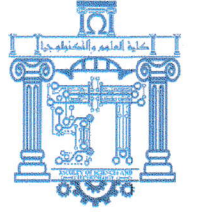




الجمهورية الجزائرية الديمقراطية الشعبية
Republique Algerienne Democratique Et Populaire
وزارة التعليم العالي والبحث العلمي



Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

جامعة العربي التبسي - تبسة

Université Larbi Tébessi – Tébessa –

Faculté des Sciences et de la Technologie

Département de Génie Electrique

MEMOIRE

Présenté pour l'obtention du diplôme de Master Académique

En : Instrumentation

Spécialité : Electronique

Par : KHALDI Imene Hanene et ZERIFI Djihene



Sujet

Systeme de Classification Automatiques des Lymphomes

Présenté et évalué, le 06/06/ 2023, par le jury composé de :

M. Mohamed SAIGAA

MCB

Président

M. Abdallah MERAOUMIA

Prof

Rapporteur

M. Lotfi HOUAM

MCB

Examineur

Promotion : 2022/2023

Remerciement

Merci et louange à Dieu Tout-Puissant d'abord pour la bénédiction de la patience et la capacité d'accomplir le travail. Louange à Dieu pour ces bénédiction.

Même exprimons nos remerciements à dieu d'abord qui nous a donné la force et la puissance pour finaliser ce mémoire En second lieu, nous tenons à remercier du fond du cœur, notre encadreur Mr Meraoumia Abdallah d'avoir dirigé ce travail avec beaucoup de patience, d'avoir toujours été présent quand on avait besoin, de nous avoir offert tous ses précieux conseils et de nous avoir encouragé et motivé à donner le meilleur de nous-même.

Nous espérons que ce travail sera à la hauteur de ses espérances Nous remercions le de juré qui a accepté de juger notre travail ainsi que tous les enseignants qui ont contribué à notre formation.

Enfin, Nous remercions aussi tous nos famille amis et collègues qui nous ont soutenu et tous ceux qui ont contribué de près ou de loin à la réalisation de ce travail.

S. Khaldi et D. Zerifi

Dédicace

Je remercie dieu de m'avoir donné la force de faire ce travail. Toutes les lettres ne sauraient trouver les mots qu'il faut. . . Tous les mots ne sauraient exprimer la gratitude, l'amour, le respect, la reconnaissance. . . Aussi, C'est tout simplement que je dédie ce modeste mémoire à :

Mes parents, Boubaker et Linda, les merveilles de mon monde, aucun mot ne pourrait décrire l'amour que je porte pour eux. Votre patience sans fin, votre compréhension et votre encouragement sont pour moi le soutien indispensable que vous avez toujours su m'apporter.

Que Dieu, le tout puissant, vous préserve, vous accorde santé, bonheur, quiétude de l'esprit et vous protège de tout mal. Mes adorables frères Hamed et Mohamed, ma sœur Asma et ma belle-sœur Kawter, sans qui ma vie serait quand même bien ennuyeuse.

En témoignage de mon amour et de ma grande affection, je vous prie de trouver dans ce travail l'expression de mon estime et mon sincère attachement. Je prie Dieu, le tout puissant, pour qu'il vous donne bonheur et prospérité.

A tous les membres des familles : Ma grand-mère, mes oncles et mes tantes, mes cousins et mes cousines, Puisse ce

travail témoigner de ma profonde affection et de ma sincère estime.

Mes amis, mes proches, et tous les gens qui me soutiennent au quotidien et en particulier à Aïmene que dieu te garde et te protège, merci pour tous ces fous rires.

My team et binôme Djihene pour avoir été une super équipier et amie pendant cette année.

A tous mes enseignants et toutes mes enseignantes, Il vous revient le mérite de m'avoir prodigué un enseignement profitable et une formation complète. Veuillez accepter mes remerciements les plus sincères.

Kaldi Iemene Hanene

Dédicace

Je remercie Dieu, le Très-Haut, le Tout-Puissant, qui m'a accordé avec sa grâce l'esprit et la religion qui dit dans la révélation décisive, "Et au-dessus de tous ceux qui ont la connaissance, l'Omniscient", qui m'a donné la force pour faire ce travail.

Toutes les lettres ne peuvent pas trouver les mots justes - tous les mots ne peuvent pas exprimer la gratitude, l'amour, le respect et la reconnaissance. De plus, je dédie simplement cet humble souvenir à :

Mes parents, Toumi et Habiba, les merveilles de mon monde,

Je dédie cette graduation à celui qui m'a encouragé à persévérer tout au long de ma vie et à l'homme le plus en vue de ma vie, « Papatí », j'espère qu'il prolongera sa vie et me verra dans les plus hautes sphères.

Et à qui je m'élève au-dessus et dont je dépends, elle est le secret de ma réussite tout au long de mon parcours universitaire, « Mamati ». Et j'espère que Dieu me la préservera.

*Mes frères bien-aimés Samí Chams al-Dín Hamza
Mouataz Billah Mountasser, sans qui ma vie resterait si terne
et dénuée de sens. Et à toute la famille de mon père. Et à
toute la famille de ma mère.*

*Et à tous mes amis, connaissances et tous ceux qui m'ont
consolé dans ce voyage, Ikram Khawla Ghada Nadjette Iman
Asmaa et tous mes amis de la cité universitaire.*

*Mon équipe et mon amie Imene Hanene pour avoir été
une collègue et une amie incroyable cette année.*

*À tous mes professeurs, vous méritez le mérite de m'avoir
donné une éducation fructueuse et une formation approfondie.
Veuillez accepter mes sincères remerciements. "en particulier
au professeur Djarí"*

*Enfin, tous ceux qui m'ont soutenu et m'ont tendu la main
dans ma carrière universitaire.*

Zerifi Djihene

Table des Matières

Remerciement	i
Dédicace	ii
Table de Matières	vi
Liste des tableaux	ix
Listes des figures	x
Introduction Générale	1
CHAPITRE 1 : DIAGNOSTIC ASSISTE PAR ORDINATEUR	
I.1 Introduction	3
I.2 Définition	3
I.2.1 Objectifs et avantages	4
I.2.2 Historique du développement du DAO	5
I.2.3 Impact sur la prise de décision médicale	7
I.3 Principe de fonctionnement	8
I.3.1 Collecte et préparation des données	8
I.3.2 Extraction de caractéristique	9
I.3.3 Entraînement de modèle (apprentissage)	9
I.3.4 Évaluation et validation du système	10
I.3.5 Intégration des résultats dans le processus de diagnostic	10
I.4 Algorithmes d'apprentissage	11
I.4.1 Réseaux de neurones	11
I.4.2 Machines à vecteurs de support	11
I.4.3 Forêts aléatoires	12
I.4.4 Systèmes experts	12
I.4.5 Méthodes de classement	12
I.5 Fusion de données multimodales	13
I.5.1 Fusion d'images médicales	13
I.5.2 Fusion de données cliniques et d'imagerie	13
I.5.3 Fusion de données génomiques et d'imagerie	13

I.5.4	Fusion de données temporelles	13
I.6	Domaines d'application	14
I.6.1	Imagerie médicale	14
I.6.2	Pathologie numérique	14
I.6.3	Analyse de données cliniques	14
I.7	Exemples de systèmes de DAO existants	15
I.7.1	CAD-RADS	15
I.7.2	CADx	15
I.7.3	MIA	15
I.7.4	DLS	15
I.8	Défis et perspectives du DAO	15
I.8.1	Limitations et problèmes éthiques liés au DAO	16
I.8.2	Tendances et développements futurs du DAO	17
I.9	Conclusion	18

CHAPITRE 2 : EXTRACTION DES CARACTERISTIQUES

II.1	Introduction	19
II.2	Objectifs et importance	19
II.3	Rôle de l'extraction de caractéristiques	19
II.3.1	Réduction de la dimensionnalité	20
II.3.2	Identification des informations pertinentes	20
II.3.3	Amélioration de la précision diagnostique	20
II.3.4	Optimisation de la prise de décision	20
II.4	Types de l'extraction de caractéristiques	20
II.4.1	Caractéristiques liées aux lignes	21
II.4.2	Caractéristiques liées aux textures	21
II.4.3	Caractéristiques liées aux Apparence	21
II.5	Méthodes d'extraction de caractéristiques	21
II.5.1	Méthodes traditionnelles	22
II.5.2	Méthodes traditionnelles apprises	22
II.5.3	Méthodes profondes	23
II.6	Méthode proposée	23
II.6.1	Fondement théorique	23
II.6.2	Caractéristique profonde	29

II.6.3	Classifier	32
II.7	Conclusion	33
CHAPITRE 3 : RESULTATS EXPERIMENTAUX		
III.1	Introduction	34
III.2	Contexte de travail	34
III.2.1	Cancer lymphatique	35
III.2.2	Sous-types de lymphome	35
III.3	Ensemble de données	37
III.3.1	Motivation	37
III.3.2	Description	38
III.4	Configurations de test	38
III.4.1	Protocole de tests	38
III.4.2	Implémentation	39
III.5	Evaluation des performances	39
III.5.1	FF-LPQ	40
III.5.2	FG-LPQ	42
III.5.3	FO-LPQ	45
III.6	Discussion	47
III.6.1	Principales conclusions	47
III.6.2	Comparaison avec les méthodes précédentes	48
III.7	Conclusion	48
Conclusion Générale		49
Annexe A	Évaluation des Performances	50
Glossaire		55
Références		57
Résumé		61

Liste des tableaux

	Tableau	Page
III.1	Performance de système sous la configuration FF-LPQ (format NG)	40
III.2	Performance de système sous la configuration FF-LPQ (format RGB)	41
III.3	Performance de système sous la configuration FG-LPQ (format NG)	43
III.4	Performance de système sous la configuration FG-LPQ (format RGB)	44
III.5	Performance de système sous la configuration FO-LPQ (format NG)	45
III.6	Performance de système sous la configuration FO-LPQ (format RGB)	46
III.7	Comparaison des performances avec des travaux récents dans l'état de l'art	48

Liste des figures

	Figure	Page
I.1	Signal électrocardiogrammes (ECG)	6
I.2	Techniques d'imagerie médicale.	6
I.3	Apprentissage automatique et le DAO	6
I.4	Apprentissage profond et le DAO	7
I.5	Collecte et préparation des données médicales	9
I.6	Extraction de caractéristiques	9
I.7	Entraînement de modèle	10
I.8	Intégration des données et des résultats dans le processus de diagnostic	10
I.9	Réseaux de neurones convolutifs	11
I.10	Machine à vecteurs de support	11
I.11	Forêts aléatoires	12
I.12	Systèmes experts	12
I.13	Algorithme de k -means	12
II.1	Méthode traditionnelle d'extraction de caractéristiques.	22
II.2	Méthode traditionnelle apprise d'extraction de caractéristiques.	22
II.3	Méthode profonde d'extraction de caractéristiques.	23
II.4	Extraction de caractéristiques basée sur le LPQ	24
II.5	Extraction de caractéristiques.	30
III.1	Leucémie Lymphoïde Chronique	35
III.2	Lymphome à cellules du manteau	36
III.3	Lymphome folliculaire	37
III.4	Sous-types de lymphome dans l'ensemble de données de test	38
III.5	Performance de système sous la configuration FF-LPQ	40

III.6	Performance de système sous la configuration FF-LPQ pour les différentes bandes de couleur.	41
III.7	Performance de système sous la configuration FG-LPQ.	43
III.8	Performance de système sous la configuration FO-LPQ.	46
A.1	Receiver Operating Characteristic (ROC)	55
A.2	Area Under the ROC Curve	56
A.3	Precision-Recall Curve	56

Introduction

G é n é r a l e

Introduction Générale

L'imagerie médicale est devenue un outil de diagnostic essentiel pour la plupart des maladies. Les informations obtenues à partir des images jouent un rôle crucial dans la prise de décision à différentes étapes des soins, telles que la détection, la caractérisation, l'évaluation de la réponse au traitement et le suivi des patients, ainsi que dans l'orientation des interventions chirurgicales et de la radiothérapie. Malgré les progrès importants réalisés dans les dispositifs d'imagerie médicale, l'analyse et l'interprétation des images capturées restent un défi majeur pour les radiologues, car toute erreur d'interprétation peut entraîner la prescription de médicaments incorrects, mettant ainsi la vie du patient en danger. En effet, en raison du manque d'expérience de nombreux radiologues et de la quantité élevée de cas examinés périodiquement, le processus de traitement devient coûteux en raison du nombre élevé d'erreurs. Par conséquent, afin d'éviter ou de réduire les erreurs de diagnostic, les radiologues ont besoin d'un moyen d'aide pour prendre les bonnes décisions.

Le diagnostic médical basé sur l'imagerie médicale a été l'un des domaines qui ont considérablement bénéficié du développement technologique. Ces dernières années, les entreprises médicales ont travaillé pour produire des dispositifs d'imagerie médicale de haute qualité offrant des résultats très précis. Ainsi, ces dispositifs sont utilisés pour détecter et classer de nombreuses maladies, telles que les cancers. Le lymphome est l'un des cancers que les chercheurs se sont efforcés de détecter à l'aide de techniques d'imagerie médicale. Le lymphome est un type de cancer qui affecte le système lymphatique, un composant essentiel de notre système immunitaire. Il existe différents sous-types de lymphome, tels que la

leucémie lymphoïde chronique (LLC), le lymphome folliculaire (LF) et le lymphome du manteau (LM), qui nécessitent une identification précise pour un traitement approprié. En effet, la classification manuelle des lymphomes peut être complexe et sujette à des variations d'interprétation entre les pathologistes, de sorte que les techniques de diagnostic informatisées peuvent jouer un rôle important dans les résultats du diagnostic. Dans ce mémoire, avant d'appliquer directement la méthode d'extraction des caractéristiques traditionnelles à l'image brute, elle sera analysée en détail à l'aide d'un ensemble spécifique de filtres pondérés par un système chaotique (carte logistique). Ces filtres extraient plusieurs modalités, qui sont ensuite combinées en un seul descripteur. Ensuite, un algorithme d'extraction de caractéristiques traditionnelles (LPQ) sera appliqué au descripteur obtenu. De plus, le classifieur optimise les paramètres des filtres à l'aide d'une optimisation basée sur un algorithme inspiré par les chauves-souris afin d'obtenir des performances optimales et un comportement d'analyse approfondi. En utilisant l'ensemble de données de l'INA, les résultats expérimentaux montrent que l'approche d'analyse de texture profonde pour les images de lymphome est significativement plus précise que les recherches précédentes.

Pour atteindre notre objectif, ce manuscrit est organisé en trois chapitres :

Dans le **premier chapitre**, nous allons d'abord introduire les concepts fondamentaux du Diagnostic Assisté par Ordinateur (DAO). Ensuite, nous aborderons un aperçu des techniques avancées utilisées en DAO, ainsi que les défis et les perspectives associés.

Dans le **deuxième chapitre**, nous allons d'abord donner un aperçu des techniques d'extraction de caractéristiques. Ensuite, nous aborderons notre technique proposée pour l'extraction de caractéristiques afin qu'elle puisse être utilisée dans une chaîne de DAO.

Le **troisième chapitre** donne les résultats de nos expériences et les analyses qui en découlent. L'objectif de ce chapitre est de fournir une évaluation objective de notre approche proposée et d'analyser sa performance dans le contexte de notre étude

Enfin, une **conclusion générale** avec des futures perspectives que nous envisagerons est donnée à la fin de cette mémoire.

Chapitre 1

DIAGNOSTIC ASSISTÉ PAR ORDINATEUR

Résumé

Diagnostic Assisté par Ordinateur (DAO) est une discipline en plein essor qui exploite la puissance de l'intelligence artificielle et de l'analyse des données pour soutenir les professionnels de la santé dans leurs processus de diagnostic. Grâce à l'utilisation d'algorithmes sophistiqués et de techniques d'apprentissage automatique, le DAO analyse de vastes ensembles de données médicales afin de fournir des suggestions et des évaluations complémentaires aux médecins, améliorant ainsi la précision et l'efficacité des diagnostics. Dans ce chapitre, nous allons d'abord introduire les concepts fondamentaux du DAO. Ensuite, nous aborderons un aperçu des techniques avancées utilisées en DAO, ainsi que les défis et les perspectives associés.

I.1 Introduction

I.2 Définition

I.3 Principe de fonctionnement

I.4 Algorithmes d'apprentissage

I.5 Fusion de données multimodales

I.6 Domaines d'application

I.7 Exemples de systèmes de DAO existants

I.8 Défis et perspectives du DAO

I.9 Conclusion

Diagnostic assisté par ordinateur

I.1 Introduction

Le Diagnostic Assisté par Ordinateur (DAO) est une technologie utilisée dans le domaine médical pour aider les professionnels de la santé à établir des diagnostics plus précis et rapides. Il s'appuie sur des algorithmes et des modèles statistiques pour analyser des données médicales telles que des images radiologiques, des résultats de laboratoire et des antécédents médicaux. Le DAO permet de détecter des anomalies, d'identifier des pathologies et de proposer des recommandations aux médecins, améliorant ainsi la précision et l'efficacité du processus de diagnostic [1]. Grâce à cette technologie, les praticiens peuvent prendre des décisions éclairées et offrir des soins de santé plus efficaces aux patients. Dans ce chapitre, nous allons d'abord introduire les concepts fondamentaux du DAO. Ensuite, nous aborderons un aperçu des techniques avancées utilisées en DAO, ainsi que les défis et les perspectives associés.

I.2 Définition

Le DAO est un domaine en pleine expansion dans le domaine médical, offrant des possibilités prometteuses pour améliorer la précision et l'efficacité des diagnostics cliniques. Le DAO désigne l'utilisation de technologies informatiques et d'algorithmes pour soutenir les professionnels de la santé dans le processus de diagnostic médical [2]. Il repose sur l'analyse et l'interprétation de données médicales telles que des images médicales, des résultats de tests ou des antécédents médicaux. Le DAO vise à améliorer la précision diagnostique en fournissant des informations complémentaires, en détectant des anomalies ou en proposant des suggestions diagnostiques aux médecins. En combinant l'expertise humaine et les capacités de l'ordinateur, le DAO permet d'accélérer le processus de diagnostic et d'améliorer la qualité des soins prodigués aux patients.

I.2.1 Objectifs et avantages

Les objectifs du DAO sont multiples et visent à améliorer le processus de diagnostic médical. L'objectif principal du DAO est d'améliorer l'exactitude, la rapidité et la fiabilité du diagnostic. En utilisant des méthodes informatiques avancées, le DAO peut aider à détecter des modèles ou des caractéristiques diagnostiques qui pourraient échapper à l'observation humaine, et à évaluer objectivement les informations diagnostiques disponibles. Cela peut conduire à une meilleure détection précoce des problèmes ou des maladies, à une réduction des erreurs de diagnostic et à une optimisation des traitements ou des interventions.

I) Objectifs de DAO: Le DAO a plusieurs objectifs, qui peuvent être résumés comme [3]:

- **Amélioration de la précision diagnostique :** Le DAO vise à aider les professionnels de la santé à obtenir des diagnostics plus précis en fournissant des informations complémentaires basées sur l'analyse des données médicales.
- **Accélération du processus de diagnostic :** En automatisant certaines tâches d'analyse et d'interprétation des données médicales, le DAO permet de gagner du temps et d'accélérer le processus de diagnostic, ce qui peut être crucial pour les patients nécessitant une réponse rapide.
- **Détection précoce des maladies :** Le DAO possède la capacité de détecter des anomalies subtiles ou de repérer les signes précurseurs de maladies, ce qui permet une intervention précoce et une prise en charge améliorée des patients. Cette fonctionnalité joue un rôle crucial dans l'identification précoce des maladies, facilitant ainsi des traitements plus efficaces et des résultats de santé positifs pour les patients.
- **Aide à la prise de décision :** Le DAO est un précieux outil d'aide à la décision pour les professionnels de la santé, en particulier lorsqu'ils sont confrontés à des cas complexes ou ambigus. Grâce à ses capacités d'analyse avancées et à ses algorithmes sophistiqués, le DAO est en mesure de fournir des suggestions et des recommandations diagnostiques précieuses. En combinant l'intelligence artificielle et l'expertise médicale, le DAO améliore la qualité des décisions cliniques, permettant aux médecins de bénéficier d'une assistance fiable et de prendre des décisions plus éclairées pour des résultats optimaux pour les patients.
- **Réduction des erreurs de diagnostic :** Grâce à son analyse objective des données médicales, le DAO contribue à minimiser les erreurs de diagnostic et à améliorer la

précision des résultats. En fournissant une évaluation complémentaire et en évitant les biais cognitifs, le DAO permet aux médecins de prendre des décisions plus éclairées, renforçant ainsi la qualité des soins prodigués aux patients. Cette collaboration entre l'intelligence artificielle et l'expertise humaine joue un rôle essentiel dans l'amélioration globale de la fiabilité des diagnostics médicaux.

