

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

BADJI MOKHTAR-ANNABA UNIVERSITY
UNIVERSITE BADJI MOKHTAR-ANNABA



جامعة باجي مختار - عنابة

Faculté des Sciences de L'Ingénieur
Département d'Informatique

THÈSE

Présentée en vue de l'obtention du diplôme de Doctorat en
Sciences

Option : Intelligence Artificielle

THÈSE

**Apprentissage automatique et Sélection de
caractéristiques pour la classification des
images médicales**

Présenté par :

M^r GHERGHOUT Youcef

Devant le jury:

Président	M ^r ZARZOUR Hafed	Pr	Université de Souk Ahras
Rapporteur	M ^{me} TLILI Yamina	Pr	Université d'Annaba
Membre	M ^{me} AZIZI Nabiha	Pr	Université d'Annaba
Membre	Mr KHEDAIRIA Soufiane	MCA	Université de Souk Ahras

Année 2024

REMERCIEMENTS

Je tiens à exprimer mes remerciements à ma directrice de thèse **Professeur Yamina TLILI**, pour son encadrement, ses conseils, ses orientations, ses encouragements et son soutien permanent.

Je tiens, également, à remercier les membres du jury qui m'ont fait l'honneur de bien vouloir évaluer mon travail, et plus précisément :

Monsieur Hafed ZARZOUR, Professeur à l'université Mohamed Chérif Messaadia -Souk Ahras-, pour l'honneur qu'elle m'a fait, en acceptant de présider ce jury.

Madame Nabiha AZIZI, Professeur à l'université Badji Mokhtar –Annaba, d'avoir accepté de faire partie de ce jury.

Monsieur Sofiane KHEDAIRIA, MCA à l'université Mohamed Chérif Messaadia -Souk Ahras-, pour avoir accepté de juger le présent travail.

Je voudrais aussi profiter de cette occasion pour remercier profondément ma mère, mes frères et mes sœurs, ainsi à tous ceux qui me sont chers en particulier les deux médecins Docteur Mosbah GHERGHOUT et son ami Docteur Abdenour MESSAADIA.

DEDICACES

A la mémoire du mon père

RÉSUMÉ

Dans cette thèse, une approche innovante de classification des images médicales repose sur l'intégration des **réseaux de files d'attente** dans les **réseaux de neurones à rétropropagation (BPNN)**, où chaque composante des couches du réseau est associée à une file d'attente dynamique. Dans cette architecture, chaque neurone ne se contente pas de stocker un seul état activé, mais conserve un historique de ses activations passées sous forme de file d'attente. L'apprentissage devient ainsi **adaptatif**, car un réseau ouvert met à jour ses couches cachées en intégrant progressivement les informations agrégées de l'autre réseau. Cela permet d'introduire une **mémoire adaptative** qui améliore l'apprentissage en intégrant l'évolution temporelle des caractéristiques extraites. L'entraînement repose sur deux BPNN fermés, chacun apprenant sur une partie de la base de données. Les états intermédiaires des neurones sont enregistrés et transférés via la file d'attente à un **réseau ouvert**, qui adapte ses poids en fonction des états cumulés. Cette structure permet une **rétropropagation enrichie** en tenant compte des tendances d'apprentissage plutôt que des mises à jour instantanées uniquement. Par ailleurs, une **sélection optimisée des caractéristiques** est réalisée pour améliorer la robustesse du modèle, en extrayant des attributs **texturaux** (GLCM, LBP, GRLM) et **morphologiques** (circularité, compacité, descripteurs de Fourier), suivie d'une réduction de dimension via **MRMR** ou **RELIEF**, qui permet d'éliminer les redondances et de maximiser la séparation des classes. L'optimisation de la fonction coût **MSE** via la descente de gradient garantit une convergence efficace du modèle. Cette approche améliore la **robustesse, la précision et la généralisation** du réseau, tout en préservant l'interprétabilité des résultats pour une application clinique fiable.

Mots clés: Apprentissage automatique; sélection de caractéristiques; imagerie médicale; Réseau de files d'attente; Reconnaissance de formes; caractéristiques texturales; caractéristiques morphologiques.

