralement entre deux classes dont les valeurs d'étiquettes sont $y_c \in [-1, 1]$. Dans le contexte de sélection de caractéristiques, on vérifie la corrélation entre les valeurs de caractéristiques et les étiquettes des échantillons. Soit X_F , le vecteur des valeurs de la caractéristique F pour tous les échantillons de la base de données et Y , le vecteur des valeurs d'étiquette correspondant, la corrélation $C(F)$ est définie par :

$$C(F) = \frac{COV(X_F, Y)}{\sqrt{VAR(X_F) VAR(Y)}} \quad (2-22)$$

Avec COV correspondant à la covariance et VAR à la variance. Ce critère peut être estimé de la manière suivante :

$$C(F) = \frac{\sum_{c=1}^C (x_c^F - \bar{x}^F)(y_c^F - \bar{y}^F)}{\sqrt{\sum_{c=1}^C (x_c^F - \bar{x}^F)^2} \sqrt{\sum_{c=1}^C (y_c^F - \bar{y}^F)^2}} \quad (2-23)$$

Avec x_c^F et \bar{x}^F les valeurs de la caractéristique F et la moyenne de ces valeurs pour les éléments de la classe c , y_c^F et \bar{y}^F les valeurs d'étiquettes des échantillons et la moyenne des étiquettes. Les valeurs de corrélations sont comprises entre -1 et 1. Plus les valeurs se rapprochent de 1, plus la corrélation est grande. Ce critère est donc aussi à maximiser.

L'information mutuelle (Guyon & Elisseeff, 2003) mesure la dépendance entre deux variables aléatoires, similairement à la corrélation. Dans le contexte de sélection de caractéristiques, ces variables sont les valeurs de caractéristiques X et les valeurs d'étiquettes des classes Y . L'information mutuelle $I(F)$ d'une caractéristique F représente le degré de dépendance au sens probabiliste et se calcule comme le rapport de l'entropie jointe des deux variables sur l'entropie des deux variables séparées :

$$I(F) = \sum_{x_F} \sum_y P(X = x_F, Y = y) \log \left(\frac{P(X = x_F, Y = y)}{P(X = x_F) P(Y = y)} \right) \quad (2-24)$$

Les deux variables sont dites indépendantes si la réalisation de l'une n'apporte aucune information sur la réalisation de l'autre, et la valeur de l'information mutuelle est alors $I(F) = 0$. Évidemment, les meilleurs caractéristiques sont celles qui présentent la plus grande dépendance par rapport à l'étiquetage des échantillons. La corrélation est un cas particulier de l'information mutuelle dans lequel la relation de dépendance entre les deux variables est strictement linéaire.

L'avantage des méthodes selon le mérite individuel est leur efficacité calculatoire et leur robustesse face au sur-apprentissage. Leur limite est que le score de chaque caractéristique est calculé indépendamment les unes des autres. Ces caractéristiques sont donc triées selon leur score de manière individuelle, sans considérer leurs interactions avec les autres caractéristiques. Or, une caractéristique peu performante seule peut présenter une bonne performance lorsqu'elle est en combinaison avec d'autres. Ces méthodes tendent donc à sélectionner des caractéristiques redondantes plutôt que complémentaires (Guyon & Elisseeff, 2003; Gu, Li, & Han, 2012).

2.1.2.2 Sélection séquentielle

Les méthodes de sélection séquentielle sont des méthodes itératives de sélection dont le but est la recherche d'un ensemble sous-optimal de caractéristiques. Elles consistent à évaluer les performances d'un modèle pour un ensemble de caractéristiques dont la taille augmente (sélection croissante) ou diminue (sélection décroissante). Les méthodes les plus simples sont la sélection séquentielle croissante (SFS, pour *sequential forward selection*) et la sélection séquentielle arrière (SBS, pour *sequential backward selection*). Des méthodes plus élaborées ont été développées afin de pallier aux problèmes des méthodes précédentes : la *sequential floating forward selection* (SFFS) et la *sequential floating backward selection* (SFBS).

La SFS commence d'abord avec un modèle vide de sous-ensemble de caractéristiques sélectionnées. À chaque itération, la meilleure caractéristique selon un certain critère de performance est choisie parmi l'ensemble des caractéristiques à sélectionner, est supprimée de cet ensemble et est ajoutée au modèle. Ce processus est ainsi réitéré jusqu'au critère d'arrêt, qui correspond généralement soit à la performance qui n'augmente plus ou à l'atteinte du maximal d'itération. La méthode est simple et rapide à appliquer, cependant elle ne prend en compte que l'interaction d'une caractéristique avec le sous-ensemble déjà sélectionné et non avec un ensemble plus large (Marill & Green, 1963).

La SBS se base sur le même principe, à la différence que la méthode commence avec tout l'ensemble de caractéristiques dans le modèle. À chaque itération, on retire du modèle la caractéristique qui donne la moins bonne performance. Cette méthode permet de trouver un meilleur sous-ensemble de caractéristiques, car elle prend en considération les interactions de chaque caractéristique avec un ensemble plus large. Cependant, la méthode SBS pose des difficultés en termes de performance calculatoire et est très difficile à appliquer pour des systèmes à très grande dimension (Whitney, 1971).

Les avantages de SFS et SBS sont donc leur simplicité d'application. La sélection peut également être adaptée au classifieur utilisé, selon le critère de performance choisi. Il pourrait, par exemple, correspondre au taux d'erreur de classification sur des données de validation à minimiser. La méthode de sélection est alors dite de type *wrapper*. Le défaut des méthodes SFS et SBS est que le parcours total de l'espace de recherche n'est pas réalisé. Ces méthodes ne garantissent donc pas de produire le sous-ensemble optimal de caractéristiques car toutes les possibilités de sous-ensembles

de caractéristiques ne sont pas examinées. C'est pour cela que les méthodes de sélection séquentielle flottante ont été proposées afin de palier à ces inconvénients (Chouaib, 2011).

La SFFS et la SFBS consistent, pour chaque itération, à appliquer l fois l'algorithme de SFS de manière à ajouter les l meilleures caractéristiques, et r fois l'algorithme de SBS de manière à supprimer les r plus mauvaises caractéristiques du modèle (Figure 2-11). Ceci est répété jusqu'au critère d'arrêt. La SFFS commence avec un ensemble vide pour le modèle de sous-ensemble de caractéristiques et la SFBS commence avec tout l'ensemble de caractéristiques dans le modèle. La dimension du sous-ensemble optimal dépend très fortement des paramètres l et r , qui ne sont à priori pas connus. La solution proposée est donc de laisser ces paramètres flottants au cours du processus afin de se rapprocher au maximum de la solution optimale (Pudil, Novovicova, & Kittler, 1994). Dans la plupart des problèmes de sélection à grande dimension, ces méthodes réalisent de meilleurs résultats que SFS et SBS. Généralement, les résultats obtenus se rapprochent de la solution optimale, tout en ayant un temps de calcul pas trop coûteux (Jain, Duin, & Mao, 2000).

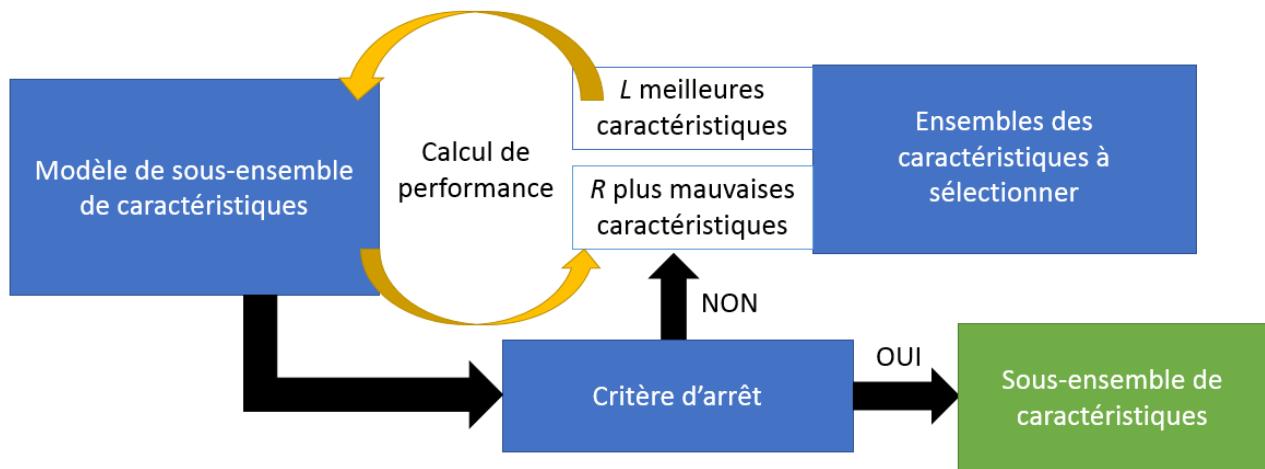


Figure 2-11: Schéma résumant les méthodes SFFS et SFBS

2.1.2.3 Sélection par séparation et évaluation

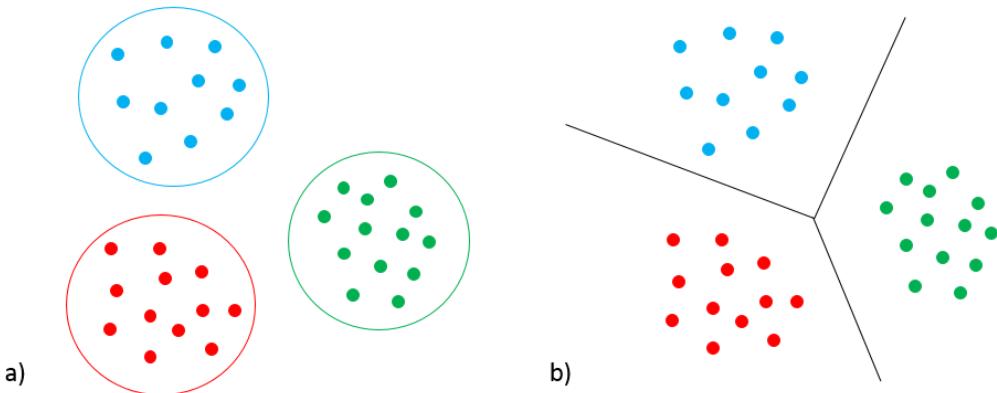
La sélection par séparation et évaluation (Somol, Pudil, & Kitler, 2004) est une méthode optimale de recherche de caractéristiques par graphes. Elle consiste à énumérer un ensemble possible de solution, de telle manière que l'énumération des mauvaises solutions est évitée, sur base de

l'analyse des propriétés du problème donné. L'algorithme comporte deux étapes : la séparation et l'évaluation. Dans la phase de séparation, le problème est divisé en un ensemble de sous-problèmes. Pour ce faire, un arbre de recherche est construit. La racine représente l'ensemble des caractéristiques et les nœuds représentent des sous-ensembles de caractéristiques. Le parcours dans l'arbre se fait alors de la racine aux feuilles. Dans la phase d'évaluation, une fonction est utilisée pour mettre une borne sur certaines solutions et décider de, soit les exclure comme mauvaises solutions, soit de les garder comme solutions potentielles. La performance dépend de la qualité de la fonction d'évaluation choisie. Lorsque des nœuds ne répondent pas au critère d'évaluation, les sous-arbres de ce nœud sont alors éliminés, évitant des recherches inutiles dans cette partie de l'arbre. Cette méthode garantit de trouver le sous-ensemble optimal qui satisfait la fonction d'évaluation répondant à la propriété de monotonie. Si le critère n'est pas monotone, des problèmes dans le parcours de l'arbre, appelés « nidification », peuvent alors subvenir. Un autre inconvénient de la méthode est sa complexité de calcul exponentielle qui devient rapidement grande lorsque le nombre de caractéristiques est grand (Jain, Duin, & Mao, 2000).

2.1.3 Modélisation de classifieur

La modélisation d'un classifieur consiste à établir un système capable de classer les échantillons dans la bonne catégorie. Dans un cadre supervisé, il se construit généralement en deux phases : la phase d'apprentissage ou d'entraînement et la phase de test. La phase d'apprentissage établit les fonctions de décision du système permettant de classifier les échantillons dans des groupes qui ont des caractéristiques similaires, mesurées sur des exemples d'entraînement. La phase de test vérifie les performances du système en lui introduisant de nouveaux exemples. La difficulté se porte sur le choix du classifieur qui dépend fortement de l'application et de la complexité des données. On peut catégoriser les classifieurs en deux grandes familles (Figure 2-12): les classifieurs à modèle génératif où on cherche à créer des ensembles et les classifieurs à modèle discriminant où on cherche des frontières de séparations.

Parmi les classifieurs de la première famille, l'approche la plus simple et la plus intuitive se base sur le concept de similarité. Des exemples de classifieurs se basant sur ce principe sont le *template matching* (Brunelli, 2009), le classifieur du plus proche de la moyenne (Tibshirani, Hastie, Narasimhan, & Chu, 2002) ou le classifieur du plus proche voisin (Cover & Hart, 1967).



D'autres classificateurs de cette famille se basent sur l'approche probabiliste. Lorsque la loi de la distribution des données est connue, les classificateurs bayésiens donnent des solutions optimales. Des modèles paramétriques sont alors utilisés, comme par exemple une distribution normale pour des caractéristiques continues, ou une distribution binomiale pour des caractéristiques binaires. Une fois le modèle paramétrique déterminé, les échantillons sont classés dans la catégorie présentant la probabilité à posteriori maximale. Les classificateurs bayésiens sont cependant très sensibles aux erreurs d'estimation de densité. D'autres classificateurs de ce type, comme les classificateurs de régression logistique, se basent sur le maximum de vraisemblance (Duda, Hart, & Stork, Non-parametric techniques, 2000).

Les méthodes non-paramétriques englobent les classificateurs des k-plus proches voisins et les classificateurs Parzen. Le premier type de classificateur se base sur l'étiquette des k plus proches voisins pour déterminer celui du nouvel échantillon. Pour les classificateurs Parzen, la densité des classes est remplacée par leur estimé en utilisant des fenêtres de Parzen. Les deux méthodes ont un paramètre libre, le nombre de voisins k ou le paramètre de lissage des noyaux Parzen, qui peut être optimisé par une estimation *leave-one-out* du taux d'erreurs (Altman, 1992).

La deuxième famille de classificateurs correspond aux classificateurs qui se basent sur la construction de frontières de décision. Le but est alors de trouver la frontière qui minimise un certain critère, comme par exemple, l'erreur de classification apparente ou l'erreur des moindres carrés entre la

sortie et les valeurs cibles. Des exemples de ce type de classificateurs sont le discriminant linéaire de Fisher (Scholkopft & Mullert, 1999), le perceptron à une couche ou multicouche (Gallant, 1990).

Dans la suite de ce chapitre, les trois classificateurs étudiés dans le cadre de ce projet sont présentés plus en détail : l'analyse linéaire discriminante, les machines à vecteurs de support et les forêts d'arbres décisionnels.

2.1.3.1 Analyse linéaire discriminante (LDA)

L'analyse linéaire discriminante fait partie des classificateurs à modèle discriminant et a pour but d'établir des fonctions linéaires discriminantes, obtenues par combinaison linéaire de caractéristiques mesurées sur les échantillons. La projection des données sur ces fonctions dispose les échantillons de manière à séparer au mieux les classes (Figure 2-13). Autrement dit, cela consiste à chercher l'espace vectoriel de faible dimension qui maximise la variance interclasse et qui minimise la variance intraclasse (Lebart, Morineau, & Piron, 1997).

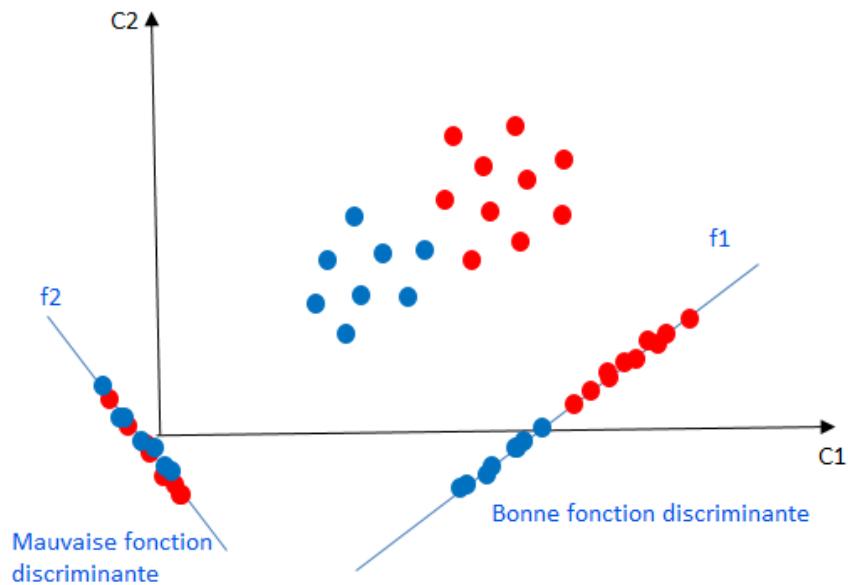


Figure 2-13: LDA pour deux classes dans l'espace de caractéristiques C1-C2 avec f_1 un exemple de bonne fonction discriminante car les deux classes sont bien séparées et f_2 un exemple de mauvaise fonction discriminante car les deux classes sont superposées.

Dans un cas à deux classes, une fonction discriminante qui est combinaison linéaire des éléments de x est défini par :

$$H(x) = w^t x + w_0 \quad (2-25)$$

Avec w le poids vecteur et w_0 le biais ou le poids seuil. Cette fonction est un hyperplan qui correspond à la frontière de décision entre les deux classes. Le classifieur détermine alors la classe selon le signe $H(x)$. Si $H(x) > 0$, le système décide d'attribuer la classe positive, sinon c'est la classe négative.

2.1.3.2 Machine à vecteurs de support (SVM)

Dans les méthodes de classification linéaire discriminante, le but est de trouver l'hyperplan qui arrive à séparer le mieux les deux classes. Cependant, ces méthodes ne proposent qu'une solution parmi d'autres et ne garantissent pas que la solution soit optimale. La machine à vecteurs de support (SVM) a pour but de rechercher la frontière qui maximise la marge entre deux classes. La marge correspond à l'espace entre les éléments les plus proches de l'hyperplan de séparation H . L'hyperplan présentant la marge la plus grande (Figure 2-14) est alors considéré comme la frontière de décision, car il permet une meilleure généralisation que les autres solutions (Steinwart & Christmann, 2008). Dans un cas linéairement séparable, on peut établir une frontière de séparation linéaire représentée par un hyperplan qui est défini par l'équation (2-25). Les vecteurs de support sont les éléments les plus proches de l'hyperplan et sont considérés comme étant sur la marge.

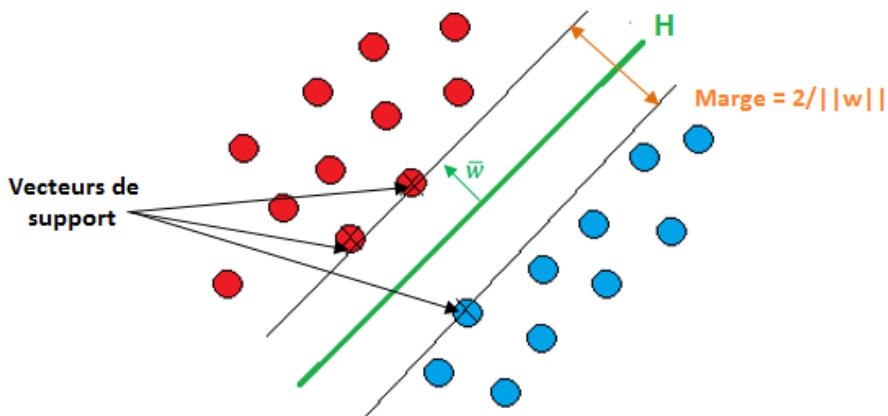


Figure 2-14: Schéma de l'hyperplan d'une machine vecteur de support séparant deux classes

Les données sont normalisées de telle manière qu'aux vecteurs de support x_1 de la classe positive (étiquette $y = 1$) et x_2 de la classe négative (étiquette $y = -1$), on a :

$$H(x_1) = w \cdot x_1 + w_0 = 1 \quad (2-26)$$

$$H(x_2) = w \cdot x_2 + w_0 = -1 \quad (2-27)$$

Les éléments x ayant la valeur de $H(x)$ supérieure ou égale à 1 appartiennent à la classe positive et les éléments x ayant la valeur de $H(x)$ inférieure ou égale à -1 appartiennent à la classe négative.