2) Avantages du DAO : Le DAO offre de nombreux bénéfices, qui peuvent être récapitulés de la manière suivante [4] :

- **Précision améliorée :** Grâce à l'analyse automatisée des données, le DAO peut fournir des informations plus précises, réduisant ainsi les risques de diagnostic erroné ou de sous-diagnostic.
- **Efficacité accrue :** En automatisant certaines tâches de diagnostic, le DAO permet aux professionnels de la santé de consacrer plus de temps à l'interaction avec les patients et à la prise de décisions cliniques.
- **Standardisation des pratiques :** Le DAO permet d'établir des protocoles et des critères de diagnostic cohérents, réduisant ainsi les variations dans les pratiques médicales et favorisant une meilleure uniformité des soins.
- **Accès à une expertise spécialisée :** Le DAO peut offrir aux professionnels de la santé un accès à une expertise spécialisée et à des connaissances médicales avancées, même dans des domaines spécifiques où ils pourraient avoir moins d'expérience.

En combinant l'intelligence artificielle (IA) et l'expertise clinique, le DAO présente de multiples avantages pour atteindre divers objectifs tant pour les professionnels de la santé que pour les patients. Il améliore la qualité des diagnostics en fournissant des analyses précises et fiables, tout en favorisant des soins plus efficaces et adaptés à chaque individu. Cette convergence entre technologie et expertise médicale ouvre de nouvelles perspectives pour des traitements plus personnalisés et une prise en charge globale et optimisée.

I.2.2 Historique du développement du DAO

Le DAO a connu une évolution remarquable au fil des décennies, bénéficiant des avancées significatives dans les domaines de l'IA et de l'informatique médicale. L'histoire du développement du DAO remonte aux années 1950, marquées par les premières tentatives visant à utiliser les ordinateurs pour soutenir le processus de diagnostic médical [5].

☑ **1950-1960 :** Les premiers travaux dans le domaine du DAO ont été basés sur l'utilisation d'ordinateurs pour l'interprétation des électrocardiogrammes (ECG). Des algorithmes ont

été développés pour détecter les anomalies cardiaques et aider les médecins dans leurs diagnostics.



Fig. I.1 : Signal électrocardiogrammes (ECG)

☑ **1970-1980** : Le développement des techniques d'imagerie médicale, telles que la tomodensitométrie (TDM) et l'imagerie par résonance magnétique (IRM), a ouvert de nouvelles perspectives pour le DAO. Des algorithmes ont été créés pour l'analyse automatisée des images médicales et la détection de lésions ou de structures anormales.



Fig. I.2 : Techniques d'imagerie médicale. (a) Tomodensitométrie (TDM) et (b) Résonance magnétique (IRM)

☑ **1990-2000** : L'émergence des techniques d'apprentissage automatique a joué un rôle majeur dans le développement du DAO. Les approches basées sur l'intelligence artificielle, telles que les réseaux de neurones et les systèmes experts, ont été largement adoptées pour améliorer la précision des diagnostics dans différentes spécialités médicales. Ces avancées technologiques ont permis d'exploiter efficacement les données médicales et d'obtenir des résultats des diagnostics plus précis et fiables.



Fig. I.3 : Apprentissage automatique et le DAO

☑ **2010-2020** : Les progrès rapides dans le domaine de l'apprentissage automatique, en particulier dans le domaine de l'apprentissage profond (*deep learning*), ont constitué une

avancée majeure pour le DAO. Les modèles de DAO basés sur l'apprentissage profond ont démontré des performances remarquables, surpassant parfois celles des experts humains dans certains domaines spécifiques. Ces avancées technologiques ont ouvert de nouvelles perspectives en termes de capacités de diagnostic et ont permis d'améliorer la précision et la fiabilité des résultats obtenus par le DAO.

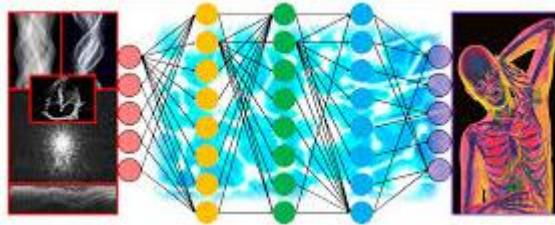


Fig. I.4 : Apprentissage profond et le DAO

Aujourd'hui, le DAO continue de se développer avec l'émergence de nouvelles technologies, telles que l'apprentissage par renforcement et l'intégration de données multimodales (fusion de données). Les progrès en matière de connectivité et de partage de données médicales ouvrent également la voie à une utilisation plus répandue du DAO dans les pratiques cliniques.

I.2.3 Impact sur la prise de décision médicale

L'impact du DAO sur la prise de décision médicale est significatif et peut apporter plusieurs avantages aux médecins. Les façons dont le DAO peut influencer la prise de décision médicale sont [6] :

Assistance à l'interprétation des données : Le DAO peut aider les médecins à interpréter les données médicales en fournissant des analyses approfondies et des informations supplémentaires. Les systèmes de DAO peuvent analyser de grandes quantités de données cliniques, telles que les résultats de tests, les images médicales et les antécédents des patients, pour identifier des modèles, des anomalies ou des relations qui peuvent être difficiles à détecter pour les humains. Cela peut aider les médecins à prendre des décisions plus éclairées et à formuler des hypothèses diagnostiques plus précises.

Aide au diagnostic : Le DAO peut jouer un rôle crucial dans l'aide au diagnostic en proposant des suggestions basées sur l'analyse des données cliniques. En utilisant des algorithmes d'apprentissage automatique et des modèles prédictifs, le DAO peut identifier des schémas, des caractéristiques spécifiques qui peuvent indiquer la présence d'une maladie ou d'un trouble médical. Cela peut aider les médecins à affiner leurs diagnostics et à réduire les erreurs de diagnostic.

☑ **Prédiction des résultats** : Les systèmes de DAO peuvent également être utilisés pour prédire les résultats des traitements ou des interventions médicales. En analysant les données cliniques, les caractéristiques des patients et les résultats passés, le DAO peut fournir des informations prédictives sur les résultats probables des traitements proposés. Cela permet aux médecins d'anticiper les résultats potentiels, de personnaliser les plans de traitement et de prendre des décisions plus informées.

☑ **Optimisation des plans de traitement** : Le DAO peut contribuer à l'optimisation des plans de traitement en fournissant des recommandations basées sur des preuves et des données probantes. En analysant les données cliniques, les résultats des essais cliniques et les connaissances médicales, le DAO peut aider à identifier les traitements les plus efficaces et les stratégies de suivi optimales pour les patients individuels. Cela permet une prise de décision plus personnalisée et une meilleure gestion des soins de santé.

☑ **Gestion des ressources et des priorités** : Le DAO peut également jouer un rôle dans la gestion des ressources et des priorités en aidant les médecins à évaluer la gravité des cas, à identifier les patients à risque élevé et à optimiser l'allocation des ressources médicales. En utilisant des modèles de prévision et d'évaluation des risques, le DAO peut contribuer à une prise de décision plus efficace en matière de planification des soins et d'affectation des ressources limitées.

Il est important de souligner que le DAO ne remplace pas le jugement clinique ni l'expertise médicale. Il doit être utilisé comme un outil d'aide à la décision pour compléter les compétences des cliniciens. Les résultats du DAO doivent être interprétés avec soin et considérés dans le contexte global du patient. La collaboration et la communication entre les médecins et les DAO sont essentielles pour une prise de décision médicale optimale.

I.3 Principe de fonctionnement

Le principe de fonctionnement du DAO repose sur l'utilisation de technologies informatiques et d'algorithmes pour analyser et interpréter les données médicales [7].

I.3.1 Collecte et préparation des données

Ces tâches sont essentielles pour le fonctionnement du DAO. Il est nécessaire de collecter de manière appropriée les données médicales requises, telles que des images médicales et des données cliniques pertinentes.



Fig. I.5 : Collecte et préparation des données médicales

Par la suite, ces données doivent être soigneusement préparées afin d'être utilisées de manière optimale dans l'apprentissage et la validation du modèle du DAO. Une collecte et une préparation adéquates des données garantissent la qualité et la fiabilité des résultats obtenus par le DAO dans ses processus d'analyse et de prise de décision.

La collecte et la préparation des données médicales constituent des éléments fondamentaux du DAO, englobant plusieurs étapes tels que le choix des sources de données, la normalisation et la standardisation des données, le prétraitement des données et l'intégration de données multimodales. Ces étapes sont essentielles pour assurer la qualité, la cohérence et l'intégrité des données utilisées dans le cadre du DAO. Elles permettent d'obtenir des résultats fiables et pertinents lors de l'analyse et de la prise de décision, contribuant ainsi à l'efficacité et à la fiabilité du DAO dans le domaine médical.

I.3.2 Extraction de caractéristiques : Le DAO utilise des techniques d'analyse automatisée pour extraire les caractéristiques ou les variables les plus pertinentes à partir des données médicales. Cela peut inclure des caractéristiques extraites à partir des images, des mesures cliniques spécifiques ou d'autres informations cliniques pertinentes.

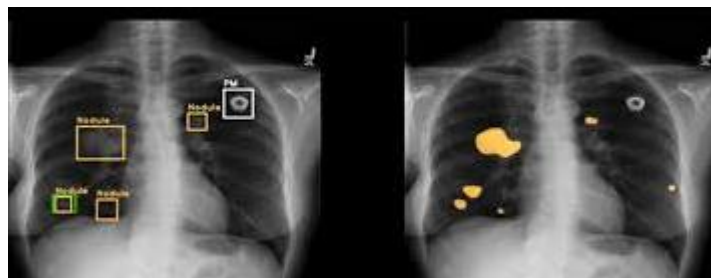


Fig. I.6 : Extraction de caractéristiques

I.3.3 Entraînement de modèle (apprentissage) : Une fois les données préparées et les caractéristiques extraites, un modèle de DAO est entraîné à l'aide d'algorithmes d'apprentissage automatique ou d'apprentissage profond. Le modèle apprend à partir de

données préalablement annotées afin de pouvoir effectuer des prédictions ou des classifications sur de nouvelles données.

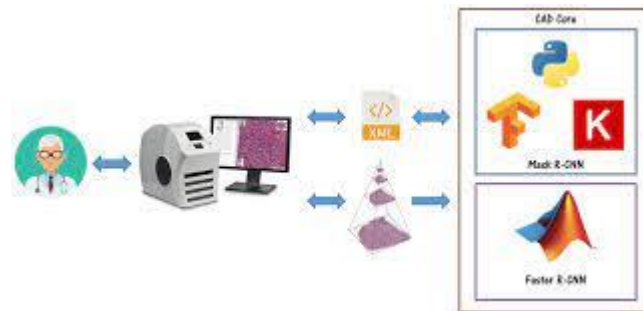


Fig. I.7 : Entraînement de modèle

I.3.4 Évaluation et validation du système

Le système de DAO est évalué en utilisant des données de validation ou des jeux de données indépendants. L'évaluation peut inclure des mesures de performance telles que la précision, la sensibilité, la spécificité, etc. Le système est ajusté et amélioré en fonction des résultats de l'évaluation.

I.3.5 Intégration des résultats dans le processus de diagnostic

Les résultats générés par le DAO (modèles) sont intégrés dans le processus de diagnostic clinique. Ils servent de support aux médecins, qui évaluent les informations fournies par le DAO en les combinant avec leur expertise clinique et leur expérience pour établir un diagnostic final.

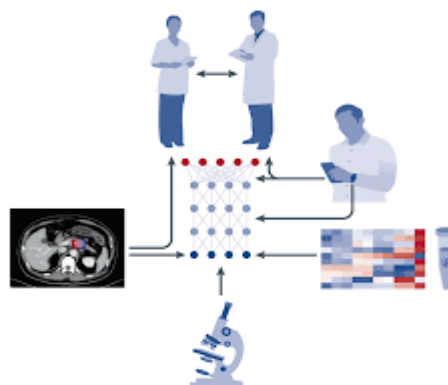


Fig. I.8 : Intégration des données et des résultats dans le processus de diagnostic

Le DAO est conçu pour compléter l'expertise humaine et aider les médecins à améliorer la précision et l'efficacité de leurs diagnostics.

I.4 Algorithmes d'apprentissage

En DAO, une variété d'algorithmes et de modèles sont utilisés pour analyser les données médicales et faciliter le processus de diagnostic. Les algorithmes et modèles couramment utilisés en DAO sont :

I.4.1 Réseaux de neurones

Les réseaux de neurones, en particulier les réseaux neuronaux profonds (*deep neural networks*) [8], sont largement utilisés en DAO. Ces modèles sont capables d'apprendre à partir de données et de reconnaître des schémas complexes dans les images médicales, les signaux physiologiques ou les données textuelles. Des architectures telles que les réseaux de neurones convolutifs (CNN) et les réseaux de neurones récurrents (RNN) sont utilisées pour la classification, la segmentation et la détection d'anomalies.

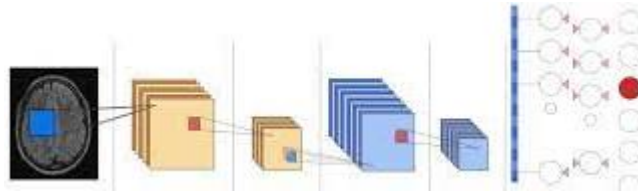


Fig. I.9 : Réseaux de neurones convolutifs

I.4.2 Machines à vecteurs de support

Les machines à vecteurs de support (SVM) [9] sont des modèles d'apprentissage automatique qui sont largement utilisés en DAO pour la classification et la détection de motifs. Les SVM sont efficaces pour traiter des données à haute dimension et sont couramment utilisés dans l'analyse d'images médicales pour la classification de tissus, la détection de lésions ou la prédiction de résultats cliniques.

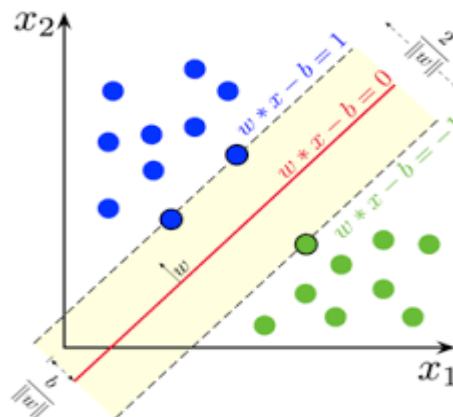


Fig. I.10 : Machine à vecteurs de support

I.4.3 Forêts aléatoires

Les forêts aléatoires (*Random Forests-RF*) [10] sont des ensembles d'arbres de décision qui sont utilisés pour la classification et la prédiction en DAO. Ces modèles sont robustes et performants dans la détection d'anomalies, la classification de maladies et la prédiction de résultats cliniques. Les RFs sont souvent utilisées en combinaison avec d'autres techniques d'apprentissage automatique pour améliorer la précision de diagnostic.

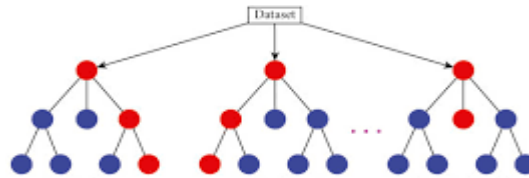


Fig. I.11 : Forêts aléatoires

I.4.4 Systèmes experts

Les systèmes experts [11] sont des modèles basés sur des règles et des connaissances expertes qui sont utilisés en DAO pour évaluer les symptômes, poser des hypothèses diagnostiques et générer des recommandations de traitement. Ces modèles utilisent des bases de connaissances médicales et des règles logiques pour guider le processus de diagnostic.

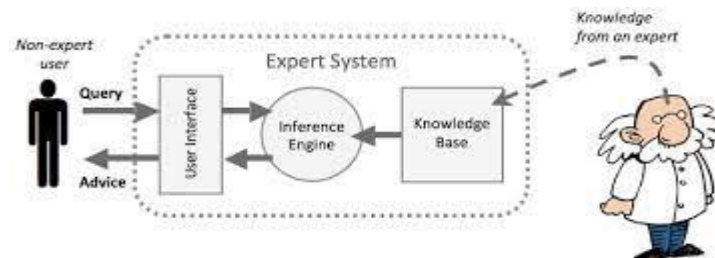


Fig. I.12 : Systèmes experts

I.4.5 Méthodes de classement

Les algorithmes de classement (*clustering*) [12], tels que le *k*-means, sont utilisés en DAO pour regrouper les données médicales similaires et identifier des sous-groupes de patients ou de caractéristiques cliniques. Ces méthodes peuvent aider à la découverte de motifs et à la personnalisation des diagnostics en fonction des profils spécifiques des patients.



Fig. I.13 : algorithme de *k*-means

Il convient de noter que ces exemples ne représentent qu'une fraction des algorithmes et modèles utilisés en DAO. De nouveaux développements et avancées sont constamment réalisés dans le domaine, notamment avec l'émergence de techniques d'apprentissage profond et d'apprentissage par renforcement. L'utilisation d'algorithmes et de modèles appropriés dépend des caractéristiques des données médicales et des objectifs spécifiques de la tâche diagnostique.

I.5 Fusion de données multimodales

La fusion de données multimodales [13] est une approche utilisée en DAO pour améliorer la précision de diagnostic en intégrant des informations provenant de différentes modalités de données médicales. Cette fusion permet de combiner les forces de chaque modalité et de fournir une vision plus complète et précise sur l'état de patient.

I.5.1 Fusion d'images médicales

Lorsque plusieurs modalités d'imagerie, telles que l'IRM, la TDM et l'échographie, sont disponibles pour un patient, la fusion de ces images peut fournir des informations plus précises sur la localisation, la taille et les caractéristiques des lésions ou des anomalies.

I.5.2 Fusion de données cliniques et d'imagerie

En intégrant des données cliniques, telles que les symptômes, les antécédents médicaux et les résultats de tests, avec les données d'imagerie, la fusion multimodale peut améliorer la précision de diagnostic. Par exemple, en combinant les résultats d'un examen d'imagerie cardiaque avec des données cliniques telles que l'âge, le sexe, les facteurs de risque et les symptômes, il est possible d'obtenir une évaluation plus précise des maladies cardiaques.

I.5.3 Fusion de données génomiques et d'imagerie

La fusion de données génomiques, telles que les mutations génétiques ou les variations du génome, avec les données d'imagerie peut permettre une meilleure compréhension des maladies à un niveau moléculaire. Par exemple, en intégrant des données d'imagerie cérébrale avec des profils génomiques, il est possible de mieux comprendre les mécanismes sous-jacents des troubles neurologiques, tels que la maladie d'Alzheimer ou la sclérose en plaques.

I.5.4 Fusion de données temporelles

Dans certains cas, la fusion de données temporelles provenant de différentes modalités peut être utilisée pour suivre l'évolution d'une maladie ou évaluer l'efficacité d'un traitement. Par

exemple, en fusionnant des données d'imagerie recueillies à différents moments avec des mesures cliniques répétées, il est possible d'observer les changements dans les lésions, la taille des tumeurs ou les paramètres physiologiques au fil du temps.

