



جامعة أبو بكر بلقايد - تلمسان

Université Abou Bakr Belkaïd de Tlemcen

Faculté de Technologie

Département de Génie Biomédical

MEMOIRE DE PROJET DE FIN D'ETUDES

Pour l'obtention du Diplôme de

MASTER en GENIE BIOMEDICAL

Spécialité : Informatique Biomédicale

Présenté par : BELHOUARI Imane et BENABDELKRIM Fatima

**UN SYSTEME D'AIDE A LA DECISION MEDICALE
POUR LA DETECTION PRECOCE DU GLAUCOME**

Soutenu le 20/09/2017 devant le Jury

Mme.	DALI YUCEF Lamia	<i>MAA</i>	Université de Tlemcen	Présidente
Mr	LAZOUNI Med El Amine	<i>MCB</i>	Université de Tlemcen	Encadreur
Mme	MEKKIOUI Nawel	<i>MAA</i>	Université de Tlemcen	Examinatrice

Année universitaire 2016-2017

Résumé

Résumé :

Le glaucome est une affection chronique potentiellement cécitante qui regroupe un ensemble de maladies neuro dégénératives caractérisées par une neuropathie optique progressive. Le glaucome est le plus souvent une affection asymptomatique, et au moins la moitié des patients qui en sont atteints ne sont pas diagnostiqués. Une détection précoce est indispensable avant l'apparition d'un déficit fonctionnel définitif.

La mise en pratique des méthodes permettent une détection précoce des signes du glaucome qui assurent un meilleur diagnostic de cette maladie.

Le prototype proposée dans ce projet de fin d'étude est d'implémenté cinq différentes techniques de classification en apprentissage supervisé sur une nouvelle base de données : les K-plus proche voisins KNN, les machines à vecteur de support SVM, les réseaux de neurones perceptron multi couche MLP et les arbres de décision (CART, Foret Aléatoire). Dans le cadre de ce travail les algorithmes sont développés et testées sur un ensemble de données (1004 données de 514 patients).

Les résultats expérimentaux montrent que les taux de classification 79.28%, 96.62%, 96.02%, 99.01%, 98.60% respectivement par KNN, SVM, MLP, CART, RF la sensibilité et la spécificité pour les cinq classifieurs 64.22 % et 90.79 %, 96.97% et 97.02%, 96.20% et 96.97%, 98.80% et 99.15%, 97.34% et 99.65%.

Les résultats obtenus sont satisfaisants et présentent la robustesse de nos systèmes proposés.

Mots clés :

Le glaucome, base de données, classification, apprentissage supervisé, KNN, SVM, MLP, Arbre de décision (CART, Forets aléatoire).

ABSTRACT:

Glaucoma is a potentially blinding chronic condition that gathers a group of neurodegenerative diseases characterized by a progressive optic neuropathy. Glaucoma is most often an asymptomatic condition, and at least half of patients concerned are undiagnosed. Early detection is essential before the onset of a definitive functional deficit.

The application of the methods allows the early detection of glaucoma symptoms which ensure a better diagnosis of this disease.

The prototype proposed in the end of study project is to implement five different classification techniques in supervised learning on a new database: k-nearest neighbors algorithm (k-NN), SVM support vector machines, MLP multilayer perceptron neural networks and decision trees (CART, Random Decision Forest). In this work, the algorithms are developed and tested on a set of data (1004 data from 514 patients).

Experimental results demonstrate that the classification rate 79.28%, 96.62%, 96.02%, 99.01%, 98.60% respectively by KNN, SVM, MLP, CART, RF sensitivity and specificity for the five classifiers 64.22% and 90.79%, 96.97 % and 97.02%, 96.20% and 96.97%, 98.80% and 99.15%, 97.34% and 99.65% respectively.

Results obtained are satisfactory and show the robustness of the proposed systems.

Keywords: **Glaucoma**, database, classification, supervised learning, KNN, SVM, MLP, Decision tree (CART, Random Forest).

Résumé

الملخص:

الزرق حالة مزمنة مسببة للعمى، تشمل مجموعة من الأمراض العصبية التي تتميز بالاعتلال التدريجي للعصب البصري، هو حالة ليس لديها اعراض في معظم الاحيان، فعلى الأقل نصف المصابين به حالاتهم غير مشخصة. لذلك الكشف المبكر للزرق ضروري قبل بداية العجز الوظيفي للعصب البصري. وعليه فان اللجوء الي وضع أساليب للكشف المبكر عن الاعراض الطبية للزرق و التي تضمن تشخيص أفضل لهذا المرض.

النموذج المقترح في مذكرة التخرج هاته: هو تطبيق خمسة خوارزميات للتصنيف مختلفة في التعلم الاشرافي على قاعدة بيانات جديدة:

SVM, MLP, KNN و شجرة القرارات (RF, CART) ففي هذا العمل، يتم تطوير الخوارزميات واختبارها على مجموعة بيانات (1004 بيانات من 514 مريض).

وقد اظهرت النتائج أن معدل تصنيف ل KNN , SVM , MLP , CART و RF على التوالي هو، 79.28٪، 96.62٪، 96.02٪، 99.01٪، 98.60٪. النتائج التي تم الحصول عليها مرضية وتظهر دقة الأنظمة المقترحة لدينا.

الكلمات المفتاحية: الزرق، قاعدة البيانات، التصنيف، التعلم الاشرافي، KNN, SVM, MLP, شجرة القرارات (RF, CART)

Remerciements

Avant tous nous remercions DIEU (الله) le tout puissant qui nous a donné aide, pouvoir et patience pour finaliser ce travail.

Un immense remerciement à notre prophète Mohammed que salut de Dieu soit sur lui صلى الله عليه وسلم

Notre haute gratitude, nos profonds respects, nos sincères remerciements et reconnaissance à notre encadreur **Mr LAZOUNI Mohamed El Amine**, professeur à l'université de Tlemcen, pour sa patience, sa compétence, sa grande rigueur scientifique et ses conseils judicieux qui nous ont été très précieux. Qu'il trouve ici l'expression de notre profonde gratitude.

Nous tenons à remercier aussi **Mme. DALI YUCEF Lamia** pour l'intérêt porté à ce travail et d'avoir accepté de présider ce jury.

Ensuite nous désirons adresser nos sincères sentiments à **Mme. MEKKIOUI Nawel** d'avoir examiné ce travail.

Nous adressons également nos vifs remerciements à tous les enseignants et les personnes qui ont contribué de près ou de loin à l'élaboration de ce travail.

Dédicace

DÉDICACE

Je dédie ce modeste travail

A :

Mes parents qui représentent pour moi l'exemple des sacrifices du dévouement, l'honnêteté et qui ont fait de moi ce que je suis devenu et qu'ils trouvent en ce mémoire l'expression de mon éternelle affection avec mon amour infini.

A :

Mes grands-mères, mes grands pères

Mes oncles

Mes tantes

Mes frères IBRAHIM et Mohamed Mustapha

Mes sœurs Khadidja et la petite Ines

Ma grande famille BELHOUARI et HACINI

A mes proches amis :

Ma fidèle et meilleure amie et sœur B. Fatima qui m'a soutenu et encouragé durant ce travail. Je n'oublierai jamais ce que tu as fait pour moi.

Toutes les lettres ne sauraient trouver les mots qu'il faut pour te dire merci mes merveilleuses amies :

L. Asma, T. Sihem, M. Amina, B. Hanane, G. Fatima, T. Fatima pour le courage qui m'a été offert au cours de ce parcours, que Dieu, Allah Le Tout Puissant, protège.

A ma très chère amie B. Chahrazed.

BELHOUARI IMANE

Dédicace

DÉDICACE

Je dédie ce travail à :

Mes parents,

Mes grands-parents,

Toute la famille,

Mes amis

Qu'ils trouvent ici l'expression de toute notre reconnaissance.

BENABDELKRIM FATIMA

Table des matières

Résumé.....	I
Remerciement.....	II
Dédicace.....	III
Table des matières.....	IV
Table des figures.....	V
Liste des tableaux.....	VI
Glossaire.....	VII
Introduction générale.....	1
Chapitre I: Aspect médical.....	3
I. Introduction	4
I.1 La description de l'œil :.....	4
1.1.1 La composition de l'œil :.....	4
II Le glaucome	6
II.1. Les différentes formes cliniques du glaucome	7
III. Facteurs de risques.....	9
IV. Epidémiologie du glaucome.....	9
V. Les examens cliniques	10
V.1. Examen du champ visuel	11
V.2. Examen de la tomographie de cohérence optique (OCT)	12
V.3. L'examen de la Gonioscopie	12
V.4. Examen du fond d'œil (ophtalmoscopie)	13
V.5. Examen de la tonométrie/Pachymétrie	14
VI. Les traitements du glaucome.....	16
VII : Conclusion	16
Chapitre II: Etat de l'art et base de données utilisée.....	17
I. Introduction	18
II. Etat de l'art	18
III. Base de données:.....	20
III.2 Base de données utilisée	21

IV. Conclusion :	28
Chapitre III: Panorama des méthodes de classification utilisées.....	29
I. I Introduction :	30
II II Les méthodes de classification utilisées :	30
II. 1 : Les k plus proches voisins :	31
II. 1 .1 : Algorithme des K-plus proches voisins.....	32
II. 2 : Les machines à vecteur de support.....	33
II. 2. 1 : Principe de fonctionnement général de la technique SVM.....	34
II. 2. 2 : Linéarité et non-linéarité.....	35
II. 3. Les réseaux de neurones	36
II. 3.1.Le neurone formel.....	37
II. 3.2. Algorithme rétro-propagation du gradient.....	39
II. 4. Les arbres de décisions	40
II. 4.1. Algorithmes d'apprentissage par arbre de décision.....	41
II.4.1.1 L'algorithme CART (classification and regression Tree).....	41
III. Les méthodes d'ensemble	42
III.1. Bagging	42
III.2. Bootstrap	42
III.3.Les forêts aléatoires	42
III.3.1 Algorithme foret aléatoire	43
V. Conclusion	44
Chapitre 04 : Résultats obtenus et discussion.....	45
I.Introduction	46
II. Les différents types de la classification	46
III La validation croisée	47
IV. Les paramètres d'évaluation	47
V. Les résultats expérimentaux obtenus	51
VI .Analyse des résultats	55
VI. Conclusion	59
Conclusion générale.....	60

Liste des figures

Liste des Figures

Chapitre01 : Aspect médical

Figure I.1.1	Schéma en coupe de l'œil.....	06
Figure II	Diminution du champ de vision par la maladie du glaucome..	07
Figure II.1	Diagramme de différentes formes du glaucome.....	08
Figure II.1.1	Glaucome selon l'âge.....	09
Figure V	Diagramme des différents examens clinique.....	10
Figure V.1	L'appareil de périmétrie et le résultat de l'exame.....	11
Figure V.2	L'appareil d'OCT et résultat de l'examen.....	12
Figure V.3	L'examen de gonioscopie.....	13
Figure V.4	L'examen du fond d'œil.....	14
Figure V.5.1	Les appareils de la tonométrie	15
Figure V.5.2	L'examen de la pachymétrie.....	15

Chapitre02 : Etat de l'art et base de données utilisée

FigureIII.2.1	Répartition de la BDD selon le sexe.....	22
FigureIII.2.2	Répartition de la BDD selon l'âge.....	22
FigureIII.2.3	Les parties touchées dans l'image CV.....	24
FigureIII.2.4	La répartition de la BDD selon les classes.....	26
FigureIII.2.5	La répartition de la classe 0 selon le sexe	27
FigureIII.2.6	La répartition de la classe 0 (patient malade) selon les catégories...	27

Chapitre03: Les techniques de classification utilisées

Figure II	Les différentes techniques appliquées sur la BDD.....	31
Figure II.1	Représentation des différentes formes pour le calcul de la distance.....	32
Figure II. 2. 1.1	Différents hyperplan possible.....	34
Figure II. 2. 1.2	Hyperplan optimal.....	34
Figure II. 2. 1.3	Choix de l'hyperplan optimal.....	35
Figure. II. 2. 2. 1	Cas linéairement et non linéairement séparable.....	36
Figure. II. 2. 2. 2	Transformation d'espace.....	36
Figure II. 3	Schéma général d'un réseau de neurone.....	37
Figure II. 3.1.1	Schéma d'un neurone formel.....	38
Figure II. 3.1.2	Schéma d'un PMC.....	39
Figure II. 4.1	Exemple d'arbre de décision pour classer les caractères	41

Chapitre 04 : Résultats et discussions

Figure III	La méthode de la validation croisée.....	47
Figure V	Le meilleur arbre donné.....	57

Liste des tableaux

Liste des tableaux

Chapitre01: Aspect médical

Tableau V.3:Description de la gradation gonioscopique	13
-------------------------------------------------------------	----

Chapitre02 : Base de données utilisée

Tableau III.2.1:Les différents descripteurs de la BDD.....	26
Tableau III.2.2:Correction du tonus oculaire selon la pachymétrie.....	26

Chapitre 04 : Résultats et discussions

Tableau IV.I: matrice de confusion.....	48
Tableau IV.II: L'architecture d'un perceptron multicouche appliqué sur la BDD.....	50
Tableau V.I: les différents résultats expérimentaux obtenus....	54
Tableau. V.II: Les paramètres utilisés pour les meilleurs résultats de chaque expérimentation.....	55
Tableau. VI.I: les meilleurs résultats obtenus avec les différentes méthodes de classification.....	56
Tableau. VI.II: Les meilleurs taux de classification pour k=9.....	56
Tableau. VI.III :Les mesures évoquées en utilisant la matrice de confusion.....	57

Glossaire

Glossaire:

AD:	Arbres de Décision
BDD:	Base de Données
CART:	Classification And Regression Trees
C/D:	Cup/Disc
CV:	Champ Visuel
DMLA:	Dégénérescence Maculaire Liée à l'Age
GCC:	Cellules Ganglionnaire
FN:	False Negative
FNR:	False Negative rate
FP:	False Positive
FPR:	False Positive Rate
GPAO:	Glaucome Primaire à Angle Ouvert
HRT :	Heidelberg RetinaTomograph
ID3:	Iterative Dichotomiser 3
IA:	Intelligence Artificielle
KNN :	k-nearest neighbors algorithm
Kppv:	K plus proches voisin
MLP :	Multi Layer Perceptron
OCT:	Optical Coherence Tomographie
OMS:	Organisation Mondiale de la Santé
PIO:	Pression à l'Intérieur de l'Œil
Pos:	le nombre d'éléments positifs dans la base.
PNeg :	le total d'éléments classés négatifs.
PPF:	Point le Plus Fin de la cornée
PPos :	le total d'éléments classés positifs
RN:	Réseau de Neurone
RF:	Random Forests

Glossaire

ROI:	Region of Interest
SD-OCT:	la tomographie par cohérence optique dans le domaine spectral
Se:	Sensibilité
Sp:	Spécificité
SVM:	Séparateurs à Vaste Marge(Support Vector Machine)
TC :	Taux de Classification
TN:	True negative
TNR:	True negative rate
TO:	Tension oculaire
TP:	True positive
TPR:	True positive rate

Introduction générale

Introduction générale :

Les yeux, avec lesquels nous voyons le monde, sont des organes sensoriels les plus complexes, délicats et sensibles de notre corps. Ils reçoivent 80% des informations extérieures. Leur rôle principal est uniquement de voir, puis transmettre ces informations au cerveau via le nerf optique. L'œil est susceptible par plusieurs maladies. Ce qui nous intéresse dans notre travail est la détection d'une maladie oculaire qui est la maladie du glaucome.

Le mot "glaucome" est le nom donné à un groupe de maladies de l'œil qui endommagent progressivement le nerf optique. Le glaucome définit par le voleur silencieux de la vision est la première cause de perte de vue irréversible et la deuxième cause de cécité dans le monde. Il existe trois types de cette maladie qui sont : le glaucome primitif c'est-à-dire débutant, secondaire c'est-à-dire avancé, et congénital c'est-à-dire à la naissance. Dans chaque type il existe deux différents types (glaucome à angle ouvert et à angle fermé). Dans le chapitre suivant nous allons voir en détail tout ce qui concerne la maladie du glaucome.

Il est nécessaire de détecter le glaucome à son premier stade pour un traitement qui préserve la vision du patient affecté. Pour cela le médecin ophtalmologue réalise plusieurs examens ophtalmologiques et différentes mesures (l'acuité visuelle, tonus oculaire et des examens depuis la cornée jusqu'à le nerf optique) afin de détecter cette maladie d'une manière précoce.

Dans ce travail, nous proposons un système d'aide à la décision médicale pour le médecin ophtalmologue dans le but de détecter cette maladie d'une manière précoce. Pour cette raison, plusieurs techniques d'intelligence artificielle ont été testées sur une nouvelle base de données collectées par Mr LAZOUNI Med El Amine en collaboration avec plusieurs médecins ophtalmologue dans le but d'un travail de recherche.

Ces techniques sont : les K plus proche voisin, les machines à vecteur de support, les perceptrons multicouches, les arbres de décision (CART et les forêts aléatoires).

Ce projet de fin d'étude se compose de quatre grandes parties. Dans la première partie, nous allons présenter un aspect médical contenant l'anatomie de l'œil, la pathologie du glaucome avec ses différents types, ses symptômes, ses facteurs de risque, les statistiques de cette maladie, les différents examens cliniques pour la détection de cette

Introduction générale

pathologie ainsi que les différents traitements possible pour arrêter l'évolution du glaucome. La deuxième partie se penchera sur une revue de la littérature des travaux sur cette maladie ainsi qu'une description détaillée de la base de données du glaucome utilisée dans ce travail. La troisième partie consacrée pour présenter les différentes techniques d'intelligence artificielle utilisées dans ce projet K-NN, SVM, MLP, CART, Foret aléatoire. Dans la dernière partie nous allons présenter nos résultats expérimentaux avec une analyse détaillée, des conclusions et les perspectives. Nous terminerons ce mémoire par une conclusion générale.

Chapitre 01 :

Aspect médical

Introduction :

Le glaucome est le nom d'une maladie oculaires dans lequel le nerf optique à l'arrière de l'œil et qui assure la liaison entre le cerveau et l'œil est lentement détruit. Dans la plupart des cas cette maladie est due à une pression intraoculaire c'est-à-dire à l'intérieur de l'œil élevé. Le glaucome est l'une des principales causes de cécité dans le monde.

Ce premier chapitre consiste la présentation de l'aspect médical de cette maladie. Il nous semble évident de présenter cette partie dans le but de mieux comprendre la maladie du glaucome. En premier lieu, on présente l'anatomie de l'œil, les différentes pathologies qui lui atteignent et plus précisément le glaucome qui fait l'objet de notre étude. On présente aussi ces types, ces symptômes ainsi que ces facteurs de risque. Nous présentons après l'épidémiologie de cette maladie et les différents examens cliniques pratiqués par les médecins ophtalmologues pour la détection de cette maladie. Nous terminons ce chapitre par une conclusion.