Afin d'être la fonction de séparation optimale, l'hyperplan doit respecter les deux contraintes données par les équations (2-26) et (2-27). Ces deux contraintes peuvent être réduites en une seule, et l'hyperplan doit donc satisfaire la condition suivante :

$$y_i(w \cdot x_i + w_0) \geq 1, \quad \forall x_i \quad (2-28)$$

Avec $y_i \in \{-1, 1\}$ les étiquettes des éléments x_i .

Une autre condition que l'hyperplan doit respecter est la maximisation de la marge, c'est-à-dire la distance entre les vecteurs de support. La distance d'un point x à l'hyperplan H est définie par :

$$d(x) = \frac{|w \cdot x + w_0|}{\|w\|} = \frac{|H(x)|}{\|w\|} \quad (2-29)$$

Comme $|H(x)| = 1$ aux vecteurs de support, la marge M a pour longueur :

$$M = d(x_1) + d(x_2) = 2/\|w\| \quad (2-30)$$

Où x_1 et x_2 sont les vecteurs de support de part et d'autres de H .

Le fait de maximiser la marge revient donc également à minimiser $\|w\|^2/2$ ou $\|w\|^2/2$. L'erreur de classification pour chaque donnée ξ_i doit être également minimisée. Une certaine tolérance à l'erreur peut être cependant acceptée via le paramètre C , qui correspond au nombre d'échantillons qu'on accepte à être sur la marge. Le critère d'optimisation devient donc :

$$\min_{w, w_0} \frac{1}{2} w^T w + C \sum_i \xi_i \quad (2-31)$$

$$tel que - (y_i(w \cdot x_i + w_0) - 1) \leq 0, \quad \forall x_i \quad (2-32)$$

Ce système consiste en un problème d'optimisation avec contraintes. La résolution du système se fait par une transformation duale. Le système est alors représenté par son Lagrangien :

$$L(w, w_0, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^l \alpha_i (y_i (w x_i + w_0) - 1) \quad (2-33)$$

Avec l le nombre d'échantillons x_i d'apprentissage dont l'étiquette est y_i , α_i les paramètres de Lagrange qui sont non-nuls pour tous les éléments qui sont vecteurs de support, c'est-à-dire ceux qui sont sur la marge. Le Lagrangien doit ensuite être maximisé pour trouver les valeurs α_i :

$$\max_{\alpha} \sum_{i=1}^l \alpha_i (y_i (w x_i + w_0) - 1) \quad (2-34)$$

$$tel que \sum_{i=1}^l \alpha_i y_i = 0 \ et \ 0 \leq \alpha_i \leq C \quad (2-35)$$

Le vecteur w optimal est alors donné par :

$$w = \sum_{i=1}^l \alpha_i y_i x_i \quad (2-36)$$

Et la fonction de séparation devient :

$$H(x) = \sum_{i=1}^l \alpha_i y_i x_i x + w_0 \quad (2-37)$$

L'étiquette y de l'élément x qu'on veut catégoriser est alors donnée par le signe de la fonction de séparation au point x :

$$y(x) = sign(\sum_{i=1}^l \alpha_i y_i x_i x + w_0) \quad (2-38)$$

Pour les problèmes qui ne sont pas linéairement séparables, un noyau k est alors utilisé. Celui-ci applique une transformation ϕ sur les données qui va les plonger dans un espace de dimension plus grande où les deux classes seraient possiblement linéairement séparables. Le problème à résoudre est le même, avec en plus l'introduction du noyau. L'équation de la frontière de décision devient donc :

$$H(x) = \sum_{i=1}^l \alpha_i y_i k(x_i, x) + w_0, \ avec \ k(x_i, x) = \phi(x_i) \cdot \phi(x) \quad (2-39)$$

L'étiquette est alors donnée par :

$$y(x) = sign(\sum_{i=1}^l \alpha_i y_i k(x_i, x) + w_0) \quad (2-40)$$

Il existe plusieurs noyaux dont deux exemples très utilisés sont présentés ici :

$$\text{Polynomial de degré } d : k(x, y) = (a x y + b)^d \quad (2-41)$$

$$\text{Noyau gaussien (rbf)} : k(x, y) = \exp(-\gamma ||x - y||^2) \quad (2-42)$$

Pour les problèmes à plusieurs classes, différentes approches peuvent être utilisées comme les méthodes *one-versus-all* ou *one-versus-one* (Hsu & Lin, A comparison of methods for multiclass support vector machines, 2002). Nous ne présentons que la méthode *one-versus-all* qui est la méthode utilisée pour les problèmes à plusieurs classes dans le présent projet. La méthode consiste à générer autant de classifieurs SVM qu'il y a de classes, lors de la phase d'apprentissage. Chaque classifieur SVM correspond à une classification binaire dont la frontière de décision sépare une classe individuellement des autres (Figure 2-15). Pour un problème à k classes, la classe de l'échantillon x correspondra à celle qui maximise $H_i(x)$:

$$y(x) = \arg \max_{i=1, \dots, k} H_i(x) \quad (2-43)$$

Avec $H_i(x)$ correspondant à l'équation de la frontière de décision entre la classe i et les autres classes (Équations (2-37) ou (2-39)).

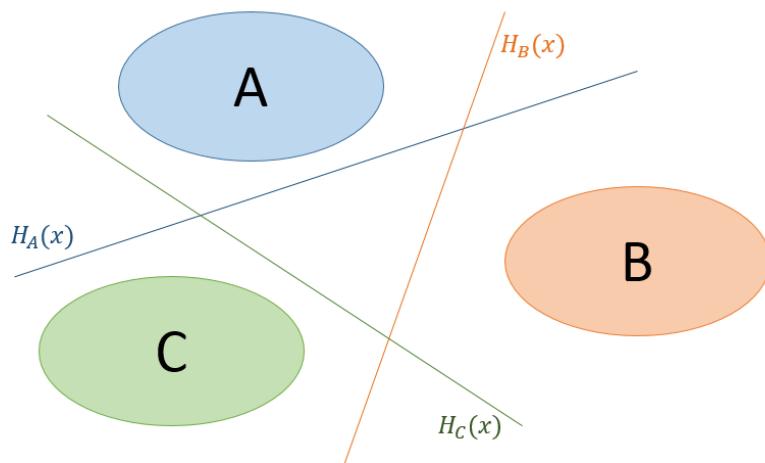


Figure 2-15: Diagramme des frontières binaires pour la méthode one-versus-all pour un problème à trois classes.

Les classifieurs SVM présentent comme avantage de ne pas être dépendant de la densité des classes, car seuls les éléments à la frontière contribuent à la modélisation du classifieur. Le choix des paramètres du modèle doit être déterminé selon la complexité des données, et le pouvoir de généralisation. Par exemple, une valeur élevée de paramètre C entraîne moins d'erreurs sur les données d'apprentissage, mais le risque de surapprentissage devient alors plus grand et le modèle classifiera moins bien de nouveaux échantillons. Le choix du noyau est également important pour avoir une bonne discrimination entre les classes. Celui qui est généralement utilisé est le noyau gaussien. Les paramètres du modèles SVM sont fixés à l'aide d'algorithmes d'optimisation, comme la recherche de grille (Hsu, Chang, & Lin, *A practical guide to support vector classification*, 2003).

2.1.3.3 Forêt d'arbres décisionnels (RF)

La forêt d'arbres décisionnels est une méthode se basant sur la génération de différents modèles et sur l'agrégation de leurs résultats pour déterminer la classe d'un élément. Les concepts exploités sont la construction de sous-espace aléatoire et d'arbres décisionnels (Breiman, *Random forests*, 2001). Le but est donc de construire des arbres de décisions (Figure 2-16) capables de déterminer l'étiquette y en fonction des variables $\bar{x} = x_1, \dots, x_n$ (où n est le nombre de caractéristiques représentant l'objet). Chaque arbre est construit en utilisant un échantillon *bootstrap* différent de la base de données. La construction du modèle est donc différente d'un arbre à un autre. L'arbre s'agrandit par la division de nœuds qui correspondent à une instance de décision. Dans les méthodes standards de constructions d'arbres décisionnels comme le *bagging* (Breiman, *Bagging predictors*, 1996), chaque nœud est divisé en prenant la meilleure séparation parmi les n variables. Dans la RF, chaque nœud est divisé en prenant le meilleur prédicteur parmi un sous-ensemble de m variables choisies de manière aléatoire. Ce processus est appelé la randomisation des variables.

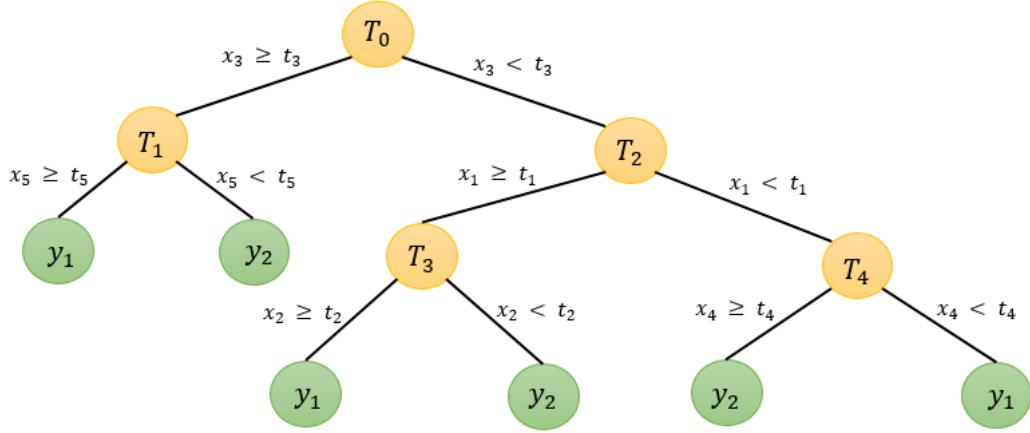


Figure 2-16: Schéma d'un arbre de décisions avec T_f les nœuds de l'arbre, x_i les prédicteurs, t_i le seuil sur le prédicteur et y_k l'étiquette des classes.

Le nombre de variables m dans un sous-ensemble est un des paramètres de la RF et un choix optimal est de fixer sa valeur à $m = \sqrt{n}$. Le meilleur prédicteur est déterminé selon un critère d'évaluation calculé sur les échantillons d'apprentissage non retenus pour construire l'arbre de décision (évaluation *out-of-bag*) et un seuil sur la variable est alors déterminé pour décider à quelle classe l'objet appartient (Liaw & Wiener, 2002). Les feuilles de l'arbre correspondent à la décision finale sur la classe de l'objet et la classification est d'autant meilleure que ses feuilles sont homogènes. Afin de mesurer cela et ainsi savoir quand arrêter la division d'un nœud, différents critères peuvent être calculés. Soient, k le nombre de classes, p_1, \dots, p_k les fréquences relatives de chaque classe à l'instance de décision T_f et k^* la classe la plus fréquente, les mesures d'hétérogénéité fréquemment utilisées sont :

- Le taux d'erreurs de classification sur l'ensemble d'apprentissage :

$$e(T_f) = 1 - p_k^* \quad (2-44)$$

- L'entropie (Quinlan, 1986; Kohavi & Quinlan, 1999) est une mesure de la quantité d'incertitude dans le nœud T_f :

$$e(T_f) = -\sum_k p_k \log(p_k) \quad (2-45)$$

- Le critère de Gini (Breiman, Friedman, Stone, & Olshen, 1984) est le taux d'erreur sur l'ensemble d'apprentissage d'un algorithme randomisé qui retournerait la classe k avec une probabilité p_k . Il correspond à une approximation de l'entropie et est très utilisé comme critère d'évaluation en RF :

$$e(T_f) = \sum_k p_k (1 - p_k) \quad (2-46)$$

Dans les modèles RF, la sélection de caractéristiques est donc implicitement réalisée et leur importance peut alors être mesurée par un des critères cités ci-dessus. D'autres mesures peuvent également utilisées comme l'importance de caractéristiques basée sur la permutation, ou la décroissance moyenne de la précision *out-of-bag* (Breiman, Manual on setting up, using and understanding random forests v3.1., 2002).

Un deuxième paramètre de la RF est le nombre d'arbres décisionnels. La classe d'un élément est alors déterminée par vote des différents arbres décisionnels (Figure 2-17) et on attribue la classe qui a été choisie par la majorité. Plus la proportion d'arbres à choisir cette classe est grande, plus la confiance associée au résultat est grande.

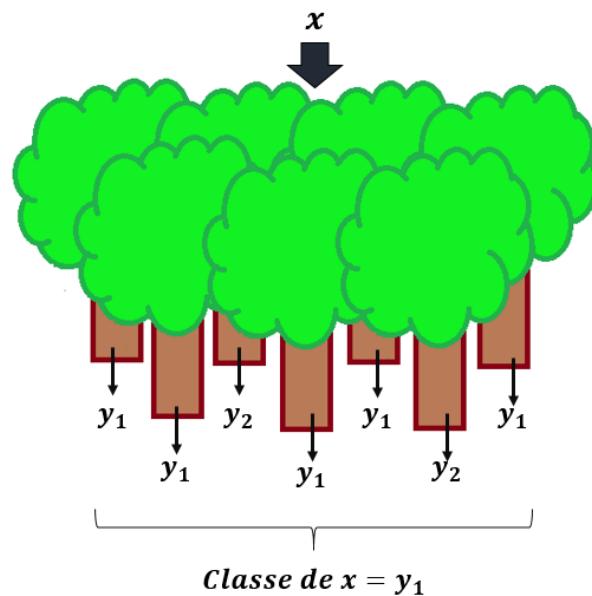


Figure 2-17: Vote des différents arbres décisionnels pour déterminer la classe y d'un élément x .

Les RF présentent comme avantage d'être facile à modéliser, puisque seuls les paramètres du nombre de caractéristiques sélectionnées à chaque nœud et le nombre d'arbre dans la forêt doivent être déterminés. Les caractéristiques sont implicitement sélectionnées et leur importance peut être également mesurée. Le défaut est que, contrairement aux SVM, la densité des classes a une certaine importance et dès lors, la classification pourrait être biaisée lorsqu'on se retrouve avec une classe moins représentée que d'autres. Cependant, pour des classifications binaires, ce problème pourrait être pallié en établissant un seuil sur la proportion d'arbres votant pour la classe la moins représentée.

2.2 Système automatique pour l'évaluation de la DMLA

Nous avons vu dans la section précédente les différentes étapes et les outils pour construire un système automatique pour la classification d'images. Nous allons maintenant présenter les travaux existant dans la littérature sur les systèmes automatiques qui sont spécifiques à l'analyse de la DMLA à partir d'images de fond d'œil.

2.2.1 Méthodes basées sur la segmentation des drusen

Dans un contexte de dépistage précoce de la DMLA, beaucoup de travaux se sont focalisés sur la segmentation de drusen. Le but est de bien les détecter et d'obtenir des informations quantitatives, comme leur taille, leur nombre et leur localisation. À partir de ces caractéristiques, l'évaluation de la DMLA se base alors sur la classification des drusen qui ont été préalablement segmentés. Différents algorithmes de segmentation ont alors été proposés, se basant sur la texture, les contours, la morphologie, le seuillage et le regroupement.

2.2.1.1 Segmentation basée sur la texture

Les drusen sont des structures plus claires sur la surface de la rétine. Selon la méthode appliquée, la texture peut alors les mettre en évidence par rapport à l'arrière-plan. Différents travaux ont, par exemple, utilisé les filtres de Gabor qui extraient l'information de texture représentée par les variations de réponse de ces filtres pour différentes orientations sur différentes surfaces (Parvathi & Devi, 2007). Vu la différence de couleur et de la distribution d'intensité entre les drusen et l'arrière-plan, leurs réponses aux filtres de Gabor sont donc différentes. D'autres travaux ont utilisé la transformée en ondelettes pour l'extraction de la texture. Les fonctions de bases orthonormales

ont souvent montré leur efficacité (Brandon & Hoover, 2003; Priya & Aruna, 2011). Les méthodes AM-FM ont permis, quant à elles, d'extraire les structures à différentes variations d'amplitude et à différentes échelles de fréquence (Barriga, et al., 2009; Agurto, et al., 2011). En effet, les lésions comme les drusen sont des structures à hautes fréquences qui se différencient des autres structures à plus basse fréquence, comme l'arrière-plan ou la macula peu vascularisée. Enfin, il existe également des techniques d'analyse de texture basées sur les informations statistiques des structures qui segmentent les drusen en utilisant des caractéristiques comme la moyenne standard d'intensité, la distribution d'intensité ou l'entropie d'une zone particulière de l'image (Kose, Sevik, Gençalioglu, Ikibas, & Kayikicioglu, 2010). Les informations de texture sont ensuite représentées sous forme de vecteur de caractéristiques et sont utilisées pour classifier les drusen, notamment pour différencier les drusen miliaires des drusen séreux (van Grinsven, et al., 2013).

2.2.1.2 Segmentation basée sur le seuillage

Le seuillage consiste à binariser l'image suivant un seuil sur l'intensité ou la fréquence. Cette binarisation définit alors deux régions : la région d'intérêt (ROI, pour *region of interest*) qui est mise à blanc et l'arrière-plan qui est mis à noir. Les différentes techniques portent alors sur le choix de ce seuil, qui peut se déterminer, soit empiriquement, soit automatiquement à partir de l'histogramme de l'image globale ou dans des zones locales. Dans le cadre de la détection de drusen, les techniques reposent principalement sur un seuillage de l'intensité des pixels. Parmi ces techniques, des travaux se sont basés sur une région d'intérêt pour déterminer empiriquement le seuil (Morgan, Cooper, Constable, & Eikelboom, 1994) et d'autres ont utilisé la méthode d'Otsu pour déterminer automatiquement le seuil (Smith, et al., 2005). Des prétraitements ont également été appliqués avant le seuillage, comme les filtres médians et filtres gaussiens (Soliz, Wilson, Nemeth, & Nguyen, 2002), les filtres homomorphiques (Rapantzikos, Zervakis, & Balas, 2003) ou des opérations morphologiques (Liang, Wong, Liu, Chan, & Wong, 2010) pour atténuer le bruit. Le but du seuillage est alors de détecter les drusen, déterminer leur taille ou leur étendue. La précision pour la détection est généralement très bonne (Rapantzikos, Zervakis, & Balas, 2003), mais le mauvais calcul de l'étendue des drusen est généralement une limite des méthodes de seuillage (Morgan, Cooper, Constable, & Eikelboom, 1994).

2.2.1.3 Segmentation basée sur le regroupement

Les techniques de regroupement consistent à assembler les objets dans des groupes à l'aide de méthodes d'apprentissage non-supervisé. Des éléments d'un même groupe doivent présenter plus de similarité entre eux qu'avec des éléments d'autres groupes. Dans le contexte de la DMLA, cela a été utilisé notamment pour l'étude des phénotypes liés aux gènes de la DMLA (Quellec, Russel, & Abràmoff, 2011) ou pour la détection de la DMLA (Hanafi, Hijazi, Coenen, & Zheng, 2010). Ces méthodes forment des groupes à partir d'un vecteur de caractéristiques généré sur la base de la distribution d'intensité ou d'histogrammes spatiaux.