I.6 Domaines d'application

Le DAO trouve des applications dans plusieurs domaines de la médecine. Les domaines principaux où le DAO est largement utilisé sont :

I.6.1 Imagerie médicale

Le DAO est couramment utilisé dans l'analyse et l'interprétation des images médicales [14], telles que les radiographies, les tomographies (Computed Tomography-CT), les images par résonance magnétique (IRM), les échographies, etc. Il aide les radiologues et les médecins à détecter des anomalies ou des maladies dans les images, à suivre l'évolution des maladies, etc. Le DAO en imagerie médicale peut être utilisé dans des domaines tels que la détection précoce du cancer et la neuro-imagerie, etc.

I.6.2 Pathologie numérique

Le DAO est également appliqué à l'analyse de lames histologiques numérisées utilisées en pathologie [15]. Il aide les pathologistes à diagnostiquer les maladies, à caractériser les tissus et les cellules, etc. Les techniques de DAO en pathologie numérique incluent la segmentation des cellules et des tissus, la classification des échantillons, la détection d'anomalies, l'estimation de la progression des maladies, etc.

I.6.3 Analyse de données cliniques

Le DAO peut être utilisé pour analyser les données cliniques [16], telles que les dossiers médicaux électroniques, les rapports cliniques, les données génomiques, etc. Il aide à l'extraction de connaissances à partir de ces données et à la prise de décisions diagnostiques. Par exemple, le DAO peut aider à la prédiction de résultats cliniques, à l'identification de facteurs de risque, à la détection précoce de maladies, à la recommandation de traitements personnalisés, etc.

Il convient de noter que le DAO est également utilisé dans d'autres domaines de la médecine, tels que la cardiologie, la pneumologie, la génétique, l'ophtalmologie, etc. Les applications du DAO continuent de se développer avec les avancées technologiques et l'exploitation de nouvelles sources de données médicales.

I.7 Exemples de systèmes de DAO existants

Les exemples ci-dessous illustrent une variété de systèmes de DAO existants dans différents domaines de la médecine. Chacun de ces systèmes est conçu pour aider les médecins à améliorer le diagnostic et la prise de décision clinique en utilisant des algorithmes et des techniques d'apprentissage automatique

I.7.1 CAD-RADS

Coronary Artery Disease Reporting and Data System (CAD-RADS) [17] est un système de DAO utilisé dans l'imagerie cardiaque pour évaluer le risque de maladies coronariennes. Il fournit une classification des résultats des angiographies coronaires basée sur des critères spécifiques.

I.7.2 CADx

(CADx) [18] est un système de DAO utilisé dans le domaine de la pathologie pour aider les pathologistes dans le diagnostic du cancer. Il utilise des techniques d'apprentissage automatique pour analyser les images histologiques et fournir des informations sur la présence ou l'absence de cellules cancéreuses.

I.7.3 MIA

Le Mammography Image Analysis (MIA) [19] est un système de DAO utilisé dans le dépistage du cancer du sein. Il utilise des algorithmes d'apprentissage automatique pour analyser les mammographies et détecter les anomalies potentielles.

I.7.4 DLS

Le Deep Learning System (DLS) [20] est un système de DAO développé par Google pour détecter les maladies oculaires à partir de photographies de la rétine. Il utilise des réseaux neuronaux convolutionnels pour identifier les signes de rétinopathie diabétique et d'autres affections oculaires.

I.8 Défis et perspectives du DAO

Il est nécessaire de développer des cadres réglementaires et des lignes directrices appropriées pour faire face aux défis et problèmes éthiques rencontrés dans les DAO. De plus, les avancées technologiques visent à améliorer la précision des diagnostics, la prise de décision médicale et les soins de santé de manière plus générale [21].

I.8.1 Limitations et problèmes éthiques liés au DAO

Le DAO présente de nombreux avantages, mais il est également confronté à des défis et des problèmes éthiques. Les défis et les problèmes éthiques couramment associés au DAO sont:

☑ **Limitations des performances** : Bien que le DAO ait fait d'énormes progrès, il existe encore des limites quant à sa précision et à sa fiabilité. Les systèmes de DAO peuvent présenter des taux d'erreur et des faux positifs/négatifs, ce qui peut entraîner des erreurs de diagnostic et une prise de décision inappropriée. Il est important de reconnaître ces limitations et de les prendre en compte lors de l'utilisation des systèmes de DAO.

☑ **Manque de transparence** : Les modèles de DAO basés sur des techniques d'apprentissage automatique, tels que les réseaux de neurones profonds, peuvent être considérés comme des boîtes noires, ce qui signifie qu'il peut être difficile de comprendre comment ils prennent leurs décisions. Cette opacité peut poser des problèmes éthiques, notamment en matière de responsabilité, de confiance et de justification des décisions prises par les systèmes de DAO.

☑ **Biais et équité** : Les systèmes de DAO sont susceptibles de refléter les biais présents dans les données d'entraînement, ce qui peut entraîner des résultats discriminatoires ou injustes, en particulier pour les populations sous-représentées. Il est essentiel de s'assurer que les systèmes de DAO sont équitables et ne renforcent pas les inégalités existantes en matière de soins de santé.

☑ **Confidentialité et sécurité des données** : L'utilisation du DAO nécessite la collecte et l'accès à de grandes quantités de données médicales sensibles. Il est crucial de garantir la confidentialité et la sécurité de ces données pour prévenir les atteintes à la vie privée et les risques de violation de données. Les protocoles de protection des données et de conformité aux réglementations doivent être mis en place pour garantir un usage responsable des informations des patients.

☑ **Relation médecin-patient** : L'introduction du DAO peut avoir un impact sur la relation entre les médecins et leurs patients. Il est important de maintenir une communication claire et transparente avec les patients, en expliquant le rôle du DAO et en répondant à leurs préoccupations. Les systèmes de DAO ne doivent pas remplacer complètement l'interaction humaine, mais plutôt servir d'outil d'aide à la décision pour les médecins.

☑ **Responsabilité et responsabilité légale** : Lorsque des erreurs de diagnostic ou des problèmes surviennent avec l'utilisation du DAO, il peut être complexe de déterminer la responsabilité et la responsabilité légale. Les questions de responsabilité doivent être clarifiées pour garantir une utilisation responsable et sûre du DAO.

I.8.2 Tendances et développements futurs du DAO

Les tendances et développements futurs du Diagnostic Assisté par Ordinateur (DAO) incluent les avancées technologiques et les évolutions dans les pratiques médicales. Voici quelques-unes des tendances clés à surveiller [22] :

☑ **Intégration de l'IA avancée** : L'utilisation de techniques d'IA avancée, telles que l'apprentissage profond (*deep learning*), continuera à jouer un rôle majeur dans le DAO. Ces approches permettent une meilleure analyse et compréhension des données médicales complexes, ainsi que des performances améliorées dans des domaines tels que la détection d'anomalies, le diagnostic et la prédiction des résultats.

☑ **Fusion de données** : La fusion de données provenant de différentes modalités, telles que les images médicales, les résultats de laboratoire et les données cliniques, est une tendance émergente. Cette approche vise à tirer parti de l'information contenue dans divers types de données pour améliorer la précision diagnostique et la prise de décision médicale.

☑ **Explication et interprétabilité des modèles de DAO** : L'explication et l'interprétabilité des modèles de DAO sont des domaines de recherche actifs. Il est crucial de comprendre comment les systèmes de DAO prennent leurs décisions afin de garantir la confiance des médecins et des patients. Les méthodes pour expliquer les prédictions du DAO et pour fournir des justifications compréhensibles sont en développement afin de rendre les modèles de DAO plus transparents et explicables.

☑ **Intégration dans les dispositifs médicaux portables et l'internet des objets (IoT)** : Les dispositifs médicaux portables et l'IoT (*Internet of Things*) sont de plus en plus utilisés pour collecter des données sur la santé des patients. L'intégration du DAO avec ces technologies permet une surveillance continue de la santé, la détection précoce des anomalies et la personnalisation des soins médicaux. Les données en temps réel provenant de ces dispositifs peuvent être utilisées pour alimenter les modèles de DAO et améliorer les capacités de diagnostic et de suivi des patients.

☑ **DAO basé sur les connaissances** : En plus de l'apprentissage à partir de données, le DAO basé sur les connaissances utilise des bases de connaissances médicales pour

améliorer les performances de diagnostics. Ces bases de connaissances comprennent des ontologies médicales, des règles d'inférence et des modèles de raisonnement qui permettent de combiner les connaissances médicales expertes avec les données cliniques pour prendre des décisions précises.

☑ **Validation et réglementation** : La validation et la réglementation des systèmes de DAO sont des enjeux importants. Il est essentiel de développer des méthodologies de validation rigoureuses pour évaluer les performances et la sécurité des systèmes de DAO. Les régulateurs travaillent également à l'élaboration de directives et de normes pour encadrer l'utilisation du DAO dans le contexte clinique.

I.9 Conclusion

Le Diagnostic Assisté par Ordinateur (DAO) est une approche innovante qui utilise l'intelligence artificielle et l'apprentissage automatique pour soutenir les médecins dans leur prise de décision médicale. Le DAO offre de nombreux avantages, notamment une précision accrue dans l'interprétation des données médicales, une aide au diagnostic plus précise, une optimisation des plans de traitement et une gestion efficace des ressources médicales. Le DAO représente une avancée significative dans le domaine de la médecine, offrant de nouvelles opportunités pour améliorer la précision de diagnostic, les plans de traitement et la gestion des soins de santé. Cependant, il est important de reconnaître que le DAO ne remplace pas le jugement clinique et l'expertise médicale. Il doit être utilisé comme un outil complémentaire, en collaboration avec les cliniciens, pour prendre des décisions éclairées et personnalisées.

Chapitre 2

EXTRACTION DE CARACTÉRISTIQUES

Résumé

L'extraction de caractéristiques est une étape clé du DAO qui vise à identifier les informations pertinentes dans les données médicales. Son objectif est de réduire la complexité des données en les transformant en représentations significatives et informatives. Dans ce chapitre, nous allons d'abord donner un aperçu des techniques d'extraction de caractéristiques. Ensuite, nous aborderons notre technique proposée pour l'extraction de caractéristiques afin qu'elle puisse être utilisée dans une chaîne de traitement DAO.

II.1 Introduction

II.2 Objectifs et importance

II.3 Rôle de l'extraction de caractéristiques

II.4 Types de l'extraction de caractéristiques

II.5 Méthodes d'extraction de caractéristiques

II.6 Méthodes proposée

II.7 Conclusion

Extraction de Caractéristiques

II.1 Introduction

L'extraction de caractéristiques en DAO [23] est une étape essentielle dans le processus de transformation des données médicales brutes en représentation plus significatives et informatives. Son objectif principal est de faciliter la prise de décision médicale en fournissant des informations pertinentes et exploitables. Cette étape joue un rôle crucial dans l'amélioration de la précision des diagnostics, de l'efficacité des traitements et des soins de santé en général. Dans ce chapitre, nous commencerons par donner un aperçu des techniques d'extraction de caractéristiques. Ensuite, nous présenterons notre technique proposée pour l'extraction de caractéristiques, qui pourra être utilisée dans une chaîne de traitement DAO.

II.2 Objectifs et importance

L'extraction de caractéristiques en DAO a pour objectif principal de transformer les données médicales brutes en représentations plus significatives et informatives. Cela permet de réduire la complexité des données, d'identifier les aspects pertinents et de fournir une base solide pour l'analyse et l'interprétation ultérieures. L'importance de cette étape réside dans sa capacité à améliorer la précision de système de diagnostic et à faciliter la prise de décision médicale. En identifiant les caractéristiques discriminatives et informatives dans les données médicales, elle aide à l'amélioration de la détection ou la classification des pathologies, de prédire les complications, et de suivre l'évolution des maladies [24].

II.3 Rôle de l'extraction de caractéristiques

L'étape d'extraction de caractéristiques joue un rôle essentiel dans la facilitation de l'analyse et de la prise de décision médicale [25].

II.3.1 Réduction de la dimensionnalité

Les données médicales peuvent contenir un grand nombre de variables, ce qui complique l'analyse et la visualisation. L'extraction de caractéristiques permet de réduire la dimensionnalité des données en identifiant les caractéristiques les plus discriminantes et informatives. Cela simplifie l'analyse et permet une meilleure compréhension des données.

II.3.2 Identification des informations pertinentes

Les données médicales peuvent contenir des informations bruitées, redondantes ou non pertinentes. L'extraction de caractéristiques permet d'identifier les aspects les plus pertinents des données en mettant en évidence les caractéristiques clés liées aux pathologies, aux symptômes ou aux résultats de tests. Cela permet de filtrer les informations essentielles et d'éliminer le bruit.

II.3.3 Amélioration de la précision diagnostique

L'extraction de caractéristiques permet de mettre en évidence les caractéristiques distinctives des différentes pathologies ou conditions médicales. En identifiant ces caractéristiques, les systèmes de DAO peuvent être plus précis dans la détection, la classification et la prédiction des maladies. Cela contribue à améliorer la précision de diagnostic et à réduire les erreurs médicales.

II.3.4 Optimisation de la prise de décision

En fournissant des informations pertinentes et exploitables, l'extraction de caractéristiques aide les médecins à prendre des décisions éclairées et personnalisées. Les caractéristiques extraites peuvent servir de base pour évaluer les risques, déterminer les traitements appropriés et prédire les résultats probables. Cela permet une prise de décision médicale plus précise et efficace.

II.4 Types de l'extraction de caractéristiques

En DAO basé sur l'imagerie médicale, différents types de caractéristiques d'image peuvent être utilisés pour représenter les données médicales et faciliter l'analyse et la prise de décision. Il est important de noter que la sélection des caractéristiques appropriées dépend de la tâche spécifique de DAO et du type de données utilisées [26].

II.4.1 Caractéristiques liées aux lignes

L'extraction de caractéristiques basée sur les lignes consiste à détecter et à représenter les lignes présentes dans une image. Les lignes peuvent être des contours, des arêtes ou des lignes plus complexes, et elles peuvent fournir des informations importantes sur la structure et la géométrie d'une image. Différentes techniques d'extraction de caractéristiques peuvent être appliquées pour capturer les caractéristiques de lignes, telles que les techniques de détection de contours, les filtres de Gabor et les techniques de seuillage.

II.4.2 Caractéristiques liées aux textures

L'extraction de caractéristiques basée sur les textures se concentre sur les motifs récurrents et les variations de luminosité présente dans une image. Les textures peuvent être des motifs réguliers, tels que des carreaux ou des motifs aléatoires, tels que des grains de sable. L'extraction de caractéristiques de textures permet de capturer des informations sur la composition et la structure locale d'une image. Différentes techniques d'extraction de caractéristiques peuvent être appliquées pour capturer les caractéristiques de textures, telles que les motifs locaux binaires (*Local Binary Patterns-LBP*), les caractéristiques d'image statistiques binarisées (*Binarized Statistical Image Features-BSIF*) et les histogrammes des gradients orientés (*Histogram of Oriented Gradients-HOG*).

II.4.3 Caractéristiques liées aux Apparence

L'extraction de caractéristiques basée sur l'apparence se concentre sur les attributs visuels globaux d'une image, tels que la couleur, la forme, la taille et la distribution spatiale des objets. Cette approche permet de représenter les caractéristiques visuelles globales d'une image, ce qui peut être utile dans des tâches telles que la reconnaissance d'objets ou la classification d'images. Différentes techniques d'extraction de caractéristiques peuvent être appliquées pour capturer les caractéristiques de textures, telles que l'analyse en composantes principales (*Principal Component Analysis-PCA*), l'analyse en composantes indépendantes (*Independent Component Analysis-ICA*) et l'analyse discriminante linéaire (*Linear Discriminant Analysis-LDA*).

II.5 Méthodes d'extraction de caractéristiques

L'extraction de caractéristiques est le processus de transformation des valeurs brutes des pixels d'une image en informations plus significatives et utiles qui peuvent être utilisées dans une application de vision par ordinateur. Pour cette raison, ces processus ont été soigneusement conçus par des spécialistes (création manuelle ou artisanales) en utilisant leurs

connaissances et leur expertise. Généralement, ces techniques sont appliquées à l'image sans connaître préalablement son contenu. Ainsi, pour améliorer leurs performances, ces techniques ont été dotées de connaissances préalables sur leur contexte grâce à l'apprentissage automatique (artisanales apprises). Ces dernières années, avec l'avènement de l'apprentissage profond, les réseaux neuronaux eux-mêmes effectuent automatiquement le travail de l'expert dans les couches de convolution.

II.5.1 Méthodes traditionnelles

Également appelée méthode artisanale ou conçue à la main (*handcrafted*), fait référence à l'extraction de caractéristiques basée sur des connaissances et des règles spécifiques définies par les experts. Dans cette approche, des descripteurs sont conçus manuellement pour capturer des informations pertinentes dans les images.

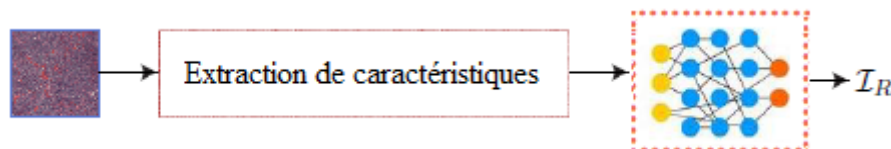


Fig. II.1 Méthode traditionnelle d'extraction de caractéristiques.

Par exemple, dans le cas de la détection de contours, des opérateurs tels que le Laplacien ou les opérateurs de gradient peuvent être utilisés pour extraire les caractéristiques de contour. La méthode artisanale permet une certaine interprétabilité et contrôle sur le processus d'extraction de caractéristiques, mais elle peut être limitée par la capacité des experts à définir des descripteurs pertinents [27].

II.5.2 Méthodes traditionnelles apprises

La méthode traditionnelle apprise (*Learned handcrafted*) combine des aspects de la méthode artisanale et de l'apprentissage automatique. Elle consiste à apprendre les descripteurs manuellement conçus à partir d'un ensemble de données d'apprentissage. Dans cette approche, les experts conçoivent initialement un ensemble de descripteurs, puis utilisent des techniques d'apprentissage automatique pour sélectionner et pondérer les descripteurs les plus pertinents.

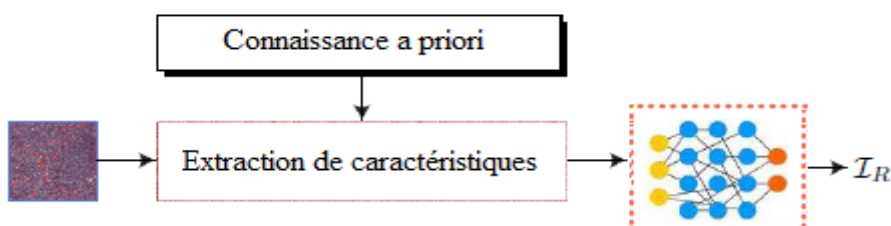


Fig. II.2 Méthode traditionnelle apprise d'extraction de caractéristiques.

Cela permet d'automatiser une partie du processus de conception de descripteurs et de les adapter aux données spécifiques [28].

II.5.3 Méthodes profondes

Les méthodes profondes (basées sur l'apprentissage profond) ont révolutionné le domaine de l'extraction de caractéristiques. Ces méthodes utilisent des architectures de réseaux de neurones profonds pour apprendre automatiquement des descripteurs à partir des données brutes. Les réseaux de neurones profonds sont capables d'apprendre des représentations hiérarchiques et abstraites des données, ce qui leur permet de capturer des caractéristiques complexes et non linéaires.

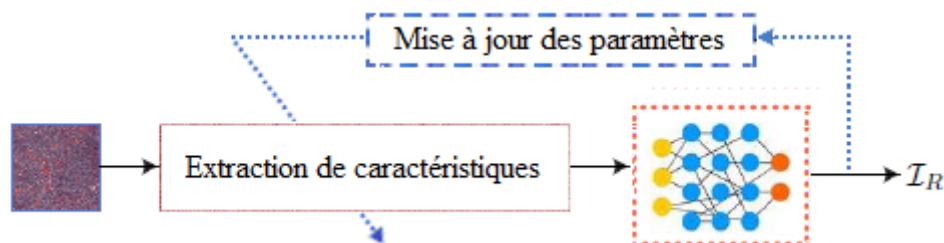


Fig. II.3 Méthode profonde d'extraction de caractéristiques.