I.1 : La description de l'œil :

L'œil est un organe sensoriel le plus complexe, délicat et fragile du notre corps humain, empli de liquide et protégé dans de profondes cavités du crâne osseux. Il est de faible volume, il pèse 7 grammes, de forme d'une sphère d'environ 24 mm de diamètre avec à l'avant de l'œil une demi-sphère la cornée de 8 mm de rayon. Chaque œil est constitué de 3 tuniques (sclérotique, choroïde, rétine) [1].

I.1.1 : La composition de l'œil :

On discerne une première tunique, la plus externe de l'œil.

La cornée : est une membrane transparente de 11 mm de diamètre qui permet la réfraction des rayons lumineux vers l'intérieur de l'œil. Elle est privée des vaisseaux sanguins. [1]

La figure (Figure I.1.1) [4] ci-dessous représente une coupe des différents composants de l'œil.

La sclérotique : (ou sclère, blanc de l'œil). De 1 à 2 mm d'épaisseur cette dernière est une membrane blanche très résistante et opaque. La sclérotique contient de nombreux vaisseaux sanguins. [1]

Chapitre 01 : Aspect médical

Ensuite se trouve **la tunique uvéale**, ou uvée qui est constituée de l'iris, la choroïde et le corps ciliaire.

L'iris : est une membrane circulaire, donne à l'œil sa couleur. Il est situé dans l'humeur aqueuse entre la cornée et le cristallin [1].

Le trabéculum : Tissu translucide formé de fibres enchevêtrées qui filtrent le liquide intraoculaire protègent et nourrissent le cristallin et la cornée (l'humeur aqueuse) lorsqu'il s'écoule hors de l'œil. Si ce filtre est altéré et que le liquide ne peut s'écouler correctement, la pression intraoculaire augmente, provoquant un glaucome [2].

La pupille : qui forme une tâche noire au centre de l'iris, à travers laquelle la lumière gagne le fond de l'œil. Une forte lumière elle se rétrécit, réduisant la quantité de lumière qui entre dans l'œil, tandis qu'en faible lumière elle s'agrandit pour admettre le plus de lumière possible [2].

La choroïde : située entre la sclérotique et la rétine, est une couche vasculaire qui assure la nutrition de l'iris et des cellules de la vision de la rétine. De plus, elle absorbe les rayons lumineux inutiles à la vision [1,2].

Le corps ciliaire : est un anneau de tissu musculaire derrière l'iris qui produit l'humeur aqueuse [1,2].

La tunique la plus interne de l'œil est **la tunique nerveuse**. Elle se compose de :

La rétine : c'est la couche sensible à la lumière grâce aux bâtonnes et les cônes et elle qui traduit les rayons lumineux en influx nerveux.

Elle possède deux types de photorécepteurs :

La macula : région centrale de la rétine, située proche de l'axe optique.

La Fovéa : région centrale de la macula où se concentrent les cônes. Cette région forme une petite dépression au centre de la rétine, où l'acuité visuelle est à son maximum. Les champs récepteurs des cônes se trouvent au centre du champ visuel. [3]

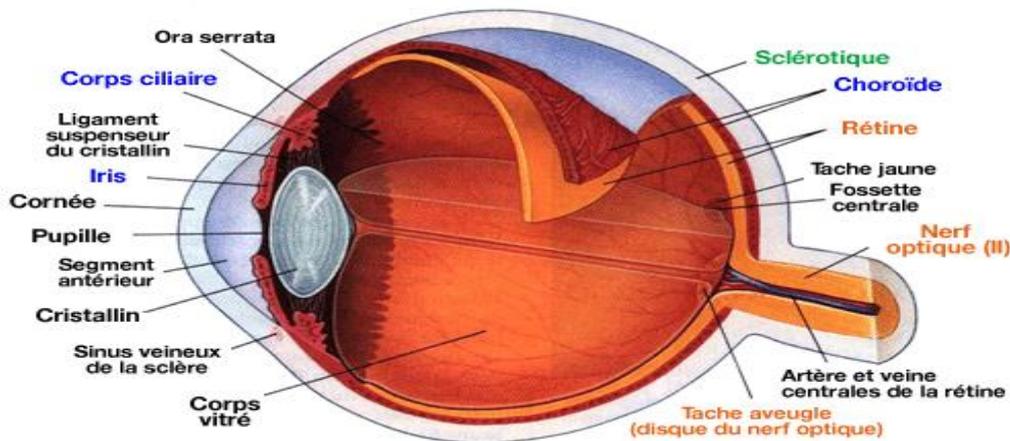
Le nerf optique : est le nerf reliant chaque globe oculaire, plus précisément au niveau de la rétine au niveau des cellules ganglionnaires jusqu'au cerveau. Le nerf optique se charge de transférer les informations rétiniennes obtenues par l'œil au cerveau qui les traitera par la suite car la vision est un ensemble constitué de deux yeux reliés à une partie du cerveau. Il enregistre, interprète et traduit en images. Il mesure 4 mm de diamètre et 5 cm de long. Cette partie de l'œil est la zone d'intérêt, car elle est la plus touchée par la neuropathie du glaucome. [3]

Ces trois tuniques cernent trois milieux transparents : l'humeur aqueuse, le cristallin et l'humeur vitrée.

L'humeur aqueuse située entre la cornée et le cristallin, est un liquide transparent qui contrôle la pression de l'œil. Ce liquide se renouvelle en 3h et permet de protéger le cristallin et de nourrir la cornée. L'humeur aqueuse est composée de 99,6% d'eau [1,2]

Le cristallin est une lentille transparente biconvexe situé en arrière de l'iris qui rassemble les rayons lumineux sur la rétine [1, 2].

L'humeur vitrée ou corps vitré est une gelée transparente qui assure la rigidité du globe oculaire et maintient la rétine contre la paroi de l'œil. Elle représente 90 % du volume oculaire et est formée de 95 % d'eau [1, 2].



(Figure I.1.1) : Schéma en coupe de l'œil

II. Le glaucome :

Il existe plusieurs anomalies associées à l'œil telle que la cataracte, la Dégénérescence Maculaire Liée à l'Age (DMLA), la rétinopathie diabétique, la myopie aiguë et le glaucome...etc.

Ces maladies provoquent une perturbation visuelle, une réduction de la vue et aussi peut mener à la cécité. Dans la partie qui suit nous présenterons la maladie du glaucome qui fait l'objet de ce mémoire de master, ces types, ces symptômes, ces facteurs de risque, l'épidémiologie de cette maladie, les différents examens ainsi que les traitements de cette maladie.

Le glaucome est une maladie oculaire caractérisée par une dégénérescence et une excavation progressive du nerf optique qui sera lentement détruit. Il est attribuable à un mauvais écoulement du liquide intraoculaire. L'accumulation de ce liquide fait

augmenter la Pression à l'Intérieur de l'Œil (PIO), ce qui peut avoir des effets néfastes sur la vue, cette neuropathie s'associe à une perte graduelle et irréversible des champs visuels périphériques reliée à la sévérité de l'atteinte des tissus. Selon Eduard Jaeger, le glaucome est un voleur silencieux car le patient perd sa vision sans la sentir au début et non récupérable après comme illustrée dans la (Figure II). (Eduard Jaeger 1854) [5]. Le glaucome est l'une des principales causes de cécité dans le monde. [6 ,7]

Le glaucome provoque plusieurs difficultés aux personnes atteintes cette maladie mais les principales difficultés sont le déplacement à cause de la réduction du champ visuel, ne peuvent plus se diriger toutes seules aussi le besoin d'un éclairage suffisant, mais certaines personnes peuvent gardées des aptitudes comme la lecture.

La figure ci-dessous [8] représente les différentes étapes de la diminution du champ de vision causé par le glaucome. A : une vision normale, B : perte légère à modérée, C. perte visuelle sévère cause de l'absence du traitement.

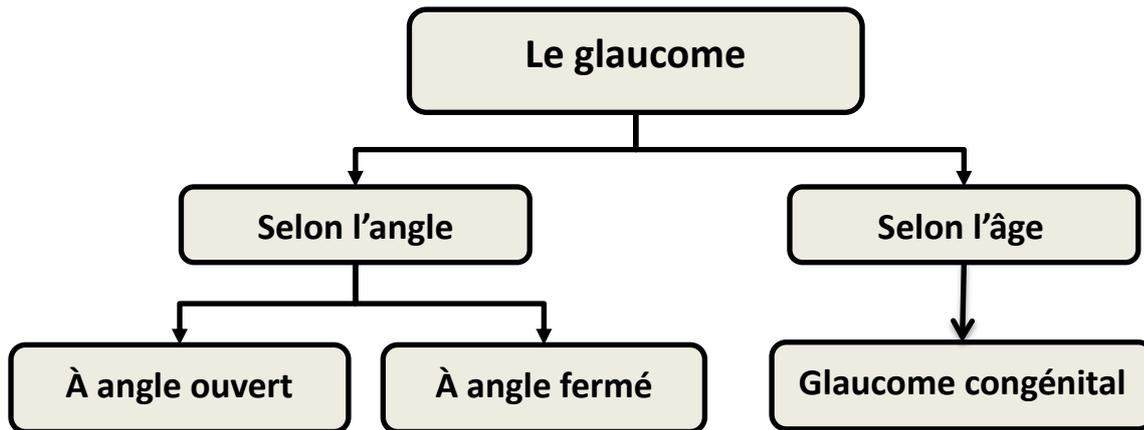


(Figure II) : Diminution du champ de vision par la maladie du glaucome

II.1. Les différentes formes cliniques du glaucome :

Le glaucome est une neuropathie optique progressive dont a plusieurs formes ou types.

Le diagramme ci-dessous présente ces différentes formes :



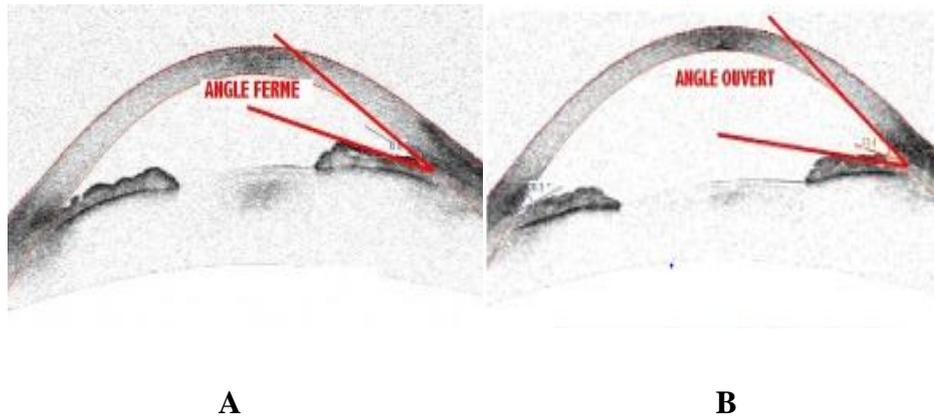
(Figure II.1) : Diagramme de différentes formes du glaucome.

Lorsque cette maladie touche les nouveaux nés, on parle d'un glaucome congénital. Ce type de glaucome est due à un défaut de développement dans l'angle de l'œil, ou qui présentent un manque de maturité dans le développement de l'angle de l'œil, un glaucome peut apparaître dans les premiers jours de vie ou plus tardivement.

Lorsque cette maladie touche les adultes. Deux différents types de glaucome sont possibles : le glaucome a angle ouvert ou le glaucome a angle fermé.

- **Le glaucome à angle ouvert** : dans ce type le trabéculum devient de moins en moins perméable et le liquide intraoculaire s'écoule trop lentement. Par conséquent, l'humeur aqueuse s'accumule dans l'œil et la pression augmente graduellement, ce qui détériore le nerf optique. Ce type est plus fréquent à 80% et il est indolore et son évolution lente [9]. la figure (Figure II.1.1.A) ci-dessous représente l'angle ouvert.
- **Le glaucome à angle fermé** : dans ce type le trabéculum s'obstrue complètement parce que l'iris bombe en avant .et alors l'élévation de la PIO est brusque et parfois douloureuse et peut détériorer le nerf optique rapidement .il est nécessaire une intervention médicale urgente. On peut voir comme signes : forte douleur oculaire, vision brutalement floue, vision de halos colorés autour des sources lumineuses, rougissement des yeux, nausées ou vomissements. On peut voir comme signes : forte douleur oculaire, vision brutalement floue, vision de halos colorés autour des sources lumineuses, rougissement des yeux, nausées ou vomissements. [9]

La figure (Figure II.1.1): ci-dessous représente l'angle irido-cornéen d'une manière fermée à gauche de la figure et ouverte à droite de la figure.



(Figure II.1.1): Glaucome selon l'angle

III. Facteurs de risques :

Le glaucome est une maladie de l'œil qui perturbe le bon fonctionnement du nerf optique qui peut toucher des gens de toutes races et de tous sexes. Cette maladie peut se déclarer chez quiconque, mais certaines personnes sont plus à risques que d'autres. [12]

- Pression intraoculaire anormalement élevée plus de 21 mmHg.
- Âgé de plus de 40 ans.
- Antécédents familiaux de glaucome.
- Descendance africaine, asiatique ou hispanique.
- Diabétique.
- Blessure oculaire antérieure.

IV. Epidémiologie du glaucome :

Selon l'Organisation Mondiale de la Santé (OMS), le glaucome est la deuxième cause de la cécité dans les pays développés, toutes personnes sont susceptibles par cette neuropathie, sa fréquence augmente avec les années. [13]

Le glaucome touche 1 à 2% de la population de plus de 40 ans et environ 10% après 70 ans.

Chapitre 01 : Aspect médical

Au niveau international environ 70 million de personnes sont affectées et plus de 6,6 millions de personnes ont perdu la vue des suites de cette maladie et en 2020, 80 millions de personnes à travers le monde seront atteintes de glaucome [12]

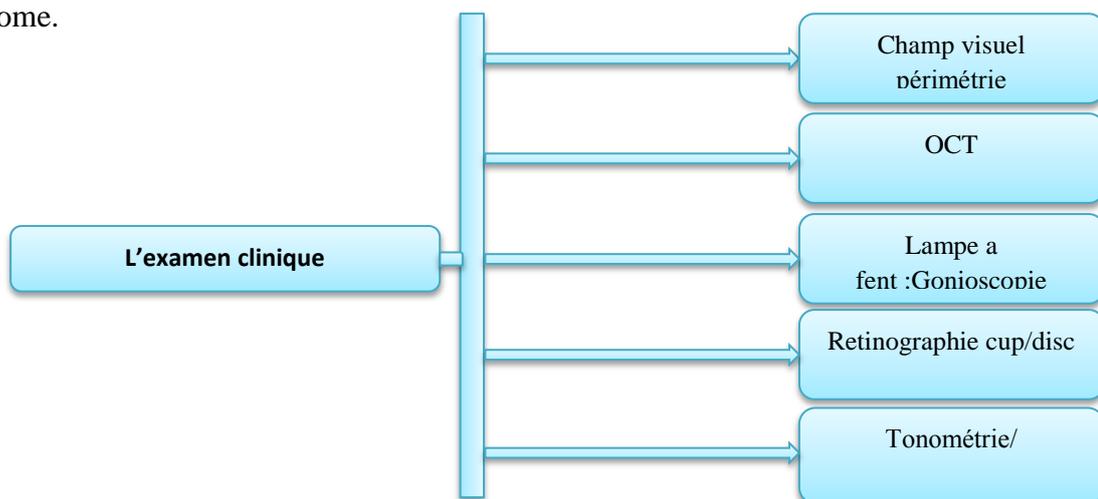
Au niveau national le glaucome est la deuxième cause de la cécité et se situe à hauteur de 4.6% de la population à risque ainsi que 500 000 algériens sont affectés et particulièrement sur la population de la région du sud du pays.

Selon une étude récente à Oued Souf 34% des résidents de cette ville sont glaucomeux. L'incidence de glaucome est :

- 3,4% dans les villes de l'Ouest.
- 3,9% à l'Est, 5,5% au Centre.
- Atteint le pic de 8% dans le sud de l'Algérie [15].

V. Les examens cliniques:

Le glaucome est une maladie sournoise et le sujet qui en est atteint ne sent rien, pouvant perdre progressivement la vue. Seuls l'examen clinique fait par l'ophtalmologiste permet de suspecter et parfois d'affirmer le diagnostic de glaucome. Il permet aussi de préciser le type. Le diagramme ci-dessous présente les examens cliniques pour dépister le glaucome.



(Figure V) diagramme des différents examens clinique.

V.1. Examen du champ visuel :

Le champ visuel est la partie de l'espace qu'un œil peut percevoir autour du point qu'il fixe sans le bougé. Cet examen se mesure pour chaque œil individuellement au moyen d'un appareil appelé « périmètre ».

L'examen du champ visuel se fait dans une pièce peu éclairée. L'appareil devant lequel le patient assis a la forme d'une coupole. Un œil est caché, l'autre fixe un point lumineux au centre d'une coupole. Des points de luminosité variable sont alors présentés aux différents endroits de la coupole. Si le patient perçoit le point sur le côté (tout en gardant sa fixation sur le point central), il appuie sur un bouton. L'appareil cherche à déterminer pour chaque point la luminosité la plus faible que le patient peut encore percevoir. Les résultats sont ensuite analysés par ordinateur. [16]

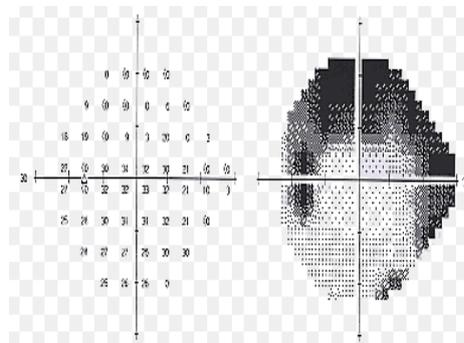
Le résultat est présenté sous forme de plusieurs cartes. Chaque carte est présentée sous forme chiffrée, et à côté, sous forme de carrés plus ou moins foncés.

La figure (Figure V.1) [17] ci-dessous montre l'appareil de la périmétrie et les résultats de l'examen du champ visuel, tel que : A. l'appareil du champ. B. résultat d'examen champ visuel.

Les carrés blancs : indique que les cellules saines, **Les carrés gris :** indique que les cellules affectées ou malade, **Les carrés noirs :** indique que les cellules mortes. L'examen du champ visuel reste l'examen le plus fiable pour la détection et le suivi de la maladie du glaucome. Son inconvénient est que la détection n'est pas précoce.



A



B

(Figure V.1) L'appareil de périmétrie et le résultat de l'examen.