2.2.1.4 Segmentation basée sur la mise en correspondance de contour ou de modèle

La détection de contour a pour but de rechercher les pixels qui forment les bords d'une région bien distincte de son voisinage. La segmentation de l'image se fait alors par la détection de changement abrupte d'intensité, par exemple via le calcul du gradient ou du Laplacien (Parvathi & Devi, 2007; Mora, Vieira, Manivannan, & Fonseca, 2011). La détection de contour fonctionne bien pour détecter les drusen miliaires, mais cette méthode se montre peu efficace pour segmenter les drusen sérieux à cause de leur apparence plus diffuse, ne permettant pas une bonne évaluation de ces lésions. La correspondance de modèle consiste à faire correspondre une partie de l'image au modèle d'un élément à détecter. Dans ce cas-ci, le but est donc de trouver les zones de l'image correspondant aux drusen. Le modèle utilisé doit donc pouvoir représenter les propriétés des drusen, comme leur bord un peu plus diffus ou leur plus haute intensité par rapport au voisinage. Des exemples de modèles pour détecter les drusen sont les modèles circulaires ou les modèles gaussiens. Ensuite, la correspondance se fait via des mesures de similarité. Ces méthodes permettent notamment la localisation des régions où se trouvent des drusen (Remeseiro, Barreira, Calvo, Ortega, & Penedo, 2009).

2.2.1.5 Discussion sur les méthodes basées sur la segmentation de drusen

La segmentation des drusen présente un très grand intérêt pour le dépistage précoce de la DMLA. Il est en effet important de dépister au plus tôt la maladie car, bien qu'il n'existe pas encore de traitement qui permette de guérir la DMLA, il est possible d'en ralentir la progression vers un stade avancé, voire l'éviter, avec les traitements actuels.

Les différentes méthodes proposées sont, pour la plupart, très efficaces pour détecter les drusen miliaires, mais présentent souvent des difficultés à détecter les drusen séreux à cause de leur forme plus diffuse. Or, ces derniers sont très importants car ils caractérisent la forme modérée de la maladie. À ce stade, le patient ne présente pas encore de cécité mais le risque de progresser vers un stade avancé est très grand.

Ensuite, la segmentation des drusen n'est pas suffisante pour un suivi complet de la maladie. En effet, certaines formes avancées avec de fortes atrophies ou hémorragies ne présentent pas de drusen et donc, ces cas-là ne sont pas pris en considération. Or, pouvoir caractériser toutes les formes de la maladie peut être très utile, notamment pour le suivi de la progression de la maladie ou dans les études cliniques dont l'objectif est le développement de traitements permettant de guérir la maladie.

2.2.2 Méthodes basées sur le calcul direct de caractéristiques

Les méthodes présentées précédemment se concentraient principalement sur le dépistage précoce de la DMLA, en analysant exclusivement les drusen. Afin de tenir compte des autres structures caractérisant la DMLA, comme les hémorragies et les GA, des méthodes calculant directement les caractéristiques de l'image de fond d'œil ont alors été proposées. Au lieu de chercher des structures dans l'image, on extrait les propriétés de l'image qui permettent de caractériser les différentes formes de la DMLA. Cette section présente quelques travaux basés sur le calcul direct de caractéristiques.

2.2.2.1 Détection de la DMLA par le calcul de LBP en multi-résolution

La méthode suivante a été proposée par *Garnier et al.*, dans le cadre de détection de la DMLA (Garnier, Hurtut, Tahar, & Cheriet, 2014). Pour ce faire, une transformée en ondelette de Lemarié est appliquée sur l'image de fond d'œil afin d'extraire l'information à différentes résolutions. À chaque décomposition, quatre coefficients sont obtenus, pour lesquels les histogrammes de signe et de magnitude du LBP sont calculés (Figure 2-18). Tous ces histogrammes sont ensuite concaténés afin de former le vecteur de caractéristiques. La classification de l'image se fait ensuite par LDA.

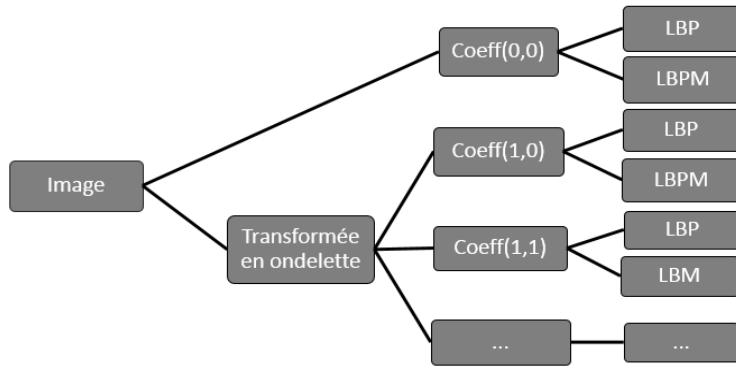


Figure 2-18: Méthode de calcul des caractéristiques LBP en multi-résolution proposée par Garnier et al (2014).

D'après l'auteur, cette méthode permet une bonne classification des images de DMLA sous différentes formes et de qualité variable, avec 95,5% en spécificité, 91,3% en sensibilité et 93,3% en précision. Cependant, la base de données n'était pas suffisamment grande pour une validation exhaustive. En ce qui concerne le choix du classifieur, la LDA permet de réduire intrinsèquement la dimensionnalité de l'espace de caractéristiques, mais suppose que la séparation des données de ce nouvel espace est linéaire. Or, il se peut que la séparation optimale entre les différentes catégories de DMLA et de non-DMLA soit plus complexe que cela. De plus, cette méthode est dépendante de la densité des classes, ce qui n'est pas pratique lorsque le nombre d'exemples d'apprentissage est petit.

2.2.2.2 Classification automatique par « sac de mots visuels »

Cette méthode proposée par *Kankanhalli et al* se base sur le concept de « sac de mots », c'est-à-dire classifier un élément sur base de mots-clés caractérisant une catégorie (Kankanhalli, Burlina, Wolfson, Freund, & Bressler, 2013). Dans le cadre de la classification de l'image, les mots-clés sont en fait des caractéristiques de contexte visuel.

Tout d'abord, un prétraitement de normalisation d'illumination et de rehaussement de contraste est appliqué sur une ROI, qui correspond au carré inscrit dans le cercle formé par la rétine. Ces images sont ensuite converties en $L^*a^*b^*$. La construction du vocabulaire de mots visuels s'effectue ensuite par le regroupement K-means appliqué sur les images de DMLA, qui sont supposées

contenir tous les types de structures caractérisant la maladie. Les différents points SURF sont alors classifiés en K catégories, qui représentent les types de points qui peuvent être rencontrés dans le cadre de la DMLA. Les centroïdes des différentes catégories constituent alors la base du vocabulaire pour construire le vecteur de caractéristiques.

Ensuite, le calcul de caractéristiques d'une image consiste à calculer les points SURF de l'image (Figure 2-19), et de construire un histogramme dont chaque intervalle correspond au nombre de points SURF les plus proches à un centroïde du vocabulaire. Ces histogrammes sont pondérés spatialement selon leur position par rapport au centre de l'image.

Enfin, une RF est modélisée à partir d'un ensemble d'apprentissage. Ce modèle est alors utilisé pour classifier de nouvelles images de fond d'œil, selon les catégories de l'AREDS (voir Section 1.3).

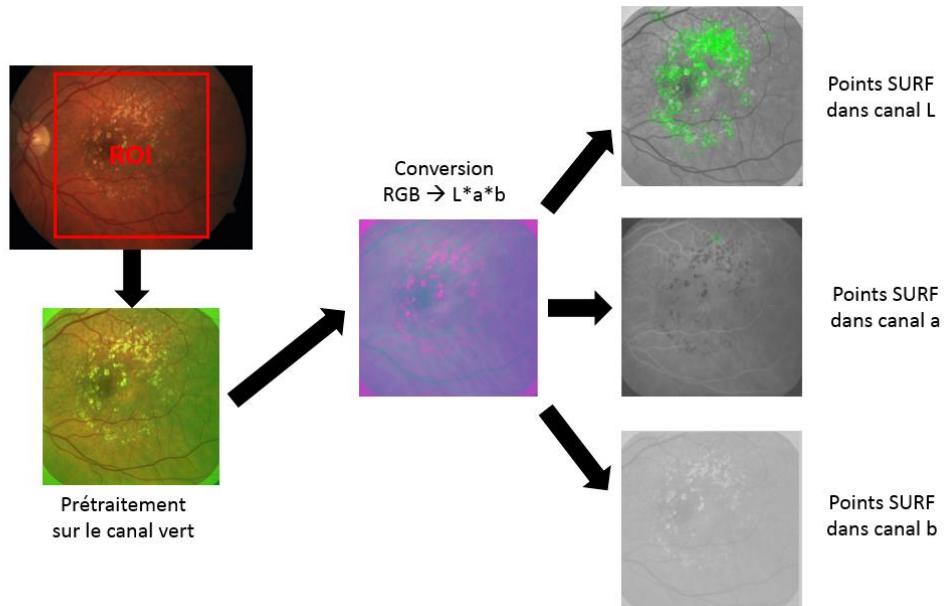


Figure 2-19: Méthode de calcul des points SURF proposée par *Kankanhalli et al* (2013).

Différents types de classification ont été réalisés pour tester la méthode. Des classifications binaires comme {1 & 2} vs {3 & 4}, {1 & 2} vs {3}, {1} vs {3} ou encore {1} vs {3 & 4} ont été effectuées pour mettre en évidence la catégorie des cas modérés qui est celle qui serait la plus importante à

identifier. Une classification à trois classes a été également réalisée pour différencier les cas non-DMLA et précoces, les cas modérés et les cas avancés. Les résultats ont montré de très bonnes performances avec une précision au-dessus de 90% pour les différentes classifications sur une grande base de données. Cependant, ces résultats ont été validés sur des images sélectionnées comme étant de bonne qualité. Or, il n'est pas garanti que la méthode soit efficace pour classifier des images de moins bonne qualité. Une autre limitation qui peut être relevée est que la pondération spatiale des histogrammes considère que la macula soit au centre de l'image. Or, ce n'est pas toujours le cas dans les photographies de fond d'œil. Dans un contexte de télémédecine, cette méthode pourrait donc moins bien performer du fait que les images sont de qualité très variable et pas nécessairement centrées sur la macula.

2.3 Objectifs et hypothèses de recherche

La revue de littérature présentée précédemment a permis d'établir les limitations des méthodes actuelles pour l'évaluation de la DMLA. La problématique est donc qu'il n'existe aucune méthode automatique qui performe une classification du degré de sévérité de la DMLA et qui soit robuste à la qualité de l'image. L'objectif général de ce projet est donc de développer et valider un système de classification automatique de la DMLA sur base d'images de résolution et de qualité variable.

Pour ce faire, plusieurs sous-objectifs doivent être réalisés :

- Constituer une base de données, où les quatre catégories de l'AREDS sont représentées par des exemples dont la qualité de l'image est variable.
- Étudier quelles sont les caractéristiques les plus pertinentes pour la classification de la DMLA et de les implémenter. En vue des résultats obtenus dans la littérature, on se base sur l'hypothèse que les caractéristiques de texture sont suffisamment discriminantes et robustes pour catégoriser la DMLA dans les quatre catégories de l'AREDS.
- Développer des méthodes de sélection de caractéristiques afin de réduire la dimensionnalité du problème et de déterminer l'importance de la contribution de chacune des caractéristiques pour la classification.
- Modéliser des classifieurs, et d'évaluer leurs performances à classifier les images de fond d'œil dans les différentes catégories et de comparer les résultats obtenus avec ceux de la littérature.

CHAPITRE 3 ARTICLE 1 : UNE CLASSIFICATION AUTOMATIQUE POUR LA DÉGÉNÉRESCENCE MACULAIRE LIÉE À L'ÂGE BASÉE SUR LES CARACTÉRISTIQUES D'IMAGES DE FOND D'ŒIL

3.1 Présentation de l'article

Ce chapitre reprend intégralement l'article de revue qui a été soumis à *IEEE Transactions on Biomedical Engineering* en juin 2015. Dans cet article dont la forme a été adaptée à celle du mémoire, sont présentées la méthodologie qui a conduit sur l'étude de l'importance des différentes caractéristiques implémentées, et de l'évaluation des performances pour différents systèmes de classification de degré de sévérité de la DMLA. Cet article fait suite à un article de conférence soumis en avril 2015 et accepté à *International Conference of Image Analysis and Recognition*, présenté à l'Annexe C. Celui-ci présente les résultats préliminaires au développement du système, notamment sur les caractéristiques les plus intéressantes individuellement et sur une comparaison entre le classifieur LDA et SVM.

3.2 An Automatic Classification for Age-related Macular Degeneration Based on Fundus Image Features

3.2.1 Introduction

Age-related macular degeneration (AMD) is the main cause of visual deficiency and irreversible blindness in the elderly in Western countries (Kasuga, Chen, & Zhang, 2011). It combines a variety of disorders affecting the macula. The early stage of AMD is asymptomatic but small lesions, called drusen, can be revealed through examination of the retina. When the size and number of drusen increase, this indicates the moderate stage during which vision is gradually lost, until legal blindness in advanced stages, which present hemorrhages (wet AMD) or geographic atrophies (late dry AMD). Currently, there is no approved treatment to cure AMD. However, treatments to slow its progression exist. For this reason, regular eye examination is required (Meleth , Raiji, Krishnadev, & Chew, 2011).

A fast and non-invasive modality often used for eye examination is fundus photography. Grading AMD's severity stage based on fundus images helps the ophthalmologist in determining the suitable treatment option.

A recommended clinical classification is the simplified classification proposed by the Age-Related Eye Disease Study (AREDS) (Age-Related Eye Disease Study Research Group (AREDS), 2005), dividing AMD stages into four categories, illustrated in Figure 3-1: non-AMD {1}, early {2}, moderate {3} and advanced {4}. However, human evaluation of fundus images is time-consuming and leads to inter- and intra-expert variability. To address this problem, automatic grading systems for AMD have been proposed for a faster and reproducible assessment.

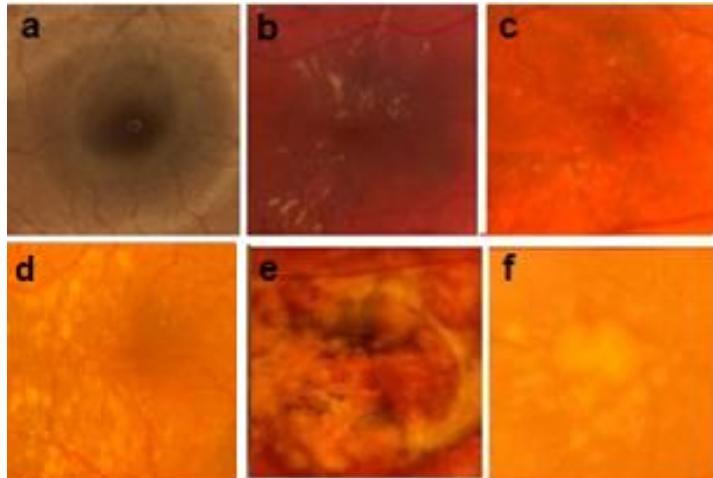


Figure 3-1: Images of macula area for different AMD categories: (a) healthy and (b) diabetic retinopathy in category {1}, (c) hard drusen in category {2}, (d) soft drusen in category {3}, (e) hemorrhages and (f) geographic atrophy in category {4}.

Previous works focus mostly on dry AMD screening based on the detection and quantification of drusen on fundus images (Kanagasingam, et al., 2014). The drusen segmentation techniques are categorized into methods based either on texture analysis, thresholding, clustering edge or template matching. A number of texture-based methods use Gabor filters (Parvathi & Devi, 2007), wavelet transform (Brandon & Hoover, 2003; Priya & Aruna, 2011), amplitude and frequency modulation (Barriga, et al., 2009; Aguerto, et al., 2011) or statistical structural information (Kose, Sevik,

Gençalioglu, Ikibas, & Kayikicioglu, 2010). The segmentation is based on the response of drusen to the applied texture method which is supposed to be different from the background. Thresholding-based methods aim to find the appropriate threshold for separating the drusen from the background. This threshold can be set empirically (Morgan, Cooper, Constable, & Eikelboom, 1994), or automatically with Otsu's method (Smith, et al., 2005). Some image preprocessing is performed before thresholding using median and Gaussian filters (Soliz, Wilson, Nemeth, & Nguyen, 2002), homomorphic filters (Rapantzikos, Zervakis, & Balas, 2003) or morphological operations (Liang, Wong, Liu, Chan, & Wong, 2010). Clustering-based methods consist in gathering into the same groups, elements which are similar to each other. These methods were used for phenotyping genes of AMD (Quellec, Russel, & Abràmoff, 2011) or for identifying AMD (Hanafi, Hijazi, Coenen, & Zheng, 2010). Drusen segmentation can also be achieved through edge detection by identifying abrupt intensity variations using gradient or Laplacian filters (Mora, Vieira, Manivannan, & Fonseca, 2011). Template matching methods use circular or Gaussian templates (Remeseiro, Barreira, Calvo, Ortega, & Penedo, 2009) to model and to detect drusen using similarity measurements. After drusen segmentation, a classification based on drusen features is performed for AMD screening or to assess the risk of progression to the advanced stage (van Grinsven, et al., 2013).

However, drusen segmentation does not provide sufficient information for a complete AMD grading. In fact, mostly in advanced stages, drusen are not observed, especially when there are large hemorrhages or atrophies. Moreover, even if these methods show high accuracy for hard drusen detection (up to 100%, with the best method), the segmentation of soft drusen, which characterize the moderate cases, is highly challenging because of their diffuse shape.

To address these limitations, AMD classification methods based on image features, without segmentation, were introduced. Kankanhalli et al. propose a method based on visual context using SURF keypoints as features and a random forest (RF) for classification (Kankanhalli, Burlina, Wolfson, Freund, & Bressler, 2013). Different binary classifications such as {1&2} vs. {3&4} or {1} vs. {3} and a trinary classification ({1&2} vs. {3} vs. {4}) were considered to discriminate the moderate cases. In fact, close attention to moderate cases is important because, even though the patient still has adequate visual acuity, there is a considerable risk of progression to a more severe stage. To our knowledge, the method in (Kankanhalli, Burlina, Wolfson, Freund, & Bressler, 2013) is the only automatic AMD grading system following a clinical classification. It shows a

good accuracy (above 90%) for AMD severity classification. However, the evaluation was conducted on 2772 images out of 11344 images available in AREDS database (24.4% of the database), selected for their good quality. Trained solely on good quality images the classifier might not be as effective on images of lower quality. In fact, poor quality images are often encountered in a telemedicine context where the acquisition conditions might not always be optimal.

Prior preliminary studies (Garnier, Hurtut, Tahar, & Cheriet, 2014; Phan, Seoud, & Cheriet, 2015) conducted by our group for the evaluation of new features demonstrated promising results with local binary pattern (LBP) in multiresolution for AMD detection. However, the validation were conducted on small datasets and the different features subsets were evaluated individually without considering any combination. Moreover, these preliminary studies aimed only at detecting images with AMD.

The aim of this paper is to go a step further and to propose an automatic system for clinical AMD severity classification robust with respect to image quality. To do so, various features based on texture, color and visual context are computed and used to classify the images according to the different AREDS categories. The validation is performed on a highly heterogeneous set of 159 fundus images. Because the only AMD grading system based on the AREDS classification in the literature is the one proposed by Kankanhalli et al. (Kankanhalli, Burlina, Wolfson, Freund, & Bressler, 2013), we re-implemented their method and evaluated it on the same database to compare the two methods' results depending on the image features used and the choice of classifier. Additionally, the robustness of the two methods to poor quality images is also evaluated.