Des modèles populaires tels que les CNNs et les RNNs sont utilisés dans les tâches de vision par ordinateur pour extraire des caractéristiques à partir d'images, de vidéos ou de séquences temporelles. Les méthodes profondes offrent souvent de meilleures performances que les méthodes traditionnelles d'extraction de caractéristiques, mais elles nécessitent généralement de grands ensembles de données et des ressources de calcul considérables pour l'apprentissage. De plus, l'interprétation des caractéristiques apprises peut-être plus difficile à comprendre que dans les méthodes traditionnelles avec ou sans apprentissage [29].

II.6 Méthode proposée

Pour décrire la texture des images dans un système DAO, nous avons proposé une méthode d'analyse profonde basée sur la convolution. Toutes les images ont été initialement prétraitées afin d'améliorer le contraste de l'image. Dans la suite, nous détaillerons les techniques utilisées dans la méthode proposée.

II.6.1 Fondement théorique

Cette partie nous permet de décrire les exigences préliminaires qui seront utilisées dans le processus d'analyse de texture proposé.

☑ **Descripteur LPQ** : La quantification de phase locale (LPQ) [30] extrait les informations de phase locale à partir d'une image floutée, en se basant sur l'invariance au flou de la réponse en phase de Fourier.

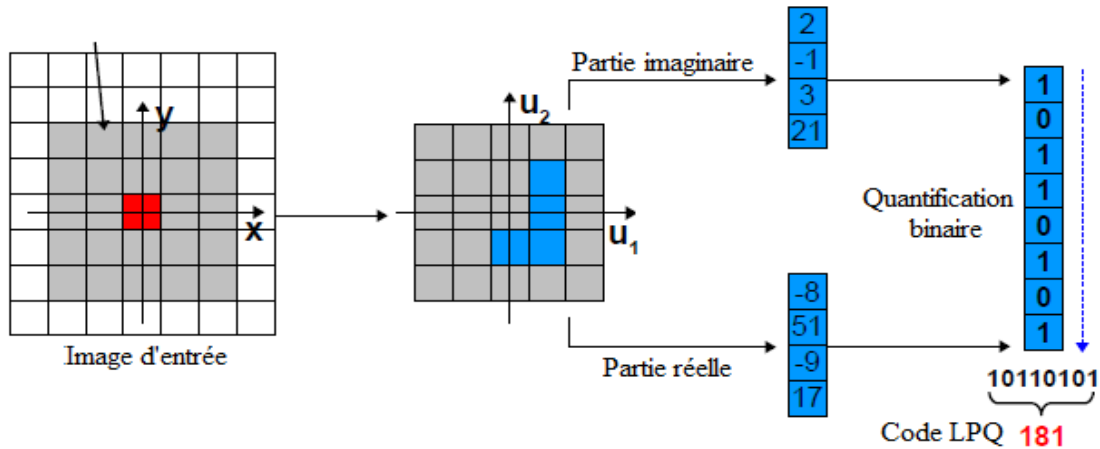


Fig. II.4 Extraction de caractéristiques basée sur le LPQ

- **Analyse de fréquence locale** : L'opérateur LPQ sur un pixel d'image est réalisé en utilisant la transformée de Fourier à court terme (STFT) sur une fenêtre de taille $M \times M$ (W_n). Ainsi, pour tous les emplacements de pixels $x = \{x_1, x_2, \dots, x_{HW}\}$ dans une image de taille $H \times W$ ($f(x)$), les patches d'image locaux dans W_n sont définis comme suit :

$$f_x(y) = f(x - y), \quad \forall y \in W_n \quad (1)$$

En appliquant la transformée de Fourier à court terme (STFT) à W_n , nous obtenons la représentation du domaine de fréquence locale :

$$\mathcal{F}_x(u) = \sum_{y_i \in W_n} f_x(y_i) e^{-j2\pi u^T y_i} \quad (2)$$

Où $i = 1, \dots, M^2$. La variable u comprend ℓ variables de fréquence (u_1, u_2, \dots, u_ℓ) qui servent de descripteurs locaux insensibles au flou. L'équation (2) peut être réécrite de la manière suivante :

$$\mathcal{F}_x(u) = \psi_u^T f_x, \quad \psi_u^T(y) = e^{-j2\pi u^T y} \quad (3)$$

Où

$$\psi_u^T = [\psi_u(y_1), \psi_u(y_2), \dots, \psi_u(y_{M^2})], \quad \text{et} \quad f_x = [f_x(y_1), f_x(y_2), \dots, f_x(y_{M^2})] \quad (4)$$

Pour trouver la valeur de LPQ, un processus de quantification de phase est effectué. En effet, si les coefficients à quantifier sont statistiquement indépendants, une quantification scalaire est effectuée directement ; sinon, les échantillons doivent être décorrelés avant la quantification.

• **Quantification de phase** : La représentation insensible au flou utilise ℓ fréquences, produisant un vecteur de caractéristiques de longueur ℓ pour chaque pixel. En pratique, la phase est quantifiée en quatre quadrants à l'aide du quantificateur suivant :

$$Q(\mathcal{F}_x(u)) = (\text{Re}\{\mathcal{F}_x(u)\} > 0) + 2(\text{Im}\{\mathcal{F}_x(u)\} > 0) \quad (5)$$

Cette quantification peut être exprimée avec 2 bits par fréquence pour chaque pixel, ce qui donne un total de 2ℓ bits ; la concaténation des codes des ℓ composantes de fréquence donne un seul mot de code. Pour obtenir une représentation insensible au flou, seules les composantes de basse fréquence sont utilisées, car elles contiennent la majeure partie de l'énergie de l'image. En pratique, les coefficients locaux \mathcal{F}_x sont calculés pour chaque pixel à quatre basses fréquences ($\ell = 4$) :

$$[u_1, u_2, u_3, u_4] = \begin{bmatrix} a & 0 & a & a \\ 0 & a & a & -a \end{bmatrix} \quad (6)$$

Le scalaire a correspond à la plus haute fréquence dont la fonction de propagation (*point spread function* (PSF)) est positive, et il est généralement calculé à partir de M (c'est-à-dire $a = \frac{1}{M}$). Enfin, un mot de code sur 8 bits ($[q_7, q_6, q_5, \dots, q_0]$) peut être obtenu pour décrire la texture locale entourant chaque pixel (W_n). Ce mot de code peut être facilement transformé en un nombre décimal (dans la plage de 0 à 255) par un simple décodage binaire :

$$\text{LPQ}_i = \sum_{j=1}^8 q_j 2^{j-1} \quad (7)$$

• **Décorrélation** : La quantification vectorielle est généralement plus efficace dans les cas où les coefficients à quantifier sont non corrélés. Pour décorrélérer les coefficients de fréquence, nous séparons d'abord les composantes réelles et imaginaires de $\mathcal{F}_x(u)$, puis les concaténons de la manière suivante :

$$\mathcal{F}_x(u) = [\mathcal{F}_x(u_1), \mathcal{F}_x(u_2), \dots, \mathcal{F}_x(u_\ell)] \quad (8)$$

$$F_x = [\text{Re}\{\mathcal{F}_x(u)\} \quad \text{Im}\{\mathcal{F}_x(u)\}] = [F_x^R \quad F_x^I] \quad (9)$$

La transformée STFT (équation 3) montre que F_x et f_x sont linéairement dépendants, donc nous pouvons écrire :

$$F_x = \Psi \cdot f_x, \quad / \quad \Psi = [\Psi_R \quad \Psi_I]^T \quad (10)$$

Et

$$\Psi_R = \text{Re}\{[\psi_{u_1}, \psi_{u_2}, \dots, \psi_{u_\ell}]\} \quad \text{Et} \quad \Psi_I = \text{Im}\{[\psi_{u_1}, \psi_{u_2}, \dots, \psi_{u_\ell}]\} \quad (11)$$

Pour décorrélérer F_x , la transformation suivante est appliquée :

$$G_x = V^T \cdot F_x \quad (12)$$

Où V est une matrice orthogonale obtenue à partir de la matrice (D) mentionnée ci-dessous :

$$D = U\Sigma V^T \quad (13)$$

D est la matrice de covariance de F_x et peut être obtenue par :

$$D = \Psi C \Psi^T \quad (14)$$

Où C est la matrice de covariance des échantillons $M \times M$ dans W_n . Après le processus de décorrélation, un quantificateur scalaire quantifie le $j^{\text{ième}}$ coefficient (g_j) de G_x :

$$q_j = \begin{cases} 1, & \text{if } g_j \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

Le mot de code obtenu est ensuite transformé en un nombre décimal en utilisant l'équation 7. Après avoir calculé les mots de code LPQ pour chaque pixel de l'image, et étant donné que la plage de LPQ est de 0 à 255, une image contenant les informations LPQ de l'image d'entrée est produite. Pour calculer un vecteur de caractéristiques (\mathcal{H}) pour l'image d'entrée en pratique, nous divisons l'image LPQ en sous-régions de taille $r \times c$ sans chevauchements et calculons les histogrammes pour chaque sous-région. Ensuite, ces histogrammes sont concaténés pour fournir un vecteur de caractéristiques LPQ pour l'ensemble de l'image d'entrée.

$$\mathcal{H} = [h_1, h_2, \dots, h_{rc}] \quad (16)$$

L'opération de division de l'image est utilisée pour conserver certaines informations sur l'arrangement spatial des motifs, car les histogrammes de l'ensemble de l'image ne tiennent pas compte de cette information.

☑ Filtre gaussien : Le filtre gaussien est une technique de filtrage utilisée en traitement d'images pour lisser les données et réduire le bruit. Il est basé sur la distribution gaussienne, également connue sous le nom de distribution normale. Un filtre passe-bas gaussien est défini à partir de la réponse impulsionnelle suivante :

$$h(x, y) = e^{\left(-\frac{x^2+y^2}{2\sigma^2}\right)} \quad (16)$$

Pour déterminer la valeur de σ , on choisit une valeur ε petite pour l'élément de H de coordonnées $(x, y) = (P, 0)$, situé au centre d'un pixel du bord. La valeur au centre $((x, y) = (0, 0))$ est 1. La valeur sur un bord doit être petite par rapport à la valeur au centre. On obtient:

$$\sigma = \frac{P}{\sqrt{-2\ln(\epsilon)}} \quad (17)$$

La seconde étape est la normalisation de la matrice H , qui consiste à diviser tous ses coefficients pour que leur somme soit égale à 1. La normalisation permet de ne pas modifier le niveau de gris moyen de l'image.

Le filtre gaussien fonctionne en calculant la moyenne pondérée des pixels voisins autour de chaque pixel de l'image. Les poids sont déterminés par la distance entre les pixels voisins et le pixel central, selon la distribution gaussienne. Plus précisément, les pixels voisins qui sont plus proches du pixel central ont des poids plus élevés, tandis que ceux qui sont plus éloignés ont des poids plus faibles [31].

☑ **Système chaos (cartes logistiques)** : Un système chaotique, également connu sous le nom de dynamique chaotique, est un système dynamique non linéaire sensible aux conditions initiales. Ces systèmes sont caractérisés par un comportement complexe, imprévisible et sensible aux petites variations dans les conditions de départ.

Les cartes logistiques sont un exemple classique de systèmes chaotiques unidimensionnels. Elles sont largement étudiées dans le domaine de la théorie du chaos et de la dynamique non linéaire. Une carte logistique est une fonction non linéaire qui modélise la croissance d'une population en fonction de sa population actuelle. Elle est définie mathématiquement par l'équation suivante :

$$x_{n+1} = \mu x_n (1 - x_n) \quad (18)$$

Où x_n est la population à l'instant n , x_{n+1} est la population à l'instant $n + 1$, et μ est le paramètre de croissance (contrôle). La valeur de μ détermine le comportement du système.

Lorsque μ est inférieur à 1, la population tend à converger vers zéro, représentant une extinction. Lorsque μ est compris entre 1 et 3, la population atteint une valeur stable, représentant une croissance régulière. Cependant, lorsque μ dépasse une certaine valeur critique (environ 3.57), le système présente un comportement chaotique. Dans la région chaotique, les populations futures oscillent de manière complexe et imprévisible, même si les conditions initiales sont très proches. De petites variations initiales peuvent conduire à des trajectoires complètement différentes au fil du temps. Cela illustre la sensibilité aux conditions initiales, une caractéristique clé des systèmes chaotiques [32].

☑ **Optimisation basée sur l'algorithme de chauve-souris (BAT algorithm)** : L'objectif de l'optimisation est de trouver la meilleure solution ou de faire fonctionner un système de la

manière la plus efficace possible. Ainsi, il s'agit du processus d'ajustement des données d'entrée (variables de décision) pour obtenir la meilleure solution (minimum/maximum) pour le problème réel. En effet, les approches méta-heuristiques surpassent les approches précises car elles explorent et sélectionnent de manière plus efficace les parties attractives de l'espace de recherche, notamment d'un point de vue computationnel. L'algorithme BAT [33] est l'une des approches méta-heuristiques inspirées par le comportement d'écholocation des chauves-souris et basées sur l'intelligence collective.

Lorsqu'il chasse, la chauve-souris émet de brèves impulsions acoustiques puissantes, puis analyse l'écho pour déterminer la taille et l'emplacement d'un obstacle ou d'une proie. En observant ce comportement, *Xin-She Yang* [34] a proposé l'algorithme BAT standard qui peut être expliqué comme suit :

- i*) Initialisation de la population de chauves-souris et affectation à chaque individu i de la position (x_i), de la vitesse (v_i), de la fréquence d'impulsion (f_i), du taux d'impulsion (r_i) et de l'intensité sonore (A_i).
- ii*) Pour la $i^{\text{ème}}$ chauve-souris, mettre à jour la meilleure position globale (x^*), f_i , v_i et x_i comme suit :

$$\begin{cases} f_i = f_{\min} + (f_{\max} - f_{\min})\beta, & \beta \in [0,1] \\ v_i^{t+1} = v_i^t + (x_i^t - x^*)f_i \\ x_i^{t+1} = x_i^t + v_i^t \end{cases} \quad (19)$$

- iii*) Une nouvelle solution est générée pour la chauve-souris si le nombre aléatoire est supérieur à r_i :

$$x_{\text{new}} = x_{\text{old}} + \varepsilon A^t \quad (20)$$

Où ε est un nombre aléatoire dans l'intervalle $[-1, 1]$, et A^t représente l'intensité sonore moyenne de toutes les chauves-souris à l'instant t .

- iv*) Acceptez la nouvelle solution si le nombre aléatoire est inférieur à A_i et $f(x_i) < f(x^*)$, et mettez à jour A_i et r_i comme suit :

$$\begin{cases} A_i^{t+1} = \alpha A_i^t \\ r_i^t = r_i^0 [1 - e^{-\gamma t}] \end{cases} \quad (21)$$

Où f désigne la fonction objectif et α et γ sont des paramètres constants choisis tels que $0 < \alpha < 1$ et $0 < \gamma$, de sorte que $t \rightarrow \infty$, $A_i^t \rightarrow 0$, et $r_i^t \rightarrow r_i^0$.

- v*) Trouver la solution optimale actuelle (x^*) en triant les chauves-souris en fonction de leur adéquation (fitness).

vi) Revenir à l'étape *ii* et une fois que le nombre maximal d'itérations est atteint, renvoyer la solution optimale.

II.6.2 Caractéristique profonde

Dans un problème de classification, et en raison de la forte corrélation entre les images des différentes classes, l'objectif est de trouver des vecteurs de caractéristiques avec une corrélation intra-classe élevée et une corrélation inter-classes faible. Généralement, dans les images de tissus pathologiques, les spécialistes rencontrent des difficultés pour détecter la maladie à l'œil nu. Les approches traditionnelles d'extraction de caractéristiques (artisanales) ne fournissent pas de vecteurs précis et distinctifs pour le diagnostic automatique.

Inspirés par les avantages de l'analyse profonde, nous proposons dans cette partie une nouvelle méthode d'extraction de caractéristiques pour la classification des images médicales. Tout d'abord, nous filtrons l'image à l'aide d'une banque de filtres gaussiens. Les coefficients de ces filtres sont pondérés avec des valeurs chaotiques. En effet, les lignes, les rides, les crêtes, les points singuliers et la texture peuvent tous être détectés dans une image en utilisant un espace de caractéristiques basé sur cette opération de filtrage. Ensuite, nous combinons toutes les images filtrées en un seul descripteur pour réduire la quantité de données, et enfin nous utilisons LPQ pour extraire le vecteur de caractéristiques à partir du descripteur. Les paramètres de la banque de filtres sont ajustés (paramètre de contrôle et valeur initiale de la carte logistique), en utilisant l'optimisation basée sur l'algorithme des chauves-souris, lors de l'apprentissage pour obtenir des vecteurs précis. La figure II.5 contient un schéma bloc du processus complet d'extraction de caractéristiques, comprenant le filtrage, la réduction des données et l'extraction du vecteur de caractéristiques.

La structure de la figure II.5 comprend trois couches principales : la couche de convolution, la couche de réduction de données (*pooling layer*) et la couche d'extraction du vecteur de caractéristiques [35]. Pour discuter de l'architecture du système, nous supposons que les images d'entrée ont les dimensions $H \times W$ et que le filtre de convolution, pour la couche de convolution est :

$$\mathcal{W}_{i,\lambda_i} = M_{\lambda_i} \cdot \text{Gauss}(k_1, k_2) \in \mathbb{R}^{k_1 \times k_2}, \quad \lambda_i = (\mu^i, x_0^i) \quad (21)$$

Où M_{λ_i} est une matrice des poids générée par la carte logistique, λ_i est représenté l'ensemble des paramètres de la carte logistique (paramètre de contrôle μ^i et valeur initiale x_0^i), et N est le nombre de filtres utilisés.

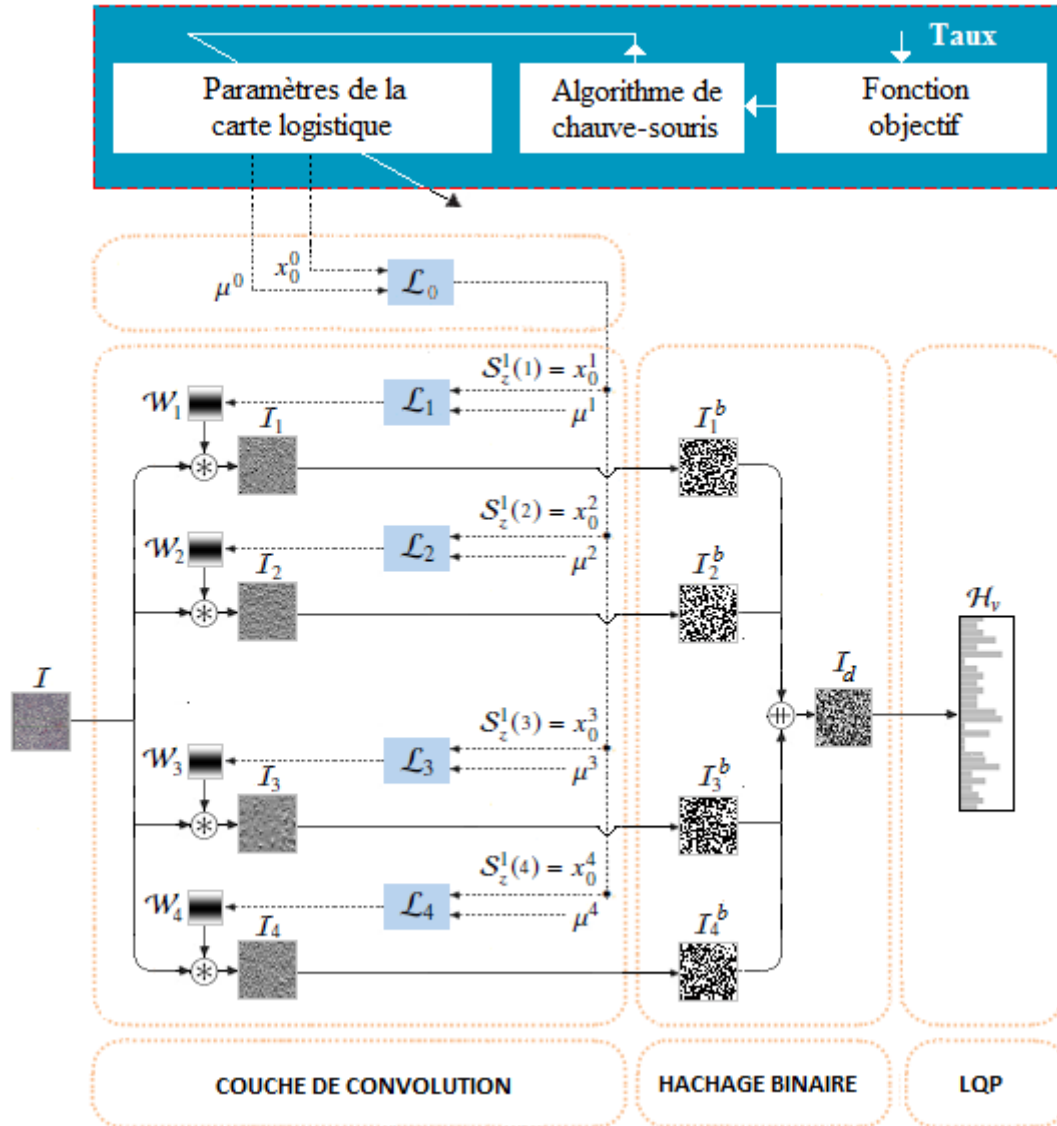


Fig. II.5 Extraction de caractéristiques. Un exemple d'une structure à une couche avec quatre filtres de convolutions différentes.