V.2. Examen de la tomographie de cohérence optique (OCT) :

La Tomographie par Cohérence Optique (OCT) est un examen permettant d'obtenir des coupes anatomiques fines par l'application d'un balayage d'une lumière laser sur la structure à examiner. Il permet alors d'avoir des coupes très précises des différentes structures de l'œil, de la cornée à la rétine et au nerf optique. L'examen n'est pas douloureux car sans contact avec l'œil, et il n'est pas éblouissant. Le patient devra uniquement suivre une cible de fixation

L'examen OCT permet d'analyser l'angle entre la cornée et le cristallin, d'évaluer l'épaisseur des fibres nerveuses, et de mesurer la taille du nerf optique. [18]

La figure (Figure V.2) [19] qui suit présente l'appareil d'OCT. Tel que :

A. Appareil d'OCT, B. Examen automatisé de la tête du nerf optique

La couleur vert indique les cellules saines, le jaune affectées, le rouge mortes.



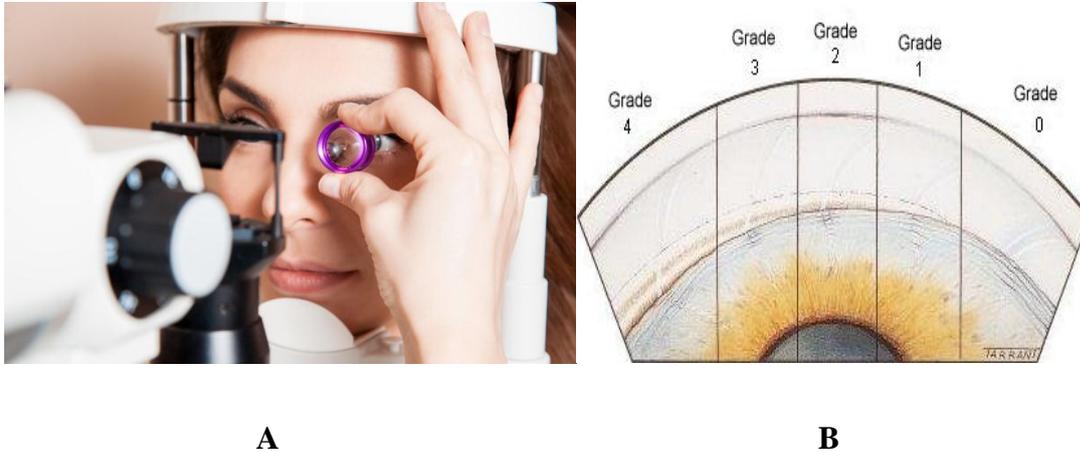
A **B**
(Figure V.2) L'appareil d'OCT et résultat de l'examen.

V.3. L'examen de la Gonioscopie :

La gonioscopie n'est rien d'autre que l'examen de l'angle de l'œil. Celui-ci n'est pas visible sans moyens auxiliaires. Après avoir anesthésié la cornée avec des gouttes, on se sert d'une loupe spéciale que l'on place à la surface de l'œil, cette loupe s'appelle une lentille de gonioscopie. L'ophtalmologue s'intéresse à la largeur de l'angle, à d'éventuelles adhérences entre l'iris et la cornée, à des néo-vaisseaux et à des dépôts pigmentaires abondants. [16]

Chapitre 01 : Aspect médical

Cet examen permet de mesurer l'angle irido cornéen. La figure [20] et le tableau qui suivent nous montre comment le médecin ophtalmologue interprète son examen de gonioscopie.



(Figure V.3) L'examen de la gonioscopie.

Grade	Angle	Description
0	0°	Fermé
<u>1</u>	< 10°	très étroit
<u>2</u>	10-20°	Etroit
<u>3</u>	20-35°	Ouvert
<u>4</u>	35-45°	grand ouvert

Tableau V.3. Description de la gradation gonioscopique

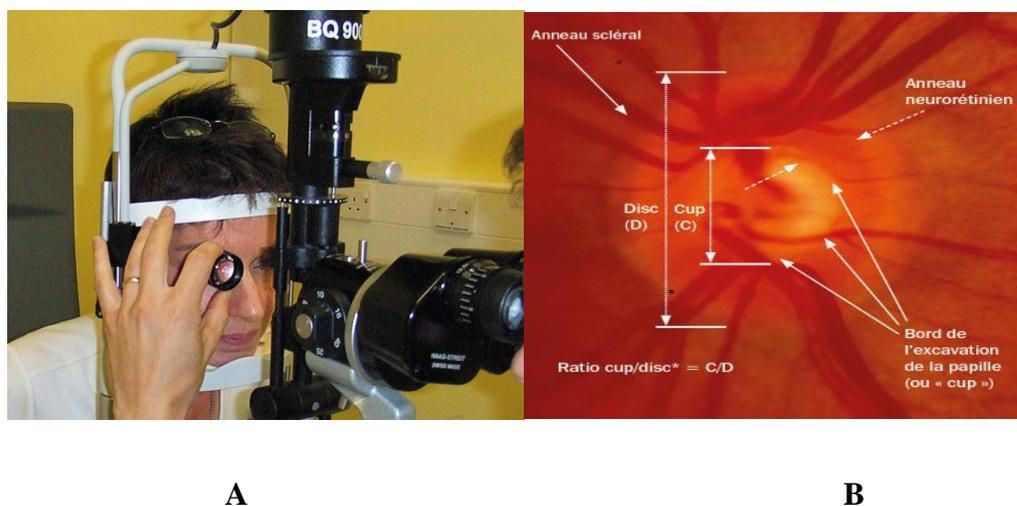
V.4. Examen du fond d'œil (ophtalmoscopie) :

Le fond d'œil est un examen ophtalmologique destiné à étudier les structures de l'œil à l'arrière du cristallin et plus particulièrement la rétine. Pour le dépistage et la surveillance d'une certaine atteinte oculaire. Il s'agit d'un examen indolore, réalisé par un ophtalmologiste avec des instruments d'optique tel qu'une grosse loupe, qu'il maintient devant l'œil. Il examine tout particulièrement le nerf optique, aussi l'état de la papille qu'est évalué selon son apparence, sa forme et d'autres paramètres. La comparaison faite toujours avec les papilles des 2 yeux entre elles. [16]

Il sert à diagnostiquer ou surveiller certaines maladies proprement oculaires comme la DMLA, le décollement de la rétine, le glaucome, etc. [21]. Il permet également de vérifier s'il y existe des atteintes de la circulation sanguine comme une hypertension artérielle ou un diabète. [21]

La figure (FigureV.4) [22] suivante présente l'examen de fond l'œil, tel que

A .L'ophtalmoscopie, B : fond de l'œil (la rétine)



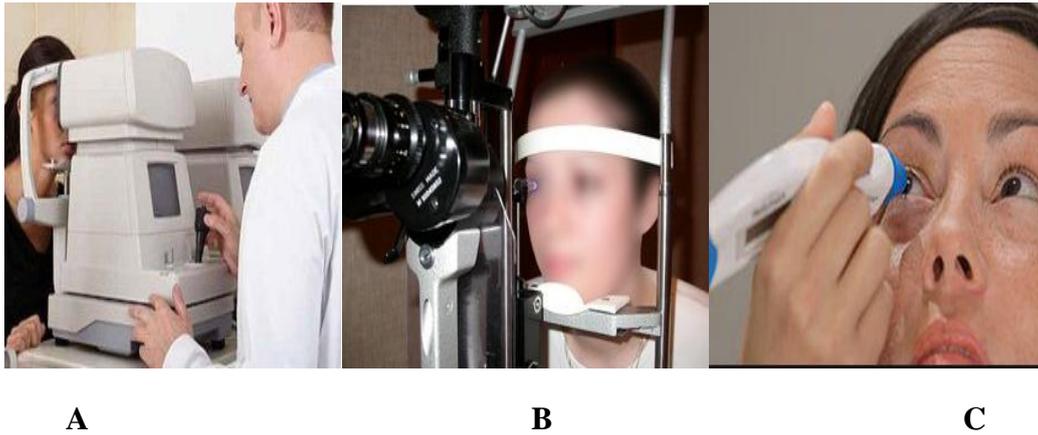
(FigureV.4) L'examen du fond d'œil.

V.5. Examen de la tonométrie/Pachymétrie :

- **La tonométrie :** C'est un examen oculaire permettant de détecter les changements de pression intervenant dans l'œil longtemps avant que des symptômes se manifestent. La méthode de tonométrie la plus répandue est la tonométrie à aplanation de Goldmann, qui a été considérée pendant des décennies comme le test de référence sur le plan international pour mesurer la pression intraoculaire (PIO).mais il existe d'autres types de la tonométrie à air et Tono Pen.[23]

La figure (Figure V.5.1) [24] suivante présente les appareils de La tonométrie

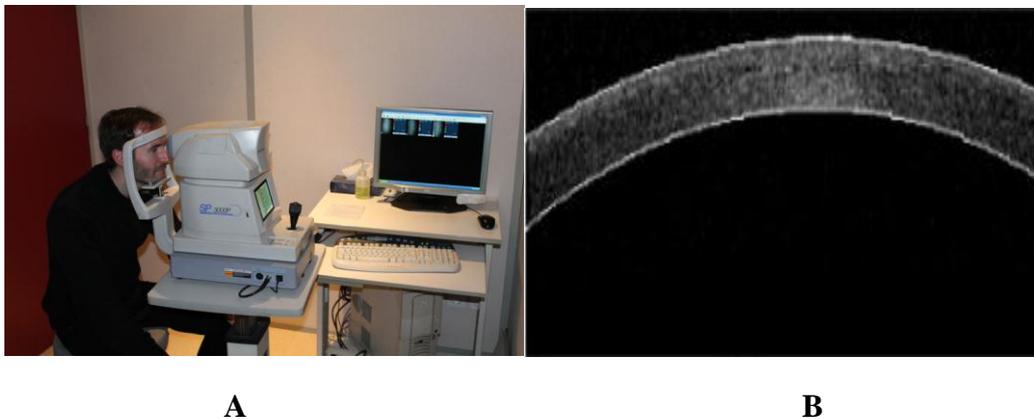
A. à aplanation, B. Tonomètre à air pulsé C. le Tono Pen



(Figure V.5.1) Les appareils de La tonométrie.

- **La pachymétrie :** C'est un examen pour la mesure de l'épaisseur de la cornée. Cette mesure est importante pour le diagnostic du glaucome. L'examen de la mesure de l'épaisseur cornéenne peut s'effectuer avec plusieurs types d'appareils :

La figure ci-dessous [26] présente A. L'examen de pachymétrie, B. le résultat de la pachymétrie. .



(Figure V.5.2) L'examen de la pachymétrie.

Il existe une relation entre la tonométrie et le pachymétrie, pour confirmer existence du glaucome ; on a

Tension oculaire (TO) réelle= TO mesuré + correction (des valeurs fixe pour chaque valeur du pachymétrie).

VI. Les traitements du glaucome :

Chapitre 01 : Aspect médical

Le glaucome est une maladie sournoise de l'œil qui affecte le nerf optique. Cette neuropathie peut se traiter mais n'est pas guérissable mais seulement de ralentir au maximum la disparition des fibres optiques car ils sont irréversibles. il existe trois possibilités pour traiter le glaucome soit : [16]

- Par des médicaments (gouttes oculaires, comprimés).
- Par la chirurgie au laser.
- Par une opération oculaire traditionnelle.

Le but principal à atteindre dans le traitement du glaucome est de faire baisser la pression intraoculaire.

VII : Conclusion :

Le glaucome constitue un problème de santé publique à l'échelle mondiale, Les pertes liées à cette maladie sont irréversibles mais la vision peut être sauvée si celui-ci est détecté suffisamment précocement pour éviter la cécité grâce aux différents examens cliniques qui permettent d'étudier l'évolution des maladies.

Cette première partie a été consacrée à la présentation de l'aspect médical : l'anatomie de l'œil, la pathologie du glaucome avec ses différents types, ses symptômes, ses facteurs de risque, les statistiques et les différents examens cliniques ainsi que les traitements.

La partie suivante traitera une base donnée de glaucome des patients avec les différents attributs.

Chapitre 02 :

Etat de l'art et base de données utilisée

I. Introduction :

La neuropathie oculaire du glaucome est parmi les premières causes de la cécité dans les pays développés. Par conséquent, la recherche d'un système à la détection pour l'aide au diagnostic du glaucome capte l'attention des chercheurs dans différents pays.

Dans ce chapitre, nous présentons un état de l'art complet des approches utilisées dans les systèmes d'aide au diagnostic médical et plus spécifiquement celles qui concernent l'aide à la détection précoce du glaucome. Nous présentons par la suite une description de notre base de données utilisée ainsi que les différents descripteurs qui la caractérise.

II. Etat de l'art :

Le glaucome est une maladie de l'œil liée à l'augmentation anormale de la pression du liquide oculaire [28], ce dernier provoque une diminution irrémédiable du champ de vision et entraîne des dommages au niveau du nerf optique. Nous allons présenter dans la partie qui suit les différents travaux qui proposent une aide au diagnostic médical pour la détection du glaucome en utilisant différents techniques de classification : les réseaux de neurones, les machines à vecteurs de supports, les K-plus proches voisins ainsi que les arbres de décisions. .

En 2006, Joseph Caprioli et al. [29] ont évalué la capacité du réseau neurone à discriminer les yeux normaux et glaucomateux en utilisant des mesures structurelles et fonctionnelles. Une base de données de 239 yeux a été utilisées (185 yeux glaucomateuse et de 54 yeux de sujets normaux). Les résultats obtenus avec un réseau de rétro-propagation de deux couches cachées sont comme suit : Taux de Classification (TC) de 88%, une Sensibilité (Se) de 90% et une Spécificité (Sp) 84% de spécificité.

Pour déterminer si la technique de réseau de neurone peut améliorer la classification entre les cas glaucomateux et les cas normaux, Christopher Bowd et al. [30], ont fait une comparaison entre les techniques de classification réseau neurone et les séparateurs à vastes marges (SVM) en utilisant deux différentes fonction noyaux (linéaire et gaussienne) sur les données du disque optique obtenus par un rétinographe Heidelberg Retina Tomograph (HRT). Une Base de Données (BDD) de 108 patients

Chapitre 02 : Etat de l'art et base de données utilisée

atteints du glaucome et 189 patients sains a été utilisée dans ce travail. Trois différents algorithmes ont été testés sur cette base de données : le perceptron multicouches Multi Layer Perceptron (MLP), les SVM linéaire et les SVM Gaussienne. Les résultats obtenus pour SVM linéaire 93.8%, SVM gaussienne 94.5%, Pour MLP 94.1%.

Un autre travail a été réalisé par A.Rajan et al. [31] dans le but de proposer une aide automatique à la détection précoce du glaucome. Les SVM et les KNN ont été testés sur 200 images OCT : 100 images de patients sains et 100 images de patients glaucomateux. Le processus de ce travail suit trois étapes importantes : Le prétraitement dans cette phase, l'image OCT doit être filtrée du bruit par le filtre médian et Region of Interest (ROI) a été extrait de l'image OCT qui contient uniquement la région de la rétine. Cela se fait par l'application du seuil d'Otsu. L'étape suivante a été l'extraction des caractéristiques l'image extrait ROI a été décomposée par l'utilisation de la transformation par ondelettes discrète. Et la dernière étape est la classification, les techniques de classification ont été utilisées tel que les SVM, les KNN.

La technique de k cross-validation a été utilisée dans ce travail et les résultats obtenus ont été comme suit :

- SVM : Un taux de classification de 91.75% une sensibilité de 93.90% et une spécificité de 87.50%.
- KNN : le taux de classification égale 90% une sensibilité de 84.40% et une spécificité de 93.80%.

Dans le même contexte, Christopher Bowd et al. [32], ont développés un système de classification par SVM sur une base de données de 164 images RNFL obtenues par balayage de polarimétrie laser (SLP). 72 sains et 92 glaucomateux avec les descripteurs de l'âge et de champ visuel. Le classifieur SVM a été formés et testés sur des mesures d'épaisseur RNFL déterminées par SLP à partir de 14 paramètres standard et 64 secteurs. Une validation croisée à dix fois a été utilisée pour former et tester le classifieur SVM. Le résultat obtenu par ce classifieur était de 94% de précision.

En 2008, F. Fink et al [33]. Ont proposé un système basé sur la technique de classification K-NN sur des images de rétine recueillies avec le Tomographie par Cohérence Optique (OCT) Heidelberg. La méthode est testée sur une base d'images de 120 patients et le taux de classification obtenu est de 91%.

Chapitre 02 : Etat de l'art et base de données utilisée

Pour le diagnostic du glaucome Tatsuya Yoshida et al [34]. ont testés les forêts aléatoires (Random Forests) sur des images de la tomographie par cohérence optique dans le domaine spectral (SD-OCT). Une BDD de 210 patients a été utilisée : 84 patients sains et 126 patients atteints d'un glaucome à angle ouvert (GAO). La méthode des forêts aléatoires a été appliquée pour discriminer les cas des yeux normaux et glaucomateux et adopte à l'aide de 151 paramètres OCT, y compris des mesures d'épaisseur de la couche de fibre nerveuse rétinienne circumpapillaire (cpRNFL), de la RNFL maculaire (mRNFL) et de la couche cellulaire ganglionnaire-couche interne pléxiforme combinée (GCIPL) . À l'aide de la méthode des forêts aléatoires en adoptant une validation croisée le résultat obtenu a été de 98.5% de précision, 92.9% de sensibilité et de 96% de spécificité.

En ce qui nous concerne, dans ce projet de fin d'étude nous allons utiliser une nouvelle base de données collectée localement composée de 1004 données et 19 descripteurs.

Sachant que cette base de données est utilisée pour la première fois, on propose l'utilisation de plusieurs techniques de classification afin d'avoir les meilleures performances possibles. Ces techniques sont : les machines à vecteurs de supports, les réseaux de neurones multi couches, les k-plus proches voisins ainsi que deux modèles des arbres de décision (CART, et forêts aléatoires).

Dans la partie qui suit, nous allons présenter une idée générale sur les bases de données ainsi qu'une présentation de notre base de données utilisée ainsi que les différents descripteurs qui la caractérise.

III. Base de données:

L'informatique est considérée comme la science du traitement automatique de l'information. Cette dernière est stockée, traitée, et préparée dans des Bases De Données (BDD). Une BDD est un ensemble permettant le stockage de grandes quantités d'informations mémorisées sur un support permanent et pouvant être restituées aux experts des domaines tel que les médecins pour le suivi, la détection automatique et la progression des maladies [3]. Les BDDs sont utilisées dans les systèmes d'aide au diagnostic médical. Pour cela la collecte des données est une tâche très importante. Le choix de ces données doit être fait en collaboration avec les médecins.

Chapitre 02 : Etat de l'art et base de données utilisée

Une base de données permet de mettre des données à la disposition d'utilisateurs afin de faciliter son utilisation comme ajouter ou supprimer une données c'est-à-dire faire une mise à jour de la BDD, faire une recherche d'une donnée et éventuellement faire une analyse ou un test dans des systèmes intelligents par exemple.

L'avantage majeur de l'utilisation d'une base de données est la possibilité de pouvoir être accédées par plusieurs utilisateurs simultanément.