The organization of the paper is as follows. In section II, the main steps of the proposed AMD classification method are described in details. The experimental setup is explained in section III. The results are presented in section IV, followed by a discussion and a conclusion.

3.2.2 Materials and methods

Fundus images acquired in a real screening context often show an uneven illumination and a poor contrast. To address these issues, a preprocessing step is required. Then, different features based on texture, color and visual context are extracted to characterize the fundus image. Next, a classifier modeling step allows us to measure the relevance of the features. Finally, two classifiers, RF and SVM, are tested with a 10-fold cross validation approach on a database of 159 fundus images.

3.2.2.1 Image Preprocessing

An image normalization is required to correct the illumination drift introduced by the geometry of the eye and the bright flash of light used by the fundus camera. Contrast enhancement is also necessary to improve the information on details in the fundus images.

To perform these preprocessing steps, we use the same methodology proposed in (Kankanhalli, Burlina, Wolfson, Freund, & Bressler, 2013). First, the region of interest (ROI) is defined as the square inscribed in the circle formed by the retina. Then, the green channel is extracted for preprocessing. A median filter with a kernel size of one fourth the ROI size is applied in order to estimate the background illumination. The filtered image is then subtracted from the green channel of the original image. Finally, the green values are multiplied by 2 for contrast enhancement and shifted by the mean of their intensity range for visualization purposes (Figure 3-2).

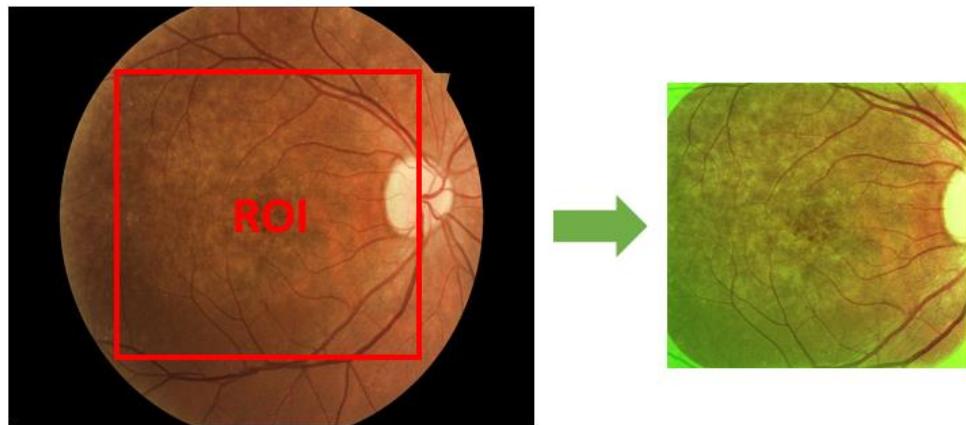


Figure 3-2: Preprocessing method: ROI corresponding to the square inscribed in the circle formed by the retina and the result of preprocessing with illumination normalization and contrast enhancement in green channel.

3.2.2.2 Feature Extraction

Several features based on color, texture and visual context are chosen because they prove to be effective in fundus image analysis. Color information is an intuitive feature, since AMD-related

structures are characterized by specific colors. The texture and local gradient information are also different depending on the state of the retina. The image features considered in this study and their parameter settings are presented in the following subsections.

3.2.2.2.1 *Color Histograms*

Blood vessels and lesions offer the highest contrast in the green channel. That is why most of the methods proposed in the literature for fundus image analysis focus solely on this channel. Still, even though the red channel is considered as saturated and with low contrast, and the blue channel as very noisy in fundus images (Walter, et al., 2007), all three color channels should be considered, especially to discriminate drusen from exudates, which are highly similar lesions but do not characterize the same disease (van Grinsven, et al., 2013). In this study, the RGB and L*a*b* spaces are used. In RGB, the red and blue channels provide additional information to the green one. The L*a*b* space is chosen because luminance and chrominance are separated and color differences can be measured using the Euclidean distance.

We compute the 8-bin histograms for each channel from both color spaces as image features. The number of bins is set to 8 because there are no improvements in the results with a larger number of bins, thus we consider it sufficient for AMD classification.

3.2.2.2.2 *Local Binary Patterns (LBP) in Multiresolution*

To obtain the multiresolution information, a Lemarié wavelet transform is used with four levels of decomposition. For each level, an approximation coefficient and three detail coefficients are computed, containing respectively the low resolution image and the high resolution details in horizontal, vertical and diagonal directions. From the original image and the 16 coefficient images, textural information is extracted using LBP. This consists in measuring the occurrence of local texture primitives, such as corners or edges. To do so, the LBP (Ojala, Pietikainen, & Maenpaa, 2002) is computed for each pixel of gray value g_c in a neighborhood of radius R and P neighbors with gray values g_p :

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \quad (3-1)$$

$$\text{With } s(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{Otherwise} \end{cases}$$

In this study, the parameters are empirically set to R=1 and P=8. The magnitude component of the LBP (Guo & Zhang, 2010) is also computed from the absolute differences of gray intensity between the central pixel and its neighbors $m_p = |g_p - g_c|$:

$$LBP M_{P,R} = \sum_{p=0}^{P-1} t(m_p, c) 2^P \quad (3-2)$$

$$\text{With } t(x, c) = \begin{cases} 1, & \text{if } x \geq c \\ 0, & \text{Otherwise} \end{cases}$$

The threshold c is set to the image mean value.

From the sign and magnitude components of LBP, two histograms are computed by measuring the occurrence of the different patterns in the image. For each RGB color channels, LBP are computed and generate a vector of 2006 features.

3.2.2.2.3 *Histograms of Oriented Gradients (HOG)*

The histogram of oriented gradient is a feature generally used for edge detection (Dalal & Triggs, 2005), but it also contains local directional information which can be used for classification.

The horizontal and vertical gradients are computed by applying a 1D point-centered derivative kernel [-1 0 1] to the color image. Then, local histograms of the four main directions are constructed by dividing the RGB color image into 16x16 cells, with 2x2 cells for block normalization. The constructed vector contains 3600 features.

3.2.2.3 **Classifier Modeling**

Two different classifiers are used in this study: a random forest and a support vector machine. We use the first classifier in order to compare our method fairly with the one in (Kankanhalli, Burlina, Wolfson, Freund, & Bressler, 2013). We use the second one to verify if the choice of classifier has an impact on the results of classification.

3.2.2.3.1 *Random Forest (RF)*

The training of a RF consists in constructing decision trees, using randomly selected training samples and features. Then, the classification of new samples is determined by aggregating the votes of the different trees (Breiman, Random forests, 2001). This method is very user-friendly since only two parameters need to be set: the number of features in the random subset at each tree

node and the number of trees in the forest (Liaw & Wiener, 2002). The first parameter is set to the square root of the total number of features. The second parameter is set to 1000 decision trees for binary classification and to 2500 decision trees for trinary classification, as in (Kankanhalli, Burlina, Wolfson, Freund, & Bressler, 2013). The feature selection step is implicitly performed in the RF training. The features relevance can be measured with different criteria (Breiman, Manual on setting up, using and understanding random forests v3.1., 2002). In this study, two criteria are used: the mean decrease in accuracy and the mean decrease in Gini index. These parameters measure the loss in the accuracy or Gini criterion on the out-of-bag samples, when the feature is removed or permuted. The larger the decrease, the more relevant is the feature.

3.2.2.3.2 Feature Selection and Support Vector Machine (SVM)

The training of an SVM consists in finding the decision boundary that maximizes the margin, that is the space between the elements nearest to the boundary (Steinwart & Christmann, 2008).

As opposed to the RF, the SVM does not perform feature selection on its own. This step therefore must be conducted before training the classifier to avoid the curse of dimensionality and overfitting.

In this study, we used two approaches for feature selection. The first one is based on the Fisher criterion, which must be maximized (Duda, Hart, & Stork, Maximum-likelihood and Bayesian parameters estimation, 2009). This criterion measures the divergence between two classes i and j based on the estimate of the mean μ and the standard deviation σ of the two classes when they are projected on the feature axis F :

$$D(F) = \frac{(\mu_i - \mu_j)^2}{\sigma_i^2 + \sigma_j^2} \quad (3-3)$$

The maximum number of features for classifier modeling is set to one tenth the number of training samples. The final number of features retained is determined based on the best SVM performance obtained by varying the number of features and testing on validation samples.

The second feature selection approach is sequential floating forward selection (SFFS) (Pudil, Novovicova, & Kittler, 1994). It consists in adding iteratively to the model, the l best features and removing the r worst features according to a given criterion which, in our case, is the SVM classification accuracy using non-optimized parameters. The values of l and r are determined by the system itself and thus, are left floating. The process is stopped when the criterion does not

improve anymore or when the maximum number of iterations is reached. This method tends to find a solution that is close to optimal.

In this study, a Gaussian kernel is chosen for the SVM because it is more efficient for systems with complex separations than a linear classifier. In addition, SVMs are useful for systems with a low number of samples because only the elements near the boundary, *i.e.* the support vectors, contribute to the SVM modeling. For classifier modeling, the parameters to be set are γ , the parameter of the Gaussian kernel and C , the number of elements accepted in the margin. These parameters are set according to a performance assessment using a grid search strategy with 10-fold cross-validation.

To consider more than two classes, we used the one-against-all approach. In the training phase, one SVM per class is constructed for discriminating the samples of that class from all the others. The label of a new observation is then determined by the SVM classifier that returns the highest value.

3.2.3 Experimental setup

The validation is conducted on a database of 159 fundus images. 94 images were provided by Diagnos Inc. in Canada using their telemedicine platform. To increase our dataset, we used 11 and 54 images respectively from the Automated Retinal Image Analysis (ARIA, United Kingdom) and Structured Analysis of Retina (STARE, United States) public databases.

Depending on the acquisition conditions, the images vary in terms of resolution and illumination which both affect the image quality (Bartling, Wanger, & Martin, 2009). Image resolution is defined as the capacity to visualize structural details and a low resolution can lead to a loss of this information. Different artefacts, illustrated in Figure 3-3, can be encountered in fundus photography: shadows, intense reflections, specular reflections, blur, haze, or arcs. In this study, an image is considered as being of poor quality if one of these artefacts is present.

An expert in ophthalmology labelled the images into four categories: non-AMD {1}, early {2}, moderate {3} and advanced {4}. These labels are considered as the reference. Table 3-1 shows the number of images in each category with the corresponding image quality.

To assess the classification performance, a 10-fold cross validation approach is used. This consists in taking one tenth of the dataset as a testing set. The rest is used for training the classifier. The prediction results from this classification are kept and the process is repeated for the other data subsets. The prediction and the reference for all images are then compared to obtain the accuracy

(proportion of well classified elements), the sensitivity (proportion of well classified positive elements, usually AMD cases) and the specificity (proportion of well classified negative elements, usually non-AMD). For a trinary classification, a confusion matrix is used to assess the performance.

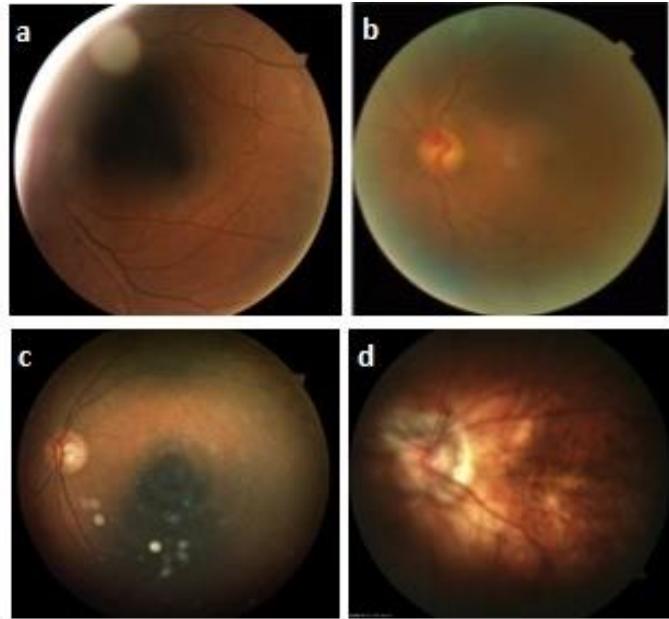


Figure 3-3: Examples of poor quality images: (a) shadows and intense reflection, (b) haze, (c) arc and specular reflections and (d) blur

Table 3-1: Number of images in each category with corresponding image quality

Category	{1} Non-AMD	{2} Early	{3} Moderate	{4} Advanced
Good Quality	44	7	7	1
Poor quality	36	8	6	50

For the performance assessment, different binary classifications and a trinary classification are performed. Because of the limited number of samples, categories {2} and {3} cannot be taken individually for classification. The different classifications tested are thus:

- {1} vs. {2&3&4}. This was performed in (Garnier, Hurtut, Tahar, & Cheriet, 2014; Phan, Seoud, & Cheriet, 2015) for AMD screening
- {1&2} vs. {3&4} and {1} vs. {3&4}. This was performed in (Kankanhalli, Burlina, Wolfson, Freund, & Bressler, 2013) for AMD severity classification, to discriminate moderate and advanced cases that require treatment from those not at risk.
- {1} vs. {2&3} and {1} vs. {4}. It allows us to verify how well each AREDS category can be distinguished from non-AMD cases.
- {1} vs. {2&3} vs. {4}. It allows us to discriminate cases for which treatment would help to maintain vision at a useful level, from non-AMD cases and from cases that are too advanced.

First, our method is compared with the one proposed in (Kankanhalli, Burlina, Wolfson, Freund, & Bressler, 2013) which we re-implemented as follows. SURF keypoints are computed on the subset of images presenting AMD, previously converted into $L^*a^*b^*$, based on the hypothesis that all AMD features (drusen and other lesions) are represented in that image subset. The keypoints are detected using ten octaves, three layers per octave and a Hessian threshold of 600. Using the SURF features (sign of Laplacian, orientation, scale of detection and strength of detected feature), a K-means clustering selects centroids on which the vocabulary is based to construct the features vector. For binary classification, K is set to 100 and for trinary classification, K is set to 300. All the parameters used to compute the SURF keypoints and to construct the vocabulary are the same as in (Kankanhalli, Burlina, Wolfson, Freund, & Bressler, 2013). Once the vocabulary is established, a histogram is constructed by measuring the occurrence of keypoints depending on the nearest centroid. Here, the features vector is not spatially dependent. The AMD classifications listed above, using a RF as classifier, are assessed with our features (LBP+Color+HOG) and with those proposed in (Kankanhalli, Burlina, Wolfson, Freund, & Bressler, 2013) (SURF), applied to the whole database.

If good quality images are selected to train a classification system, its efficiency will not be guaranteed for processing images of variable quality, e.g. in a telemedicine context. Thus, an assessment of the two methods using only good quality images for training and only poor quality images for testing is also performed so as to compare their robustness to variations in image quality.

Finally, an assessment is performed using SVM to evaluate the impact of the choice of the classifier on the results. The two feature selection approaches are used to validate which image features are the most discriminant.

Our overall approach for performance assessment aims at allowing us to determine the best solution for robust AMD classification.

3.2.4 Results

3.2.4.1 RF Classification

The results of binary classification assessment on the whole database are shown in Table 3-2 and the results for trinary classification assessment are shown in Table 3-3. The prediction of every system is compared to the reference established by the clinical expert. The systems using LBP+Color+HOG always features perform good classification, their accuracy ranging between 74.8% and 90.8% and outperform those using SURF features, whose accuracy range between of 49.1% and 72.2% on the same dataset.

Table 3-4 and Table 3-5 show the results for the same classifiers trained with good quality images and tested with poor quality images. Due to the small number of good quality images in category {4}, 14 poor quality images in this category were used for training. The results show a considerable decrease of the sensitivity for SURF systems, with values reduced to 0.0%-4.0%. On the other hand, the values of specificity, sensitivity and accuracy for LBP+Color+HOG systems stay much more stable in the classification of poor quality images.

Table 3-2: Binary classification assessment with RF

Category test	Feature sets	Specificity	Sensitivity	Accuracy
{1} vs. {2&3&4}	LBP+Color+HOG	85.0%	73.4%	79.3%
	SURF	68.8%	35.4%	52.2%
{1&2} vs. {3&4}	LBP+Color+HOG	92.6%	79.7%	87.4%
	SURF	87.4%	12.5%	57.2%
{1} vs. {2&3}	LBP+Color+HOG	97.5%	17.9%	76.9%
	SURF	96.3%	3.6%	72.2%
{1} vs. {3&4}	LBP+Color+HOG	90.0%	81.3%	86.1%
	SURF	85.0%	15.6%	54.2%
{1} vs. {4}	LBP+Color+HOG	92.5%	88.2%	90.8%
	SURF	92.5%	7.8%	58.8%

Table 3-3: Trinary classification assessment with RF

Category test	Feature sets	Accuracy	Confusion matrix		
{1} vs. {2&3}	LBP+Color+HOG	74.8%	91.3%	75.0%	11.8%
			0.00%	3.6%	0.0%
vs. {4}	SURF	49.1%	8.7%	21.4%	88.2%
			92.5%	96.4%	90.2%
			0.0%	0.0%	2.0%
			7.5%	3.6%	7.8%

Table 3-4: Binary classification assessment with RF trained with good quality images and tested on poor quality images

Category test	Feature sets	Specificity	Sensitivity	Accuracy
{1} vs.	LBP+Color+HOG	83.8%	94.0%	87.7%
{2&3&4}	SURF	98.8%	4.0%	62.3%
{1&2} vs.	LBP+Color+HOG	71.1%	88.1%	79.3%
{3&4}	SURF	98.0%	0.0%	50.6%
{1} vs. {2&3}	LBP+Color+HOG	91.9%	14.3%	70.6%
	SURF	100%	0.0%	72.6%
{1} vs. {3&4}	LBP+Color+HOG	88.8%	97.6%	91.8%
	SURF	98.8%	2.6%	65.6%
{1} vs. {4}	LBP+Color+HOG	78.4%	77.8%	78.1%
	SURF	100%	0.0%	50.7%

Table 3-5: Trinary classification assessment with RF trained with good quality images and tested on poor quality images

Category test	Feature sets	Accuracy	Confusion matrix		
			85.0%	25.0%	15.7%
	LBP+Color+HOG	79.9%	5.0%	60.7%	2.0%
{1} vs. {2&3}			10.0%	14.3%	82.3%
vs. {4}			100%	50.0%	70.6%
	SURF	68.6%	0.0%	50.0%	0.0%
			0.0%	0.0%	29.4%

3.2.4.2 RF Classification with Threshold

It can be seen in Table 3-2 that the sensitivity for labelling category {2&3} versus {1} is very low, regardless of the features used. The reason for this is the unbalanced proportion of samples in the different categories. To address this issue, a classification can be performed with a threshold based on the proportion of decision trees voting for a given category. Figure 3-4 shows the sensitivity (TPR, for true positive rate) and the specificity (1-FPR, for false positive rate) curves as a function of the threshold on the proportion of trees voting for category {2&3}.