Dans ce qui suit, nous examinerons le fonctionnement de notre méthode d'extraction de caractéristiques pour fournir un vecteur de caractéristiques précis afin d'améliorer les performances du système de classification utilisé dans un système DAO.

☑ **Couche de convolution** : L'objectif principal de la couche de convolution est d'améliorer les caractéristiques dominantes de l'image d'entrée. En plus de paramètre du filtre (paramètres de la carte logistique), sa taille (variance) joue également un rôle important, car elle détermine la taille des voisins impliqués dans le calcul du coefficient de chaque pixel. Les sorties de cette couche sont obtenues par un processus de convolution 2D de toutes les images de la base d'apprentissage (Ψ_{train}) et des filtres $\mathcal{W}_{i,\lambda_i}$:

$$\hat{I}_{ji} = I_j * \mathcal{W}_{i,\lambda_i}, \quad i \in [1..N], \quad \text{et} \quad j \in [1..N_{\text{train}}] \quad (22)$$

Où N_{train} représente le nombre d'images dans Ψ_{train} , le symbole $*$ représente le processus de convolution 2D, et \hat{I}_{ji} sont les images de sortie filtrées.

☑ **Couche de réduction** : Lorsque N filtres sont appliqués aux images, N images supplémentaires (filtrées) sont produites, ce qui entraîne une augmentation significative des données. Cette couche réduit la quantité de données tout en conservant les caractéristiques de l'image. En effet, étant donné que les caractéristiques pertinentes ont des valeurs élevées et positives, le seuillage linéaire peut être utilisé pour localiser ces caractéristiques. Dans notre travail, nous utilisons la fonction de seuillage linéaire suivante :

$$I_{ji}^b(x, y) = \begin{cases} 0 & \text{si } \hat{I}_{ji}(x, y) < \tau_b \\ 1 & \text{si } \hat{I}_{ji}(x, y) \geq \tau_b \end{cases}, \quad i \in [1..N] \quad (23)$$

Où τ_b est le seuil de binarisation. Dans notre travail, nous avons fixé ce seuil à 0 car les coefficients de l'image filtrée ($\hat{I}_i(x, y), \forall x, y$) avaient la même probabilité d'être négatifs ou positifs. Ensuite, les N images binarisées ($I_{ji}^b|_{i=1..N}$) sont converties en une image à valeurs entières. Par conséquent, le code binaire N entourant chaque pixel est transformé selon la formule suivante :

$$I_{jd}(x, y) = \sum_{i=1}^N I_{ji}^b(x, y) \cdot 2^{(i-1)} \quad (24)$$

En conséquence, nous obtiendrons une seule image à laquelle nous appliquerons une méthode traditionnelle d'extraction de caractéristiques pour extraire ses caractéristiques.

☑ **Couche d'extraction de vecteur de caractéristiques** : À cette couche, une méthode traditionnelle d'extraction de caractéristiques (LPQ) est appliquée à l'image analysée (I_j) pour extraire le vecteur de caractéristiques (\mathcal{V}).

$$\mathcal{V}_j = \mathcal{F}_{\text{LPQ}}(I_{jd}) \quad (25)$$

La méthode LPQ possède une variété de paramètres qui peuvent être modifiés pour contrôler la taille et la précision du vecteur de caractéristiques. En effet, la précision de ce vecteur affecte l'efficacité du système de diagnostic ; c'est pourquoi nous avons sélectionné et quantifié ses coefficients en utilisant la formule suivante lors de la phase d'entraînement.

$$\tilde{\mathcal{V}}_j = \mathcal{F}_{\text{QNTZ}}(v_j(x)) = \begin{cases} v_j(x) & \text{if } v_j(x) \geq \lambda_0 \cdot \rho_v \\ 0 & \text{ailleurs} \end{cases}, \quad \tilde{\mathcal{V}}_j \in \mathbb{R}^{1 \times L}, \quad x \in [1..L] \quad (26)$$

Et

$$v_j = \mathcal{F}_{\text{SLCT}}(\mathcal{V}_j) \quad (27)$$

Où v_j représente le vecteur de caractéristiques réorganisé selon les coordonnées obtenues par la méthode de sélection de Fisher [36], ce qui donne un vecteur avec un nombre prédéfini de coordonnées (L) dans un ordre décroissant d'importance, ρ_v est la valeur moyenne de v_j et λ_0 est une valeur prédéfinie.

☑ **Optimisation** : Pour créer une banque de filtres pondérés, il est nécessaire de fournir deux paramètres (μ^i, x_0^i). Dans notre travail et pendant la phase d'apprentissage, nous utilisons l'algorithme d'optimisation de chauve-souris pour trouver les meilleures valeurs de performance du système (μ^0, x_0^0).

$$[\mu^{0best}, x_0^{0best}] = \mathcal{F}_{\text{OPT}}(S_{\text{taux}}) \quad (28)$$

Où S_{taux} , qui est le taux de classification du système, représente la fonction objective [37].

II.6.3 Classifier

Dans les problèmes de classification, l'apprentissage automatique discriminant permet de trouver une fonction capable d'identifier correctement les étiquettes des instances. En tant que l'une des techniques les plus importantes en apprentissage automatique discriminant, le classifieur des Machines à Vecteurs de Support (SVM) [38] trouve l'hyperplan (Eq. 29) qui sépare correctement deux classes avec la marge maximale.

$$g(x) = w^T x + b \quad (29)$$

L'apprentissage d'un SVM a été formulé comme un problème d'optimisation avec contrainte qui minimise la fonction objective suivante :

$$J(w, b, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \quad (29)$$

Avec les contraintes suivantes :

$$\begin{cases} y_i [x_i w_i + b] \geq 1 - \xi_i, & i = 1, 2, \dots, N \\ \xi_i \geq 0 \end{cases} \quad (30)$$

Où C représente le paramètre de contrôle de la marge. La fonction lagrangienne pour SVM est construite en ajoutant une somme pondérée des contraintes à la fonction objectif :

$$\mathcal{L} = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i - \sum_{i=1}^N \alpha_i [y_i (x_i w_i + b) - 1 + \xi_i] - \sum_{i=1}^N \mu_i \xi_i \quad (31)$$

Après l'exécution des processus pertinents, nous obtenons w et b , qui déterminent l'orientation optimale de l'hyperplan :

$$w = \sum_{i=1}^N \alpha_i y_i x_i, \quad \text{et} \quad b = \frac{1}{N_s} \sum_{s \in S} (y_s - \sum_{m \in S} \alpha_m y_m x_m \cdot x_s) \quad (32)$$

Où S représente l'ensemble des indices des vecteurs de support ($i / \alpha_i > 0$), et x_s représente le point de données qui satisfait $\sum_{i=1}^N \alpha_i y_i = 0$.

En utilisant les vecteurs de support, qui sont les échantillons les plus proches de l'hyperplan de décision, les SVM peuvent classer de nouveaux points en les évaluant par rapport à cet hyperplan. Cela permet de généraliser la classification à de nouvelles données après l'apprentissage initial. Cependant, chaque nouveau point x_{new} est classé en évaluant la formule suivante :

$$y_{\text{new}} = \text{sgn}(w_i \cdot x_{\text{new}} + b) \quad (33)$$

Les SVM sont basées sur des principes mathématiques solides et ont été largement étudiées dans la littérature. Elles sont populaires en raison de leur capacité à traiter des problèmes de classification complexes et à obtenir de bonnes performances prédictives.

II.7 Conclusion

L'extraction de caractéristiques en DAO est un domaine de recherche prometteur qui joue un rôle crucial dans de nombreux domaines, tels que l'architecture, l'ingénierie, la médecine, etc. En surmontant les défis actuels et en exploitant les tendances futures, l'extraction de caractéristiques en DAO continuera de progresser et de fournir des solutions plus avancées pour l'analyse, la modélisation et la compréhension des données en DAO. Dans notre proposition, avant d'appliquer directement la méthode d'extraction de caractéristiques traditionnelle à l'image brute, celle-ci est soigneusement analysée à l'aide d'un ensemble spécifique de filtres gaussiens pondérés. Ensuite, un algorithme d'extraction de caractéristiques traditionnelles (LPQ) est appliqué au descripteur obtenu. De plus, le classifieur optimise les paramètres des poids de filtres à l'aide d'une méthode d'optimisation basée sur un algorithme inspiré des chauves-souris afin d'obtenir des performances optimales et un comportement d'analyse profonde.

Chapitre 3

RESULTATS

EXPERIMENTAUX

Résumé

L'extraction de caractéristiques est utilisée dans le domaine du traitement et de l'analyse d'images pour extraire des informations importantes. Dans notre travail, nous avons proposé une méthode d'extraction de caractéristiques avec des paramètres ajustables automatiquement avec le classifieur. Dans ce chapitre, nous présentons les résultats de nos expériences et les analyses qui en découlent. L'objectif de ce chapitre est de fournir une évaluation objective de notre approche proposée et d'analyser sa performance dans le contexte de notre étude.

III.1 Introduction

III.2 Contexte de travail

III.3 Ensemble de données

III.4 Configurations de test

III.5 Évaluation des performances

III.6 Discussion

III.7 Conclusion

Résultats Expérimentaux

III.1 Introduction

Le lymphome, également appelé cancer lymphatique [39], est un type de cancer qui affecte les cellules du système lymphatique responsables de la production et de la circulation des lymphocytes, qui sont des cellules immunitaires essentielles. Ce type de cancer est généralement diagnostiqué par des tests tels qu'un examen physique, des analyses de sang, des biopsies de ganglions lymphatiques ou d'autres tissus, ainsi que des tests d'imagerie tels qu'un scanner ou une radiographie. Dans ce contexte, de nombreux chercheurs ont développé des techniques d'analyse d'images médicales assistées par ordinateur pour aider au diagnostic de cette maladie. Malheureusement, le plus grand défi réside dans la similarité des images des trois types de cette maladie, rendant difficile leur classification. Pour cela, nous allons tenter de tester notre méthode sur la classification des trois types de ce lymphome. En effet, dans ce chapitre, nous présentons les résultats de nos expériences et les analyses qui en découlent. L'objectif de ce chapitre est de fournir une évaluation objective de notre approche proposée et d'analyser sa performance dans le contexte de notre étude.

III.2 Contexte de travail

La méthode proposée dans ce travail vise à analyser et classifier les images, et elle peut être testée et ajustée dans n'importe quel contexte. Dans cette étude, nous examinerons cette méthode dans le domaine de l'analyse et de la classification des images médicales, plus précisément dans le domaine de l'analyse d'images de lymphomes. Le système proposé peut aider les médecins dans le processus de diagnostic et de suivi du lymphome en fournissant

une analyse automatisée des images de lymphomes et en contribuant à la détection et à la caractérisation des lésions liées à cette maladie. Il convient de noter que notre système est spécifiquement dédié à la tâche de classification des types de lymphome. Il peut donc contribuer à la caractérisation des différents sous-types de lymphome en analysant les caractéristiques morphologiques et les motifs observés dans les images médicales [40].

III.2.1 Cancer lymphatique

Le système lymphatique joue un rôle crucial dans la réponse immunitaire de l'organisme. Il aide à protéger le corps contre les infections en détruisant les agents pathogènes tels que les bactéries et les virus. En outre, il aide à prévenir l'accumulation de liquide dans les tissus corporels, en transportant l'excès de liquide et les déchets à travers les vaisseaux lymphatiques pour élimination. Les lymphocytes peuvent être exposés à de nombreuses maladies, parmi lesquelles le cancer (cancer lymphome) est la plus dangereuse. Lorsqu'il infecte ces cellules, le cancer peut représenter une menace sérieuse pour l'ensemble du corps en attaquant la principale ligne de défense de corps humain. Le lymphome est un type de cancer qui se forme dans les cellules du système lymphatique, qui fait partie du système immunitaire. Le système lymphatique est responsable de la production et du transport des lymphocytes, des cellules immunitaires qui aident à combattre les infections et les maladies. Le cancer lymphome survient lorsque les cellules lymphatiques se développent de manière anormale et se multiplient de manière incontrôlée, formant des tumeurs dans les ganglions [41].

III.2.2 Sous-types de lymphome

Il existe de principalement trois sous-types de lymphomes, qui sont classés en fonction des caractéristiques spécifiques des cellules lymphocytaires et de leur comportement clinique[42]:

☑ **CLL** : La leucémie lymphoïde chronique (*Chronic Lymphocytic Leukemia - CLL*) est un type de lymphome à cellules B qui se caractérise par l'accumulation de lymphocytes B matures et anormaux dans la moelle osseuse, le sang et les ganglions lymphatiques.

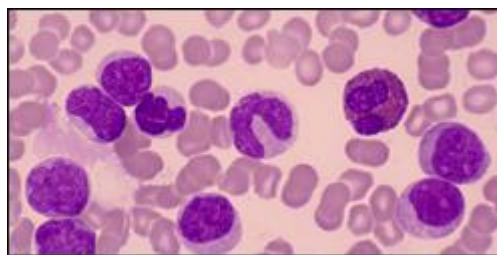


Fig III.1 Leucémie Lymphoïde Chronique

La CLL est généralement une maladie à évolution lente, ce qui signifie que les cellules cancéreuses se développent progressivement et peuvent souvent passer inaperçues pendant des années. Elle est souvent diagnostiquée chez des personnes âgées, mais peut également toucher des personnes plus jeunes.

Les symptômes de la CLL peuvent varier, mais ils peuvent inclure une fatigue persistante, une augmentation des ganglions lymphatiques, une augmentation de la taille de la rate et du foie, des infections fréquentes, des saignements anormaux et une perte de poids inexplicée. Le diagnostic de la CLL est généralement confirmé par une analyse sanguine et une biopsie de la moelle osseuse. Les facteurs pronostiques, tels que le stade de la maladie, la présence de certaines anomalies chromosomiques et les caractéristiques des cellules cancéreuses, sont pris en compte pour déterminer le meilleur plan de traitement.

☑ **MCL** : Le lymphome à cellules du manteau (*Mantle Cell Lymphoma - MCL*) est un type de lymphome non hodgkinien agressif, qui se développe à partir des lymphocytes B dans la zone du manteau des ganglions lymphatiques. Le MCL est généralement diagnostiqué chez des personnes âgées, avec un âge médian de diagnostic autour de 60 ans.

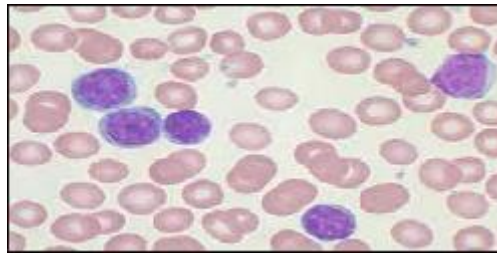


Fig III.2 Lymphome à cellules du manteau

Les symptômes du MCL peuvent inclure une augmentation des ganglions lymphatiques, des sueurs nocturnes, une fatigue persistante, une perte de poids involontaire, de la fièvre et une augmentation de la taille de la rate et du foie. En raison de sa nature agressive, le MCL a tendance à se propager rapidement à d'autres parties du corps, y compris la moelle osseuse, les ganglions lymphatiques distants, le tube digestif et le système nerveux central.

Le diagnostic du MCL est généralement confirmé par une biopsie d'un ganglion lymphatique ou d'un autre site affecté. Des tests d'imagerie, tels que la tomodensitométrie (CT scan) et la tomographie par émission de positrons (PET scan), peuvent également être utilisés pour évaluer l'étendue de la maladie.

- **FL** : Le lymphome folliculaire (*Follicular Lymphoma-FL*) est un sous-type de lymphome non hodgkinien. Il se caractérise par la croissance anormale de cellules B dans les follicules lymphoïdes, qui sont des structures du système lymphatique.

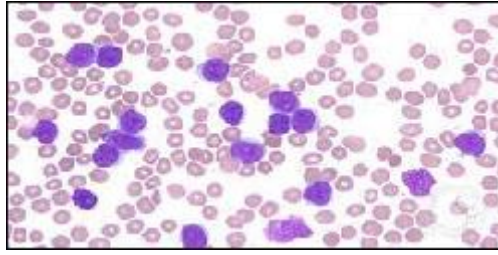


Fig III.3 Lymphome folliculaire

Les symptômes du FL peuvent inclure une augmentation indolore des ganglions lymphatiques, une fatigue persistante, une perte de poids involontaire, des sueurs nocturnes, des épisodes de fièvre inexplicables et parfois des démangeaisons cutanées.

Le diagnostic du FL implique généralement une combinaison d'examen médicaux, de tests sanguins, de biopsies et d'imagerie. Des analyses sanguines peuvent être effectuées pour détecter des marqueurs spécifiques du lymphome. La confirmation du diagnostic est généralement obtenue par une biopsie des ganglions lymphatiques affectés ou d'autres tissus. Des tests d'imagerie tels que la CT ou la tomographie PET peuvent être utilisés pour évaluer l'étendue de la maladie et déterminer le stade du lymphome.

III.3 Ensemble de données

Il est clair que ces types de systèmes nécessitent une évaluation à l'aide d'une base de données typique et reconnue afin de mesurer leur niveau de performance et de déterminer les paramètres appropriés de la méthode d'extraction des caractéristiques.

III.3.1 Motivation

Cet ensemble de données vise à répondre au besoin d'identifier trois sous-types de lymphome : CLL, FL et MCL. Actuellement, des sondes spécifiques sont utilisées pour distinguer de manière fiable ces sous-types, mais cela entraîne des coûts supplémentaires et des exigences matérielles. Cependant, des pathologistes experts spécialisés dans ces types de lymphomes ont obtenu des résultats prometteurs en utilisant la coloration à l'Hématoxyline-Eosine (H&E) pour différencier ces sous-types, ce qui suggère que des approches basées sur l'intelligence artificielle peuvent être utilisées. Cet ensemble de données a été créé pour refléter des situations réelles et contient donc des échantillons préparés par différents pathologistes dans différents sites. Ils ont également sélectionné des échantillons présentant un degré de variation de coloration plus important que ce que l'on pourrait normalement attendre.

III.3.2 Description

L'ensemble de données est composé de 374 images de taille 1388x1040 pixels. Ces images sont réparties en 113 pour la classe CLL, 139 pour la classe FL et 122 pour la classe MCL.