ABSTRACT

In this thesis, an innovative approach to medical image classification is based on the integration of **queue networks** into **backpropagation neural networks (BPNN)**, where each component of the network layers is associated with a dynamic queue. In this architecture, each neuron does not just store a single activated state, but keeps a history of its past activations in the form of a queue. Learning thus becomes **adaptive**, because an open network updates its hidden layers by progressively integrating the aggregated information from the other network. This allows the introduction of **an adaptive memory** that improves learning by integrating the temporal evolution of the extracted features. Training is based on two closed BPNNs, each is learning on a part of the database. The intermediate states of the neurons are recorded and transferred via the queue to an **open network**, which adapts its weights according to the cumulative states. This structure allows for **enriched backpropagation** by taking into account learning trends rather than only instantaneous updates. Furthermore, **an optimized feature selection** is performed to improve the robustness of the model, by extracting **textural** (GLCM, LBP, GRLM) and **morphological** (circularity, compactness, Fourier descriptors) attributes, followed by a dimension reduction via **MRMR** or **RELIEF**, which allows to eliminate redundancies and maximize class separation. The optimization of the **MSE** cost function via gradient descent ensures efficient convergence of the model. This approach improves **the robustness, accuracy and generalization** of the network, while preserving the interpretability of the results for a reliable clinical application.

Keywords: Machine learning; Feature selection; Medical imaging; Queuing network; Pattern recognition; Textural features; Morphological features.

الملخص

في هذه الأطروحة، يعتمد النهج المبتكر لتصنيف الصور الطبية على دمج شبكات الانتظار في الشبكات العصبية ذات الانتشار الخلفي (BPNN)، حيث يرتبط كل مكون من طبقات الشبكة بقائمة انتظار ديناميكية. في هذا البناء، لا يقوم كل عصبون بتخزين حالة نشطة واحدة فقط، بل يحتفظ أيضًا بسجل لتنشيطاته السابقة في شكل قائمة انتظار. وبهذا يصبح التعلم متكيفًا، لأن الشبكة المفتوحة تقوم بتحديث طبقاتها المخفية من خلال دمج المعلومات المجمعة من الشبكة الأخرى تدريجيًا. وهذا يسمح بتقديم الذاكرة التكيفية التي تعمل على تحسين التعلم من خلال دمج التطور الزمني للميزات المستخرجة. يعتمد التدريب على شبكتين BPNN مغلقتين، حيث تتعلم كل منهما على جزء من قاعدة البيانات. يتم تسجيل الحالات الوسيطة للخلايا العصبية ونقلها عبر قائمة الانتظار إلى شبكة مفتوحة، والتي تتكيف مع أوزانها وفقًا للحالات التراكمية. يتيح هذا الهيكل إمكانية الانتظار إلى شبكة الغني من خلال مراعاة اتجاهات التعلم بدلاً من التحديثات اللحظية فقط. علاوة على ذلك، يتم إجراء اختيار ميزة محسّن لتحسين متانة النموذج، عن طريق استخراج السمات النسيجية (GRLM، GLCM، LBP) والشكلية (الدائرية، الاكتناز، موصوفات فوربيه)، متبوعة بتخفيض الأبعاد عبر MRMR أو RELIEF، مما يساعد إزالة التكرار وتعظيم الفصل بين الطبقات. يضمن تحسين دالة تكلفة MSE عبر الانحدار التدريجي التقارب الفعال للنموذج. يعمل هذا النهج على تحسين قوة الشبكة ودقتها وتعميمها، مع الحفاظ على إمكانية تفسير النتائج للتطبيق السريري الموثوق.

الكلمات المفتاحية: التعلم الآلي؛ اختيار الميزة؛ التصوير الطبي؛ شبكة طوابير؛ التعرف على الأشكال؛ الخصائص النسيجية؛ الخصائص المورفولوجية.