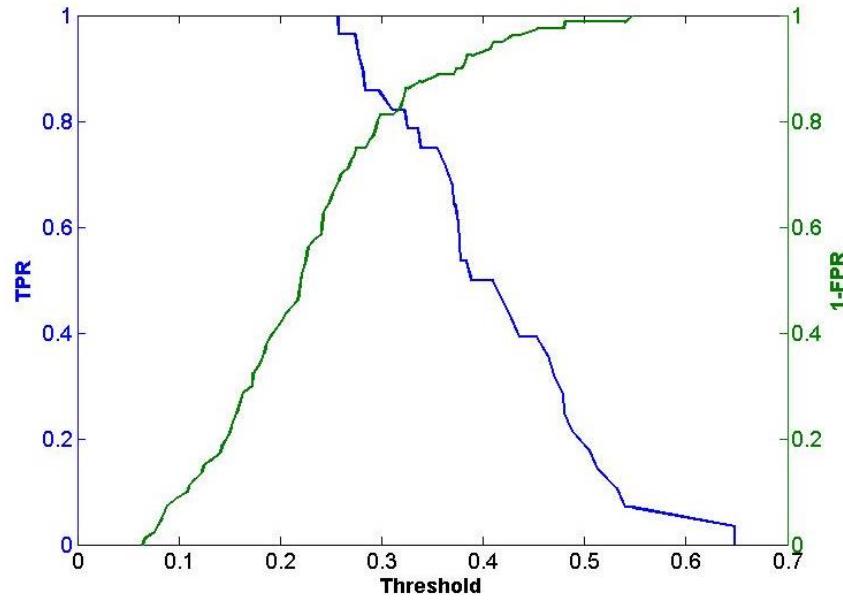


Figure 3-4: TPR and 1-FPR for classification {1} vs {2&3} for different thresholds using LBP+Color+HOG features.

The optimal threshold to obtain a good compromise between sensitivity and specificity is set to 0.3. The result of this classification is shown in Table 3-6 and reveals a much better sensitivity at the cost of a lower specificity.

Table 3-6: Binary classification with threshold assessment with RF

Category test	Feature sets	Specificity	Sensitivity	Accuracy
{1} vs. {2&3}	LBP+Color+HOG	81.3%	82.1%	81.5%

3.2.4.3 Feature Relevance

We also want to know which features are the most relevant among the different ones computed. The feature relevance results, as measured with mean decrease in accuracy and in Gini Index, are nearly the same for all the tested classifications. Both these tests yield the result that LBP in the green channel are the most relevant feature set.

To illustrate this, Figure 3-5 shows the ten most relevant features for one of the binary classification ($\{1 \& 2\}$ vs $\{3 \& 4\}$), as measured by the mean decrease in accuracy.

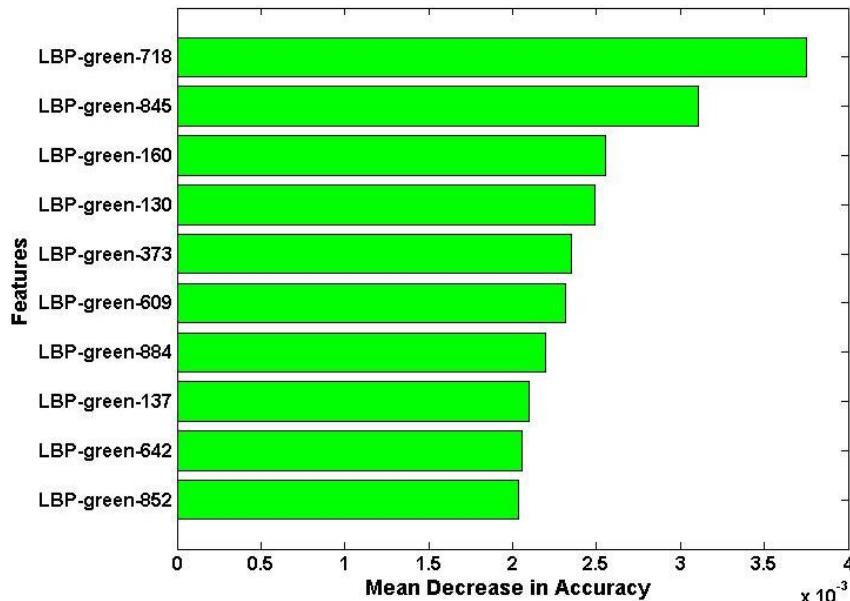


Figure 3-5: The ten most relevant features for $\{1 \& 2\}$ vs. $\{3 \& 4\}$ classification as measured by mean decrease in accuracy.

The results of the feature relevance test for the trinary classification case ($\{1\}$ vs. $\{2 \& 3\}$ vs. $\{4\}$), as measured by the Gini index, also show that LBP features contribute the most to AMD classification (Figure 3-6).

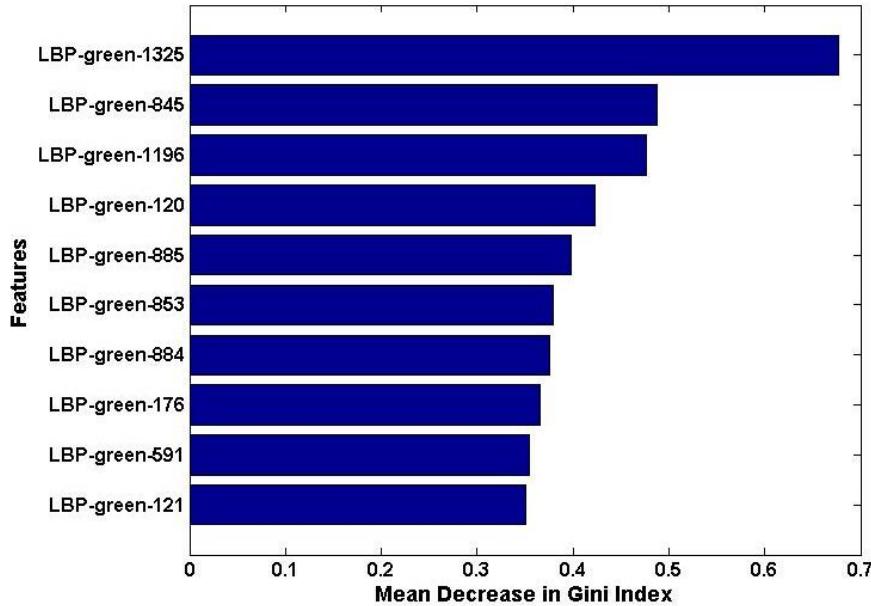


Figure 3-6: The ten most relevant features for $\{1\}$ vs. $\{2 \& 3\}$ vs. $\{4\}$ classification as measured by mean decrease in Gini index.

3.2.4.4 SVM Classification

The results obtained with the SVM classifier and the features selected with Fisher criterion are shown in Table 3-7 for binary classification and Table 3-8 for trinary classification. The selected features are all LBP features applied to the green channel. This confirms the results of the feature relevance measurements obtained with RF, *i.e.* that LBP features are the most suitable for this application. Comparison between RF and SVM results in terms of accuracy do not show a clear difference. This highlights the fact that the main strength of our method comes from the choice of features and not from the classifier. However, we can notice an improvement in sensitivity for $\{1\}$ vs. $\{2 \& 3\}$ binary classification. In the trinary classification case, the labelling of elements in $\{2 \& 3\}$ is also better with the RF classifier.

Table 3-7: Binary classification assessment with SVM and features selected with Fisher criterion

Category test	Selected features	Specificity	Sensitivity	Accuracy
{1} vs. {2&3&4}	LBP-green (4)	87.5%	79.8%	83.7%
{1&2} vs. {3&4}	LBP-green (7)	90.5%	78.1%	85.5%
{1} vs. {2&3}	LBP-green (4)	87.5%	92.9%	88.9%
{1} vs. {3&4}	LBP-green (3)	92.5%	82.8%	88.2%
{1} vs. {4}	LBP-green (4)	92.5%	86.8%	90.1%

Rem.: Number in () represents the number of selected features.

Table 3-8: Trinary classification assessment with SVM and features selected with Fisher criterion

Category test	Feature sets	Accuracy	Confusion matrix		
{1} vs. {2&3}	LBP-green	81.0%	35.7%	11.8%	
vs. {4}	(14)	76.1%	12.5%	50.0%	3.9%
		7.5%	14.3%	84.3%	

Rem.: Number in () represents the number of selected features.

Table 3-9 and Table 3-10 present the results of classification performance with the SVM and the features selected using SFFS. In this experiment, the SFFS method only selected one feature from the LBP set applied to the green channel. The selected feature was not necessarily the same between tests. It is thereby seen that a single LBP feature is strongly discriminating for AMD classification. However, the confusion matrix in Table 3-10 reveals that a single feature is not enough to recognize category {2&3} in the 3-class problem.

Table 3-9: Binary classification assessment with SVM and features selected with SFFS

Category test	Selected features	Specificity	Sensitivity	Accuracy
{1} vs. {2&3&4}	LBP-green (1)	87.5%	81.0%	84.3%
{1&2} vs. {3&4}	LBP-green (1)	81.1%	89.3%	82.4%
{1} vs. {2&3}	LBP-green (1)	82.5%	89.3%	84.3%
{1} vs. {3&4}	LBP-green (1)	87.5%	85.9%	86.8%
{1} vs. {4}	LBP-green (1)	91.3%	92.1%	91.6%

Rem.: Number in () represents the number of selected features.

Table 3-10: Trinary classification assessment with SVM and features selected with SFFS

Category test	Feature sets	Accuracy	Confusion matrix		
			88.8%	82.1%	11.8%
{1} vs. {2&3} vs. {4}	LBP-green (1)	73.0%	0.0%	0.0%	0.0%
			11.2%	17.9%	88.2%

Rem.: Number in () represents the number of selected features.

3.2.5 Discussion

The validation revealed that the best results for AMD classification are obtained with LBP in multi-resolution applied to the green channel. These features were considered as the most relevant for AMD classification and were favoured by both feature selection approaches. The present work confirms that these features are robust with respect to image quality, such as suggested in our prior studies (Garnier, Hurtut, Tahar, & Cheriet, 2014; Phan, Seoud, & Cheriet, 2015), and extends these last works from AMD detection to AMD severity classification. Even with small learning samples, the systems using our approach achieves between 73.0% and 91.6% in classification accuracy, which is especially good considering the large proportion of poor quality images in the database. By comparison, the AMD severity classification system proposed in (Kankanhalli, Burlina, Wolfson, Freund, & Bressler, 2013) only obtains accuracy between 49.1% and 72.2% on the same dataset.

There are certain differences between the results obtained in this study, using the method proposed in (Kankanhalli, Burlina, Wolfson, Freund, & Bressler, 2013) and the results reported in (Kankanhalli, Burlina, Wolfson, Freund, & Bressler, 2013). First, the database on which the classifications were performed is quite different. Our database contains a large proportion of poor quality images, whereas only good quality images were selected in (Kankanhalli, Burlina, Wolfson, Freund, & Bressler, 2013). Furthermore, the number of samples in each category is lower compared to their database. As a result, the vocabulary is built with fewer training examples in our

cases. This can affect the system modeling and the classification results. One of the limitations of their method resides in the parameter setting. For the histogram generation, the K-means clustering parameter represents the number of visual words present in AMD images. But the chosen value may not represent the optimal number of AMD characteristic structures. Another limitation is the spatially dependent histogram generation which supposes that the macula is in the center of the image. This may not always be the case because fundus photography can capture the retina under different viewing angles.

The main strength of our method is its robustness versus image quality. In a telemedicine context, where acquisition conditions are not strictly controlled, to only select good quality images is not adequate for AMD evaluation because we want a maximum of cases to be handled. To demonstrate the robustness to image quality, we assess the classification systems performance by training them on good quality images and testing them on poor quality ones. While the systems using SURF features failed to classify poor quality images, the systems using LBP still performed well, with their specificity, sensitivity and accuracy of the same order as the results obtained with the whole database. Thus, our method presents better results in the presence of image quality variation.

As regards the classification performance of our method, we can notice that the category {4} is well differentiated from category {1}, with a sensitivity of 77.8%-92.1%. On the other hand, it was not possible to evaluate how the system distinguishes individually the categories {2} and {3} because of the low number of samples in these categories. The RF is a class density-dependent classifier. Thus, elements from an under-represented class are harder to label correctly with this model. A classifier such as SVM performs a better identification of category {2&3} with a sensitivity of 89.3%-92.9%, because this model is not density-dependent. However, in the 3-class problem, the percentage of well-classified elements in that category is still low. Thus, further validation with more samples in category {2} and {3} must be conducted, in order to better evaluate the strength of the LBP features.

Among the misclassified images, we noticed several common patterns. Misclassified images were often fundus images presenting the retina with visible choroidal vessels (Figure 3-7.a and Figure 3-7.b). Cases of diabetic retinopathy from category {1} were also often misclassified (Figure 3-7.c). This disease can in fact be hard to differentiate from AMD, even for an expert, because of their similar manifestations in the retina. But it is still important to discriminate these two diseases

because they are handled differently. Other mistakes were made when the image is too degraded (Figure 3-7.d), because, for example, of a large shadow or reflection. In such cases, the texture information cannot be extracted correctly. Preprocessing methods to remove artefacts may solve this last problem.

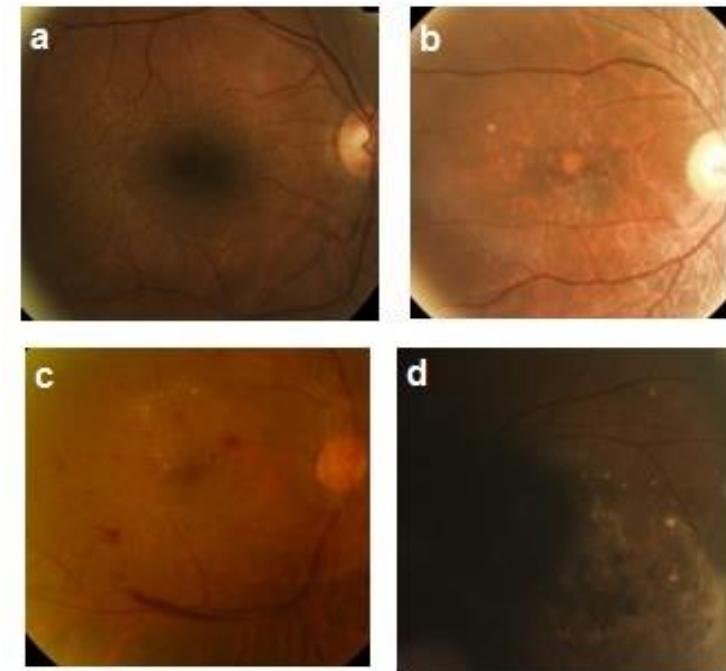


Figure 3-7: Examples of misclassified images with our method: (a) early case with dark macula, (b) moderate case with choroidal vessels in background, (c) non-AMD cases with diabetic retinopathy, (d) advanced case with shadow on the macula

In this study, we used the same preprocessing as (Kankanhalli, Burlina, Wolfson, Freund, & Bressler, 2013) for purposes of comparison, but some improvements could be made to this step. The background illumination is estimated with a median filter, but the convolution with an image of high resolution has a high computational cost. This aspect could be improved by using spectral filtering instead. Also, a local analysis focused on the macular area is a possible solution to improve the results.

Another limitation of our validation is that the AMD reference grading is provided by only one expert. To be unbiased with respect to inter- and intra-expert variability, different experts would

label the training samples. The reference label would then be determined by consensus and the difficulty in evaluating given images could also be measured by the number of experts in disagreement.

3.2.6 Conclusion

We have developed and validated an automatic clinical classification system for AMD. The validation of our method revealed promising results in terms of robustness to image quality and accuracy for different AMD severity classification tasks. Our experimental results highlight the discriminating strength of LBP features over other tested features, whether the classifier is a RF or a SVM. Further validation must be conducted on a database containing more samples in each category in order to confirm these results. Nevertheless, the proposed approach represents an important step toward providing a reliable AMD diagnostic tool for patient monitoring and for clinical trials.

CHAPITRE 4 DISCUSSION GÉNÉRALE

Dans ce chapitre, nous présentons une discussion générale portant sur les différentes étapes de la méthode proposée ainsi que sur les résultats présentés au chapitre 3 et à l'annexe C. Nous abordons également les avantages et les limites de la méthode proposée pour la classification automatique de la DMLA.

Le prétraitement utilisé a permis de corriger les défauts d'illumination et rehausser le contraste, que ce soit sur le canal vert (voir Chapitre 3) ou sur les trois canaux (voir Annexe C). Afin de ne pas introduire de biais dans la comparaison de notre méthode de classification à celle de Kankanhalli *et al.*, nous avons opté pour le même prétraitement, à savoir l'estimation de l'illumination de l'arrière-plan par filtrage médian. Néanmoins, la convolution de l'image par un filtre médian non linéaire assez large est très conteuse en temps de calcul. Une solution envisageable est de réduire la taille de l'image au risque de perdre des informations de détail pour les images de haute résolution. Une autre est de recadrer l'image sur une ROI plus petite autour de la macula. Enfin, des améliorations sur le filtrage en lui-même peuvent également être appliquées, en passant par exemple dans le domaine spectral. Appliquer le prétraitement sur les trois canaux et une normalisation des couleurs pourraient faire ressortir le pouvoir discriminant d'autres caractéristiques se basant sur les autres canaux de couleur, comme le LBP ou les histogrammes de couleurs.

Nous avons aussi démontré qu'une analyse se concentrant uniquement sur la macula donnait de meilleurs résultats qu'une analyse sur toute la rétine (voir Annexe C). Une analyse locale restreinte à la ROI est donc meilleure qu'une analyse globale pour classifier les images. Lors de ce projet, la zone maculaire a été segmentée de manière manuelle, mais des méthodes automatiques de détection de la macula pourraient facilement établir la ROI afin de rendre l'ensemble du processus complètement automatique.

Notre étude sur les caractéristiques a montré que celles contribuant le plus à la classification de la DMLA sont les caractéristiques de LBP calculées sur le canal vert. Les LBP démontrent un fort pouvoir discriminant malgré une base d'images avec peu d'exemples d'apprentissage et une large proportion d'images de mauvaise qualité. Lorsqu'on compare avec les caractéristiques employées par Kankanhalli *et al.*, ces dernières requièrent une bonne base d'apprentissage et des images de test de bonne qualité, sinon les performances de classification diminuent drastiquement. De plus,

cette étude a démontré la robustesse à la qualité de l'image lorsque les LBP en multi-résolution sont utilisés comme caractéristiques, comme cela est suggéré dans les travaux de Garnier *et al.* Ce projet a, en plus, étendu ce dernier travail d'une méthode de détection de la DLMA à une méthode d'évaluation du degré de sévérité de la DMLA. Les autres caractéristiques étudiées lors de ce projet semblent également intéressantes à étudier, et un prétraitement adéquat pourrait relever leur pouvoir discriminant.

Comme la base de données ne présente pas un grand nombre d'images, il était important de passer par une étape de sélection de caractéristiques afin d'éviter la malédiction de la dimensionnalité et le surapprentissage. Les méthodes utilisées pour la sélection de caractéristiques dans ce projet sont la sélection par mérite individuel selon le critère de Fisher et la méthode SFFS. La première méthode est une méthode peu couteuse en calcul et permet de relever rapidement les meilleures caractéristiques qui discriminent au mieux les classes. La limitation de cette méthode est de choisir des caractéristiques redondantes et non complémentaires. La deuxième tente de chercher le sous-ensemble optimal de caractéristiques qui offrent la meilleure classification sur des échantillons de validation. Cette dernière méthode semble avoir des limites lorsque le nombre de caractéristiques est trop important, surtout en temps de calcul et sur le résultat obtenu, avec un sous-ensemble réduit à une caractéristique. Dans les problèmes de classification binaire, les deux méthodes présentaient des résultats similaires, malgré un sous-ensemble différent de caractéristiques sélectionnées pour les deux méthodes. Ceci peut s'expliquer par une grande corrélation entre les caractéristiques sélectionnées par Fisher qui sépare les deux classes de la même façon que la caractéristique sélectionnée par SFFS. De plus, la caractéristique sélectionnée par SFFS faisait généralement partie des caractéristiques sélectionnées par Fisher. Dans le cas du problème à trois classes, il s'est avéré que les caractéristiques sélectionnées par Fisher ont présenté de meilleurs résultats pour la classification que celles sélectionnées par SFFS, du fait qu'un plus grand nombre de caractéristiques ont été sélectionnées. Pour ce cas-ci, avoir plus d'exemples d'apprentissage permettrait de travailler avec un sous-ensemble de caractéristiques plus grand et possiblement une meilleure classification. Dans cette étude, l'analyse en composantes principales n'a pas été abordée afin de conserver le sens de chaque caractéristique et d'en étudier la pertinence. Mais, cette dernière méthode pourrait être appliquée afin d'avoir un ensemble de caractéristiques moins corrélées.