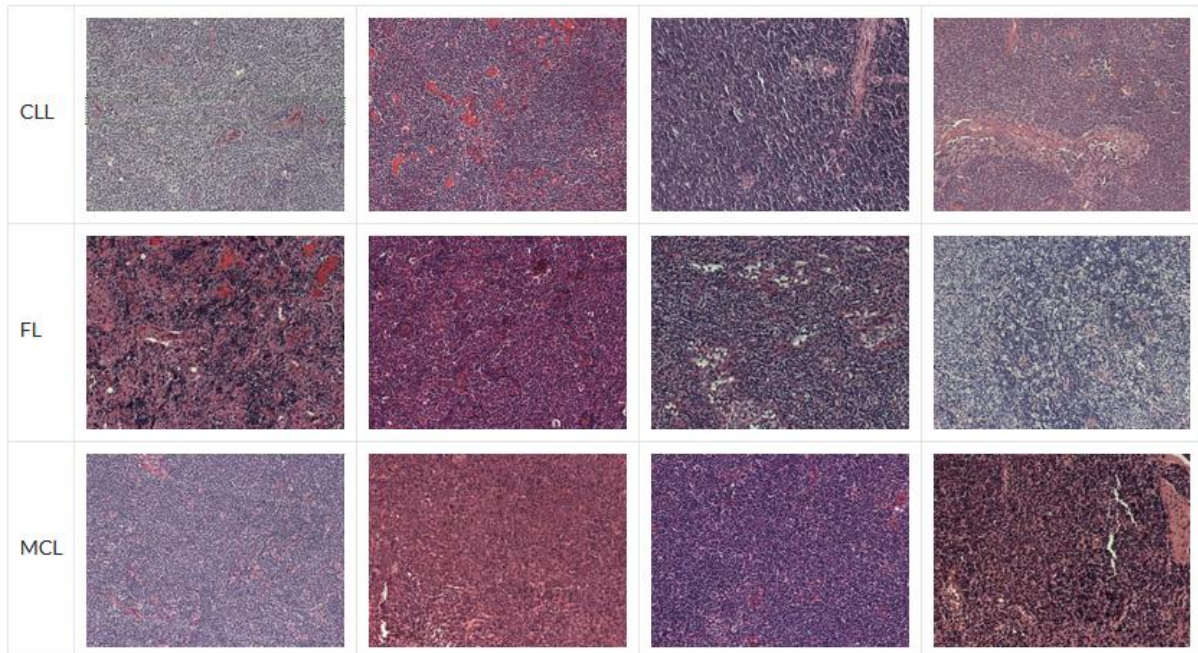


Fig III.4 Sous-types de lymphome dans l'ensemble de données de test

Les données ont été collectées à partir de différentes sources afin de créer une cohorte représentative comprenant des variations courantes de coloration et de numérisation. La figure III.4 présente quelques exemples d'images représentant les trois sous-types de lymphome de cet ensemble de données [43].

III.4 Configurations de test

Avant de commencer les expériences d'évaluation du système, il est important de clarifier le protocole des tests ainsi que le matériel utilisé pour mener les différentes expériences.

III.4.1 Protocole de tests

Tout d'abord, il convient de noter que notre système est conçu pour la classification d'un sous-type à la fois (un contre tous). Cette base de données a été divisée en deux principales galeries. La première galerie (ensemble d'apprentissage) contient 80% de la base de données, soit 300 images, tandis que la deuxième galerie (ensemble de validation) contient 20% de la base de données, soit 74 images. En raison du petit nombre d'images, nous avons utilisé une validation croisée à *10-fold* pour la phase d'apprentissage. L'ensemble de données de 300 images a été divisé de manière aléatoire en 10 sous-ensembles, et 10 cycles d'apprentissage et

de test ont été réalisés. Après avoir entraîné le classifieur et déterminé les paramètres optimaux de la méthode d'extraction de caractéristiques proposée, le meilleur modèle a été sélectionné et utilisé pour classer les 74 images aveugles restantes.

III.4.2 Implémentation

Toutes les expériences ont été réalisées avec le logiciel Matlab 2018a sur un ordinateur portable LENOVO modèle Legion Y520 équipé d'un processeur Intel(R) Core(TM) i7-7700HQ fonctionnant à 2,8 GHz et de 16 Go de mémoire vive (DRAM). En plus des nombreuses routines personnalisées développées, nous avons également utilisé les fonctions intégrées de Matlab.

III.5 Évaluation des performances

Dans cette section, nous évaluerons de manière exhaustive la méthode proposée, de l'étape de l'utilisation d'une méthode traditionnelle à l'étape d'optimisation (sélection des paramètres optimaux de la carte logistique maîtresse pour les différents filtres). Pour cela, nous diviserons les expériences en trois parties :

- **FF-LPQ (Filtering-Free & LPQ)** : Cette configuration est considérée comme traditionnelle car la méthode d'extraction de caractéristiques ne comporte pas d'interaction avec le classifieur. Dans ce cas, la méthode d'extraction de caractéristiques ne comprend pas de couche de filtrage, c'est-à-dire que l'algorithme LPQ est directement appliqué sur l'image.
- **FG-LPQ (Filtering-Gaussien & LPQ)** : Cette configuration est également considérée comme traditionnelle car la méthode d'extraction de caractéristiques ne comporte pas d'interaction avec le classifieur. Dans ce cas, la méthode d'extraction de caractéristiques comprend une couche de filtrage et l'algorithme LPQ est appliqué sur l'image de sortie de la combinaison des images filtrées.
- **FO-LPQ (Filtering-Optimization & LPQ)** : Cette configuration est considérée comme profonde car la méthode d'extraction de caractéristiques présente une interaction avec le classifieur. Dans ce cas, les paramètres de la méthode d'extraction de caractéristiques sont réglés par le classifieur.

III.5.1 FF-LPQ

L'algorithme LPQ est une méthode traditionnelle d'extraction de caractéristiques. Avant de tester notre propre méthode d'extraction de caractéristiques, nous appliquons d'abord LPQ à l'image brute afin de le comparer ultérieurement avec notre méthode. Ainsi, la fenêtre et le

nombre de BIN de LPQ sont fixés à 5×5 et 9 respectivement. Nous allons explorer les trois sous-types de lymphome (CLL, FL et MCL) pour déterminer le taux de classification de chacun. Pour cela, nous avons incorporé les images en format niveau de gris (format NG) dans un modèle de diagnostic de lymphome basé sur la technique d'extraction de caractéristiques LPQ (modèle sans couche de filtrage). La figure III.5 illustre les résultats pour les trois sous-types de lymphomes examinés.

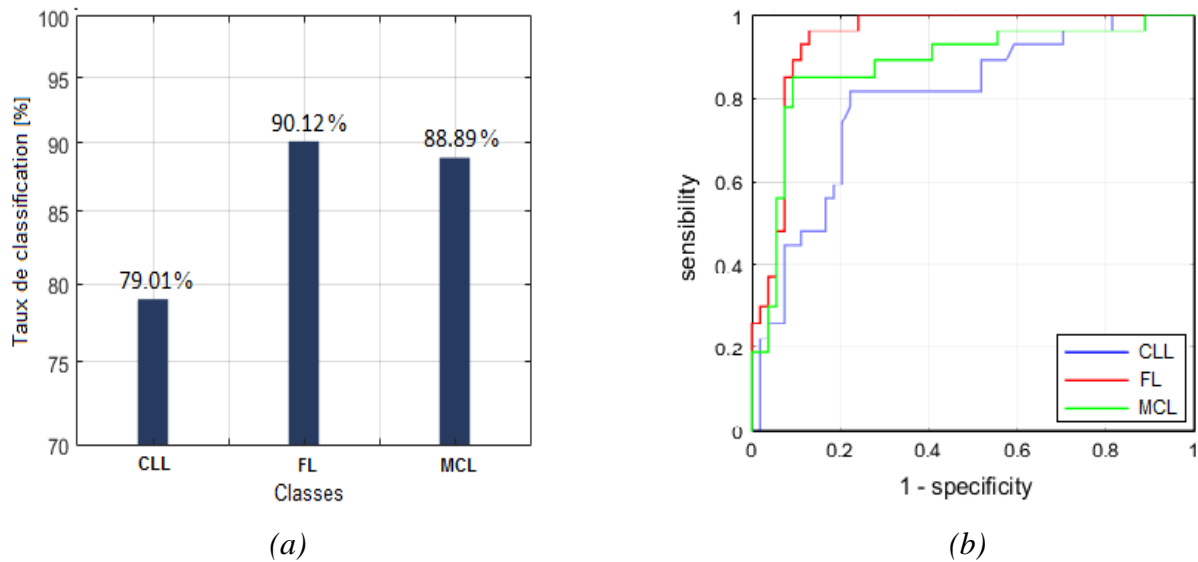


Fig III.5 Performance de système sous la configuration FF-LPQ. (a) Taux de classification et (b) Courbes ROC

Les deux graphiques de cette figure illustrent respectivement le taux de classification (*accuracy-ACC*), Fig. III.5.(a), et la courbe caractéristique de fonctionnement (*Receiver Operating Characteristic-ROC*), Fig. III.5.(b). Par conséquent, à partir des différents graphiques de la figure III.5, on peut tirer deux observations majeures :

- D'après la Fig. III.5.(a) et en termes de ACC, la méthode LPQ présente une performance très acceptable et compétitive par rapport à l'état de l'art.
- D'après la Fig. III.5.(b), le sous-type de lymphome FL a montré la plus grande efficacité, avec un taux de 90.12%.

Le tableau III.1 illustre les résultats expérimentaux pour tous les sous-types de lymphome (CLL, FL et MCL), avec toutes les métriques mesurant la performance du système.

Tableau III.1 Performance de système sous la configuration FF-LPQ (format NG)

Lymphome	ACC	λ_0	AUC	N_f	TP	FN	TN	FP	T_d	
CLL	79.01	0.10	79.49	1	42	5	12	22	1.0802	
NG	FL	90.12	0.75	94.17	151	48	2	6	25	-0.1249
MCL	88.89	0.20	87.93	161	49	4	5	23	0.1722	

Par exemple, pour le sous-type de lymphome FL, les performances en termes de (ACC, Aire sous la courbe ROC – (Area Under the ROC Curve (AUC))) atteignent respectivement (90.12%, 94.17%) au seuil de décision (T_d) de -0.1249, et une matrice de confusion de M_c (TP, FP, FN, TN) = (42, 5, 12, 22) avec un nombre de caractéristiques (N_f) égal à 151 et un coefficient de réglage (λ_0) égal à 0.75.

Les images des trois sous-types de lymphome (CLL, FL et MCL) dans la base de données sont en couleur (format RGB). Par conséquent, la configuration peut être appliquée individuellement sur chaque bande de couleur séparément. La figure III.6 ainsi que le Tableau III.2 illustrent les résultats des tests pour les différents sous-types de lymphome.

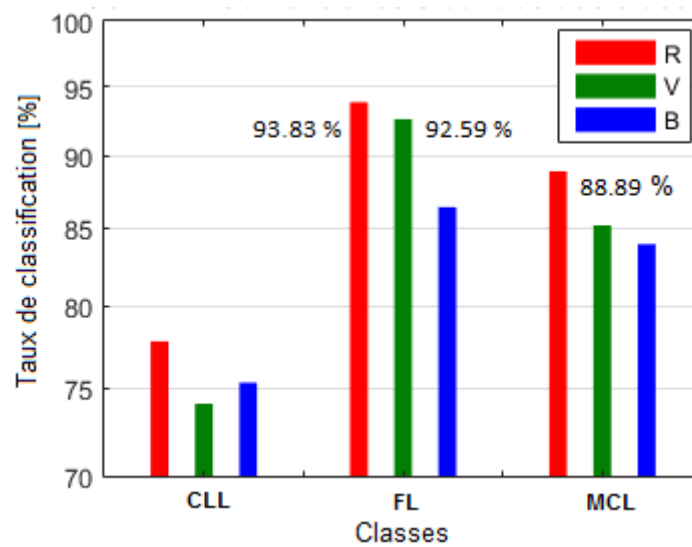


Fig III.6 Performance de système sous la configuration FF-LPQ pour les différentes bandes de couleur.

Tableau III.2 Performance de système sous la configuration FF-LPQ (format RGB)

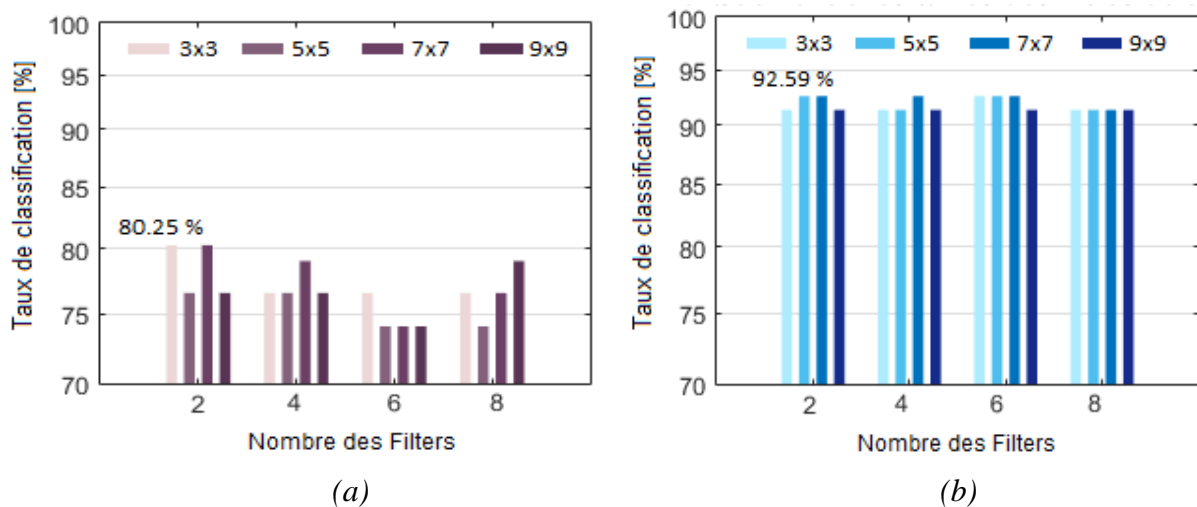
Lymphome	ACC	λ_0	AUC	N_f	TP	FN	TN	FP	T_d
R	77.78	0.10	82.44	1	41	5	13	22	0.9734
FL	93.83	0.15	94.72	86	51	2	3	25	-0.7716
MCL	88.89	0.55	86.15	116	49	4	5	23	-0.0773
G	74.07	0.10	77.71	1	41	8	13	19	0.9146
FL	92.59	0.40	93.00	186	52	4	2	23	0.1475
MCL	85.19	0.65	86.76	101	47	5	7	22	0.4291
B	75.31	1	82.17	6	39	5	15	22	0.9797
FL	86.42	0.7	83.75	31	51	8	3	19	1.7784
MCL	83.95	0.4	84.22	51	46	5	8	22	-1.5378

En observant et en analysant la figure III.6 et le Tableau III.2, nous pouvons constater que dans tous les cas, la bande rouge a considérablement amélioré les performances du système. Dans le meilleur scénario (bande rouge de FL), le système peut fonctionner avec des valeurs (ACC, AUC) de (93.83%, 94.72%) à $T_d = -0.7716$ et une matrice de confusion $M_c = (51, 2, 25, 3)$ avec des valeurs (N_f, λ_0) de (86, 0.15).

III.5.2 FG-LPQ

Dans cette section, nous allons examiner les performances du système de diagnostic avec notre méthode d'extraction de caractéristiques. En effet, les trois sous-types de lymphome (CLL, FL et MCL), en format NG, seront filtrés à l'aide d'un ensemble de filtres pour obtenir de nombreux descripteurs. Après avoir combiné ces descripteurs, la méthode LPQ sera ensuite appliquée à la combinaison obtenue pour former un vecteur de caractéristiques. Chaque sous-type de lymphome est analysé à l'aide de N filtres ($N \in [2, 4, 6, 8]$) de quatre tailles distinctes ($k_1 \times k_2 \in \{3 \times 3, 5 \times 5, 7 \times 7, 9 \times 9\}$). Notamment, les valeurs initiales de μ_0 et x_0^0 pour la carte logistique maître (\mathcal{L}_0) ont été choisies de manière aléatoire.

Dans le système basé sur FG-LPQ, les valeurs de μ_0 et x_0^0 ont été assignées aléatoirement pour chaque combinaison de nombre de filtres (N) et taille de filtre ($k_1 \times k_2$). En effet, étant donné que la variation du nombre/taille du filtre (N et $k_1 \times k_2$) produit différentes représentations de caractéristiques, nous pouvons expérimentalement sélectionner une combinaison ($N, k_1 \times k_2$) qui améliore la précision du système. Afin de comprendre l'effet de ces paramètres, la figure III.7 montre les résultats de l'ACC en fonction du nombre de filtres (N) et de toutes les tailles de filtres ($k_1 \times k_2$) pour les trois sous-types de lymphome.



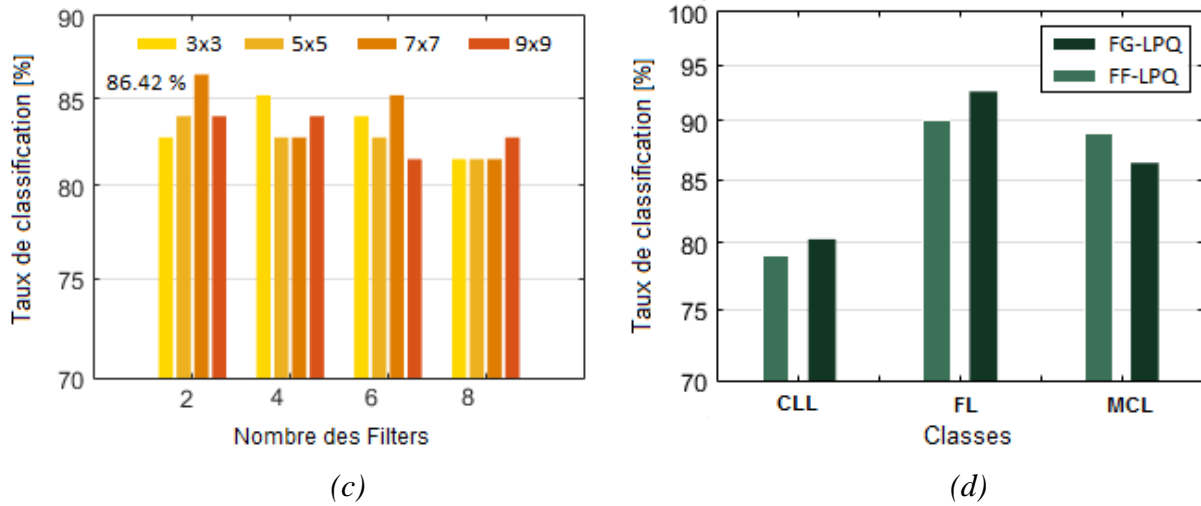


Fig III.7 Performance de système sous la configuration FG-LPQ. (a) Taux de classification de CLL, (b) Taux de classification de FL, (c) Taux de classification de MCL et (d) Comparaison des performances.

Cette figure démontre clairement que les moyennes tailles du filtre donnent de meilleurs résultats lorsque le nombre de filtres diminue, tandis que les grandes tailles donnent de meilleurs résultats lorsque le nombre de filtres augmente. Ainsi, le meilleur ACC a été généré par le sous-type FL (voir Fig. III.7.(b)) avec deux filtres de taille 7×7 . De manière évidente, deux filtres de taille 7×7 avec une matrice de confusion de $M_c = (51, 3, 3, 24)$ et des valeurs (N_f, λ_0) de $(116, 0.5)$ constituent la configuration optimale en termes de temps de traitement. Heureusement, il est évident que l'incorporation de la couche de filtrage entraîne une amélioration de 6.00% et 25.00% des performances du système par rapport à la configuration sans filtrage pour CLL et FL respectivement (voir Fig. III.7.(d)). Enfin, il convient de noter que ces résultats sont liés à certaines valeurs des paramètres du filtre (μ_0, x_0) , ce qui suggère qu'ils peuvent être améliorés en faisant varier ces valeurs de paramètres. Le Tableau 3 résume les résultats expérimentaux globaux pour le meilleur cas obtenu pour tous les trois sous-types de lymphome.