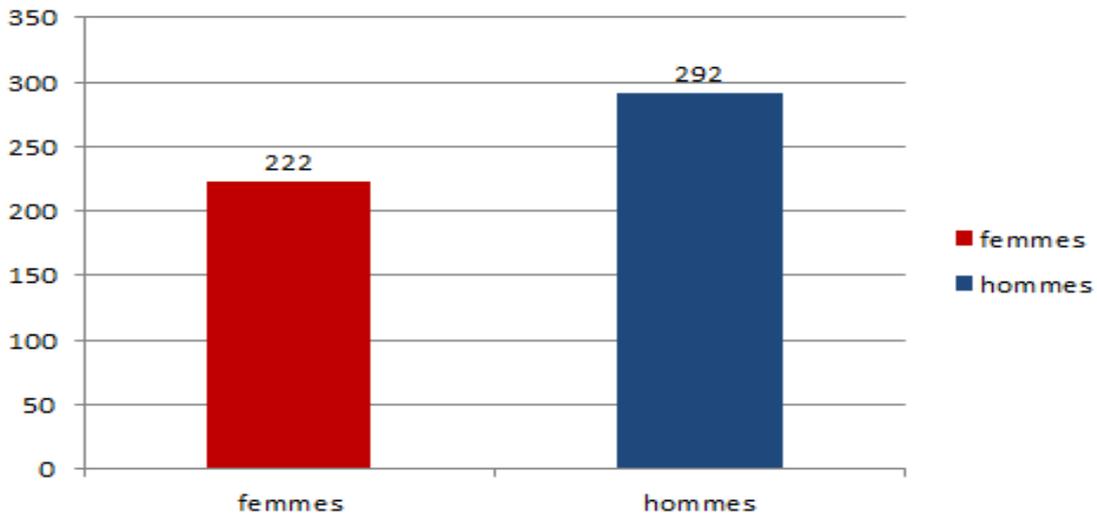
Dans ce mémoire de master nous allons étudiées et testés une nouvelle base de données qui traite la maladie du glaucome. Cette BDD a été collectée par MR LAZOUNI Mohammed El Amine durant deux années en collaborations avec plusieurs médecins ophtalmologues au niveau de plusieurs structures dans le cadre de la réalisation d'un travail scientifique sur le glaucome en collaboration avec plusieurs établissements Algériens privés et hospitaliers.

En effet, la classification des données du glaucome est une tâche compliquée en raison de l'analyse de plusieurs données de source différentes (Champ visuel, OCT, examen du fond 'œil...etc.).

III.2 Base de données utilisée :

La BDD utilisée dans ce travail contient les données des patients de différentes wilaya (centre, ouest, est et sud). Elle contient les données de 514 patients (490 patients avec les données des deux yeux et 24 patients avec les données d'un seul œil (patient monophthalme) ce qui donne une BDD de 1004 cas étudiés, un nombre suffisant pour faire le test de l'intelligence artificielle.

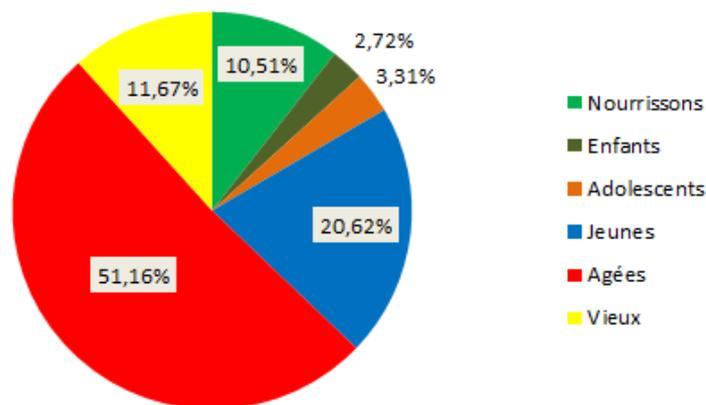
La figure II.1 qui suit présente la répartition des patients collectés dans notre BDD selon le sexe et l'âge.



(FigureIII.2.1) Répartition de la BDD selon le sexe.

La figure (FigureIII.2.1) montre que les deux sexes sont présents dans notre base de données et que le nombre d'hommes est légèrement supérieur à celui des femmes.

La figure (figureIII.2.2) montre que toutes les catégories d'âge sont incluses dans notre base de données.



(figureIII.2.2) Répartition de la BDD selon l'âge.

Selon la figure II.2, la catégorie la plus dominante dans notre BDD est la catégorie des patients âgés. C'est-à-dire les patients dont l'âge varie entre 40 et 65ans. Comme indiquer dans la littérature, l'âge est un facteur parmi les facteurs du glaucome.

Chapitre 02 : Etat de l'art et base de données utilisée

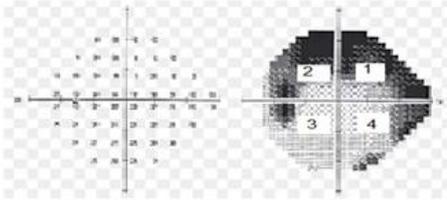
La présence de la catégorie des nourrissons dans notre BDD exprime que le glaucome peut affecter aussi les nouveau-nés. Cette catégorie est définie par le glaucome congénital.

Les différents descripteurs de la BDD utilisés dans ce projet de fin d'étude ont été choisis et validés par les différents médecins ophtalmologues qui ont collaborés à la collecte de cette BDD.

Notre travail se focalise à la proposition d'un système intelligent capable d'aider les médecins ophtalmologues à détecter le glaucome d'une manière précoce. Le tableau III.2.1 qui suit résume les 20 différents descripteurs utilisés dans notre BDD. Les patients collectés dans la BDD sont classés en deux classes malade ou non malade.

L'âge	Nourrissons/Enfants/Adolescents/ Jeunes/Agé/Vieux.
Sexe	Hommes ou Femmes.
L'œil	Gauche, droit
Tonométrie à aplanation	C'est la mesure de la tension intraoculaire en utilisant le tonomètre à aplanation. C'est la méthode la plus précise. Son inconvénient est qu'elle est invasive. La valeur normale de la pression intraoculaire est inférieure à 21 mmhg. Lorsque cette valeur est supérieure à 21mmhg, un glaucome chronique à angle ouvert peut être présent chez le patient. Dans quelque cas, la maladie peut être présente malgré une tension intraoculaire normale (glaucome sans tension). Dans notre base de données la valeur de la tonométrie à aplanation varie entre 9 à 46 mmhg.
Tonométrie jet d'air	C'est la mesure de la TO par la tonométrie à jet d'air. C'est une mesure non invasive et simple à prendre. Lorsque le médecin trouve cette valeur supérieure à 21 mmhg. Il vérifie sa mesure par le tonomètre à aplanation. Dans notre base de données la TO prise par le tonomètre à jet d'air varie entre 10 à 49 mmhg.
Pachymétrie	C'est la mesure de l'épaisseur cornéenne. Cette valeur est prise par un appareil qu'on appelle Pachymètre. La mesure de la pachymétrie a une forte relation avec la TO. Dans notre base de données la pachymétrie varie entre 251 à 675 microns.
Le point le plus fin de la cornée (ppf)	Cette mesure correspond au point le plus fin de la cornée. Dans la base de données ce point varie entre 243 et 672 microns.

Chapitre 02 : Etat de l'art et base de données utilisée

<p>Le rapport cup/disc (C/D)</p>	<p>Le rapport cup/disc correspond à la mesure du rapport cup / disc du nerf optique. Cette valeur est normale lorsqu'elle est inférieure à 0,3. Lorsque cette valeur est supérieure à 0.3 un glaucome peut être suspecté.</p> <p>La valeur C/D varie entre 0.14 à 0.76 dans notre BDD.</p>																								
<p>Champ visuel (CV)</p>	<p>L'altération du champ visuel est marquée par des niveaux de gris différents. Le blanc correspond aux cellules saines. Le noir correspond aux cellules morte ce qu'on appelle scotome négatif. Quant à la couleur grise avec ces différents niveaux, elles font références à des cellules pathologiques. Dans notre base de données les valeurs de CV variée entre 0 et 10. L'image du CV est divisée en 4 parties comme le montre la figure qui suit. Le tableau qui suit nous résume les parties touchées.</p> <div style="text-align: center;">  </div> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>Valeur</th> <th>0</th> <th>1</th> <th>2</th> <th>3</th> <th>4</th> <th>5</th> <th>6</th> <th>7</th> <th>8</th> <th>9</th> <th>10</th> </tr> </thead> <tbody> <tr> <td>Région affectée</td> <td>Aucune Région affectée</td> <td>Région 1</td> <td>Région 2</td> <td>Région 3</td> <td>Région 4</td> <td>Région 2 et 3</td> <td>Région 1,2et 3</td> <td>Région 3 et 4</td> <td>Région 1,3et 4</td> <td>Région 2,3 et 4</td> <td>Région 1, 2,3 et 4</td> </tr> </tbody> </table> <p style="text-align: center;">(figureIII.2.3) les parties touchées dans l'image CV</p>	Valeur	0	1	2	3	4	5	6	7	8	9	10	Région affectée	Aucune Région affectée	Région 1	Région 2	Région 3	Région 4	Région 2 et 3	Région 1,2et 3	Région 3 et 4	Région 1,3et 4	Région 2,3 et 4	Région 1, 2,3 et 4
Valeur	0	1	2	3	4	5	6	7	8	9	10														
Région affectée	Aucune Région affectée	Région 1	Région 2	Région 3	Région 4	Région 2 et 3	Région 1,2et 3	Région 3 et 4	Région 1,3et 4	Région 2,3 et 4	Région 1, 2,3 et 4														
<p>Cellules ganglionnaire (GCC)</p>	<p>La destruction des cellules ganglionnaire permet une détection précoce du glaucome.</p> <p>Dans notre base de données les valeurs de GCC peut prendre les valeurs suivantes :</p> <ul style="list-style-type: none"> - 0 : GCC non altéré (présence que des cellules saines). - 1 : GCC (présence que des cellules pathologiques). - 2 : (présence des cellules pathologique plus quelque cellule morte). - 3 : (La majorité des cellules sont mortes). 																								
<p>Correction de la Tension oculaire</p>	<p>La tension oculaire a une relation très importante avec la pachymétrie par la correction. Chaque valeur de la pachymétrie a une valeur de correction de la tension oculaire dans le but de l'obtention de la TO réelle.</p> <p style="text-align: center;">$TO\ réelle = TO\ mesurée + \text{valeur de correction par rapport à la pachymétrie.}$</p>																								

Chapitre 02 : Etat de l'art et base de données utilisée

	Cette correction varie dans notre base de données entre : -6 à +7 selon la valeur de la pachymétrie. Le tableau (Tableau 3) qui suit nous montre la correction du tonus oculaire selon la valeur de la pachymétrie.
Correction Tonométrie a aplanation selon la pachymétrie.	TO réelle = TO à aplanation + correction selon la pachymétrie.
Correction Tonométrie a aplanation selon le point le plus fin.	TO réelle = TO à aplanation + correction de la pachymétrie selon le point le plus fin.
Correction Tonométrie jet d'air selon la pachymétrie	TO réelle = TO à jet d'air + correction selon la pachymétrie.
Correction Tonométrie jet d'air selon le point le plus fin.	TO réelle = TO à jet d'air + correction de la pachymétrie selon le point le plus fin.
Hérédité	Oui, Non.
Acuité visuelle	La mesure de l'acuité visuelle de loin et de près. Cette valeur varie d'un patient à un autre et d'un œil à l'autre œil pour le même patient. L'acuité visuelle varie entre 0 et 10.
Gonioscopie	Cet examen permet de mesurer l'angle irido cornéen. Le degré de l'ouverture ou la fermeture et le grade soit 0, 1, 2, 3, 4 <ul style="list-style-type: none"> • Grade 0 (0°) : aucune structure n'est visible, l'angle est fermé, • Grade 1 (10°) : seul l'anneau de Schwalbe est visible : la fermeture est probable. • Grade 2 (20°) : le trabéculum est visible : la fermeture est possible. • Grade 3 (25-35°) : l'éperon est visible : la fermeture est improbable. • Grade 4 (35-40°) : la bande ciliaire est visible : la fermeture est

Chapitre 02 : Etat de l'art et base de données utilisée

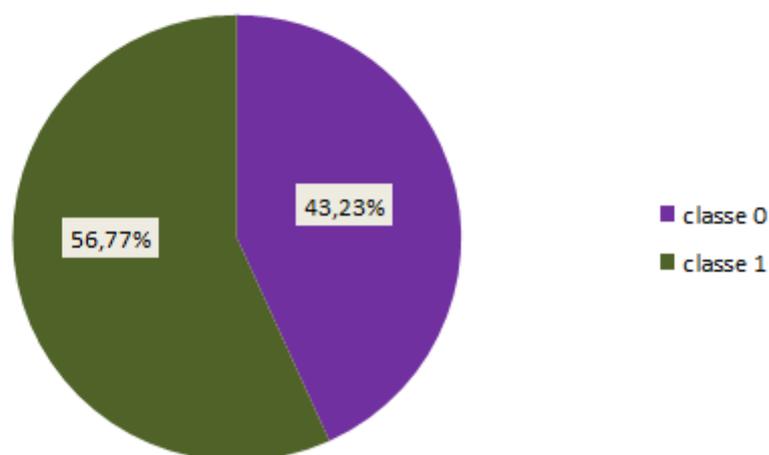
	impossible (angle ouvert).
Classe	Patients Malade, Non malade.

Tableau III.2.1. Les différents descripteurs de la BDD.

Pachymétrie	Correction de tonus oculaire (PIO)	Pachymétrie	Correction de tonus oculaire (PIO)
445	7	555	-1
455	6	565	-1
465	6	575	-2
475	5	585	-3
485	4	595	-4
495	4	605	-4
505	3	615	-5
515	2	625	-6
525	1	635	-6
535	1	645	-7
545	0		

Tableau III.2.2. Correction du tonus oculaire selon la pachymétrie

(figure III.2.4) montre la répartition de nos patients collectés selon les classes (malades et non malades).

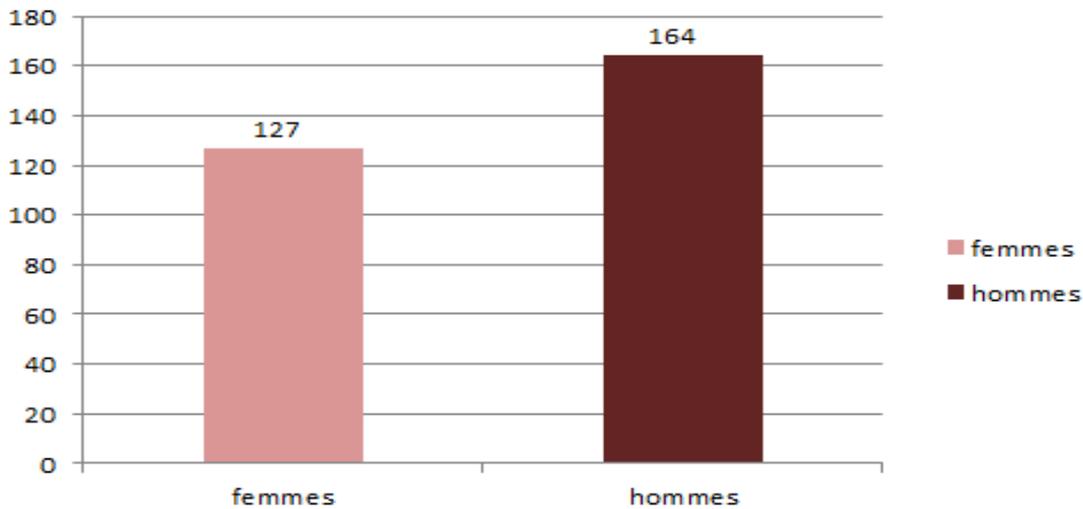


(figure III.2.4) la répartition de la BDD selon les classes.

Chapitre 02 : Etat de l'art et base de données utilisée

Selon la figure III.2.4 le nombre de patients non malade (non glaucomateux) est légèrement plus important que les patients malades (glaucomateux).

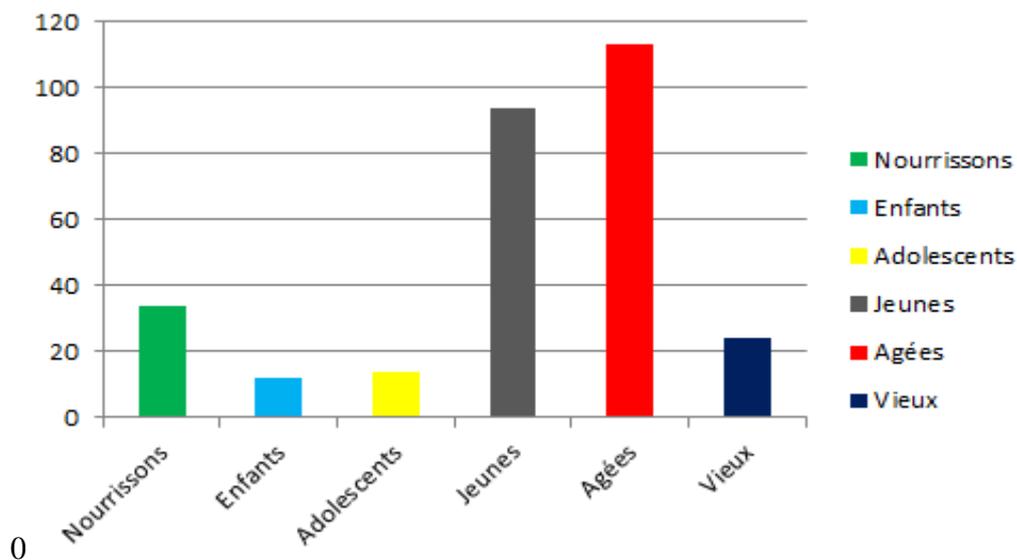
La figure (figureIII.2.5) montre la répartition de la classe 0 (malade) selon le sexe.



(figureIII.2.5) La répartition de la classe 0 selon le sexe.

La figure III.2.5 présente les nombres de patients malades selon le sexe. Le nombre des hommes appartenant à la classe 0 (patient malade) est plus grand que le nombre des femmes.

La figure (figureIII.2.6) montre la répartition des patients malades selon la catégorie âge.



(figureIII.2.6) La répartition de la classe 0 (patient malade) selon les catégories

Le pourcentage des nourrissons appartenant à la classe 0 malade est de 11.81% ce qui exprime l'existence du glaucome congénital dans notre base de données.

IV. Conclusion :

Cette partie a été concernée pour la description de la base de données du glaucome utilisée dans ce projet de fin d'étude. Elle est composée de 514 patients chacun d'entre eux est représentés par 19 paramètres.

Le but de ce travail est la réalisation d'un système intelligent capable d'aider les médecins ophtalmologues dans la détection d glaucome. Pour cela, plusieurs méthodes intelligentes vont être appliquées dans ce travail.

Les différentes techniques de classifications utilisées dans ce projet de fin d'étude pour aider au dépistage précoce du glaucome seront présentées dans le chapitre qui suit.