Les classifieurs étudiés lors de ce projet sont la LDA, la RF et les SVM. La séparation entre les différentes catégories de DMLA ne semble pas être linéaire. En effet, une classification par LDA

ne donne pas d'aussi bons résultats que les deux autres modèles (voir Annexe C). La RF a pour avantage d'être facile à modéliser sans devoir passer par une sélection de caractéristiques, qui est faite implicitement dans la méthode. De ce fait, on peut également mesurer l'importance des caractéristiques qui confirmait également que les LBP appliqués sur le canal vert étaient les meilleures caractéristiques. Les SVM ont pour avantage de ne pas être dépendant de la densité des échantillons, en ne prenant que les éléments à la frontière pour modéliser le système. Les éléments dans les catégories moins représentées, comme les cas précoce et modérés, étaient donc mieux classifiés par rapport à la méthode RF. Le désavantage est que la modélisation est plus complexe car il faut passer d'abord par une sélection de caractéristiques, bien choisir le noyau et optimiser les paramètres du classifieur SVM. Outre les classifications impliquant des catégories peu représentées, les résultats obtenus pour les différentes classifications de DMLA étaient plutôt similaires pour les deux classifieurs, ce qui démontrait que la principale force du système provenait des caractéristiques plutôt que du choix du classifieur.

En établissant les différentes classifications binaires, nous avons pu étudier comment les systèmes proposés arrivaient à distinguer les différentes catégories de DMLA les unes par rapport aux autres. Notre méthode arrive à bien distinguer les cas négatifs des cas avancés, avec 70,3% de spécificité, 92% de sensibilité et 82,8% de précision avec le classifieur RF et 91,3% de spécificité, 92,1% de sensibilité et 91,6% de précision pour le classifieur SVM dont les caractéristiques ont été sélectionnées par SFFS. En ce qui concerne les autres catégories, les catégories de cas précoce et modérés sont plus difficiles à distinguer avec le classifieur RF. Ceci pourrait éventuellement être expliqué par le fait que la distribution des échantillons d'apprentissage dans les différentes classes n'est pas équilibrée. En effet, lorsqu'on modifie le seuil de classification à 0.3 au lieu de 0.5, on obtient une augmentation de la sensibilité pour détecter les cas précoce et modérés. Les méthodes avec SVM arrivent à bien différencier ces catégories par rapport aux cas de non-DMLA avec 87,5% de spécificité, 92,9% de sensibilité et 88,9% de précision, avec les caractéristiques sélectionnées par Fisher et 82,5% de spécificité, 89,3% de sensibilité et 84,3% de précision, avec les caractéristiques sélectionnées par SFFS. Ces résultats semblent très prometteurs, vu le petit nombre d'images dans la base et la grande proportion d'images de basse qualité. Cependant, des difficultés se présentent encore pour le problème proposé à trois classes. Une étude sur la distinction individuelle des cas précoce et modérés devrait être également effectuée mais notre base de données actuelle ne le permet pas.

Dans la catégorie des cas non-DMLA, les images mal classifiées sont généralement des images de rétinopathie diabétique, peu importe la qualité de l'image (Figure 4-1.a). Lorsqu'il y a des hémorragies, ces images sont plus associées à des cas avancés. Ces erreurs sont plutôt légitimes dans le sens où les deux maladies se ressemblent fortement et posent également des difficultés pour le clinicien. Mais comme le traitement de ces deux pathologies est différent, il est important de pouvoir les distinguer. Dans la catégorie des cas précoce et modérée, des erreurs de classification correspondent à des images où les vaisseaux de la choroïde dans le fond de la rétine sont visibles (Figure 4-1.c et Figure 4-1.d). Dans la catégorie des cas avancés, les erreurs correspondent à des images où une ombre importante cache une partie de la macula, et où les hémorragies commencent à cicatriser, leur donnant une couleur beaucoup plus sombre (Figure 4-1.e). Une erreur a également été notée sur une image de fond d'œil centrée sur le disque optique, mais avec une macula bien visible (Figure 4-1.f).

Enfin, une dernière limitation de notre méthode se trouve dans l'évaluation des images pour l'étiquetage qui a été réalisée par un seul expert. Les images de fond d'œil devraient être étiquetées par plusieurs experts afin d'éviter les biais sur la variabilité inter- et intra-expert. Un consensus devra alors être établi pour déterminer la catégorie de l'image. De plus, des avis divergents sur l'étiquetage permettront de mesurer la difficulté de l'image à être évaluée. Dans le cadre de ce projet, la qualité de l'image est également établie par un seul évaluateur. L'évaluation de la qualité pose également une autre problématique du fait qu'elle soit assez subjective. Définir des normes pourrait standardiser l'évaluation de la qualité sur base de critères bien déterminés ou bien utiliser des méthodes automatiques d'évaluation de la qualité.

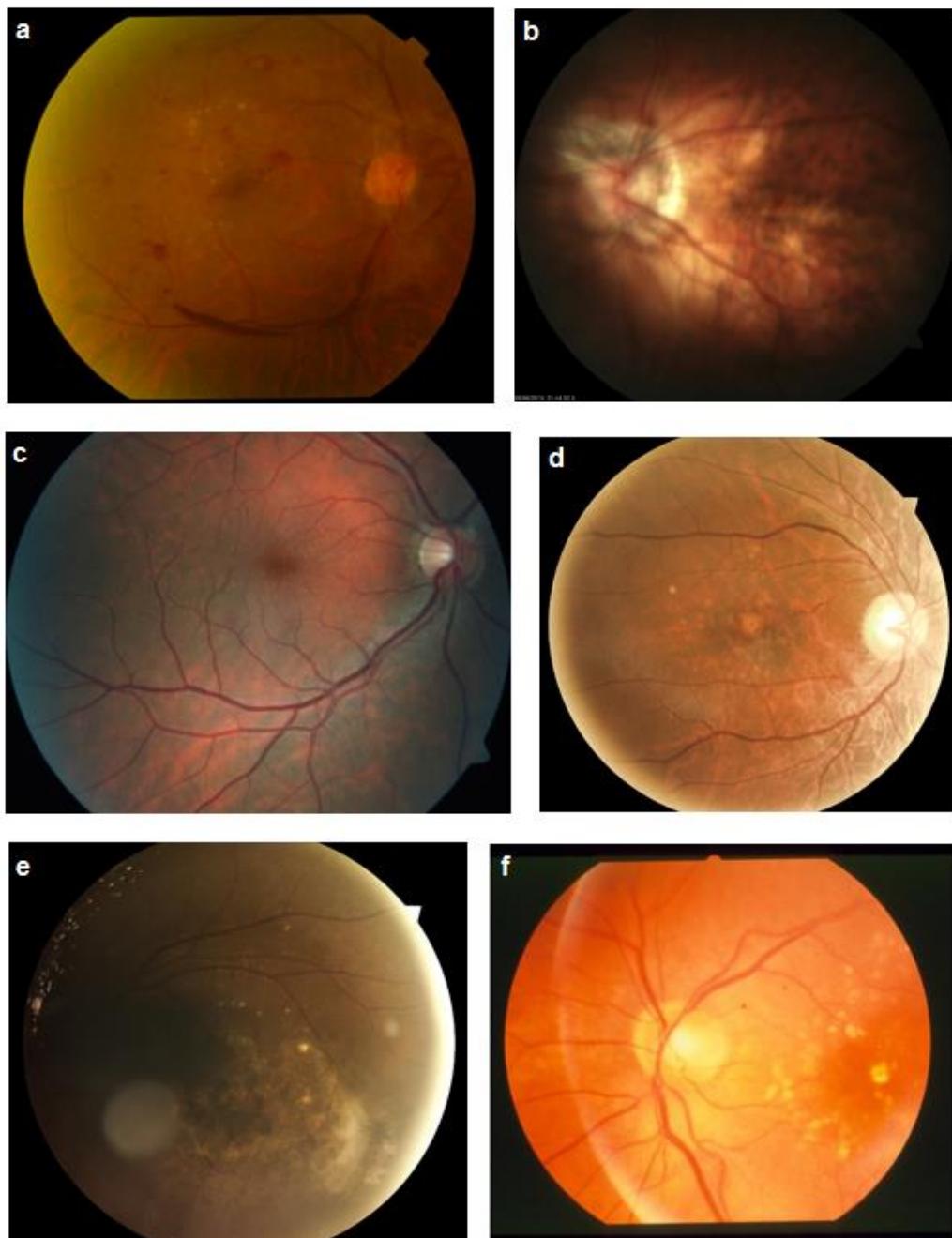


Figure 4-1: Exemples d'images mal classifiées : (a) rétinopathie diabétique détectée comme DMLA, (b) cas de non-DMLA avec du flou et des vaisseaux de la choroïde détecté comme DMLA, (c) cas précoce et (d) cas modéré avec des vaisseaux de choroïde visibles détectés comme non-DMLA, (e) cas avancé avec une ombre et des réflexions lumineuses et (f) cas avancé où l'image est centrée sur le disque optique détectés comme non-DMLA.

CONCLUSION

Une nouvelle méthode de classification de la DMLA à partir d'images de fond d'œil selon le degré de sévérité de la maladie a été développée et validée lors du présent projet. Nous avons travaillé avec une base de données contenant 159 images de fond d'œil représentant les différents degrés de sévérité de la maladie, catégorisées selon la classification simplifiée de l'AREDS, et de qualité très variable, avec une grande proportion d'images de mauvaise qualité. À partir de cela, nous avons étudié différentes caractéristiques. Une évaluation sur les performances de classification de ces caractéristiques de manière individuelle a permis de relever celles qui semblaient avoir le plus de potentiel pour la classification de la DMLA et de fixer leurs paramètres. De cette étude, nous avons relevé que les histogrammes de couleur, les LBP en multi-résolution et les HOG étaient les plus intéressants. L'étape de sélection de caractéristiques a permis une étude sur la combinaison des différentes caractéristiques et de mesurer ainsi leur importance les unes par rapport aux autres. Nous en avons conclu que les caractéristiques les plus discriminantes pour catégoriser la DMLA étaient les LPB en multi-résolution calculés sur le canal vert, ce qui confirme donc notre hypothèse de recherche qui est que les caractéristiques de texture sont suffisamment discriminantes et robustes pour catégoriser la DMLA dans les quatre catégories de l'AREDS. Après cela, nous avons modélisé différents systèmes de classifications de la DMLA, afin d'évaluer comment la méthode classifiait les images de fond d'œil dans les différentes catégories. Nous en avons conclu que les images de non-DMLA et de cas avancés étaient bien classifiées, mais que le système présentait plus de difficultés avec les images de cas précoce et modérée à cause du petit nombre d'échantillons dans ces catégories. Les différents classifieurs testés sont la LDA, la RF et les SVM. Nous avons pu observer que les résultats entre la RF et les SVM étaient similaires et que la force du système provenait donc principalement des caractéristiques. Nous remarquons cependant une légère amélioration avec les SVM, surtout pour des classifications impliquant les catégories de cas précoce et modérés. Les valeurs de précision, de sensibilité et de spécificité pour les différentes classifications obtenues par une approche de validation croisée en 10 partitions étaient très satisfaisantes en regard de la base de données et la qualité des images.

Notre système a donc montré son efficacité et sa robustesse dans la classification de la DMLA et a mieux performé que les méthodes proposées dans la littérature, lorsqu'on les appliquait sur la même base de données. Nous avons ainsi proposé le premier système d'évaluation du degré de

sévérité de DMLA robuste à la qualité de l'image. Cette dernière qualité est très importante dans un contexte de télémédecine où les conditions d'acquisitions sont très variées.

Les limitations relevées se portent principalement sur la base de données, ne contenant pas suffisamment d'échantillons dans certaines catégories pour une validation exhaustive du système proposé. L'étiquetage des images servant de référence a été effectué par un seul expert, ne rendant pas le système complètement non-biaisé aux variabilités inter- et intra-expert. Au niveau du système de classification, on constate des erreurs de classification notamment pour des images trop fortement dégradées, des images avec des vaisseaux visibles de la choroïde et des cas de rétinopathie diabétique.

Une validation plus approfondie avec une base de données contenant plus d'échantillons dans chaque catégorie et évaluée par plusieurs experts permettrait de confirmer les résultats obtenus et de vérifier également si les cas précoce et modérés sont bien identifiés individuellement par rapport aux autres cas.

De futurs travaux peuvent porter sur les différents aspects à améliorer dans le système. Tout d'abord, le prétraitement peut être amélioré avec un recadrage automatique de l'image sur la zone maculaire, et un filtrage sur les trois canaux dans le domaine spectral. Ceci permettrait d'accélérer cette étape et d'améliorer les performances de classification. L'application du prétraitement sur les trois canaux permettrait également de mettre en évidence d'autres caractéristiques se basant sur les canaux rouge et bleu, et ainsi compléter les LBP calculés sur le canal vert dans la discrimination des différentes catégories de la DMLA. D'autres prétraitements peuvent également être rajoutés, notamment pour enlever certains artefacts, comme les réflexions spéculaires ou les ombres importantes sur la macula.

Des travaux pour pouvoir différencier les cas de DMLA des cas de rétinopathie diabétique pourraient également être menés, car ces maladies sont très semblables sur les images de fond d'œil. Ces deux maladies doivent être distinguées car elles ne se traitent pas de la même façon. Une combinaison d'une segmentation de lésions claires et notre méthode pourrait être envisagée.

Avec une base de données plus complète, un système capable de classifier les images de fond d'œil dans les quatre catégories de l'AREDS, de manière séparée, pourrait alors être développé avec la même méthodologie et ainsi pouvoir établir une classification clinique qui soit automatique et robuste. Lorsque ce dernier sera amélioré et validé, le système de classification automatique du

degré de sévérité de la DMLA permettra d'établir un diagnostic fiable et trouvera son utilité dans le suivi complet du patient atteint de DMLA et dans l'évaluation de l'effet des traitements proposés pour guérir la maladie lors d'études cliniques.

BIBLIOGRAPHIE

- Abràmoff, M. D., Garvin, M. K., & Sonka, M. (2010). Retinal imaging and image analysis. *IEEE reviews in biomedical engineering*, 3, 169-208.
- Age-Related Eye Disease Study Research Group (AREDS). (2005). The Age-Related Eye Disease Study severity scale for age-related macular degeneration. *Archives of Ophthalmology*, 123, 1484-1498.
- Agurto, C., Barriga, E., Murray, V., Nemeth, S., Crammer, R., Bauman, W., . . . Soliz, P. (2011). Automatic detection of diabetic retinopathy and age-related macular degeneration in digital fundus images. *Retina*, 52 (8), 5862-5871.
- Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3), 175-185.
- Barakat, M. R., & Madjarov, B. (2004). Automated drusen quantitation for clinical trials. *Investigative Ophthalmology and Visual Science*, 45, 3017.
- Barriga, E. S., Murray, V., Agurto, C., Pattichis, M. S., Russel, S., Abramoff, M. D., & Soliz, P. (2009). Multi-scale AM-FM for lesion phenotyping on age-related macular degeneration. *Computer-Vased Medical Systems, 22nd IEEE International Symposium*, (pp. 1-5).
- Bartling, H., Wanger, P., & Martin, L. (2009). Automated quality evaluation of digital fundus photographs. *Acta ophthalmologica*, 87, 643-647.
- Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3), 346-359.
- Brandon, L., & Hoover, A. (2003). Drusen detection in a retinal image using multi-level analysis. Dans *Medical Image Computing and Computer-Assisted Intervention-MICCAI* (pp. 618-625). Springer Berlin Heidelberg.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123-140.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(2), 5-32.
- Breiman, L. (2002). *Manual on setting up, using and understanding random forests v3.1*. Department University of California Berkeley, CA, USA. Récupéré sur https://www.stat.berkeley.edu/~breiman/Using_random_forests_V3.1.pdf

- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and Regression Trees*. (éd. 1e Édition). Chapman and Hall/CRC.
- Brown, M., & Lowe, D. (2002). Invariant features from interest point groups. *British Machine Vision Conference*, 1, 656-665.
- Brunelli, R. (2009). *Template matching techniques in computer vision: theory and practice*. Wiley.
- Cheng, H. D., Jiang, X. H., Sun, Y., & Wang, J. (2001). Color image segmentation: advances and prospects. *Pattern Recognition*, 34(12), 2259-2281.
- Chiang, A., Witkin, A. J., Regillo, C. D., & Ho, A. C. (2011). Fundus imaging of age-related macular degeneration. Dans A. C. Ho, & C. D. Regillo, *Age-related macular degeneration diagnosis and treatment* (pp. 39-64). Springer.
- Chouaib, H. (2011). Sélection de caractéristiques : méthodes et applications (Thèse de doctorat, Université Paris Descartes, Paris, France).
- Chu, A., Sehgal, C. M., & Greenleaf, J. F. (1990). Use of gray value distribution of run lengths for texture analysis. *Pattern Recognition Letters*, 11(6), 415-419.
- Coleman, H. R. (2011). Modifiable Risk Factors of Age-Related Macular Degeneration. Dans C. A. Ho, & C. D. Regillo, *Age-related macular degeneration diagnosis and treatment* (pp. 15-22).
- Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21-27.
- Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *Computer Vision and Pattern Recognition 2005. Conference on IEEE Computer Society*, 1, pp. 886-893. San Diego.
- Dasarathy, B. V., & Holder, E. B. (1991). Image characterizations based on joint gray level-run length distributions. *Pattern Recognition Letters*, 12(8), 497-502.
- DeHoog, E., & Schwiegerling, J. (2009). Fundus camera systems: a comparative analysis. *Applied optics*, 48(2), 221-228.
- Duda, R. O., Hart, P. E., & Stork, D. G. (2000). Non-parametric techniques. Dans *Pattern Classification* (éd. 2e Edition, pp. 44-47). USA: Wiley-Interscience.

- Duda, R. O., Hart, P. E., & Stork, D. G. (2009). Maximum-likelihood and Bayesian parameters estimation. Dans *Pattern Classification* (éd. 2e Édition, pp. 84-159). USA: Wiley-Interscience.
- Fasih, M., Langlois, J. M., & Cheriet, F. (2014). Retinal image quality assessment using supervised classification (Mémoire de maîtrise, École Polytechnique de Montréal, Montréal, QC).
- Fischbarg, J. (2005). *The biology of the eye* (Vol. 10). Elsevier.
- Gallant, S. I. (1990). Perceptron-based learning algorithms. *IEEE Transactions on Neural Networks*, 1(2), 179-191.
- Galloway, M. M. (1975). Texture analysis using gray level run lengths. *Computer graphics and image processing*, 4(2), 172-179.
- Garnier, M., Hurtut, T., Tahar, H. B., & Cheriet, F. (2014). Automatic multiresolution age-related macular degeneration detection from fundus images. *SPIE Medical Imaging* (pp. 903532-903532). International Society for Optics and Photonics.
- Gu, Q., Li, Z., & Han, J. (2012). Generalized fisher score for feature selection. *arXiv Preprint arXiv*, 1202, 3725.
- Guo, Z., & Zhang, D. (2010). A completed modeling of local binary pattern operator for texture classification. *IEEE Transactions on Image Processing*, 19(6), 1657-1663.
- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, 1157-1182.
- Hanafi, M., Hijazi, A., Coenen, F., & Zheng, Y. (2010). Retinal image classification for the screening of age-related macular degeneration. *SGAI International Conference on Artificial Intelligence*, (pp. 325-338).
- Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3(6), 610-621.
- Haute Autorité de la Santé. (2012). *Dégénérescence maculaire liée à l'âge: prise en charge diagnostique et thérapeutique. Recommandation de bonne pratique*. France.
- Havlicek, J. P. (1996). AM-FM Image Models (Thèse de doctaoral, The University of Texas, Austin, TX).