Tableau III.3 Performance de système sous la configuration FG-LPQ (format NG)

Lymphome	ACC	λ_0	AUC	N_f	T_d	Filtres			
						Taille	Nombre	$[\mu^0, x_0^0]$	
CLL	80.25	0.1	83.27	1	1.0176	3×3	2	[0.95 3.90]	
NG	FL	92.59	0.5	93.96	116	0.2437	7×7	2	[0.71 3.72]
MCL	86.42	0.35	89.44	131	-0.7681	7×7	2	[0.67 3.62]	

De même, dans le système basé sur la configuration FG-LPQ, les trois bandes de couleur de chaque sous-type de lymphome ont été examinées. Le Tableau III.4 présente les résultats obtenus pour le meilleur cas des trois sous-types de lymphome.

Tableau III.4 Performance de système sous la configuration FG-LPQ (format RGB)

Lymphome	ACC	λ_0	AUC	N_f	T_d	Filtres			
						Taille	Nombre	$[\mu^0, x_0^0]$	
R	CLL	79.01	0.1	83.23	1	1.1101	3×3	2	[0.95 3.90]
	FL	95.06	0.3	96.50	161	-1.3069	7×7	2	[0.71 3.72]
	MCL	83.95	0.2	73.66	141	0.5598	7×7	2	[0.67 3.62]
G	CLL	80.25	0.1	82.17	1	1.0235	3×3	2	[0.95 3.90]
	FL	93.83	0.35	96.71	141	0.4685	7×7	2	[0.71 3.72]
	MCL	82.72	0.7	88.82	196	-0.7631	7×7	2	[0.67 3.62]
B	CLL	67.90	0.1	69.51	1	-1.7648	3×3	2	[0.95 3.90]
	FL	86.42	0.7	87.38	111	-0.1521	7×7	2	[0.71 3.72]
	MCL	79.01	0.15	72.29	51	0.3993	7×7	2	[0.67 3.62]

Les résultats expérimentaux de ce tableau montrent que les performances de notre système se sont généralement améliorées lorsque nous avons analysé les images par bande de couleur. Dans le cas de FL en utilisant la bande rouge, il est évident qu'en utilisant 2 filtres de taille 7×7 , nous avons obtenu une amélioration de 87.31% des performances du système. Dans ce cas, le système de diagnostic fonctionne avec un taux de classification (ACC) de 95.06% et une AUC de 96.50%. La matrice de confusion résultante est $M_c = (52, 2, 2, 25)$, tandis que les valeurs de (N_f, λ_0) sont (161, 0.3). Enfin, les résultats présentés dans cette section démontrent la faisabilité de l'ajout de la couche de filtrage, ce qui améliore considérablement les performances du système, permettant à l'ACC de dépasser 95%. Enfin, il convient de mentionner qu'il est possible d'améliorer ces performances, car la technique LPQ contient de nombreux paramètres qui, avec une sélection minutieuse, peuvent être utilisés pour améliorer l'efficacité du système de diagnostic des sous-types de lymphome.

III.5.3 FO-LPQ

Dans les expériences précédentes, les paramètres de filtre (μ^0, x_0^0) ont été sélectionnés de manière aléatoire pour notre méthode d'extraction de caractéristiques. Dans cette section, nous allons tenter de déterminer les valeurs optimales pour ces paramètres afin d'obtenir les meilleures performances. Dans ces tests, nous ne nous limitons pas à l'optimisation des

meilleures configurations mentionnées précédemment ; nous chercherons plutôt les valeurs optimales en utilisant toutes les configurations disponibles, c'est-à-dire pour des valeurs de N allant de 2 à 8 et les quatre tailles de filtres utilisées précédemment. Par conséquent, nous limitons la plage de recherche de paramètres (μ^0, x_0^0) à la plage de 0 à 1. Les résultats expérimentaux optimaux obtenus après l'exécution de l'algorithme d'optimisation (*bat algorithm*) sont présentés dans le Tableau 5.

Tableau III.5 Performance de système sous la configuration FO-LPQ (format NG)

Lymphome	ACC	λ_0	AUC	N_f	T_d	Filtres			
						Taille	Nombre	$[\mu^{0best}, x_0^{0best}]$	
CLL	81.48	0.1	84.05	1	1.0282	3×3	2	[0.90 3.90]	
NG	FL	95.06	0.55	96.91	141	0.1312	7×7	2	[0.35 3.82]
MCL	88.89	0.4	90.26	141	0.1747	9×9	2	[0.36 3.73]	

Dans ce tableau et en comparaison avec les résultats précédents, on peut facilement constater que le processus d'optimisation a amélioré les performances du système de diagnostic de Lymphome d'environ 6.23%, 3.34% et 18.19% respectivement pour CLL, FL et MCL. Il est clair que dans les tous les cas, 2 filtres ont donné les meilleurs résultats.

Dans le cas de CLL et pour les paramètres optimaux du filtre, 2 filtres de taille 3×3 , $(\mu^0, x_0^0) = (0.90, 3.90)$, le système fonctionne avec des valeurs (ACC, AUC) de (81.48%, 84.05%) pour $T_d = 1.0282$ et une matrice de confusion de $M_c = (44, 5, 10, 22)$, avec des valeurs (N_f, λ_0) de (1, 0.10). Dans le cas de Lymphome FL, les paramètres optimaux du filtre, 2 filtres de taille 7×7 , étaient $(\mu^0, x_0^0) = (0.35, 3.82)$ et le système de diagnostic de lymphome peut fonctionner avec des valeurs (ACC, AUC) de (95.06%, 96.91%) pour $T_d = 0.1312$ et une matrice de confusion de $M_c = (52, 2, 2, 25)$ avec des valeurs (N_f, λ_0) de (141, 0.55). Finalement, Dans le cas de MCL, les paramètres optimaux du filtre, 2 filtres de taille 9×9 , étaient $(\mu^0, x_0^0) = (0.36, 3.73)$ et le système de diagnostic de lymphome peut fonctionner avec des valeurs (ACC, AUC) de (88.89%, 90.26%) pour $T_d = 0.1747$ et une matrice de confusion de $M_c = (48, 3, 6, 24)$ avec des valeurs (N_f, λ_0) de (141, 0.40).

Pour démontrer l'importance du processus d'optimisation, nous avons comparé les résultats optimaux à ceux obtenus sans optimisation dans la figure III.8. Ces deux figures montrent l'impact de la procédure d'optimisation sur les performances du système.

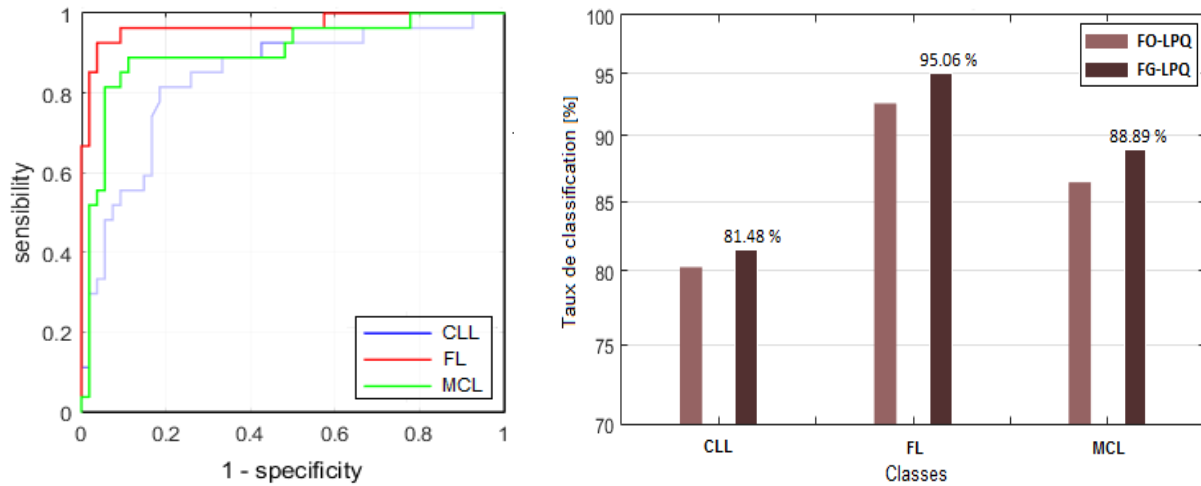


Fig III.8 Performance de système sous la configuration FO-LPQ. (a) Courbe ROC de trois classes et (b) Comparaison des performances

Dans cette section également, nous avons essayé d'appliquer notre méthode (FO-LPQ) aux bandes de couleurs de chaque sous-type de lymphome séparément, et les résultats expérimentaux sont présentés dans le Tableau III.6.

Tableau III.6 Performance de système sous la configuration FO-LPQ (format RGB)

Lymphome	ACC	λ_0	AUC	N_f	T_d	Filtres			
						Taille	Nombre	$[\mu^{0best}, x_0^{0best}]$	
R	CLL	85.19	0.1	85.36	1	1.0548	3 × 3	2	[0.90 3.90]
	FL	93.83	0.6	96.30	136	-1.8818	7 × 7	2	[0.35 3.82]
	MCL	79.01	0.65	77.50	106	-0.9861	9 × 9	2	[0.36 3.73]
G	CLL	75.31	0.85	79.42	6	0.9861	3 × 3	2	[0.90 3.90]
	FL	96.30	0.55	99.59	141	0.7131	7 × 7	2	[0.35 3.82]
	MCL	82.71	0.7	77.02	86	1.6832	9 × 9	2	[0.36 3.73]
B	CLL	79.01	0.1	78.74	1	1.3349	3 × 3	2	[0.90 3.90]
	FL	88.89	0.25	93.07	156	-0.3158	7 × 7	2	[0.35 3.82]
	MCL	82.72	0.9	81.69	196	0.0952	9 × 9	2	[0.36 3.73]

D'après ce tableau, on peut clairement voir qu'en général, la méthode d'optimisation peut également contribuer davantage à l'amélioration des performances du système de diagnostic de lymphome. Il convient également de rappeler qu'un autre algorithme d'optimisation peut être utilisé, tel que l'algorithme génétique (GA) [44] ou l'optimisation par essaim de particules (PSO) [45], et rechercher dans une large gamme pour améliorer les paramètres du filtre.

III.6 Discussion

Dans cette section, nous allons tenter de présenter et discuter des principales conclusions et fournir quelques recommandations. De plus, une étude comparative avec des recherches antérieures pertinentes sera présentée.

III.6.1 Principales conclusions

Ce travail explore l'utilisation de l'analyse de texture pour distinguer les trois sous-types de lymphome. Nous avons constaté que l'application directe des techniques d'extraction de caractéristiques traditionnelles telles que LPQ aux images brutes ne donne pas de résultats encourageants, contrairement à l'application de ces techniques aux descripteurs dérivés de l'analyse approfondie de l'image. De nos résultats expérimentaux, on peut déduire que :

- Dans le contexte pratique, l'utilisation de la procédure de filtrage permettra d'éliminer les caractéristiques non pertinentes. Pour les images de lymphome, la texture de l'image est la caractéristique clé. Par conséquent, il est conseillé d'appliquer des filtres capables d'améliorer le contraste de ces caractéristiques afin de faciliter une extraction efficace de leurs informations.
- Il est essentiel d'établir une relation entre le classifieur et la méthode d'extraction de caractéristiques. Dans ce contexte, le classifieur agit en tant qu'expert qui sélectionne les paramètres de la méthode d'extraction de caractéristiques afin d'améliorer la performance du système de diagnostic (comportement profond).
- Il convient de souligner que l'analyse individuelle de chaque bande de couleur dans une image, étant donné qu'elle contient trois bandes de couleur (format RGB), peut améliorer les performances du système de diagnostic de lymphome en exploitant les données supplémentaires fournies par chaque bande de couleur.

Pour conclure, il est important de noter qu'il reste encore de nombreux aspects à améliorer dans ce système. Par exemple, l'adoption de techniques telles que la fusion de données et l'analyse pyramidale pourrait permettre d'optimiser davantage les performances du système.

III.6.2 Comparaison avec les méthodes précédentes

Dans le domaine du développement des systèmes de reconnaissance de formes, il est primordial de comparer les résultats obtenus avec ceux des recherches précédentes. Afin d'assurer une comparaison équitable, seules les études ayant utilisé le même ensemble de données seront sélectionnées. Dans cette section, nous réalisons une étude comparative avec

quelques travaux récents afin de démontrer l'efficacité de notre méthode proposée par rapport aux méthodes existantes. Les résultats de ces travaux sont résumés dans le Tableau III.7.

Tableau III.7 Comparaison des performances avec des travaux récents dans l'état de l'art

Papier	Méthode	Classifieur	Taux de classification [%]		
			CLL	FL	MCL
[46]	Transform-Based Global Features	WND-5	91.00	90.00	75.00
		BBN	77.00	85.00	52.00
		RBF	85.00	88.00	78.00
[47]	/	ANN	79.68	81.34	79.23
FO-LPQ	Chaos-LPQ	SVM	81.48	95.06	88.89

Ce tableau met clairement en évidence l'efficacité du système de diagnostic de sous-types de lymphome proposé, car notre système dépasse les méthodes d'extraction de caractéristiques répertoriées dans le tableau en termes d'ACC (jusqu'à 95.06 %), ce qui permet son utilisation dans le diagnostic assisté par ordinateur (CAD).

III.7. Conclusion

Les résultats de l'évaluation du système de diagnostic de sous-types de lymphome a démontré son efficacité et sa précision dans la classification et la discrimination des différents sous-types de cette maladie. Les résultats obtenus ont montré des performances prometteuses, avec un taux de précision élevé et une valeur AUC significative. Cela indique que le système est capable de fournir des informations diagnostiques fiables et précises, ce qui peut avoir un impact positif sur le processus de prise de décision clinique. Notre système de diagnostic de sous-types de lymphome peut offrir un outil précieux pour les médecins, leur permettant d'améliorer la précision du diagnostic et de fournir un traitement adapté à chaque patient. Il ouvre également la voie à de futures recherches et développements dans le domaine de l'intelligence artificielle et de l'analyse d'images médicales, visant à améliorer la prise en charge des patients atteints de lymphome et à contribuer aux avancées médicales.

Conclusion Générale

Conclusion

Les besoins croissants en radiologues sont principalement dus à l'expansion rapide de l'imagerie médicale grâce aux avancées technologiques. Cette demande accrue a entraîné une augmentation de la charge de travail des radiologues, ce qui peut malheureusement entraîner des erreurs de diagnostic en raison de la surcharge de travail des spécialistes. Récemment, l'intégration de l'intelligence artificielle (IA) dans la pratique clinique de l'imagerie médicale a joué un rôle important dans l'amélioration de la précision de diagnostic et de l'efficacité. En plus d'améliorer la précision de diagnostic, ces approches comblent le fossé entre les médecins moins expérimentés et ceux qui ont plus d'expérience, ainsi qu'entre les généralistes et les spécialistes. Le diagnostic des sous-types de lymphome est l'un des aspects des diagnostics essentiels nécessitant l'utilisation de l'imagerie médicale. Dans cette étude, nous présentons une méthode basée sur l'intelligence artificielle pour détecter automatiquement les sous-types de lymphome en analysant des images microscopiques.

Dans ce mémoire, nous avons développé un système de classification des trois sous-types de lymphome en utilisant des caractéristiques traditionnelles extraites de descripteurs obtenus grâce à une analyse profonde de l'image de lymphome à l'aide d'un ensemble de filtres pondérés par des systèmes chaotiques. Pour l'extraction des descripteurs, nous avons utilisé une méthode traditionnelle bien connue appelée LPQ. Afin d'obtenir des performances élevées, nous avons utilisé une méthode d'optimisation basée sur l'algorithme de chauves-souris pour déterminer les paramètres optimaux des filtres. En fait, tous ces tests ont conduit à d'excellentes performances (avec un meilleur ACC égal à 95,06 %), surpassant ainsi plusieurs études précédemment publiées. Dans les travaux futurs, notre objectif est de combiner ces caractéristiques avec des données cliniques afin de développer un modèle multimodal. De plus, nous chercherons à utiliser ou à développer des caractéristiques traditionnelles plus distinctives afin d'améliorer la précision du diagnostic.

Évaluation des Performances

A.1 Introduction

L'évaluation des performances d'un système de classification consiste à mesurer l'efficacité et la précision du système dans la tâche spécifique de classification des images. Cela implique généralement de comparer les résultats produits par le système avec les véritables étiquettes ou classes des images. Cette évaluation permet de quantifier les erreurs de classification du système, d'identifier les forces et les faiblesses du système, et de prendre des décisions éclairées sur son utilisation ou son amélioration. Les mesures couramment utilisées pour évaluer les performances de ces types de systèmes incluent les métriques utilisées pour mesurer les performances ainsi que les courbes de performance.

A.2 Mesures de performance

La matrice de confusion est une représentation tabulaire qui permet d'évaluer les performances d'un modèle de classification en comparant ses prédictions avec les véritables étiquettes de classe. Elle est particulièrement utile pour analyser les erreurs de classification et comprendre les tendances de confusion entre les classes. La matrice de confusion est généralement présentée sous la forme d'un tableau avec les classes réelles en ligne et les classes prédites en colonne. Chaque cellule de la matrice représente le nombre d'échantillons appartenant à une classe réelle spécifique et prédits comme appartenant à une classe particulière. La matrice de confusion est généralement représentée comme suit :

	Classe Positive	Classe Négative
Classe Positive	Vrai positif	Faux négatif
Classe Négative	Faux positif	Vrai négatif

Les termes de la matrice sont définis comme suit :

- **Vrai positif (True Positive, TP)** : Les échantillons qui sont réellement de la classe positive et qui sont correctement prédits comme tels par le modèle.
- **Faux négatif (False Negative, FN)** : Les échantillons qui sont réellement de la classe positive mais qui sont incorrectement prédits comme étant de la classe négative par le modèle.
- **Faux positif (False Positive, FP)** : Les échantillons qui sont réellement de la classe négative mais qui sont incorrectement prédits comme étant de la classe positive par le modèle.
- **Vrai négatif (True Negative, TN)** : Les échantillons qui sont réellement de la classe négative et qui sont correctement prédits comme tels par le modèle.

En utilisant cette matrice, on peut calculer plusieurs mesures de performance telles que:

1) Taux de vrais positifs (True Positive Rate-TPR) : également appelé sensibilité ou rappel, est une mesure de performance qui évalue la proportion d'échantillons positifs réellement détectés parmi tous les échantillons positifs réels.

$$FPR = \frac{TP}{TP + FN} \quad (1)$$

2) Taux de faux positifs (False Positive Rate-FPR) : représente la proportion d'échantillons négatifs qui sont incorrectement classés comme positifs par le système. En d'autres termes, le FPR mesure le pourcentage d'erreurs de classification où le système prédit à tort une classe positive alors que l'échantillon réellement appartient à une classe négative.

$$FPR = \frac{FP}{TP + FN} \quad (2)$$

3) Taux de vrais négatifs (True Negative Rate-TNR) : également appelé spécificité, est une mesure de performance qui évalue la proportion d'échantillons négatifs correctement identifiés parmi tous les échantillons négatifs réels.

$$FPR = \frac{FN}{TP + FN} \quad (3)$$

4) Taux de faux négatifs (False Negative Rate-FNR) : représente la proportion d'échantillons positifs qui sont incorrectement classés comme négatifs par le système. En

d'autres termes, le FNR mesure le pourcentage d'erreurs de classification où le système prédit à tort une classe négative alors que l'échantillon réellement appartient à une classe positive.

$$\text{FNR} = \frac{\text{FN}}{\text{TP} + \text{FN}} \quad (4)$$

Ces deux taux sont souvent en conflit et peuvent être ajustés en modifiant les seuils de classification du modèle. L'augmentation du seuil de décision conduit généralement à une diminution du FPR, mais peut entraîner une augmentation du FNR, et vice versa. Il est donc important de trouver un compromis approprié en fonction des exigences spécifiques du problème de classification.

5) Exactitude (accuracy-ACC) : C'est une mesure de performance couramment utilisée qui évalue la proportion d'échantillons correctement classés par rapport à l'ensemble total d'échantillons.