Chapitre 03 :

Panorama des méthodes de classification utilisées

I. Introduction :

Actuellement l'informatique à un rôle indispensable dans notre vie courante. Dans le secteur de la santé cette spécialité est omni présente et plus particulièrement dans les systèmes intelligents c'est-à-dire ceux qui proposent une aide au diagnostic médical.

Différents types de la classification sont utilisées dans les systèmes d'aide au diagnostic médicales. Ces techniques de classification permettent de classer diverse pathologies en fonction de l'existence ou l'absence des classes. Si ce dernier existe, on parle d'une classification supervisée, sinon d'une classification non supervisée

Dans notre travail, les différentes méthodes de classifications utilisées sont en mode supervisé. Ces méthodes sont : les K plus proches voisins, les machines à vecteur de support, les réseaux de neurones, les arbres de décisions (CART et Foret aléatoire).

Dans ce chapitre nous allons présenter une description théorique des différentes techniques de classification utilisées dans notre travail

II. Les méthodes de classification utilisées :

L'intégration d'Intelligence Artificielle (IA) au sein des hôpitaux et des centres de recherche est utilisée quotidiennement sous forme des systèmes intelligents capable d'aider les médecins dans leurs routines cliniques.

Le diagramme ci-dessous représente des différents classifieurs appliqués sur notre BDD. Le choix de ces techniques ce justifie par son large utilisations par les différents chercheurs.

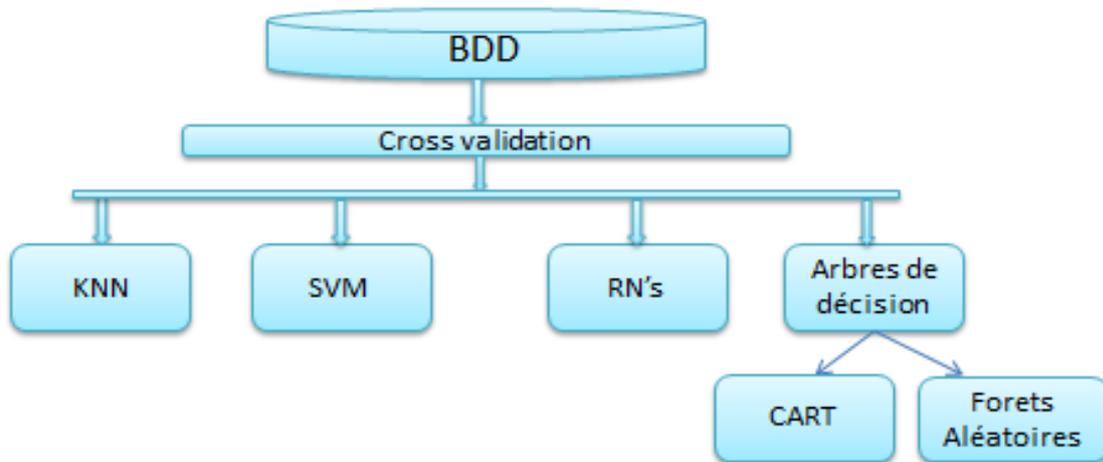
Les K plus proches voisins est une approche connue par sa simplicité et par le très nombre de paramètre a régler. Elle offre aussi des bon résultats pour quelques bases de données.

Les machines à vecteur de support sont connues par leurs puissances de traitements des grandes BDDs et par leur traitement des cas linéairement séparable et non linéairement séparable.

Chapitre 03 : Panorama des méthodes de classification utilisées

Les réseaux de neurones : sont inspirés du système nerveux. Ils sont conçus pour reproduire certaines caractéristiques des mémoires biologiques par le fait qu'ils sont capables d'apprendre et capables de mémoriser l'information (grande puissance d'apprentissage).

Les arbres de décision de type CART et les forêts aléatoires sont connues par leurs interprétabilité. Cette notion est trop demandée par les experts.



(La figure II) Les différentes techniques appliquées sur la BDD.

Afin de proposer un system intelligent, un apprentissage doit être réalisé. Dans notre travail nous avons utilisés la technique de validation croisée ou K-cross validation.

La validation croisée est une technique qui est couramment utilisée en apprentissage automatique pour déterminer les meilleurs paramètres d'un algorithme d'apprentissage. On partitionne l'ensemble d'entraînement en K sous-ensembles disjoints qui sont généralement de même taille. La procédure pour déterminer les hyper paramètres consiste à exécuter l'algorithme d'apprentissage K fois et à chaque fois un ensemble de données de test est mis à coté pour déterminer le risque du classificateur génère par l'algorithme d'apprentissage sur les autres données qui ont servies d'entraînement. [37]

II. 1 : Les k plus proches voisins :

C'est une approche très simple et directe. Elle permet de stocker les données d'apprentissage. Son principe est le suivant :

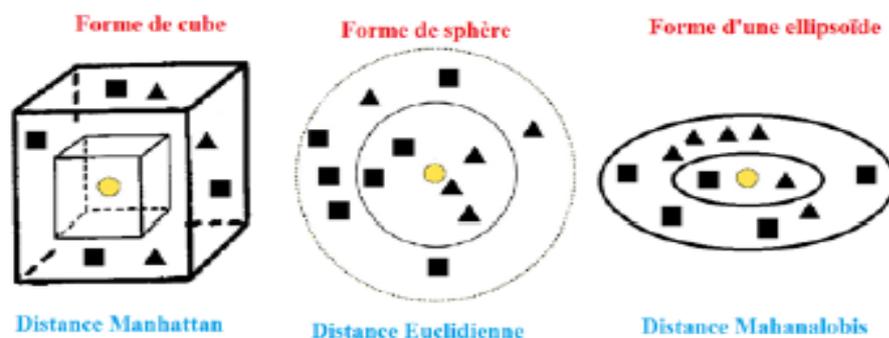
Chapitre 03 : Panorama des méthodes de classification utilisées

Une donnée de classe inconnue est comparée à toutes les données stockées. On choisit pour la nouvelle donnée la classe majoritaire parmi ses K plus proches voisins (le temps de calcul peut donc être long pour les grandes bases de données) au sens d'une distance choisie. Afin de trouver les K plus proches d'une donnée à classer, on doit choisir un type de distance. Il existe plusieurs distances utilisées dans les K plus proches voisins mais la plus utilisée est la distance euclidienne.

L'algorithme réalise un calcul de distance entre tous les individus. Suite à cela, chaque objet est classé dans le groupe où se trouvent ses K plus proches voisins. Les K plus proches voisins sont des individus de l'ensemble des données d'entraînement. Le nombre des K voisins est choisi par l'utilisateur selon l'application, et dans majorité des cas, le k choisi est impair.

Les K -NN considèrent l'estimation de la densité de probabilité comme une fonction purement locale. Un volume $V(x)$ centré sur un point x est déterminé de façon à ce qu'il englobe k observations de l'ensemble d'apprentissage.

La définition de la notion de plus proches voisins est liée au choix d'une distance. Ce choix influence la forme du volume $V(x)$. Cette forme est un cube dans le cas de la distance de Manhattan, une sphère avec la distance euclidienne, et enfin une ellipsoïde avec la distance de Mahalanalobis. La (Figure II. 1) qui suit résume les trois cas de figure vue précédemment. [35]



(Figure II. 1). Représentation des différentes formes pour le calcul de la distance

II. 1 .1 : Algorithme des K-plus proches voisins

Le principe général de la méthode des k-ppv consiste à rechercher parmi l'ensemble d'apprentissage T , contenant l'ensemble des individus et leurs classes d'affectation, un nombre k d'individus parmi les plus proches possibles de l'individu à classer. Puis, l'individu est affecté à la classe majoritaire parmi ces k individus trouvés.

Le nombre k est fixé a priori par l'utilisateur Si $k = 1$, alors l'individu est affecté à la classe du plus proche voisin de l'ensemble T . [36]

II. 2 : Les machines à vecteur de support

Les machines à vecteur support sont aussi appelées : Séparateurs à Vaste Marge présentent une technique de classification dont l'apprentissage est supervisé. Cette technique a été introduite par le russe Vladimir Vapnik en 1995.

Les SVM ont été à l'origine utilisées pour la classification binaire et la régression. Aujourd'hui, elles sont utilisées par de nombreux chercheurs et dans différents domaines tel que l'aide au diagnostic médical qui fait l'objet de notre étude, le marketing, la biologie etc. La raison du succès de cette technique est son adaptation rapide aux bases de données à grandes dimensions, son nombre réduit de paramètre à régler, et ses bons résultats en pratique.[35]

La technique SVM, permet de résoudre des problèmes linéairement et non linéairement Séparable. Elle repose sur deux notions : celle de marge maximale et celle de fonction noyau. Elles permettent de résoudre des problèmes de discrimination non linéaire.

La marge est la distance entre la frontière de séparation et les échantillons les plus proches appelés vecteurs support. Dans un problème linéairement séparable les SVM trouvent une séparatrice qui maximise cette marge. Dans le cas d'un problème non linéaire on utilise une fonction noyau pour projeter les données dans un espace de plus grande dimension où elles seront linéairement séparables. [35]

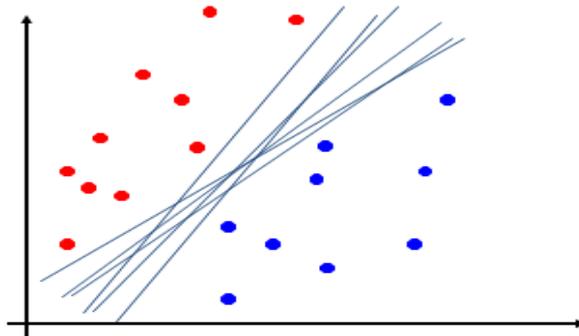
II. 2. 1 : Principe de fonctionnement général de la technique SVM

Si on prend deux classes par exemple (une classe positifs et une classe négatifs), le but des SVM est de faire la séparation entre ces deux classes.

La méthode cherche alors l'hyperplan qui sépare les exemples vrais positifs à des autres vrais négatifs, en garantissant que la marge entre le plus proche des positives et des négatives soit maximale.

Le classificateur linéaire qui effectue une séparation de données entre les classes est appelé hyperplan.

(La figure II. 2. 1) qui suit. On détermine différents hyperplans possible qui séparent deux classes.

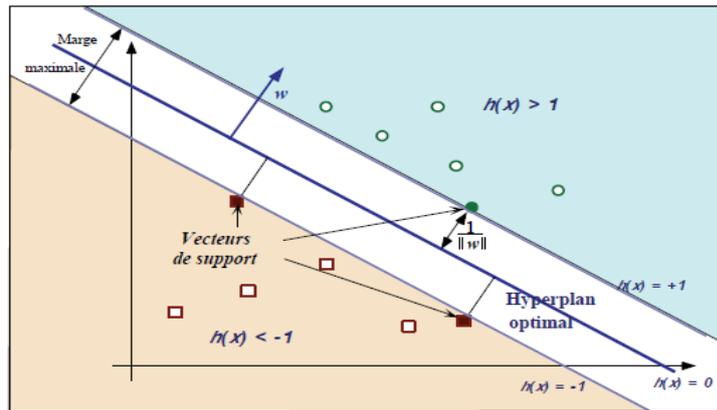


(La figure II. 2. 1.1). Différents hyperplan possible

Les échantillons les plus proches des deux classes sont appelés les vecteurs de support.

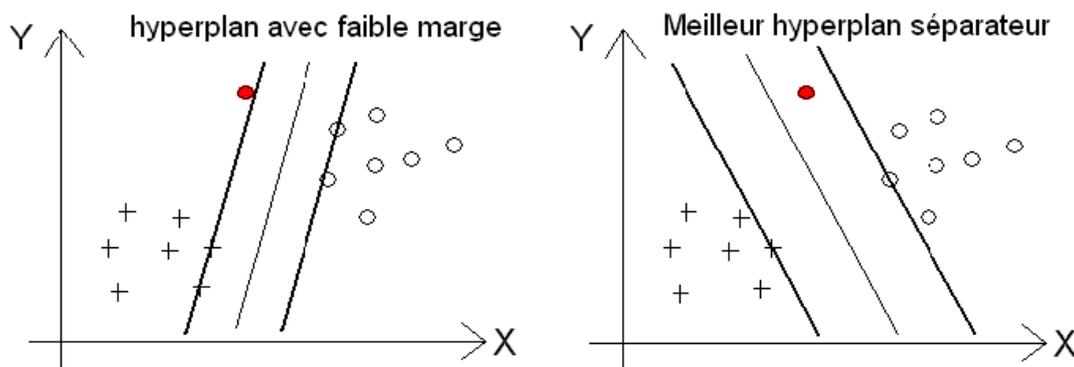
Ces derniers sont utilisés pour la détermination de l'hyperplan. On observe multitude d'hyperplans valide qui séparent les deux classes. Cette méthode doit choisir un hyperplan optimal. L'hyperplan optimal séparant les points de deux classes est celui qui passe au milieu de ces classes, c'est-à-dire dont la distance aux points les plus proches est maximale

L'hyperplan optimal est perpendiculaire au segment de droite le plus court joignant un exemple d'apprentissage à l'hyperplan



(La figure II. 2. 1.2).Hyperplan optimal

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

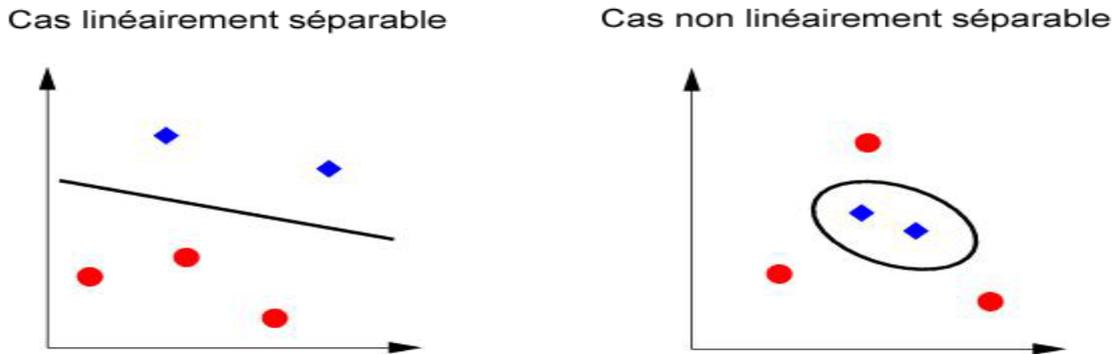


(La figure II. 2. 1.3).Choix de l'hyperplan optimal

II. 2. 2 : Linéarité et non-linéarité

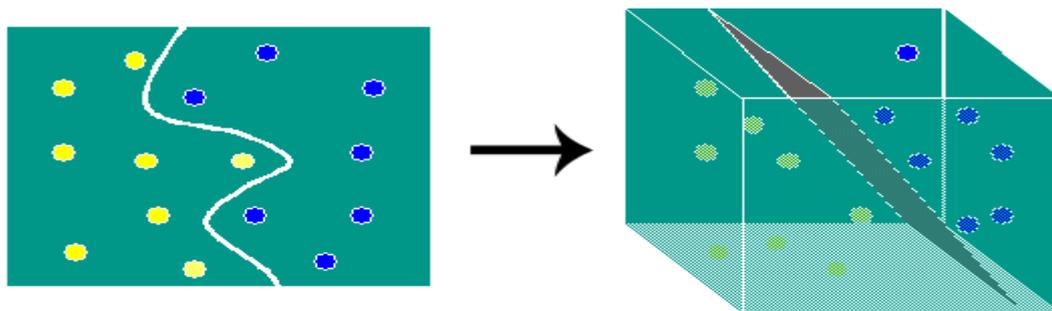
Parmi les modèles des SVM, on constate les cas linéairement séparable et les cas non linéairement séparable. Les premiers sont les plus simples des SVM car ils permettent de trouver facilement le classificateur linéaire. Dans la plupart des problèmes réels il n'y a pas de séparation linéaire possible entre les données, le classificateur de

marge maximale ne peut pas être utilisé car il fonctionne seulement si les classes de données d'apprentissage sont linéairement séparables. [35]



(Figure. II. 2. 2 .I) .Cas linéairement et non linéairement séparable

Pour surmonter les inconvénients des cas non linéairement séparable, l'idée des SVM est de changer l'espace des données. La transformation non linéaire des données peut permettre une séparation linéaire des exemples dans un nouvel espace. On va donc avoir un changement de dimension. Cette nouvelle dimension est appelée espace de redescription. En effet, intuitivement, plus la dimension de l'espace de redescription n'est grande, plus la probabilité de pouvoir trouver un hyperplan séparateur entre les exemples est élevée. [38] Ceci est illustré par le schéma suivant :



(Figure. II. 2. 2 .II).Transformation d'espace

II. 3. Les réseaux de neurones :

Un réseau de neurones est un système informatique inspiré de l'organisation des cellules du cerveau humain. Ce système comporte plusieurs unités de calcul simples, généralement organisées en couches et qui fonctionnent en parallèle. L'approche neuronale utilise des modélisations issues de la recherche biologique afin de concevoir des architectures informatiques pouvant accomplir des tâches spécifiques, telles que

l'apprentissage et la reconnaissance des formes. En informatique, on appelle réseau de neurone un ensemble d'entités (les neurones) interconnectées. Dans la grande majorité des cas les neurones sont en fait des fonctions calculées par un programme informatique, mais ils sont parfois réalisés sur des circuits électroniques

Les neurones sont caractérisés par un état d'excitation qui dépend de celui des neurones situé en amont, ainsi que de la force des liens qui les relient.