- Hsu, C. W., & Lin, C. J. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 13(2), 415-425.
- Hsu, C. W., Chang, C. C., & Lin, C. J. (2003). A practical guide to support vector classification.
- Jain, A. K., Duin, R. P., & Mao, J. (2000). Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1), 4-37.
- Jonisch, J., & Shah, G. (2011). Diagnosis of age-related macular degeneration. Dans A. C. Ho, & C. D. Regillo, *Age-related macular degeneration diagnosis and treatment* (pp. 22-38). Springer.
- Kanagasingam, Y., Bhuiyan, A., Abràmoff, M. D., Smith, R. T., Goldschmidt, L., & Wong, T. Y. (2014). Progress on retinal image analysis for age related macular degeneration. *Progress in retinal and eye research*, 38, 20-42.
- Kankanhalli, S., Burlina, P. M., Wolfson, Y., Freund, D. E., & Bressler, N. M. (2013). Automated classification of severity of age-related macular degeneration from fundus photographs. *Investigative Ophthalmology & Visual Science*, 54(3), 1789-1796.
- Kasuga, D. T., Chen, Y., & Zhang, K. (2011). Genetics of Age-Related Degeneration. Dans A. C. Ho, & C. D. Regillo, *Age-related Macular Degeneration Diagnosis and Treatment* (pp. 1-14). Philadelphia: Springer.
- Kohavi, R., & Quinlan, R. (1999). Decision tree discovery. Dans *Handbook of Data Mining and Knowledge Discovery* (pp. 267-276). University Press.
- Kose, C., Sevik, U., Gençalioglu, O., Ikibas, C., & Kayikicioglu, T. (2010). A statiscical segmentation method for measuring age-related macular degeneration in retinal fundus images. *Journal of Medical Systems*, 34, 1-13.
- Laud, K., Mukkamala, S. K., Brue, C., & Slakter, J. S. (2011). The future of non-neovascular age-related macular degeneration. Dans A. C. Ho, & C. D. Regillo, *Age-related macular degeneration diagnosis and treatment* (pp. 119-134). Springer.
- Lebart, L., Morineau, A., & Piron, M. (1997). *Statistique exploratoire multidimensionnelle*. Paris: Dunod.

- Liang, Z., Wong, D. W., Liu, J., Chan, K. L., & Wong, T. Y. (2010). Towards automatic detection of age-related macular degeneration in retinal fundus images. *Proceedings of 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (pp. 4100-4103). Buenos Aires.
- Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R news*, 2(3), 18-22.
- Lowe, D. (1999). Recognition from Local Scale-Invariant Features. *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, 2, pp. 1150-1157.
- Maragos, P., & Bovik, A. C. (1995). Image demodulation using multidimensional energy separation. *Journal of the Optical Society of America A*, 12(9), 1867-1876.
- Marill, T., & Green, D. M. (1963). On the effectiveness of receptors in recognition systems. *IEEE Transactions on Information Theory*, 9, 11-17.
- Meleth , A. D., Raiji, V. R., Krishnadev, N., & Chew, E. Y. (2011). Therapy of nonexudative age-related macular degeneration. Dans A. C. Ho, & C. D. Regillo, *Age-related macular degeneration Diagnosis and treatment* (pp. 65-78). Springer.
- Mora, A. D., Vieira, P. M., Manivannan, A., & Fonseca, J. M. (2011). Automated drusen detection in retinal images using analytical modelling algorithms. *Biomedical Engineering Online*, 10(59), 1-15.
- Morgan, W. H., Cooper, R. L., Constable, I. J., & Eikelboom, R. H. (1994). Automated extraction and quantification of macular drusen from fundal photographs. *Australian and New Zealand Journal of Ophthalmology*, 22(1), 7-12.
- Murray, V., Rodriguez, P., & Pattichis, M. S. (2010). Multiscale AM-FM demodulation and image reconstruction methods with improved accuracy. *IEEE Transactions on Image Processing*, 19(5), 1138-1152.
- Niemeijer, M., van Ginneken, B., Russel, S. R., Suttorp-Schulten, M. S., & Abràmoff, M. D. (2007). Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis. *Investigative ophthalmology & visual science*, 48(5), 2260-2267.

- Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), 971-987.
- Organisation Mondiale de la Santé . (2014). *Maladies Oculaires Prioritaires*. Récupéré sur <http://www.who.int/blindness/causes/priority/fr/index8.html>
- Parvathi, S. S., & Devi, N. (2007). Automatic drusen detection from colour retinal images. *International Conference on Conference on Computational Intelligence and Multimedia Applications* (pp. 377-381). Tamil Nadu, India: INSPEC.
- Penha, F. M., & Rosenfeld, P. J. (2011). Management of neovascular AMD. Dans A. C. Ho, & C. D. Regillo, *Age-related Macular degeneration diagnosis and treatment* (pp. 79-98). Springer.
- Phan, T. V., Seoud, L., & Cheriet, F. (2015). Towards an automatic classification of age-related macular degeneration. *International Conference of Image Analysis and Recognition*, 9164 [in press]. Niagara Falls.
- Porter, R., & Canagarajah, N. (1997). Robust rotation-invariant texture classification: wavelet, Gabor filter and GMRF based schemes. *IEE Proceedings - Vision, Image and Signal Processing*, 144(3), 180-188.
- Priya, R., & Aruna, P. (2011). Automated diagnosis of Age-related macular degeneration from color retinal fundus images. *Electronics Computer Technology (ICECT), 2011 3rd International Conference*, (pp. 227-230).
- Pudil, P., Novovicova, J., & Kittler, J. (1994). Floating search methods in feature selection. *Pattern Recognition Letters*, 15(11), 1119-1125.
- Quellec, G., Russel, S. R., & Abràmoff, M. D. (2011). Optimal filter framework for automated instantaneous detection of lesions in retinal images. *IEEE Transaction on Medical Imaging*, 30(2), 523-533.
- Quinlan, R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81-106.

- Rapantzikos, K., Zervakis, M., & Balas, K. (2003). Detection and segmentation of drusen deposits on human retina: potential in diagnosis of age related macular degeneration. *Medical Image Analysis*, 7(1), 95-108.
- Reed, T. R., & Hans du Buf, J. M. (1993). A review of recent texture segmentation and feature extraction techniques. *CVGIP: Image understanding*, 57(3), 359-372.
- Remeseiro, B., Barreira, N., Calvo, D., Ortega, M., & Penedo, M. G. (2009). Automatic drusen detection from digital retinal images: AMD prevention. Dans *Computed Aided Systems Theory-EUROCAST* (pp. 187-194). Berlin Heidelberg: Springer.
- Santallier, M., Péchereau, J., & Péchereau, A. (2008). *Anatomie pour les écoles d'orthoptie, v1.0.* (A. Péchereau, & J. Péchereau, Éds.) Nantes.
- Scholkopft, B., & Mullert, K. R. (1999). Fisher discriminant analysis with kernels. *Proceedings of the 1999 IEEE Signal Processing Society Workshop Neural Networks for Signal Processing IX*, (pp. 23-25). Madison, WI, USA.
- Smith, R. T., Chan, J. K., Nagasaki, T., Ahmad, U. F., Barbazetto, I., Sparrow, J., & Merriam, J. (2005). Automated detection of macular drusen using geometric background leveling and threshold selection. *Archives in Ophthalmology*, 123(2), 200-206.
- Soliz, P., Wilson, M. P., Nemeth, S. C., & Nguyen, P. (2002). Computer-aided methods for quantitative assessment of longitudinal changes in retinal images presenting with maculopathy. Dans *Medical Imaging* (pp. 159-170). International Society for Optics and Photonics.
- Somol, P., Pudil, P., & Kitler, J. (2004). Fast branch and bound algorithms for optimal feature selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(7), 900-912.
- Steinwart, I., & Christmann, A. (2008). *Support vector machines*. Springer.
- Tang, X. (1998). Texture information in run-length matrices. *IEEE Transactions on Image Processing*, 7(11), 1602-1609.

- Tibshirani, R., Hastie, T., Narasimhan, B., & Chu, G. (2002). Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proceedings of the National Academy of Sciences*, 99(10), 6567-6572.
- van Grinsven, M. J., Chakravarty, A., Sivaswamy, J., Theelen, T., van Ginneken, B., & Sanchez, C. I. (2013). A Bag of Words approach for discriminating between retinal images containing exudates or drusen. *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium*, (pp. 1444-1447).
- van Grinsven, M. J., Lechanteur, Y. T., van de Ven, J. P., van Ginneken, B., Theelen, T., & Sanchez, C. I. (2013). Automatic age-related macular degeneration detection and staging. *SPIE Medical Imaging* (pp. 8600M-86700M). International Society for Optics and Photonics.
- Vedaldi, A. (2006). An Implementation od SIFT Detector and Descriptor. University of California.
- Walter, T., Massin, P., Erginay, A., Ordonez, R., Jeulin, C., & Klein, J. C. (2007). Automatic detection of microaneurysms in color fundus images. *Medical image analysis*, 11(6), 555-566.
- Whitney, A. W. (1971). A direct method of nonparametric measurement selection. *IEEE Transactions on Computer*, 20, 1100-1103.

ANNEXE A – ANATOMIE DE L’ŒIL

Les structures anatomiques de l’œil et de la rétine sont explicitées dans cette section, afin de comprendre le contenu des images de fond d’œil. Un schéma simplifié de l’œil humain est présenté sur la Figure A-1. L’œil est l’organe servant à la vision, via un ensemble de tissus qui permet de capter la lumière, de la convertir et de la transmettre vers le cerveau qui va ensuite traiter l’information visuelle. L’œil se présente sous la forme d’un globe oculaire qui est constitué de différentes couches. En couche externe, se trouve dans la partie antérieure la cornée qui est la partie transparente laissant passer la lumière, et le reste constitue la sclérotique (« blanc de l’œil »). Dans la couche moyenne, la choroïde constituée de vaisseaux sanguins entoure l’œil et permet de l’alimenter en nutriments et en oxygène. Dans la partie antérieure de cette couche, se trouve l’iris qui donne la couleur des yeux et dont le centre est percé par un trou circulaire, la pupille. Le cristallin est un tissu en forme de disque qui se déforme de manière à focaliser la lumière sur la rétine. Enfin, dans la couche interne, se trouvent les tissus de conversion de la lumière au niveau de la rétine. En son centre se trouve la macula, la partie la plus sensible de l’œil, car c’est à cet endroit que la densité des cellules photoréceptrices est la plus grande. Le centre de la macula correspond à la fovea qui est une zone sensible aux couleurs et sert à la précision de la vision. Le disque optique, appelé également « tache aveugle », n’est pas une zone sensible à la lumière mais c’est dans cette zone que les fibres nerveuses se rejoignent pour former le nerf optique (Santallier, Pechereau, & Pechereau, 2008).

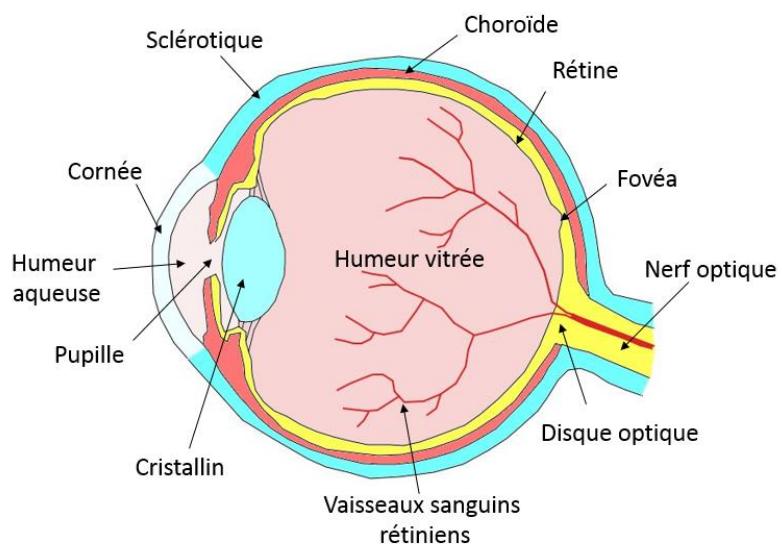


Figure A-1. Anatomie de l’œil humain

La rétine, illustrée sur la Figure A-2, est donc la partie sensible à la lumière et est constituée de différentes couches qui sont présentées dans l'ordre d'activation par la lumière. La lumière est d'abord captée au niveau de l'épithélium pigmentaire, séparée de la choroïde par la membrane de Bruch. L'épithélium pigmentaire joue un rôle de support aux cellules photoréceptrices : stockage et métabolisme de la vitamine A, production de cytokines pour la protection de la cellule et fonction de transport et de barrière. Les cellules de cet épithélium contiennent un pigment de mélanine qui capte la lumière incidente. Ceci conduit à la production de protéines qui enclenchent l'activation des photorécepteurs. Il existe deux types de photorécepteurs : les bâtonnets et les cônes. Les bâtonnets peuvent s'activer à l'aide d'une faible intensité lumineuse et permettent ainsi l'évaluation de la luminosité. Les cônes sont les photorécepteurs qui permettent de voir la couleur. Il en existe trois chez l'homme qui sont sensibles à différentes longueurs d'onde correspondant au rouge, au vert et au bleu. Les photorécepteurs convertissent l'information visuelle en signal neuronal. Ensuite, s'ensuivent un réseau de cellules nerveuses (cellules bipolaires et cellules ganglionnaires) qui sont le point de départ des fibres optiques et qui se rejoignent au nerf optique pour conduire ensuite l'information jusqu'au cerveau (Fischbarg, 2005).

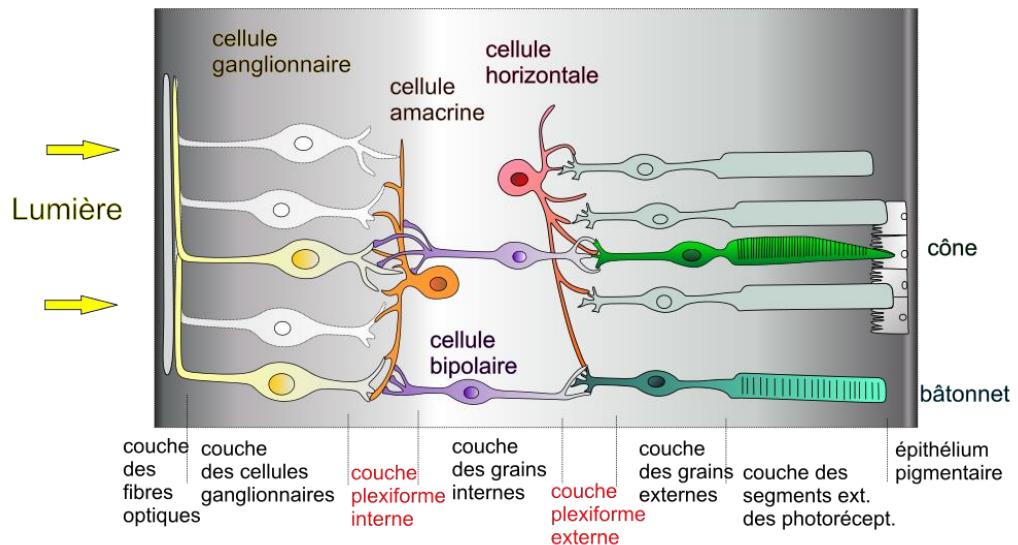


Figure A-2. Schéma simplifié de la rétine (Pancrat, 2011). Tiré de

<http://commons.wikimedia.org/wiki/File:Retina.svg>

ANNEXE B – CARACTÉRISTIQUE DE LA MATRICE DE LONGUEUR DE PASSE EN NIVEAU DE GRIS

Cette annexe présente les différentes caractéristiques qui peuvent être calculées à partir de la GLRLM.

Soit M le nombre de niveaux de gris, N la longueur de ligne maximale, n_r le nombre total de lignes et n_p le nombre de pixels dans l'image, des caractéristiques de la GLRLM peuvent être alors calculées (Galloway, 1975; Chu, Sehgal, & Greenleaf, 1990; Dasarathy & Holder, 1991):

- 1) Short Run Emphasis (SRE) :

$$SRE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{p(i,j)}{j^2} \quad (1)$$

- 2) Long Run Emphasis (LRE)

$$LRE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N p(i,j) j^2 \quad (2)$$

- 3) Gray-Level Nonuniformity (GLN)

$$GLN = \frac{1}{n_r} \sum_{i=1}^M \left(\sum_{j=1}^N p(i,j) \right)^2 \quad (3)$$

- 4) Run Length Nonuniformity (RLN)

$$RLN = \frac{1}{n_r} \sum_{j=1}^N \left(\sum_{i=1}^M p(i,j) \right)^2 \quad (4)$$

- 5) Run Percentage (RP)

$$RP = \frac{n_r}{n_p} \quad (5)$$

- 6) Low Gray-Level Run Emphasis (LGRE)

$$LGRE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{p(i,j)}{i^2} \quad (6)$$

7) High Gray-Level Run Emphasis (HGRE)

$$HGRE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N p(i,j) i^2 \quad (7)$$

8) Short Run Low Gray-Level Emphasis (SRLGE)

$$SRLGE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{p(i,j)}{i^2 j^2} \quad (8)$$

9) Short Run High Gray-Level Emphasis (SRHGE)

$$SRHGE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{p(i,j) i^2}{j^2} \quad (9)$$

10) Long Run Low Gray-Level Emphasis (LRLGE)

$$LRLGE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{p(i,j) j^2}{i^2} \quad (10)$$

11) Long Run High Gray-Level Emphasis (LRHGE)

$$LRHGE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N p(i,j) i^2 j^2 \quad (11)$$

ANNEXE C – ARTICLE DE CONFÉRENCE : TOWARDS AN AUTOMATIC CLINICAL CLASSIFICATION OF AGE-RELATED MACULAR DEGENERATION

Cette annexe présente l'article qui a été soumis à la conférence *International Conference on Image Analysis and Recognition 2015*, soumis en avril 2015. Cet article présente les résultats préliminaires sur la méthode que nous avons proposée. Le but était d'identifier les caractéristiques qui permettraient une classification robuste pour la DMLA. Les résultats ont également été comparés par rapport à la méthode utilisant les LDA.

Towards an Automatic Clinical Classification of Age-related Macular Degeneration.

Thanh Vân Phan¹, Lama Seoud^{1, 2} and Farida Cheriet¹

¹École polytechnique de Montréal, Canada
 {thanh-van.phan, lama.seoud, farida.cheriet}@polymtl.ca
²DIAGNOS inc., Canada

Abstract. Age-related macular degeneration (AMD) is the leading cause of visual deficiency and irreversible blindness for elderly individuals in Western countries. Its screening relies on human analysis of fundus images which often leads to inter- and intra-expert variability. With the aim of developing an automatic grading system for AMD, this paper focuses on identifying the best features for automatic detection of AMD in fundus images. First, different features based on local binary pattern (LBP), run-length matrix, color or gradient information are computed. Then, a feature selection is applied for dimensionality reduction. Finally, a support vector machine is trained to determine the presence or absence of AMD. Experiments were conducted on a dataset of 140 fundus images. A classification performance with an accuracy of 96% is achieved on preprocessed images of macula area using LBP features.