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (5)$$

L'exactitude donne une idée globale de la performance du système, mais elle peut être trompeuse dans des situations où les classes ne sont pas équilibrées.

6) Sensibilité (Sensitivity-SEN) : est également appelée rappel ou taux de vrais positifs, est une mesure de performance qui évalue la capacité d'un système à détecter correctement les échantillons positifs.

$$\text{SEN} = \text{FPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (6)$$

7) Spécificité (Specificity-SPE) : est une mesure de performance qui évalue la capacité d'un système de reconnaissance de formes à identifier correctement les échantillons négatifs.

$$\text{SPE} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (7)$$

8) Précision (Precision-PRC) : est une mesure de performance qui évalue la proportion d'échantillons correctement prédits comme positifs parmi tous les échantillons prédits comme positifs.

$$\text{PRC} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (8)$$

9) Rappel (Recall-REC) : également appelé sensibilité, est une mesure de performance qui évalue la proportion d'échantillons positifs réellement détectés parmi tous les échantillons positifs réels.

$$\text{REC} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (9)$$

10) F1-score : est une mesure de performance qui combine à la fois la précision et le rappel d'un système. Il est souvent utilisé lorsque l'on souhaite évaluer de manière équilibrée la capacité du système à prédire à la fois les vrais positifs et les vrais négatifs.

$$\text{F1 - score} = 2 \left(\frac{\text{PRC} \times \text{REC}}{\text{PRC} + \text{REC}} \right) \quad (10)$$

A.2 Courbes de performance

Les courbes de performance sont des outils graphiques utilisés pour évaluer et comparer les performances d'un système de classification.

1) Receiver Operating Characteristic (ROC): La courbe ROC est un outil graphique couramment utilisé pour évaluer et visualiser les performances d'un modèle de classification binaire. Elle représente la relation entre le TPR et le FPR à différents seuils de classification. Pour construire une courbe ROC, le modèle de classification est utilisé pour prédire les étiquettes de classe et générer les scores de prédiction pour chaque échantillon. Ensuite, en variant le seuil de classification, on calcule le TPR et le FPR correspondants à chaque seuil. Ces valeurs sont ensuite tracées sur un graphique avec le TPR en fonction du FPR.

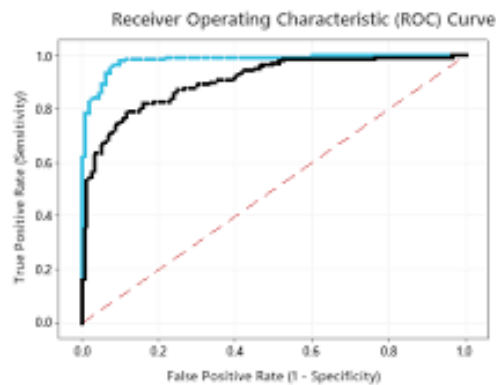


Fig. A.1 Receiver Operating Characteristic (ROC)

La courbe ROC permet de visualiser la performance du modèle sur un continuum entre sensibilité (TPR) et spécificité (1 - FPR). Idéalement, on souhaite maximiser le TPR tout en minimisant le FPR. Ainsi, un modèle avec une courbe ROC se rapprochant du coin supérieur gauche du graphique est considéré comme ayant de meilleures performances.

En plus de la courbe ROC, une mesure couramment utilisée pour résumer les performances du modèle est l'Aire sous la courbe ROC (Area Under the ROC Curve - AUC). L'AUC mesure l'aire totale sous la courbe ROC et fournit une mesure agrégée de la performance du modèle.

Un modèle avec une AUC proche de 1 est considéré comme très performant, tandis qu'une AUC proche de 0,5 indique une performance équivalente à une classification aléatoire.

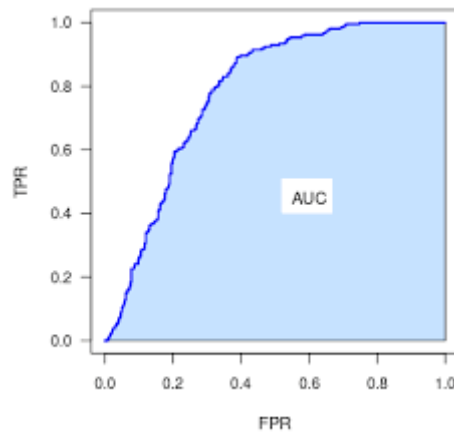


Fig. A.2 Area Under the ROC Curve

Il convient de noter que l'AUC est utilisée principalement pour les problèmes de classification binaire, où il y a deux classes distinctes.

2) Precision-Recall Curve : La courbe précision-rappel est un outil d'évaluation de la performance d'un système de classification binaire, qui met l'accent sur la précision et le rappel (sensibilité) du modèle. La courbe précision-rappel est construite en traçant les valeurs de précision en fonction des valeurs de rappel à différents seuils de classification.

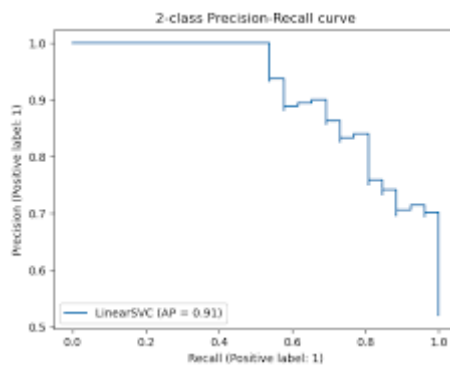


Fig. A.3 Precision-Recall Curve

La courbe précision-rappel permet d'évaluer comment le modèle se comporte à différents seuils de classification et fournit des informations sur le compromis entre la précision et le rappel. Une courbe précision-rappel idéale se rapproche du coin supérieur droit du graphique, indiquant à la fois une précision élevée et un rappel élevé.

Glossaire

Les termes suivants, classés dans l'ordre alphabétique, sont utilisés dans le texte.

ACC	: Accuracy
ANN	: Artificial Neural Network
AUC	: Area Under the Curve
BAT	: Basic Attention Token
BBN	: Bayesian Belief Network
BSIF	: Binarized Statistical Image Features
CAD	: Computer-Aided Diagnosis
CADx	: Computer-Aided Diagnosis and Decision Support
CAD-RADS	: Coronary Artery Disease Reporting and Data System
CLL	: Chronic Lymphocytic Leukemia
CNN	: réseau de Neurones Convolutif
CT	: Computed Tomography
DAO	: Diagnostic Assisté par Ordinateur
DLS	: Deep Learning System
FF-LPQ	: Filtering-Free & Local Phase Quantization
FG-LPQ	: Filtering-Gaussien & Local Phase Quantization
FL	: Follicular Lymphoma
FN	: False Negative
FNR	: False Negative Rate
FO-LPQ	: Filtering-Optimization & Local Phase Quantization
FP	: False Positive
FPR	: False Positive Rate
GA	: Genetic Algorithm
HOG	: Histogram of Oriented Gradients
H&E	: Hematoxylin and Eosin
IA	: Intelligence Artificielle

ICA	: Independent Component Analysis
IoT	: Internet of Things
IRM	: Imagerie par Résonance Magnétique
LBP	: Local Binary Patterns
LDA	: Linear Discriminant Analysis
LPQ	: Quantification de Phase Locale
MCL	: Mantle Cell Lymphoma
MIA	: Mammography Image Analysis
NG	: Niveau de Gris
PCA	: Principal Component Analysis
PET	: Positron Emission Tomography
PRC	: Precision
PSF	: Point Spread Function
PSO	: Particle Swarm Optimization
RBF	: Radial Basis Function
REC	: Recall
RF	: Random Forests
RGB	: Red-Green-Blue
RNN	: réseau de Neurones Récurrent
ROC	: Receiver Operating Characteristic
SEN	: Sensitivity
SPE	: Specificity
STFT	: Short-Time Fourier Transform
SVM	: Machines à Vecteurs de Support
TDM	: Tomodensitométrie
TN	: True Negative
TNR	: True Negative Rate
TP	: True Positive
TPR	: True Positive Rate
WND-5	: Weighted Neighbor Distance-5

Bibliographies

- [1] Aramaki, E., Shirota, Y., & Tomoko, O. (2020), “Applications of artificial intelligence to medical diagnosis”, *International Journal of Automation Technology*, Vol 14, No 1, pp. 11-28.
- [2] McPhee, Stephen J., et al. *Current Medical Diagnosis & Treatment 2021*. McGraw-Hill Education, 2020.
- [3] Rajkomar, A., Dean, J., & Kohane, I. (2019). “Machine learning in medicine”. *New England Journal of Medicine*, Vol 380, No 14, pp. 1347-1358.
- [4] McKinney S. M., et al. (2020). “International evaluation of an AI system for breast cancer screening”. *Nature*, Vol 577, No 7788, pp. 89-94..
- [5] Shortliffe, E. H., & Sepúlveda, M. J. (2018). “ Clinical Decision Support in the Era of Artificial Intelligence”. *JAMA*, Vol 320, No 21, pp.2199-2200.
- [6] Kohli, M., & Geis, J. R. (2017). “Future directions in radiology: Artificial intelligence”. *Journal of the American College of Radiology*, Vol 14, No 7, pp.862-865.
- [7] BISHOP, Christopher M. et NASRABADI, Nasser M. “*Pattern recognition and machine learning*”. New York : springer, 2006.
- [8] Esteva, A., Kuprel, B., Novoa, R. A., et al. (2017). “Dermatologist-level classification of skin cancer with deep neural networks”. *Nature*, Vol 542, No 7639, pp.115-118.
- [9] Noble, W. S. (2006). “What is a support vector machine?” *Nature Biotechnology*, Vol 24, No 12, pp.1565-1567.
- [10] Breiman, L. (2001). “Random forests”. *Machine Learning*, Vol 45, No 1, pp.5-32.
- [11] Shortliffe, E. H., & Buchanan, B. G. (2018). “A model of inexact reasoning in medicine”. *Artificial Intelligence*, Vol 62, No 1-2, pp.41-70.
- [12] von Luxburg, U. (2007). “A tutorial on spectral clustering”. *Statistics and Computing*, Vol 17, No 4, pp.395-416.
- [13] Xue, Z., et al. (2019). “Multi-modal medical image fusion using deep learning”. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* (pp. 1020-1023). IEEE.
- [14] Choi, Y. H., & Choi, M. (2018), “Artificial intelligence in medical imaging: A

- radiomics perspective”, *Korean Journal of Radiology*, Vol 19, No 4, pp. 939-951.
- [15] Louis, D. N., et al. (2020). “Computational pathology: A path ahead”. *Archives of Pathology & Laboratory Medicine*, Vol 144, No 6, pp.706-716.
- [16] Chartrand, G., et al. (2017). “Deep learning: A primer for radiologists”. *Radiographics*, Vol 37, No 7, pp.2113-2131.
- [17] Cury R. C., et al. (2016). “CAD-RADS™ Coronary Artery Disease – Reporting and Data System. An expert consensus document of the Society of Cardiovascular Computed Tomography (SCCT), the American College of Radiology (ACR), and the North American Society for Cardiovascular Imaging (NASCI)”. *Journal of Cardiovascular Computed Tomography*, Vol 10, No 4, pp.269-281.
- [18] Cruz-Roa A., et al. (2014). “Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent”. *Scientific Reports*, Vol 7, No 46450.
- [19] Kooi T., et al. (2017). “Large scale deep learning for computer aided detection of mammographic lesions”. *Medical Image Analysis*, Vol 35, pp.303-312.
- [20] Gulshan V., et al. (2016). “Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs”. *JAMA*, Vol 316, No 22, pp.2402-2410.
- [21] D. S. López-Carrasco et al.(2020). "Artificial intelligence in medical imaging: Beyond the black box," *Radiologia*, Vol. 62, No. 4, pp. 326-337.
- [22] T. H. Wong et al. (2019). “Current use and future potential of artificial intelligence in dermatology”, *Journal of the American Academy of Dermatology*, Vol. 81, No. 1, pp. 58-66.
- [23] Liang Shen, Hong Qin, Fei Shi, Qian Wang, Pingkun Yan. (2017). "Feature Extraction in Computer-Aided Diagnosis: A Review". *IEEE Access*, vol. 5, pp. 6425-6439
- [24] Bishop, C. M. (2006). “Pattern Recognition and Machine Learning. Springer.
- [25] Manikandan, M., & Vijayachitra, S. (2017)”. Feature Extraction Techniques and their Applications in Medical Image Processing. *Journal of Medical Systems*, Vol 41, No 7, pp.109.
- [26] Chandrasekhar, V., Fieguth, P. W., & Clausi, D. (2014). “A Survey on Feature Extraction Techniques in Pattern Recognition ”. *Pattern Recognition*, Vol 46, No 3, pp.791-808.

- [27] Goshtasby, A., & Pieczynski, D. P. (2016). "Feature Extraction Techniques and Their Applications in Medical Imaging". *Machine Vision and Applications*, Vol 27, No 4, pp.439-469.
- [28] Ullah, S. S., Tariq, A., Raza, S. S., & Shahzad, S. (2019). "Traditional Feature Learning Methods". *International Journal of Advanced Computer Science and Applications*, Vol 10, No 3, pp.350-356.
- [29] Guo, Y., Zhang, L., Zhang, Y., & Wang, L. (2016). "Deep Feature Learning : A Survey". *IEEE Transactions on Big Data*, Vol 2, No 1, pp.80-94.
- [30] Pietikäinen, M., Hadid, A., Zhao, G., & Ahonen, T. (2011). "Local Phase Quantization for Blur-Insensitive Image Matching and Recognition". *Image and Vision Computing*, Vol 29, No 5, pp.317-330.
- [31] Xu, L., Lu, C., Xu, Y., & Jia, J. (2011). "Image Smoothing via L0 Gradient Minimization". *ACM Transactions on Graphics (TOG)*, Vol 30, No 5, pp.174.
- [32] May, R. (1976). "The Logistic Map and the Route to Chaos". *Nature*, Vol 26, No 5560, pp. 459-467.
- [33] Yang, X. (2010). "A New Metaheuristic Bat-Inspired Algorithm". In *Nature Inspired Cooperative Strategies for Optimization (NISCO)* (pp. 65-74).
- [34] Yang, X. (2010). "Nature-Inspired Metaheuristic Algorithms". Luniver Press.
- [35] Goodfellow, I., Bengio, Y., & Courville, A. (2016). "Deep Learning". MIT Press.
- [36] Fisher, R. A. (1936). "The use of multiple measurements in taxonomic problems". *Annals of Eugenics*, Vol 7, No 2, pp.179-188.
- [37] Duda, R. O., Hart, P. E., & Stork, D. G. (2012). "Pattern Classification". John Wiley & Sons.
- [38] Hastie, T., Tibshirani, R., & Friedman, J. (2009). "The Elements of Statistical Learning : Data Mining, Inference, and Prediction". Springer Science & Business Media.
- [39] Anas Younes, Loretta J. Nastoupil, and Michael Wang. (2019). "Lymphoma: Diagnosis and Treatment". *New England Journal of Medicine*, Vol 380, No 8, pp. 741-754.
- [40] S. Kevin Zhou, Dinggang Shen, Terry M. Peters. (2020). "Handbook of Medical Image Computing and Computer-Assisted Intervention". Academic Press, pp. 3-20.
- [41] Dr Laurence Michel. (2019). "Cancer Lymphatique : Comprendre et traiter le lymphoma". Editions Eyrolles

- [42] Steven H. Swerdlow, Elias Campo, Nancy Lee Harris, Harald Stein, Reiner Sieber. (2016). "Lymphomas: Classification, Diagnosis, and Treatment". Springer.
- [43] <http://www.andrewjanowczyk.com/use-case-7-lymphoma-sub-type-classification/>
- [44] Mitchell, M. (1998). An Introduction to Genetic Algorithms. MIT Press.
- [45] Kennedy, J., & Eberhart, R. (2001). Swarm Intelligence. Morgan Kaufmann.
- [46] Nikita V. Orlov, Wayne Chen, D. Mark Eckley, Tomasz Macura, Lior Shamir, Elaine S. Jaffe, Ilya G. Goldberg (2010), "Automatic Classification of Lymphoma Images with Transform-Based Global Features", IEEE Trans Inf Technol Biomed., Vol. 14, No. 4, pp. 1003–1013.
- [47] Walsh, C.D., Taylor, N.K. (2021). Evolution of Convolutional Neural Networks for Lymphoma Classification. In: Arabnia, H.R., Deligiannidis, L., Shouno, H., Tinetti, F.G., Tran, QN. (eds) Advances in Computer Vision and Computational Biology. Transactions on Computational Science and Computational Intelligence. Springer, Cham, pp 3–26.

Résumé : Récemment, le diagnostic automatisé des maladies basé sur des images médicales est devenu un élément essentiel des logiciels de pathologie numérique. L'analyse de texture est couramment utilisée pour résoudre ce problème. La plupart des recherches dans ce domaine utilisent des méthodes artisanales pour extraire directement les caractéristiques des images malgré la corrélation substantielle entre les images de différentes classes, ce qui explique les résultats limités. Dans cette mémoire, la méthode d'extraction de caractéristiques artisanales (par exemple, LPQ) sera appliquée à un ensemble de descripteurs obtenus à partir d'une analyse profonde des images de texture du lymphome à l'aide d'une banque de filtres pondérés. De plus, le classifieur ajuste automatiquement les paramètres des filtres pondérés à l'aide d'une optimisation basée sur un algorithme inspiré des chauves-souris pour atteindre un comportement d'analyse profonde et des performances optimales. En utilisant une base de données typique des sous-types de lymphome, nos résultats expérimentaux révèlent une amélioration significative par rapport aux techniques artisanales récentes, aboutissant à une excellente performance de 95,66%.

Mots clés : DAO, Extraction de caractéristiques, LPQ, Système chaotique, Optimisation, Fusion des données.

Abstract: Recently, automated disease diagnosis based on medical images has become an integral component of digital pathology packages. Texture analysis is commonly used to address this issue. Most research in this context uses handcrafted methods to directly extract image features despite the substantial correlation between images of different classes, which explains the limited results. In this dissertation, the handcrafted feature extraction method (e.g., LPQ) will be applied to a set of descriptors obtained from a deep analysis of lymphoma texture images using a weighted filter bank. Additionally, the classifier automatically adjusts the weighted filter settings using the bat based-inspired algorithm optimization to achieve deep analysis behavior and optimal performance. Using a typical database of lymphoma subtypes, our experimental results reveal a significant improvement over state-of-the-art handcrafted techniques, resulting in an excellent performance of 95.66%.

Index term: CAD, Feature Extraction, LPQ, Chaotic System, Optimization, Data Fusion.

ملخص: مؤخرًا، أصبح التشخيص الآلي للأمراض بناءً على الصور الطبية جزءًا أساسيًا من حزم الباثولوجيا الرقمية. تُستخدم تحليلات النسيج عمومًا للتعامل مع هذه المسألة. يستخدم معظم البحث في هذا السياق طرقًا مصممة يدويًا لاستخراج مباشر لملامح الصور على الرغم من الارتباط الكبير بين الصور من فئات مختلفة، مما يفسر النتائج المحدودة. في هذه الأطروحة، تم تطبيق طريقة استخراج مميزات مصممة يدويًا (مثل LPQ) على مجموعة من الوصفات المستخلصة من تحليل عميق لصور نسيج المفومية باستخدام مجموعة فلاتر مرجحة. بالإضافة إلى ذلك، يقوم المصنف تلقائيًا بضبط إعدادات فلاتر المرجحة باستخدام طريقة تحسين مستوحاة من خوارزمية الخفافيش لتحقيق سلوك التحليل المعمق والأداء الأمثل. باستخدام قاعدة بيانات للصور المفومية النمذجية، تكشف نتائجنا التجريبية عن تحسين كبير على التقنيات المصممة يدويًا المتطورة، مما يؤدي إلى أداء ممتاز بنسبة 95.66%.

الكلمات المفتاحية : CAD، استخراج الميزات ، LPQ، النظام الفوضوي ، التحسين ، دمج البيانات..