Le schéma ci-dessous représente la structure générale d'un réseau de neurone. [39]

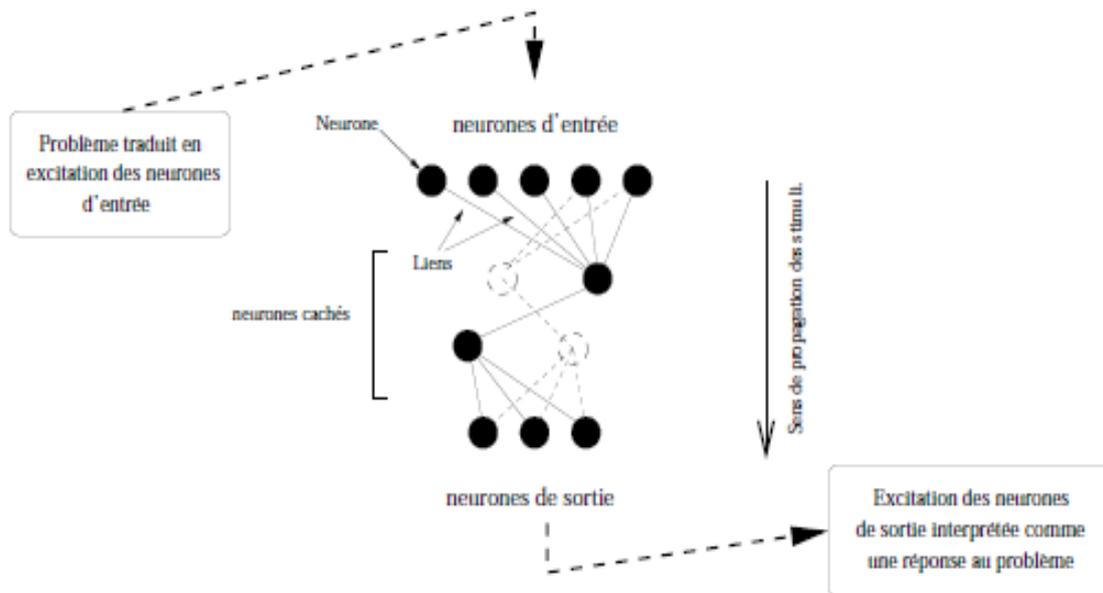
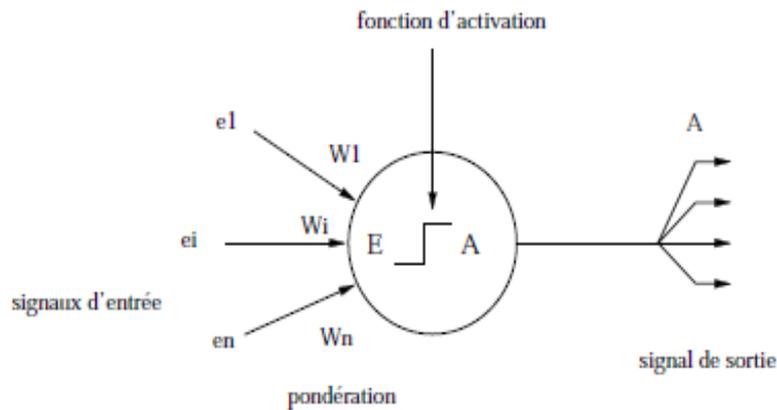


Figure (II. 3). Schéma général d'un réseau de neurone.

II. 3.1. Le neurone formel

Un neurone formel est une fonction algébrique non linéaire et bornée, dont la valeur dépend de paramètres appelés coefficients ou poids. Les variables de cette fonction sont habituellement appelées "entrées" du neurone, et la valeur de la fonction est appelée sa "sortie". Un neurone est donc avant tout un opérateur mathématique, dont on peut calculer la valeur numérique par quelques lignes de logiciel. On a pris l'habitude de représenter graphiquement un neurone comme indiqué sur la Figure (II. 3.1.1). [40]



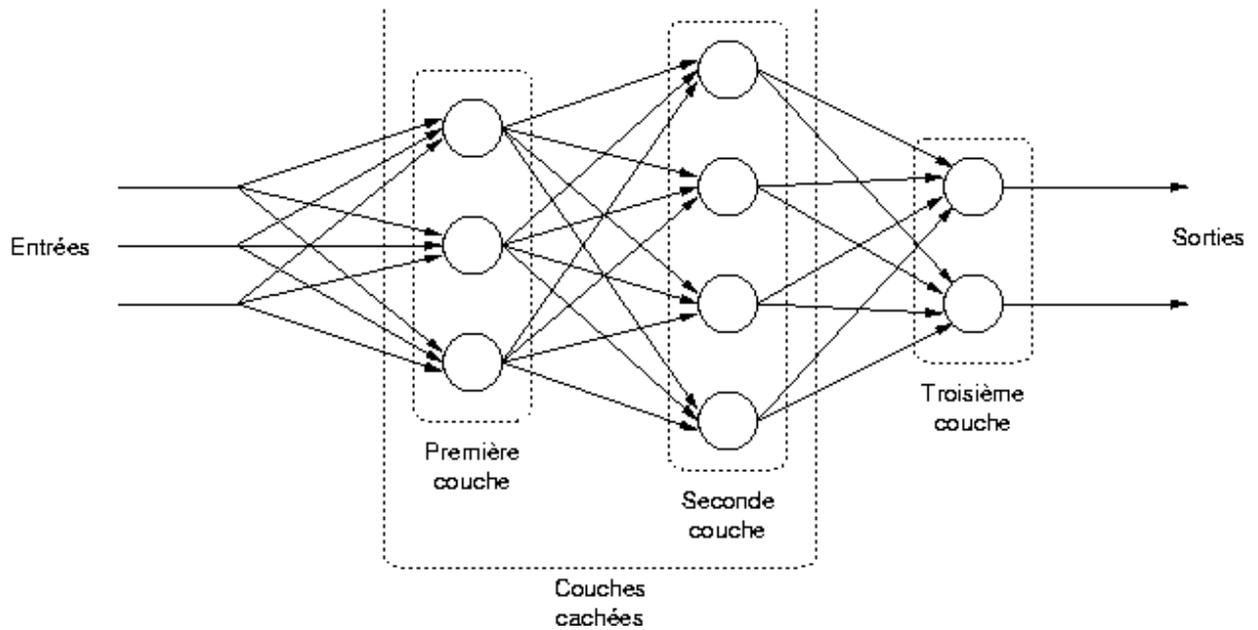
(Figure II. 3.1.1) .Schéma d'un neurone formel

Les neurones formels sont dotés de caractéristique inspirée de celles des neurones biologiques

1. Les neurones sont des éléments simples et fortement connectés (même si les neurones artificiels sont beaucoup plus simples).
2. Les connexions entre les neurones permettent de déterminer la fonction du réseau. Ces connexions, connues sous le nom de poids synaptiques, servent à stocker les connaissances acquises.
3. Les connaissances sont extraites de leur environnement via un processus d'apprentissage.

L'apprentissage d'un réseau de neurones formels consiste à déterminer les poids w_i optimaux suivant le problème à résoudre. Rappelons que l'architecture du réseau est déjà déterminée à ce stade.

Le perceptron multicouche est un réseau comportant plusieurs couches, chaque neurone d'une couche étant totalement connecté au neurone de la couche suivante comme le montre la figure (Figure II. 3.1.2) qui suit. La structure est relativement simple : une couche d'entrée, une couche de sortie et une ou plusieurs couches cachées [41]



(Figure II. 3.1.2) Schéma d'un PMC

Les nœuds d'entrée: elle recevra les données source que l'on veut utiliser pour l'analyse. Dans le cas de l'aide au diagnostic médical, cette couche recevra les caractéristiques des symptômes

Les nœuds cachés: est une couche cachée, en ce sens qu'elle n'a qu'une utilité essentielle pour le réseau de neurones et n'a pas de contact direct avec l'extérieur. Les fonctions d'activations sont en général non linéaires sur cette couche mais il n'y a pas de règle à respecter. Le choix de sa taille n'est pas implicite et doit être ajustée.

Les nœuds de sortie : est appelée couche de sortie. Elle donne le résultat obtenu après traitement des données d'entrée de la première couche. Dans le cas de l'aide au diagnostic médical, cette couche donne le diagnostic. Sa taille est directement déterminée par le nombre de classe ou de types de maladies. [41]

II. 3.2. Algorithme rétro-propagation du gradient

L'algorithme d'apprentissage pour le réseau de neurone multi couches est le suivant :

1. Initialisation des poids du réseau
2. Présentation du vecteur d'apprentissage à l'entrée du réseau (couche d'entrée)

3. Calcul du vecteur de sortie S : le système propage les activités neuronales (état excité ou inhibé des neurones) à travers le réseau. Chaque neurone calcule la somme pondérée de ses entrées et transmet le résultat par une fonction de type sigmoïde pour produire sa valeur de sortie. Le vecteur de sortie S est le résultat du calcul de la dernière couche (couche de sortie).
4. Calcul de l'erreur. On compare alors les valeurs de sortie actuelles S avec les valeurs désirées de R appelé vecteur de référence (ou encore vecteur cible).
On en déduit l'erreur J , définie comme la somme des carrés des différences observée sur chaque unité de la couche de sortie.
5. Rétro propagation de l'erreur : l'algorithme de rétro propagation du gradient permet récursivement de rétro propager l'erreur de la couche de sortie vers les couches cachées jusqu'à la première couche du réseau.
6. Modification des poids : les poids de chaque neurone sont modifiés, soit à chaque présentation d'un vecteur d'apprentissage (méthode du gradient stochastique), soit après cumul de l'erreur pour un certain nombre de vecteurs d'apprentissage (méthode du gradient standard).
7. Critère d'arrêt : seuil d'erreurs ou nombre de présentation d'exemple .ect.
8. Retour à l'étape 1 [35,41]

II. 4. Les arbres de décisions :

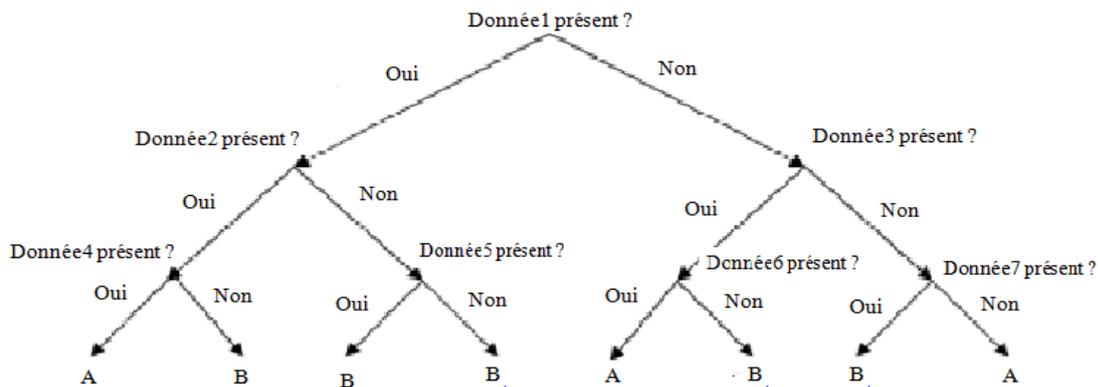
Les arbres de décision (AD) sont des techniques très employées dans les cas où il est important de trouver des règles afin de définir la classification. Les caractéristiques de la structure des arbres permettent une représentation interprétable des résultats. Un arbre de décision correspond à une structure arborescente formée par des nœuds internes, il y a une trajectoire unique qui débute avec le nœud racine. La trajectoire correspond à une règle de décision). Les algorithmes les plus représentatifs des arbres de décision sont ID3 (Quinlan, 1986), C4 (Quinlan, 1993) et CART. Elle connue par sa simplicité avec laquelle nous pouvons modéliser le problème, sa représentation graphique, sa lisibilité et la rapidité par laquelle nous pouvons catégoriser un document expliquent la popularité actuelle des arbres de décision. [42,37]

II. 4.1. Algorithmes d'apprentissage par arbre de décision

L'apprentissage de règles par arbres de décision consiste à commencer par construire un arbre de décision à partir des données de l'ensemble d'apprentissage, puis à traduire cet arbre en un ensemble de règles. On aura donc autant de règles que des feuilles dans l'arbre.

Pendant la phase de test ; les AD sont des méthodes d'induction utilisant une structure d'arbre pour classer les données. Chaque nœud de l'arbre correspond à un test sur un attribut. Une question qui se pose est celle du choix de l'attribut racine .pour y répondre, on établit un test statistique, qui détermine à quel point tel attribut peut discriminer l'ensemble d'apprentissage. [42]

Dans notre travail, deux techniques des arbres de décision sont utilisées qui sont l'algorithme CART et les forêts aléatoires.



(Figure II. 4.1) Exemple d'arbre de décision pour classer des caractères.

II.4.1.1 L'algorithme CART (classification and regression Tree):

Algorithme CART signifie Classification And Regression Tree, il désigne une méthode statique élaboré par (L.Breiman, J.H.Friedman, R.A.Olshen et C.J.Stone en1984[Aur06]), qui construit des prédicateurs par arbre Aussi bien en régression qu'en classification, il utilise un arbre binaire(toujours deux branches par nœud non feuilles).

Le principe est de diviser récursivement et le plus efficacement possible les exemples de l'ensemble d'apprentissage par des tests définis a l'aide des attributs jusqu'à ce que l'on obtienne des sous-ensembles d'exemples ne contenant (presque) que des exemples appartenant a une même classe.[43]

III. Les méthodes d'ensemble :

Le principe général des méthodes d'ensemble est de construire une collection de prédicteurs, pour ensuite agréger l'ensemble de leurs prédictions. Dans un cadre de classification, l'agrégation revient par exemple à faire un vote majoritaire parmi les classes fournies par les prédicteurs.

Les méthodes d'ensemble se reposent sur une construction aléatoire d'une famille de modèle « Bagging » et sur une construction adaptative, déterministe ou aléatoire d'une famille de modèle « Boosting ».[43]

III.1. Bagging :

Le mot bagging est la contraction des mots Bootstrap et Aggregating, inventé par Breiman en 1996 il est efficace pour corriger le manque de robustesse des arbres de décision. Il consiste à la construction d'une règle de base sur m échantillons bootstrap différents on en modifie les prédictions, et donc on construit à terme une collection de prédicteurs variés, qui sont ensuite agrégés par un vote (majoritaire des résultats des modèles) ou une moyenne des estimations c'est-à-dire que cette étape permet d'obtenir un prédicteur performant et à combiner plusieurs calssfieurs. [43]

III.2. Bootstrap :

Le Bootstrap est un principe de re-échantillonnage statistique traditionnellement utilise pour l'estimation de grandeurs ou de propriétés statistiques. L'idée du bootstrap est d'utiliser plusieurs ensembles de données re-échantillonnes à partir de l'ensemble des données observées et ceci a l'aide d'un tirage aléatoire avec remise. Ainsi chaque classifieur élémentaire de l'ensemble entrainera sur un des échantillons bootstrap de sorte qu'ils soient tous entraines sur un ensemble d'apprentissage différent. L'agrégation de ces classifieurs permet d'obtenir un prédicteur plus performant

III.3. Les forêts aléatoires :

La technique des forêts aléatoires sont une méthode statistique non-paramétrique aux performances exceptionnelles. Elles sont été introduites par Breiman (2001) et sont de plus en plus utilisées pour traiter de nombreux et divers jeux de données réelles. [44]

Les forêts aléatoires sont composées (comme le terme "forêt" l'indique) d'un ensemble d'arbres décisionnels. Ces arbres se distinguent les uns des autres par le sous-

échantillon de données sur lequel ils sont entraînés. Ces sous-échantillons sont tirés au hasard (d'où le terme "aléatoire") dans le jeu de données initial.

II.4.1.2.1 Algorithme forêt aléatoire :

Forêt Aléatoire est une méthode introduite par Leo Breiman en 2001. Elle est aussi un prédicteur constitué d'un ensemble de classifieurs élémentaires de type arbres de décision. Dans les cas spécifiques des modèles CART (arbres binaires), Breiman propose une amélioration du bagging avec un algorithme d'induction de forêts aléatoires pour Random Forest qui utilise le principe de randomisation. [28]

L'induction des arbres se fait sans élagage et selon l'algorithme CART, toutefois, au niveau de chaque nœud, la sélection de la meilleure partition, basée sur l'index de Gini, s'effectue uniquement sur un sous-ensemble d'attributs de taille préfixée sélectionnée aléatoirement depuis l'espace original des caractéristiques. La prédiction globale de la forêt aléatoire est calculée en prenant la majorité des votes de chacun de ses arbres. Cet algorithme appartient à la famille la plus large des forêts aléatoires définies comme suit par Breiman. [28]

III.3.1L'algorithme des forêts aléatoires :

- Construction d'une forêt aléatoire à N arbres :
- Pour i de 1 à N ,
- Construire un échantillon bootstrap des données
- Construire un arbre CART maximal à partir de cet échantillon bootstrap, tel que:
- A chaque nœud, on sélectionne la meilleure découpe sur la base de K variables choisies aléatoirement parmi les P variables d'entrée
- Fin pour; [28]

De nombreux modèles de forêts aléatoires ont été créés qui correspondent à autant de manières d'incorporer de l'aléatoire dans les arbres. À titre d'exemple, on peut citer :

- le Tree Bagging introduit de l'aléatoire dans l'échantillon initial sélectionnant certains points plutôt que d'autres et laisse grandir l'arbre jusqu'à ce que chaque nœud comporte un unique élément [44]

- le random subspace consiste à sélectionner à chaque nœud K variables de manière aléatoire et, parmi celles-ci, à choisir celle qui minimise un certain critère.

-la Random Forest qui consiste pour à mélanger le CART, le bagging et le random subspace: pour chaque arbre, on tire un échantillon à partir de l'échantillon initial ; à chaque nœud, on choisit aléatoire K variables et on prend, parmi celles-ci, celle qui minimise le critère de l'algorithme CART. On laisse grandir l'arbre jusqu'à ce qu'il n'y ait plus qu'un seul élément dans chaque nœud. [44]

-le Random Select Split qui sélectionne les K meilleures séparation et qui en choisit une parmi celles-ci de manière aléatoire. La position de la coupure est également calculée de manière aléatoire. [44]

V. Conclusion :

Dans ce chapitre nous avons présenté l'aspect théorique des différentes techniques de classification utilisées dans notre travail. Le choix de ces techniques se justifie par leurs résultats obtenus par les différents chercheurs dans la littérature scientifique et plus précisément dans la détection du glaucome.

Dans le chapitre suivant nous allons présenter l'application de ces techniques sur notre base de données et nous discuterons et analyserons les différents résultats obtenus.

Chapitre 04 :

Résultats obtenus et

discussion

I. Introduction :

Après la présentation du principe de fonctionnement théorique des différentes approches de classification dans le chapitre précédent, nous présenterons dans ce chapitre les différents types de classifications, les différents paramètres d'évaluation ainsi que l'implémentation de ces méthodes sur notre nouvelle base de données collectée récemment par MR LAZOUNI en utilisant le logiciel MATLAB V 2014a.

Nous présenterons aussi les résultats expérimentaux obtenus par les tests effectués à chaque classifieur et nous analysons aussi les différents paramètres d'évaluations utilisés dans ce travail qui sont le taux de classification, la sensibilité, la spécificité, ainsi que la matrice de confusion après avoir appliqué la technique k-cross validation comme technique d'apprentissage.

II. Les différents types de la classification :

D'une manière générale l'objectif de la classification est d'identifier les classes auxquelles appartiennent des objets à partir de traits descriptifs (attributs, caractéristiques, features). [45]

Ils existent trois type de la classification : supervisée, non supervisée et semi supervisée.

La classification supervisée :

Si les classes sont prédéterminées et les exemples connus, le système apprend à classer selon un modèle de classement ; on parle alors d'apprentissage supervisé .Un expert doit préalablement étiqueter des exemples. Le processus se passe en deux phases. Lors de la première phase (hors ligne, dite d'apprentissage), il s'agit de déterminer un modèle des données étiquetées. [46,47]

La classification non-supervisé :

Si le nombre de classes et leur nature n'ont pas été prédéterminés, et les exemples non étiquetés on parle d'apprentissage non supervisé ou clustering , dans ce cas l'expert ne les a pas encore classés.