Keywords: Age-related macular degeneration, automatic grading system, fundus photography, support vector machine, texture analysis

1. Introduction

Age-related macular degeneration (AMD) is an eye disease leading to progressive degeneration of the macula. It is the main cause of visual deficiency and irreversible blindness in elderly individuals in Western countries [1]. Although it is asymptomatic in early stages, central vision is gradually lost until legal blindness in advanced stages. Even though there is currently no cure to AMD, treatments for slowing its progression exist and thus, regular eye examination is required.

Fundus photography is a common imaging modality used for eye examination. It is a fast and non-invasive modality that allows direct visualisation of structures of the retina. Based on fundus photography, grading AMD's severity stages, illustrated in Fig., helps in determining specific and optimal treatment. A recommended clinical classification is the simplified AREDS classification [2], dividing AMD cases into four categories: non-AMD, early, moderate and advanced. However, human evaluation of retinal images is time-consuming and leads to inter- and intra-expert variability. To address this problem, automatic grading systems for AMD have been proposed for a faster and reproducible assessment.

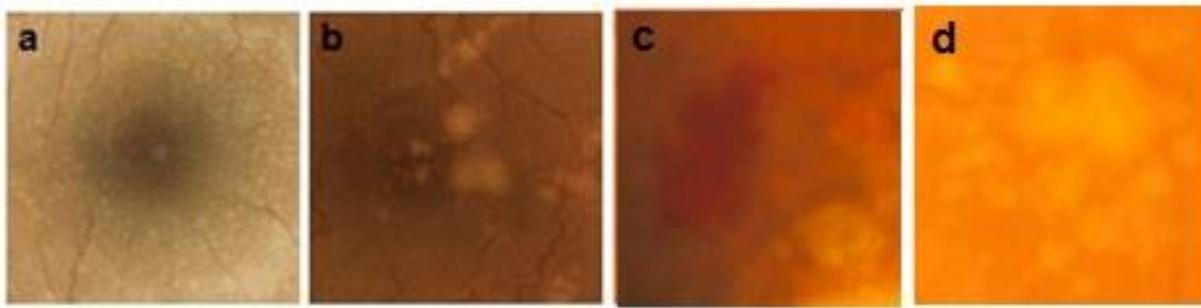


Fig. 1. Macula images of AMD stages: early with hard drusen (a), moderate with soft drusen (b), advanced with hemorrhages (c) and advanced with geographic atrophy (d)

Previous work on AMD focuses mostly on drusen segmentation. Drusen are early signs of AMD and depending on their number, size and position, a severity stage can be determined [3]. However, it is not sufficient for a complete AMD grading because drusen are not present in some advanced cases with large hemorrhages or geographic atrophies, which will then not be detected. Thus, methods with features directly computed on fundus images, such as visual context [4] or texture [5], were proposed to characterize the different forms of AMD. Generally, AMD vs. non-AMD classification is performed for AMD screening. Other binary classifications such as non-AMD vs. moderate cases or non-AMD and early cases vs. moderate and advanced cases are also considered in [4] to highlight moderate cases on which we must pay attention, because the patient still has a good visual acuity but there are high risks to progress to an advanced stage. These methods show a good accuracy (>90%) for binary classifications of good quality images. However, an automatic system performing AMD screening which is sufficiently robust to image quality does not exist.

The aim of this project is to identify the best set of image features that allows a robust AMD classification. The considered features are based on texture, color and gradient information and a support vector machine (SVM) with Gaussian kernel is used for classification.

2. Materials and methods

In this method, a classical preprocessing is first applied to assure robustness to image quality, which is highly variable depending on the acquisition system. Features are then extracted from the preprocessed images and are submitted to a procedure of feature selection to use only the most relevant ones for classification.

2.1. Processing

The fundus images come from different acquisition systems via telemedicine. For a non-biased comparison of these images, a classical preprocessing is first applied. For each color channel of the image, a large median filter with a kernel size of one-fourth the image size is applied in order to estimate the background illumination. The median filtered image is subtracted from the original image color channel, and then, the result is multiplied by 2 for contrast enhancement and the mean of the intensity range was added for the sake of visualization. An image with illumination normalization and contrast enhancement is obtained (Fig. 2c).

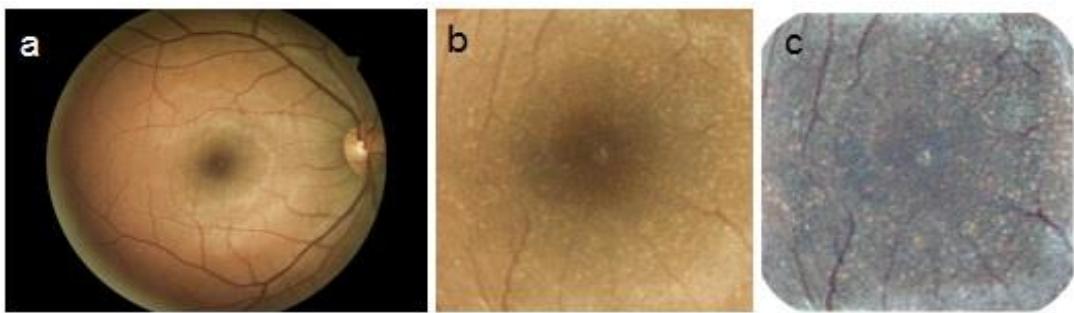


Fig. 2. Image type for each database: retina images (a), macula images (b) and preprocessed macula images (c)

2.2. Feature extraction

This section presents the different features considered for image classification.

2.2.1. Local binary pattern (LBP) features. Multiresolution information is used to analyze the images on different scales. In this study, it is obtained through Lemarié wavelet

transform with four levels of decomposition. For each level, an approximation coefficient containing low resolution information and three detail coefficients containing high resolution information are obtained. With the original image, there are 17 images on which textural information are extracted. LBP is commonly used for texture analysis and shows its efficiency in many classification problems [5]. It consists in measuring the occurrence of local textures primitives, such as corners or edges. To do so, the sign of LBP [6] is computed for each pixel of grey value g_c in a neighborhood of radius R and P neighbors of gray value g_p :

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \quad (0-1)$$

$$\text{With } s(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{Otherwise} \end{cases}$$

In this study, the parameters are empirically set to $R=1$ and $P=4$ or 8 . The magnitude of LBP [7] is also computed from the absolute differences of gray intensity between the central pixel and the neighbors $m_p = |g_p - g_c|$:

$$LBPM_{P,R} = \sum_{p=0}^{P-1} t(m_p, c) 2^p \quad (0-2)$$

$$\text{With } t(x, c) = \begin{cases} 1, & \text{if } x \geq c \\ 0, & \text{Otherwise} \end{cases}$$

The threshold c is set to the image mean value. From the sign and magnitude of LBP, two histograms are computed by measuring the occurrence of the different patterns in the image. The features vector is then constructed by their concatenation. This method is applied separately on each color channel (red, green and blue).

2.2.2. Grey level run-length matrix (GLRLM). GLRLM is another simple method for texture analysis. For a given image and a given direction, the GLRLM $P(i, j)$ measures the number of runs in the image of a grey level i and a run-length j [8]. As there are many zeros in the matrix, it is generally represented by 11 scalars derived from the GLRLM. The features vector is obtained by concatenation of the 11 features computed in the four principal directions (0° , 45° , 90° and 135°). Again, the method is applied separately on each color channel.

2.2.3. Color histograms. Generally, works on fundus images analysis only use green channel because blood vessels and lesions are more visible on it [3, 4, 5]. Because structures

characterizing AMD have different colors, the information from the other color channels must also be considered. In this study, red and blue channels are used with the green one. Once all the color plans are extracted, 4 and 8 bins histograms are computed on each channel. The features vector corresponds to the concatenation of the three color histograms.

2.2.4. Histograms of oriented gradients. HOG [9] is a feature used for edge detection, but can be used as directional information for classification. Indeed, images with lesions should show more gradient in a specific direction. The method consists in calculating the magnitude and the direction of the gradient for each pixel. The gradient images in horizontal G_x and in vertical G_y are obtained by applying 1D point centered derivatives kernels [1 0 -1], on the image. The gradient's magnitude and direction are computed from G_x and G_y . Then, the image is divided in 16x16 cells and histograms of 4 and 8 directions are locally constructed by measuring the number of pixels in a certain direction, weighted by the gradient's magnitude.

2.3. Feature selection

Generally, feature extraction leads to a large number of features compared to the number of samples. To avoid the curse of dimensionality and data overfitting, a sequential floating forward selection (SFFS) [10] is used to find the optimal subset of features. It consists in applying iteratively sequential forward selection (SFS) and sequential backward selection (SBS). SFS begins with an empty model of features subset and successively adds the best feature which, when combined to the previously selected features maximizes a criterion function. In this study, we considered a non-optimized Gaussian kernel SVM classification performance, with parameters set on $C=1$ and $\gamma=1$, on the validation set as the criterion function. SBS starts with all the features computed on the image and consists in deleting the worst feature at each iteration according to the same criterion function. In SFFS, it starts with an empty model. Then, at each iteration, the SFS algorithm is applied l times and the SBS algorithm is applied r times. The l and r parameters are determined by the system itself and thus they are left floating. The process is repeated until the maximum number of iteration is achieved or until no more improvement of the criterion function, whichever comes first. This method tends to find a close to optimal solution.

2.4. Classifier modeling

To determine the presence or absence of AMD, an SVM classifier with Gaussian kernel [11] is chosen because it is efficient for small samples and for more complex separation than a linear classifier. The decision boundary function is computed using the elements x_i of a learning set and their label y_i :

$$H(x) = \sum_{i=1}^l \alpha_i y_i k(x_i, x) + b \quad (0-3)$$

$$\text{With } k(x, y) = \exp(-\gamma \|x - y\|^2)$$

$k(x, y)$ is the Gaussian kernel, l the number of elements in the learning set, α_i and b coefficients from margin optimisation. The classifier is then optimized with γ , the kernel parameter and C , a tolerance parameter for elements to be on the margin. The optimal parameter values are chosen according to the performance assessment using 10 folds cross-validation strategy. In the testing stage, a new element x is classified with a label y depending on its relative position to the decision boundary:

$$y(x) = \text{sign}(H(x)) \quad (0-4)$$

2.5. Validation method

Most of the fundus images used in this project is provided by the telemedicine platform of DIAGNOS Inc. (Canada). Because it has been acquired in a real screening context and using various cameras, the images are highly heterogeneous in terms of quality, resolution, illumination and contrast. To complete this private database, images from public databases such as Automated Retinal Image Analysis (ARIA, United Kingdom) and Structured Analysis of Retina (STARE, United States) are added. These databases contain images with low resolution and low quality, such as blurred images or bad illumination. This preliminary study focuses solely on AMD vs. non-AMD classification. All the images have been labelled by a clinical expert. Overall, the considered dataset is composed of 80 images with different stages of AMD and 60 images without AMD. Another dataset is derived from the first one. It is obtained by manually identifying and segmenting the macula area when visible. This second dataset is thus composed of 76 macula images with AMD and 49 without AMD. A third dataset is established from the same macula images with preprocessing.

The three datasets are divided in a learning set for system modeling and in a testing set for performance assessment. The retinal learning set is composed of 115 images, 65 of which are with AMD with a large proportion of advanced cases and some early and moderate cases, and 50 of which are non-AMD images with healthy and other eye diseases cases. The macula learning sets (with and without preprocessing) are composed each of the same 100 macula images, 63 of which are with AMD, and 37 without AMD. The testing datasets are composed of 15 AMD and 10 non-AMD cases for retina images, 13 AMD and 12 non-AMD for macula images and 13 AMD and 12 non-AMD for preprocessed macula images. These images were selected to represent the different AMD severity stages and the different image quality levels.

3. Results and discussion

For each testing set, the images are represented using each of the features vectors described in Section 2.3. Tables 1, 2 and 3 show the features sets that achieved the best classification performance for each testing dataset. The performance assessment is based on the accuracy (proportion of well classified elements), the sensitivity (proportion of well classified AMD elements) and the specificity (proportion of well classified non-AMD elements). For ends of comparison, the last row corresponds to the results obtained for the method proposed in [5] and applied on the same datasets.

Table 1. Best performance on retina images

Features set	Parameters	Sensitivity	Specificity	Accuracy
LBP4 (blue)	C=10, $\gamma = 1$	93.33%	40%	72%
HOG8	C=10, $\gamma = 0.1$	86.67%	50%	72%
LBP4 (red)	C=10, $\gamma = 1$	86.67%	50%	72%
Garnier [5]	LDA	80%	60%	72%

LBP4 = local binary pattern with 4 neighbors, HOG8 = histograms of oriented gradient with 8 directions, LDA = linear discriminant analysis.

Table 2. Best performance on macula images

Features set	Parameters	Sensitivity	Specificity	Accuracy
LBP8 (blue)	$C=1, \gamma=1$	100%	75%	88%
LBP8 (green)	$C=50, \gamma=1$	92,31%	83,33%	88%
HOG4	$C=100, \gamma=1$	100%	66,67%	84%
Garnier [5]	LDA	100%	66,67%	84%

LBP8 = local binary pattern with 8 neighbors, HOG4 = histograms of oriented gradient with 4 directions, LDA = linear discriminant analysis.

Table 3. Best performance on preprocessed macula images

Features set	Parameters	Sensitivity	Specificity	Accuracy
LBP8 (green)	$C=1, \gamma=1$	100%	91,67%	96%
HOG4	$C=1, \gamma=1$	100%	75%	88%
Color8	$C=1, \gamma=1$	92.31%	83.33%	88%
Garnier [5]	LDA	69,23%	58,33%	62%

LBP8 = local binary pattern with 8 neighbors, Color8: Histograms of color with 8 bins, HOG4 = histograms of oriented gradient with 4 directions, LDA = linear discriminant analysis

In Fig. 3, examples of misclassified images are illustrated. For the retina dataset, the best accuracy achieved is 72%, which is similar to the one reported in [5]. Misclassifications were noted on images with bad quality, in particular with reflections or with non-visible macula. When the analysis is focused only on the macula area, the results are improved with 88% accuracy for the best features (LBP applied on blue and green channels). These features combined with an SVM classifier perform better than the method used in [5]. Here, misclassification were noted on

advanced AMD images with large hemorrhages and on images with reflections. When the features are extracted on preprocessed macula images, the performance increases significantly. Common misclassification with HOG and color features are noted on early AMD images with visible reflections. The best result, with 96% accuracy, is obtained with LBP applied on the green channel of the preprocessed images. In this case, the only misclassified image contains exudates, retinal lesions highly similar to drusen. This result is promising but must be confirmed with further validation.

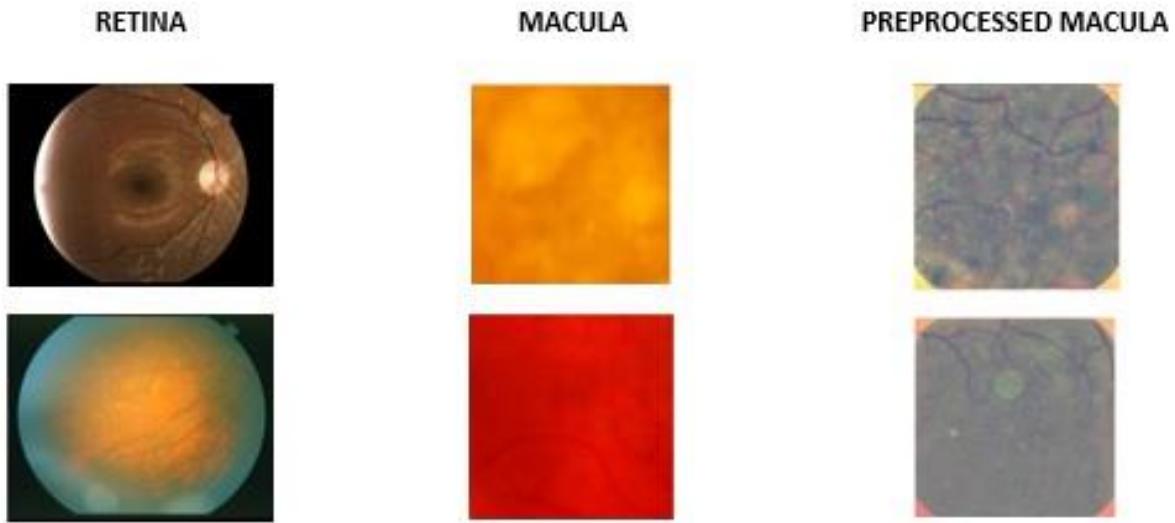


Fig. 3. Examples of misclassification for each testing dataset

The method with LBP features on preprocessed images shows robustness to image quality with a good classification of bad quality images, that the method proposed in [5] could not perform. Some examples are illustrated in Fig. 4.



Fig. 4. Examples of good classification of bad quality images: bad illumination (a), low resolution (b) and visible reflections (c)

4. Conclusion

The proposed automatic classification system for AMD showed promising results, with good classification and robustness to image quality. The conducted experiments demonstrate the effectiveness of features based on texture with a Gaussian kernel SVM for this application. The preliminary results highlights the discriminative strength of a local analysis of fundus images using LBP computed on preprocessed images. The proposed method outperforms previous published work applied on the same dataset. In future work, we plan on evaluating a combination of the features sets used in this study. Moreover, an extensive validation with a more complete dataset will be conducted in order to develop an automatic AMD grading system with the four different stages. Once improved and validated, this system should allow a reliable AMD diagnosis.

Référence

1. Kasuga, D. T., Chen, Y., Zhang K.: Genetics of Age-Related Degeneration. In: Ho, C. A., Regillo C. D. (eds.), *Age-related Macular Degeneration Diagnosis and Treatment*, pp. 1-14 (2011)
2. Age-Related Eye Disease Study Research Group (AREDS): The Age-Related Eye Disease Study Severity Scale for Age-Related Macular Degeneration. *Arch Ophthalmol*, 123, 1484-1498 (2005).
3. van Grinsven, M. J., Lechanteur, Y. T., van de Ven, J. P., van Ginneken, B., Hoyng, C. B., Theelen, T., Sánchez, C. I.: Automatic drusen quantification and risk assessment of age-related macular degeneration on color fundus images. *Investigative ophthalmology & visual science*, 54(4), 3019-3027 (2013)
4. Kankanhalli, S., Burlina, P. M., Wolfson, Y., Freund, D. E., Bressler, N. M.: Automated classification of severity of age-related macular degeneration from fundus photographs. *Investigative ophthalmology & visual science*, 54(3), 1789-1796 (2013)
5. Garnier, M., Hurtut, T., Tahar, H. B., Cheriet, F.: Automatic multiresolution age-related macular degeneration detection from fundus images. In: *SPIE Medical Imaging*, International Society for Optics and Photonics pp. 903532-903532 (2014)
6. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7), 971-987 (2002)
7. Guo, Z., Zhang, D.: A completed modeling of local binary pattern operator for texture classification. *Image Processing, IEEE Transactions on*, 19(6), 1657-1663 (2010)
8. Tang, X.: Texture information in run-length matrices. *Image Processing, IEEE Transactions on*, 7(11), 1602-1609 (1998)

9. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Computer Vision and Pattern Recognition 2005 IEEE Computer Society Conference on , Vol. 1, pp. 886-893 (2005)
10. Pudil, P., Ferri, F. J., Novovicova, J., Kittler, J.: Floating search methods for feature selection with nonmonotonic criterion functions. In: Proceedings of the Twelveth International Conference on Pattern Recognition, IAPR (1994)
11. Steinwart, I., Christmann, A.: Support vector machines. Springer Science & Business Media (2008)