L'algorithme doit découvrir par lui-même la structure plus ou moins cachée des données. Le partitionnement de données, en anglais (data clustering), est un algorithme d'apprentissage non supervisé. [46,47]

La classification semi-supervisé :

Il s'agit d'une forme d'apprentissage à mi-chemin entre supervisé et non-supervisé, qui utilise un faible nombre de données avec étiquettes pour labelliser un grand nombre de données non étiquetées est un processus généralement aisé. C'est très différent quand celles-ci doivent être annotées, puisqu demandant très souvent l'intervention humaine.

Cette forme d'apprentissage est donc un excellent compromis lorsque l'on cherche à faire de la classification mais que l'obtention d'un ensemble complètement annoté est difficile, Dans ce travail, le type d'apprentissage utilisé est de type supervisé car les différentes classes ont été définies.

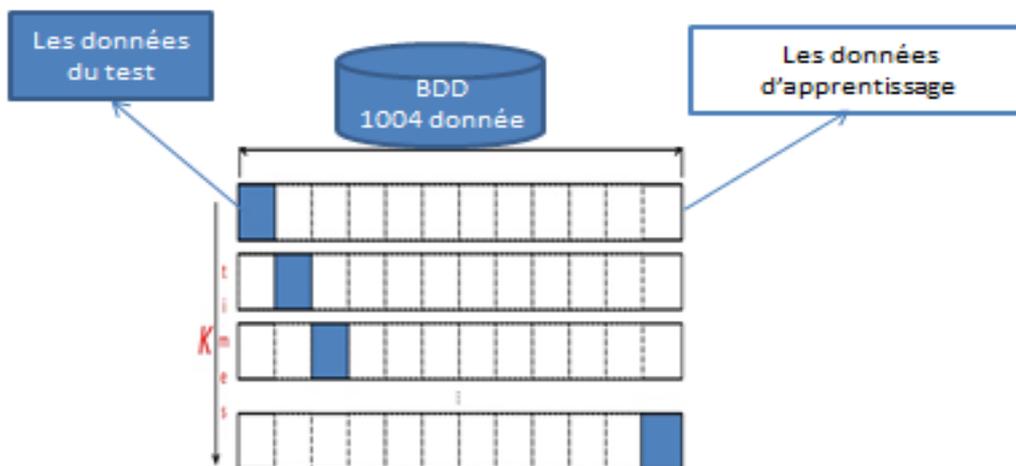
III La validation croisée :

La validation croisée constitue une méthode standard d'apprentissage et de test. Elle permet de mesurer les performances d'un modèle.

La méthode K- cross validation, est très utile lorsque le nombre de données à traiter n'est pas vraiment important. Son principe de fonctionnement est résumé comme suit :

- Approche largement utilisée pour estimer l'erreur de test.
- Les estimations peuvent être utilisées pour sélectionner le meilleur modèle et pour donner une idée d'erreur de test du modèle final choisi.
- L'idée est de diviser au hasard les données en K de taille égale des pièces, puis on laisse la partie k pour le test. K - 1 partie (combinée), consacré pour l'apprentissage.
- Ceci est fait à tour de rôle pour chaque partie $k = 1, 2, \dots, K$, jusqu'à l'obtention de résultats combinés. [48]

La figure ci-dessous montre la méthode de la validation croisée puis les données d'apprentissage et les données du test.



(Figure. III) La méthode de la validation croisée

IV. Les paramètres d'évaluation :

Dans cette partie, nous allons présenter les différents paramètres scalaires qui permettent d'évaluer et analyser les performances des systèmes de classification de notre travail.

Matrice de confusion :

Une matrice de confusion ou tableau de contingence sert à évaluer la qualité d'une classification. Elle est obtenue en comparant les données classées avec des données de référence qui doivent être différentes de celles ayant servi à réaliser la classification. [49]

Une matrice de confusion permet de faire la différenciation des erreurs selon chaque classe en vue d'évaluer un classifieur. Le tableau qui suit nous montre un modèle d'une matrice de confusion.

	Décision Positifs	Décision Négatifs	
Étiquette Positifs	Vrai Positifs, TP	Faux Négatifs, FN	Pos
Étiquette Négatifs	Faux positifs, FP	Vrai Négatifs, TN	Neg
Total	PPos	PNeg	N

(Tableau IV.I) matrice de confusion.

Pos : le nombre d'éléments positifs dans la base.

Neg : le nombre d'éléments négatifs dans la base.

PPos : le total d'éléments classés positifs.

PNeg : le total d'éléments classés négatifs.

A partir d'une matrice de confusion, on peut calculer les taux suivants :

- **Taux de bonne classification**

C'est l'indicateur le plus évident permettant d'évaluer les performances d'un système de classification.

Cette valeur est calculée facilement, elle correspond au nombre de patients bien classés sur le nombre total des patients.

$$\text{Taux de classification} = \frac{\text{Nombre de patients bien classés}}{\text{Nombre total des patients}} \quad (\text{IV.1})$$

- **La sensibilité et la spécificité :**

La sensibilité et la spécificité sont deux notions utilisées en probabilité médicale. Elles sont un outil de calcul utilisé pour établir des tests de dépistage. Elles permettent d'évaluer la valeur d'un test. Ces deux notions sont dépendantes. La spécificité exclut la maladie tandis que la sensibilité l'affirme.[51]

-La Sensibilité (Se) : proportion de vrais positifs parmi les personnes à dépister. [52] ou bien la sensibilité désigne la probabilité d'obtenir un test positif sur un individu porteur de la maladie [51]

$$\text{La sensibilité} = \frac{\text{vrais positifs}}{(\text{vrais positifs} + \text{faux négatifs})} \quad (\text{IV.2})$$

-La Spécificité (Sp) : proportion de vrais négatifs chez les non-malades. [52] ou encore la spécificité désigne la probabilité d'obtenir un test négatif sur un individu non porteur de la maladie. [51]

$$\text{la spécificité} = \frac{\text{vrais négatifs}}{(\text{vrais négatifs} + \text{faux positifs})} \quad (\text{IV.3})$$

- Le taux de faux positifs (false positive rate) :

$$\text{taux de faux positifs} = \frac{\text{faux positifs}}{(\text{vrais positifs} + \text{faux négatifs})} \quad (\text{IV.4})$$

- Le taux de faux négatifs (false negative rate) :

$$\text{taux de faux négatifs} = \frac{\text{faux négatifs}}{(\text{vrais négatifs} + \text{faux positifs})}$$

(IV.5)

On propose dans ce travail d'appliquer cinq différentes techniques de classification : K- plus proche voisins (Kppv), les machines à vecteurs de supports (SVM), réseaux de neurones (MLP), les arbres de décision CART et les forêts aléatoires(RF) en aspect supervisé, dans le but de tester et valider notre base de données. Pour cela on a réalisé plusieurs expérimentations correspondant à des changements de paramètres ou de structures pour chaque classifieur, et par la suite nous présenterons les résultats obtenus dans les tableaux énumérés ci-dessous. (V.1, V.2)

- En ce qui concerne la technique des K- plus proche voisins (Kppv), nous avons changé le nombre de K c'est-à-dire les voisinages ainsi que le type de distance utilisée. Quatre différentes distances ont été testées qui sont : euclidean, cityblock, cosine et corrélation.
- Pour le classifieur les machines à vecteurs de supports (SVM) nous avons varié la fonction Kernel : linear, quadratic, polynomial, rbf, mlp.
- Pour le réseau de neurone (MLP) : nous avons fixé certains paramètres comme :
 - Le pas d'apprentissage 0.05.
 - L'erreur minimale 0.001.
 - Le nombre d'époque 1000.

Puis changé d'autres paramètres comme le nombre des couches cachées et le nombre de neurones des couches cachées (c'est-à-dire l'architecture du réseau) ainsi que les fonctions d'activation et les fonctions du transfert.

	Base de données
	RN's (MLP)
Erreur minimale	0.001
Le nombre d'époque	1000
La fonction de transfert	Tansig/logsig/purelin
La fonction d'apprentissage	Trainlm/trainingd/trainingda.

(Tableau IV.II) L'architecture d'un perceptron multicouche appliqué sur la BDD.

Chapitre04 : Résultats obtenus et discussion

- Pour le classifieur arbre de décision CART nous avons varié le pruning on/ off et le niveau de pruning ou la hauteur de l'arbre.

V. Les résultats expérimentaux obtenus :

Le tableau qui suit nous résume les meilleurs résultats obtenus pour chaque système après la réalisation de plusieurs expérimentations.

	Cross validation K=2				Cross validation K=3			
	<i>TC</i>	<i>SE</i>	<i>SP</i>	<i>Matrice de confusion</i>	<i>TC</i>	<i>SE</i>	<i>SP</i>	<i>Matrice de confusion</i>
K-NN	77.0916	58.5777	91.2438	225 180 50 520	77.3864	65.4511	86.5388	284 150 77 493
SVM	96.2151	95.1616	97.0200	413 21 17 553	96.1155	95.4163	96.6752	414 20 19 551
MLP	95.7171	93.8630	97.0344	408 26 17 553	95.4183	93.3231	97.0218	405 29 17 553
CART	98.4064	97.3214	99.2806	422 12 5 565	97.6119	95.2703	99.4652	421 13 7 563
Foret Aléatoire	97.5100	96.3105	98.4167	418 16 9 561	97.7097	96.5454	98.6026	419 15 8 562
	Cross validation K=4				Cross validation K=5			
	<i>TC</i>	<i>SE</i>	<i>SP</i>	<i>Matrice de confusion</i>	<i>TC</i>	<i>SE</i>	<i>SP</i>	<i>Matrice de confusion</i>
K-NN	78.4861	60.8014	92.2151	263 171 45 525	78.8876	60.3350	92.9249	262 172 40 530

Chapitre04 : Résultats obtenus et discussion

SVM	96.2151	95.4138	96.8410	414 20 18 552	96.0144	95.4642	96.5738	414 20 20 550
MLP	94.2231	92.2584	95.6246	401 33 25 545	95.0214	92.5630	96.8402	402 32 18 552
CART	98.4064	97.1963	99.3056	423 11 5 565	99.0050	98.7952	99.1525	419 15 11 559
Foret Aléatoire	98.1076	96.8083	99.1526	420 14 5 565	98.1090	96.7851	99.1633	420 14 5 565
	Cross validation K=6				Cross validation K=7			
	<i>TC</i>	<i>SE</i>	<i>SP</i>	<i>Matrice de confusion</i>	<i>TC</i>	<i>SE</i>	<i>SP</i>	<i>Matrice de confusion</i>
K-NN	78.5887	63.4970	90.2675	274 160 55 515	79.0862	63.7270	90.7009	277 157 53 517
SVM	96.2159	95.8203	96.3751	416 18 20 550	95.9173	95.2579	96.2781	414 20 21 549
MLP	95.0214	92.5630	96.8402	402 32 18 552	95.5191	93.2688	97.1673	405 29 16 554
CART	97.6048	95.8904	98.9362	421 13 9 561	97.9021	96.5517	98.8235	423 11 9 561
Foret Aléatoire	97.6113	96.6120	98.4535	419 15 9 561	98.2073	97.0705	99.1518	421 13 5 565
	Cross validation K=8				Cross validation K=9			
	<i>TC</i>	<i>SE</i>	<i>SP</i>	<i>Matrice de confusion</i>	<i>TC</i>	<i>SE</i>	<i>SP</i>	<i>Matrice de confusion</i>

Chapitre04 : Résultats obtenus et discussion

K-NN	79.2833	64.2231	90.7613	279 155 53 517	78.4955	62.0318	91.3208	267 167 49 521
SVM	95.7190	94.9870	96.0592	413 21 21 548	95.6197	95.9540	95.3455	416 18 26 544
MLP	95.3167	92.6527	97.3702	402 32 16 554	96.0121	95.4301	96.5456	414 20 20 550
CART	97.9021	96.5517	98.8235	423 11 9 561	98.1982	98.2143	98.1818	424 10 9 561
Foret Aléatoire	98.2073	97.0705	99.1518	421 13 5 565	98.6049	97.3428	99.6495	422 12 2 568
	Cross validation K=10				Cross validation K=15			
	<i>TC</i>	<i>SE</i>	<i>SP</i>	<i>Matrice de confusion</i>	<i>TC</i>	<i>SE</i>	<i>SP</i>	<i>Matrice de confusion</i>
K-NN	78.7000	59.7588	93.4196	258 176 38 532	78.0899	59.7588	92.5861	257 177 43 527
SVM	96.3139	96.1762	96.4979	417 17 20 550	96.2114	95.7360	96.3402	416 18 20 550
MLP	94.5218	93.2567	95.2527	414 20 26 544	96.0154	95.2049	96.9725	412 22 18 552
CART	99	97.6744	100	423 11 7 559	97.0149	96.4286	97.4359	423 11 9 561
Foret Aléatoire	98.1079	97.0631	98.9022	421 13 6 564	98.4065	97.2352	99.3274	422 12 4 566

Chapitre04 : Résultats obtenus et discussion

	Cross validation K=20				Cross validation K=25			
	<i>TC</i>	<i>SE</i>	<i>SP</i>	<i>Matrice de confusion</i>	<i>TC</i>	<i>SE</i>	<i>SP</i>	<i>Matrice de confusion</i>
K-NN	78.7824	62.4735	91.1613	273 161 52 518	78.7756	63.1853	91.4067	270 164 49 521
SVM	96.0137	95.5915	96.2968	415 20 20 549	96.6220	96.2739	97.0210	417 14 20 553
MLP	95.6098	94.5671	96.4296	411 23 21 549	95.6171	94.0704	96.6383	411 023 21 549
CART	98	100	96	423 11 6 564	97.5000	94.1176	100	423 11 7 564
Foret Aléatoire	98.1039	96.8060	98.8769	421 13 6 564	98.4049	97.2574	99.2839	422 12 4 566

(Tableau. V.1) les différents résultats expérimentaux obtenus.

Le tableau précédent résume tous les résultats expérimentaux obtenus : Taux de classification, la sensibilité, la spécificité et la matrice de confusion par les différents classifieurs après l'application de la validation croisée. Les résultats colorés sont les meilleurs pour chaque expérimentation.

Le **(Tableau.V.II)** suivant résume les paramètres utilisés pour l'obtention des meilleurs résultats des cinq différentes techniques utilisées dans ce travail.

K- Cross validation	K-NN		SVM	MLP			CART
	Distance	Nombre de voisins	Fonction Kernel	Fonction de transfert	Fonction d'activation	Architecture	Prune

k=2	Correlation	7	Linear	Tansig	Trainlm	10 12 3	On
k=3	Cosine	3	Linear	Tansig	Traingd	8 9 10	On
k=4	Cosine	7	Linear	Logsig	Trainlm	9 9 9 9	Off
k=5	Correlation	9	Linear	Logsig	Trainlm	9 9 9	Off
k=6	Correlation	9	Linear	Logsig	Trainlm	9 9 9	Off
k=7	Correlation	5	Linear	Purelin	Trainlm	8 9 10	On
k=8	Cosine	5	Linear	Purelin	Trainlm	8 9 10	On
k=9	Cosine	7	Linear	Tansig	Trainlm	10 12 13	On
k=10	Correlation	11	Linear	Tansig	Trainlm	10 12 13	On
k=15	Cosine	11	Linear	Tansig	Trainlm	8 9 10	Off
k=20	Correlation	7	Linear	Tansig	Trainlm	8 9 10	Off
k=25	Cosine	7	Linear	Tansig	Trainlm	8 9 10	Off

(Tableau.V.II) Les paramètres utilisés pour les meilleurs résultats de chaque expérimentation

VI Analyse des résultats :

Dans ce travail, nous avons réalisé plusieurs expérimentations (k=2 jusqu'à 25). Pour une cross validation supérieur à 25, l'exécution sera lente et parfois impossible (l'algorithme affiche une erreur pour la majorité des techniques de classifications utilisées).

Le tableau (Tableau. VI.I) suivant présente les meilleurs résultats obtenus avec les différentes méthodes de classification parmi toutes les expérimentations réalisées.

Chapitre04 : Résultats obtenus et discussion

	TC	SE	SP	K-cross validation
KNN	79.28%	64.22%	90.76%	8
SVM	96.62%	96.27%	97.02%	25
MLP	96.01%	95.20%	96.97%	9
CART	99.01%	98.80%	99.15%	5
RF	98.60%	97.34%	99.64%	9

(Tableau. VI.I) les meilleurs résultats obtenus avec les différentes méthodes de classification.

Dans le tableau suivant (VI.2) on observe bien que dans notre base de données figurent les meilleurs taux de classification avec un K=9

	TC
KNN	78.50%
SVM	95.62%
MLP	96.01%
CART	98.20%
RF	98.60%

(Tableau.VI.II) les meilleurs taux de classification pour k=9

En comparant et analysant le tableau précédent avec celui des meilleurs résultats nous remarquons bien que : MLP et les forêts aléatoires demeurent les excellents taux par rapport à toutes les autres expérimentations.

- KNN une différence de 0.78%.

-SVM une différence de 1%.

- CART une différence de 0.81 %.

En utilisant plusieurs techniques, on constate que le meilleur compromis est de choisir l'expérimentation K=9.

Le tableau qui suit résume les différents taux (sensibilité, spécificité, taux de faux négatifs et positifs) pour toutes les techniques de classification lors de l'expérimentation k=9.

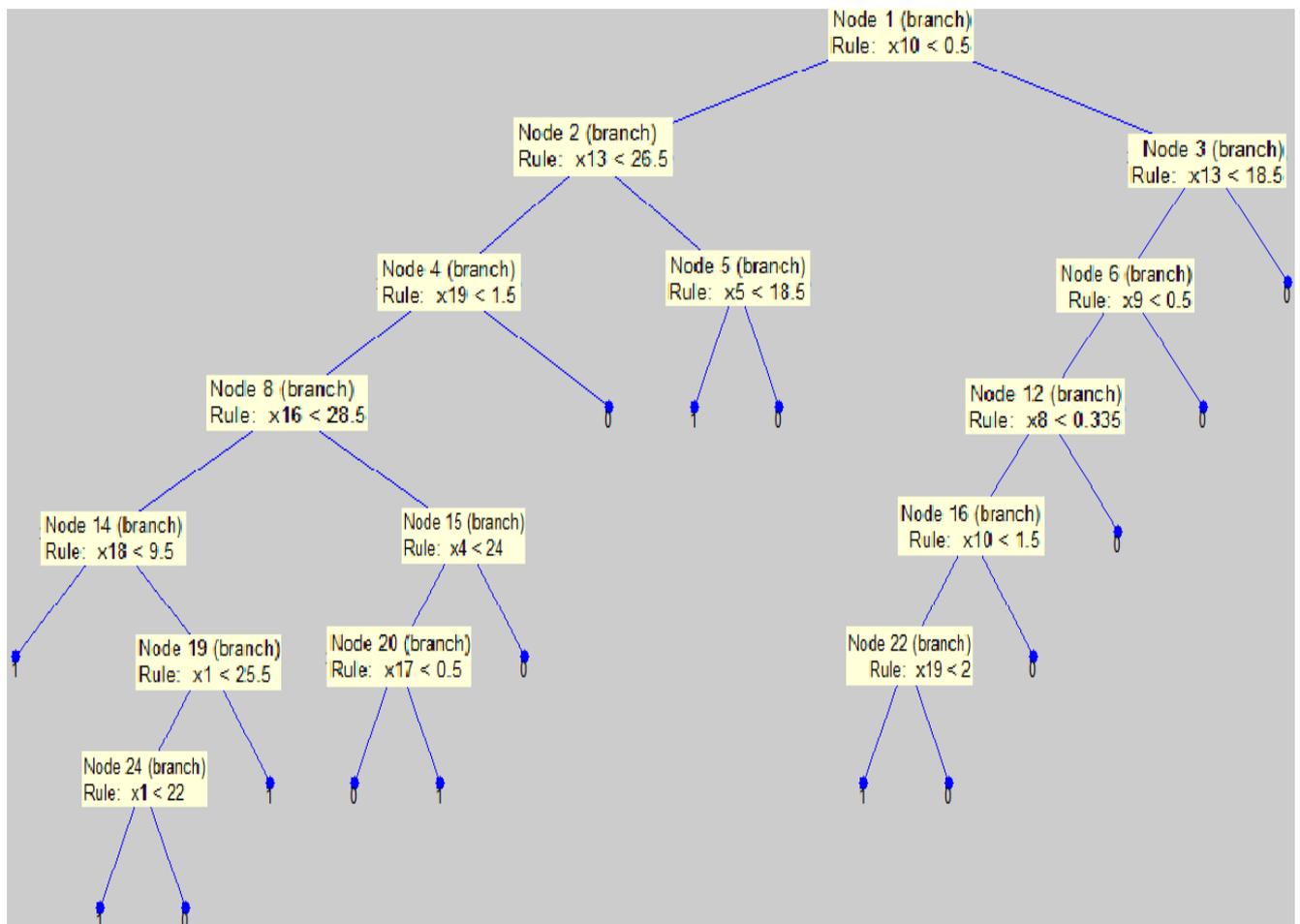
Chapitre04 : Résultats obtenus et discussion

	Sensibilité	Spécificité	Taux de faux négatifs	Taux de faux positifs
KNN	91.40%	61.52%	29.30%	11.29%
SVM	95.43%	95.85%	3.15%	5.99%
MLP	96.50%	95.39%	3.51%	4.61%
CART	98.42%	97.60%	1.75%	2.07%
RF	96.49%	96.08%	2.10%	0.46%

(Tableau. VI. III) Les mesures évoquées en utilisant la matrice de confusion.

A partir du tableau précédent on peut éliminer le classifieur KNN parce qu'il donne le plus faible taux de classification (ne dépasse pas les 80%) d'un côté et le taux de faux négatifs le plus élevé (29.30%).

La technique CART offre une interprétabilité. Nous allons voir l'arbre obtenue pour confirmer est ce que notre algorithme ressemble au raisonnement de l'expert. La figure suivante est une représentation de l'arbre obtenue lors de l'expérimentation k=9.



(Figure.VI) le meilleur arbre donné.

L'interprétation de l'arbre :

L'arbre obtenu offre plusieurs règles d'interprétation. Plusieurs répétitions sont présentes du fait que notre base de données contient 8 descripteurs différents de la tension intra oculaire. Ces descripteurs ne peuvent pas être supprimés car ils ont été choisis par les experts.

Pratiquement la totalité des règles sont juste mais on va prendre que 6 règles qui sont :

- Si la valeur des Cellules ganglionnaires est <0.5 et si la correction du tonus oculaire a applanation par la pachymétrie est <26.5 et la valeur de gonioscopie <1.5 et la correction du tonus oculaire jet d'air par le point le plus fin X16 <28.5 et si l'acuité visuel <9.5 alors la classe 1 non malade.
- Si la valeur des Cellules ganglionnaires est <0.5 et si la correction du tonus oculaire a applanation par la pachymétrie est ≥ 26.5 et la valeur du tonus oculaire jet d'air ≥ 18.5 alors la classe 0 malade.
- Si la valeur des Cellules ganglionnaires est <0.5 et si la correction du tonus oculaire a applanation par la pachymétrie est <26.5 et la valeur de gonioscopie <1.5 et la correction du tonus oculaire jet d'air par le point le plus fin ≥ 28.5 et tonus oculaire par applanation <24 et si hérédité existe alors la classe 0 malade.
- Si la valeur des Cellules ganglionnaires est <0.5 et si la correction du tonus oculaire a applanation par la pachymétrie est <26.5 et la valeur de gonioscopie <1.5 et la correction du tonus oculaire jet d'air par le point le plus fin ≥ 28.5 et tonus oculaire par applanation <24 et si hérédité n'existe pas alors la classe 1 non malade.
- Si la valeur des Cellules ganglionnaires est <0.5 et si la correction du tonus oculaire a applanation par la pachymétrie est <26.5 et la valeur de gonioscopie <1.5 et la correction du tonus oculaire jet d'air par le point le plus fin ≥ 28.5 et tonus oculaire par applanation <24 et si hérédité n'existe pas alors la classe 0 malade.
- Si la valeur des Cellules ganglionnaires est <0.5 et si la correction du tonus oculaire a applanation par la pachymétrie est <26.5 et la valeur de gonioscopie <1.5 et la correction du tonus oculaire jet d'air par le point le plus fin <28.5 et si l'acuité visuel ≥ 9.5 et si l'âge <25 et <22 alors la classe 1 non malade.

D'après la comparaison entre cet arbre avec les seuils donnés par un expert ophtalmologiste on constate que cette technique offre une bonne interprétabilité.

La performance des forêts aléatoires (RF) sur notre base de données a révélé d'excellentes performances au niveau de 500 arbres, la méthode d'ensemble RF est plus puissante par rapport aux

autres techniques de classifications. Cette décision est confirmée par le nombre d'expérimentations effectuées, par les taux de classification, sensibilité, spécificité et matrice de confusion obtenus.

VII.Conclusion :

Dans ce projet de fin d'étude, plusieurs techniques de classification ont été testées sur une nouvelle base de données collectée localement.

Ces techniques sont : les k-plus proches voisins, les machines à vecteurs de supports, les réseaux de neurones multicouches, les arbres de décision CART, ainsi que les forêts aléatoires.

Pour l'ensemble des expérimentations, les résultats obtenus sont très satisfaisants pour l'ensemble des techniques de classifications sauf pour les k-plus proches voisins.

Plusieurs critères d'évaluation ont été utilisés dans ce travail : le taux de classification, la sensibilité, la spécificité et la matrice de confusion.

Les meilleurs résultats ont été obtenus par l'algorithme CART avec un taux de classification de 99.01% et la méthode d'ensemble forêts aléatoires (RF) avec un taux de classification de 98.60% et ces résultats ont été confirmés par plusieurs critères d'évaluation.

L'analyse de l'interopérabilité de l'arbre de décision confirme la théorie pour la détection du glaucome.

Conclusion générale

Conclusion générale :

Les systèmes d'aide à la décision médicale sont des applications informatiques dont le but est de fournir aux cliniciens une aide à la décision. Aujourd'hui, beaucoup de chercheurs proposent des systèmes automatiques intelligents capable de donner un meilleur diagnostic, par la suite aider les médecins dans leurs routines cliniques

L'objectif de ce projet de fin d'étude entre dans le cadre du développement d'un système intelligent destiné aux ophtalmologues afin d'assurer d'une manière efficace la détection précoce du glaucome. La réalisation de ce travail porte sur l'utilisation d'une nouvelle base de données collectée localement dans le but d'un travail de recherche par plusieurs médecins ophtalmologues.

Le glaucome est une neuropathie optique progressive responsable d'une détérioration irréversible du champ visuel qui peut évoluer jusqu'à la cécité définitive. La maladie du glaucome est la deuxième cause de cécité dans le monde après la maladie de la cataracte.

Dans la première partie de cette étude, nous avons présenté les termes médicaux qui ont une relation avec la maladie du glaucome afin de ce familiarisé avec le domaine d'ophtalmologie. Les différents éléments constituant l'œil, la pathologie du glaucome ainsi que les différents examens de dépistage et les traitements.

La deuxième partie a été consacrée à la présentation d'un état de l'art ainsi qu'une présentation de notre nouvelle base de données utilisée.

Le but de la troisième partie est de donner l'aspect théorique des différentes techniques de classification utilisées qui sont : les K les plus proches voisins, les machines à vecteur de support, les réseaux de neurones, CART et les forêts aléatoires.

Les résultats obtenus sont très satisfaisants en temps d'exécution et le taux de classification pour l'ensemble des techniques sauf pour les k-plus proches voisin dont le taux de classification n'a pas dépassé 80%,.

Les meilleurs résultats ont été obtenus avec la technique des forêts aléatoire avec un taux de 98.60% ainsi que la technique CART avec un taux de 99.01% , la technique SVM avec un taux de 96.62%.

Conclusion générale

A partir de ces résultats nous pouvons dire que les arbres de décision CART et Forêts aléatoires forment des techniques fiables dans notre travail.

Les perspectives futures dans ce travail sont en premier lieu l'augmentation de la taille de la base de données.

En second lieu, utilisation de système de vote majoritaire.

La dernière est la création d'une interface.

Bibliographie :

- [1] : <http://www.larousse.fr/encyclopedie/medical/>
- [2] : Dr. BENOMAR Oussama, docteur en pharmacie, thèse de doctorat, prise en charge et évaluation du traitement médicamenteux du glaucome primitif à angle ouvert, soutenue : Le 13 juillet 2006.
- [3] : Mme. FEROUJ Amel .analyse des images couleur du fond d'œil pour l'aide au diagnostic en ophtalmologie application à la détection des pathologies rétiniennees.
. Thèse Pour obtenir le titre de Docteur en Electronique Biomédicale. En 2014
P 26 à 46.
- [4] : <https://www.google.dz/search?q=anatomie+de+l%27oeil>
- [5] : Lindsay I Smith. Principal Components Analysis.2002.
- [6] : H. Bresson-Dumont, A. Le Callonnec, G. Peigné, P.Y. Santiago, X. Zanlonghi. Glaucome. Glaucome, Des cas cliniques pour y voir plus clair, Simple Médecine, Paris, 2004. ISBN 2-915739-00-5
- [7] : Directeur Professeur Yves MATILLON. Glaucome et basse vision. Diplome universitaire de basse vision par PRIGENT Cindy . LYON, le 23 mai 2014. N° 102
- [8] : [https://www.google.dz/imgres?imgurl=http%3A%2F%2Ffondationglaucomequebec.com%](https://www.google.dz/imgres?imgurl=http%3A%2F%2Ffondationglaucomequebec.com%2F)
- [9] : Dr. Jacques Laloum, chirurgien ophtalmologiste, spécialiste du glaucome, Fondation Rothschild: chirurgie du glaucome, paris <http://www.glaucomes.fr>
- [10] : Philippe Berros. John Contrath. Eve Hadjadj & Dan Milea. livre OPHTALMOLOGIE. Imprimé en France –JOUVE, 18, rue Saint-Denis,75001 PARIS N 273575H.Dépot légal : Octobre 1999
- [11] : <http://www.chu-lyon.fr/web/3942>
- [12] : <http://www.glaucomaresearch.ca/fr/index.shtml>
- [13] : www.rfi.fr/emission/20160509-le-glaucome
- [14] : www.ouest-france.fr
- [15] : le président de l'association algérienne de lutte contre la cécité aussi professeur en ophtalmologie et chef de service de cette spécialité au CHU BéniMessous Mr , Mohamed Tahar Nouri et professeur Rachid Garout l'Hôpital central de l'armée, article glaucome : deuxième cause de cécité dans le pays à Oued Souf, page1. Publié le 20/02/2017

: <http://www.libertealgerie.com>

[16] : [Dr. med. P.W. Hasler et Prof. Dr. S. Orgül, avec la collaboration de: Dr. med. H.Vogten et Mme D. Haegeli

Les illustrations proviennent de: «J. Flammer: Glaukom. 2. Überarbeitete Auflage 2001.

<http://swissglaucome.ch/CMS/fr-FR/>

[17] : <http://www.glaucomes.fr>

[18] : Dr. Mamisoa ANDRIANTAFIKA est spécialisé en chirurgie réfractive (myopie , astigmatisme , hypermétropie et presbytie), chirurgie. Centre d'ophtalmologie Bruxelles. lundi, 15 avril 2013 02:29

[19] : <http://www.leglaucome.fr/consulter/les-examens/>

[20] : <http://www.setyoursights.com/fr/glaucome>

[21] : Dr.Jérôme MADAR. Ancien Interne des Hôpitaux de. Marseille, Paris, Lyon.Centre d'Ophtalmologie La Ciotat - Jérôme Madar

[22] : <http://www.cehjournal.org>

[23] : Colleen M. Story. Reviewed by George Krucik, MD. Healthline. La tonométrie
Published on 25 juillet 2012

<http://fr.healthline.com/health>

[24]: -<http://www.leglaucome.fr> (A, B) ---<http://www.medicalexpo.fr>

[25] : EBC EUROPE équipe les cabinets d'ophtalmologistes, les cliniques et les hôpitaux en instruments de diagnostic, de traitements et de chirurgie.

<http://www.ebc-europe.com>

[26]: <http://www.centre-paufique.fr/equipement-medical>

[27] : Quinze-Vingts. Centre Hospitalier National d'Ophtalmologie. 28, rue de Charenton 75 571. PARIS. www.15-20.fr

[28]: Settouti Nesma, Bechar Mohammed El Amine. Les Forêts Aléatoires en Apprentissage Semi-Supervisé (Co-Forest) pour la segmentation des images rétiniennes. Biomedical Engineering Laboratory, Tlemcen University Algeria.<hal-01294064> 2015.

[29]: Joseph Caprioli i, M.D, Dr. Brigatti, Neural Networks to Identify Glaucoma With Structural and Functional Measurements. Department of Ophthalmology and Visual

Science, Yale University School of Medicine, New Haven. Volume 121, Issue 5, Pages 511–521.mai 1996.

[30] : Christopher Bowd, Kwokleung Chan, Linda M. Zangwill, Michael H. Goldbaum, Te-Won Lee, Terrence J. Sejnowski, Robert N. Weinreb. Comparing Neural Networks and Linear Discriminant Functions for Glaucoma Detection Using Confocal Scanning Laser Ophthalmoscopy of the Optic Disc. *Investigative Ophthalmology & Visual Science*, Vol.43, 3444-3454, November 2002.

[31] : Rajan, Ramesh.G.P, Comparative Study of Glaucomatous Image Classification Using Optical Coherence Tomography, Department of Electronics and Communication Engineering, St,peter's University, *Int. J. Pharm. Sci. Rev. Res.*, 36(1); Article No. 49, Pages: 277-280, Avadi, Tamilnadu, India. Février 2016.

[32] : Christopher Bowd, Felipe A. Medeiros, Zuohua Zhang, Linda M. Zangwill, Jiucang Hao, Te-Won Lee, Terrence J. Sejnowski, California; the, Robert N. Weinreb, Michael H. Goldbaum, Relevance Vector Machine and Support Vector Machine Classifier Analysis of Scanning Laser Polarimetry Retinal Nerve Fiber Layer Measurements. *Investigative Ophthalmology & Visual Science* , Vol.46, 1322-1329. doi:10.1167/iovs.04-1122, Avril 2005.

[33] : F. Fink, K. Worle, P. Gruber, A. M. Tome, J. M. Gorriz - Saez, C. G. Puntonet, E. W. Lang. ICA analysis of retina images for glaucoma classification, *Engineering in Medicine and Biology Society*, 2008. EMBS 2008. 30th .

[34] : Yoshida T, Iwase A, Hirasawa H, Murata H, Mayama C, Araie M, et al. Discriminating between Glaucoma and Normal Eyes Using Optical Coherence Tomography and the ‘Random Forests’ Classifier. *PLoS ONE* 9(8): e106117. doi:10.1371/journal.pone.0106117. 2014

[35] Mr Lazouni Mohammed El Amine . These de Doctorat pour obtenir le grade de Docteur en Electronique Biomédicale Spécialité : Aide Au Diagnostic Pour Un Médecin Anesthésiste Réanimateur. Le 13 Mai 2014

[36] CHAMROUKHI Faicel. Classification supervisée : Les K-plus proches. Université du Sud Toulon .2012-2013

[37]. Brice ZIRAKIZA Forêts Aléatoires PAC-BAYÉSIENNES Mémoire Maîtrise en informatique Maîtrès sciences (M.Sc.) Québec, Canada © Brice ZIRAKIZA, 2013

[38]. *MOHAMADALLY HasanFomani. Machines à Vecteurs de Support our Séparateurs à Vastes. 16 janvier 2006*

[39]. <http://www.grappa.univ-lille3.fr/polys/apprentissage/sortie005.html>

[40]. <https://fr.wikipedia.org/wiki/Perceptron>

[41] Mous Amina et Mostafi Meriem .Mémoire de projet de fin d'étude Pour obtenir le Diplôme de MASTER en Génie Biomédical Spécialité: Electronique biomédicale: Reconnaissance des données kératocône par une méthode intelligente Soutenu. Le 30juin2013

[42]<http://mapdata.science/~stpierre/reseaux-neuronaux/node13.html>

[43] Md. Kalache Soumia Melle. Kouloughli Asma .Mémoire de projet de fin d'étude Pour obtenir le grade de Master En Informatique Spécialité : Les Forêts Aléatoires Floues.02Juillet2013

[44] . *Gérard Biau et de Jean-Philippe. Mémoire de diplôme de l'ENS :Apprentissage et forêts aléatoires.14 octobre 2012*

[45]Marine Campedel, Classification supervisée, Telecom Paris école nationale des télécommunications, Avril 2005. www.tsi.enst.fr/~campedel

[46]Classification supervisée, Aperçu de quelques méthodes avec le logiciel R, [www.math.univangers.fr/~labatte/enseignement%20UFR/.../classificationsupervisee.13 novembre 2012](http://www.math.univangers.fr/~labatte/enseignement%20UFR/.../classificationsupervisee.13novembre2012)

[47]https://fr.wikipedia.org/wiki/Apprentissage_automatique. 3 juin 2017

[48]https://lagunita.stanford.edu/c4x/HumanitiesScience/StatLearning/asset/cv_boot.pdf

[49]GIRARD M.C. & GIRARD C.M. Traitement des données de télédétection. Paris, **pages 328 à 333.**] 1999

[50]YANNICK Oufella, Évolution du concept de front ROC et combinaison de classifieur, Master Génie Electrique et informatique industrielle, université de Rouenfr des sciences et techniques, 2 septembre 2008

[51]Sensibilité et spécificité – Définition, Docteur Pierrick HORDÉ, Sante-Médecine.
<http://sante-medecine.journaldesfemmes.com/faq/23726-sensibilite-et-specificite-definition> . Juin 2014

[52]Gilbert Saporta, Conservatoire National des Arts et Métiers, Sensibilité, spécificité, courbe ROC etc, Ecole pasteur-Cnam de santé publique, octobre 2012