TABLE DES MATIERES

Introduction générale	1
CHAPITRE 1	4
Généralités sur les algorithmes d'apprentissage automatique.....	4
1.1. Introduction.....	4
1.2. Estimation approximative des fonctions de densité de probabilités	5
1.2.1. Les distributions exponentielles de densité.....	5
1.2.2. Estimation : du maximum de vraisemblance, biais et variance	5
1.2.3. Estimation paramétrique de l'estimateur de Bayes	9
1.2.4. Les approches semi-paramétriques (Clustering hiérarchique).....	12
1.2.5. Algorithme de maximisation des attentes.....	13
1.2.6. Classification et régression non paramétrique	15
1.3. Réduction de dimension par sélection de caractéristiques.....	18
1.3.1. Définitions de la sélection de caractéristiques	20
1.3.2. Processus de sélection de caractéristiques	20
1.3.4. Approches de sélection de caractéristiques	24
1.4. Algorithmes de classification et de reconnaissance de formes.....	27
1.4.1. La reconnaissance de formes : définition et approches	27
1.4.2. La classification : définition et algorithmes.....	28
1.4.2.1. Apprentissage basée sur les instances : K Plus proche Voisin.....	28
1.4.2.2. Approches arborescentes de classification.....	29
1.4.2.3. Approches basées sur les règles d'association.....	30
1.4.2.4. Approches bayésiennes.....	31
1.4.2.5. Approches à Noyaux pour la classification : SVM.....	32
1.4.2.6. Approches d'ensemble	34
1.4.2.7. Algorithmes évolutionnaires	35
1.4.2.8. Optimisation de l'essaim de particules (PSO).....	36
1.5. Conclusion.....	36
CHAPITRE 2.....	38
Reconnaissance Des images médicales et Sélection de Caractéristiques.....	38
2.1. Introduction	38
2.2. Prétraitement, analyse, et amélioration de qualité des images médicales.....	38
2.2.1. Manipulation individuelle des pixels des images médicales.....	38
2.2.2. Représentation et filtration des images médicales dans le domaine spatial	40
2.2.2.1. Filtration des images médicales dans le domaine spatial	41
2.2.3. Représentation et filtration des images médicales dans le domaine fréquentiel.....	42
2.2.3.1. Transformation de Fourier Rapide (FFT)	42
2.2.3.2. Filtration des images médicales dans le domaine fréquentiel.....	42
2.2.3. Filtre de diffusion anisotropique	43
2.3. Analyse et quantification bidirectionnelle morphologique et texturale des images médicales.....	44
2.3.1. Quantification de la Forme	45
2.3.1.1. Quantification par des mesures globales de forme	45
2.3.1.2. Les moments spatiaux.....	46
2.3.1.3. Mesures de Distance Radiale, code de chaînes et les descripteurs de Fourier... 48	
2.3.1.4. Codes de chaîne et les Descripteurs de Fourier	50
2.3.1.5. Amincissement (Thinning).....	51
2.3.2. Quantification des textures.....	51
2.3.2.1. Moments statistiques.....	52
2.3.2.2. Mesures de la matrice de cooccurrence.....	52
2.3.2.3. Matrice de cooccurrence des niveaux de gris (GLCM).....	52

2.3.2.4. Matrice des longueurs de plage de niveaux de gris (Gray level run lengths matrix GRLM).....	52
2.3.2.5. Analyse de texture avec des motifs binaires locaux (LBP : Local Binary Patterns)	53
2.4. Segmentation d'images médicales dans le domaine spatial	54
2.4.1. Objectifs, et techniques de segmentation d'images médicales.....	54
2.4.2. Prétraitement d'image et seuillage	55
2.4.3. Régions en croissance (Region Growing).....	56
2.4.4. Techniques de segmentation basées sur les bords.....	56
2.4.5. Fuzzy C-Means et son application en imagerie médicale.....	57
2.5. État de l'art de classification et de prédiction des images médicales	58
2.5.1. Classification et prédiction des images des seins.....	58
2.6. Conclusion.....	62
CHAPITRE 3	63
Modélisation, Sélection et Agrégation des Caractéristiques par les systèmes de files d'attente	63
3.1. Introduction	63
3.2. Modélisation et Abstractions pour l'Évaluation des Performances : Vers une Compréhension du Système Réel	63
3.2.1. Évaluation de la performance de modélisation	65
3.2.2. Phases et activités du processus d'évaluation des performances de modélisation	66
3.3. Définitions et notation des systèmes de réseaux de files d'attente	68
3.3.1. Réseaux mono classe	68
3.3.2. Réseaux multi classes	70
3.3.3. Mesures de performance	72
3.3.3.1. Réseaux mono classe	72
3.3.3.2. Réseaux multiclassés.....	75
3.3.3.3. Réseaux de file d'attente de forme de produit.....	76
3.4. Pourquoi les réseaux de files d'attentes	79
3.4.1. Plateformes de catégorisation des méthodes de sélection des caractéristiques	79
3.4.1.1. Une plateforme préliminaire	80
3.4.1.2. Une plateforme générale.....	80
3.4.1.3. Plateforme unifiée	81
3.4.2. Aspects proposés pour améliorer la catégorisation.....	82
3.4.3. Critères d'analyse d'évaluation individuelles des algorithmes de sélection des caractéristiques	86
3.4.4. Application de la plateforme unifiée sur les algorithmes existants	89
3.5. Conclusion et perspectives	90
CHAPITRE 4	92
Approche de Sélection de Caractéristiques pour la Reconnaissance de Masses Mammographiques.....	92
4.1. Introduction	92
4.2. Pourquoi les réseaux de file d'attente sont importants :.....	94
4.3. Intégration des algorithmes de sélection de caractéristiques dans les réseaux de files d'attente	95
4.4. Modélisation et optimisation du BPNN par un réseau de file d'attente	97
4.5. Une Approche Complète pour la Sélection de Caractéristiques en Reconnaissance de la Masse Mammaire	98
4.6. La méthode proposée	99
4.6.1. Pré-classification des mammographies basée sur des règles d'associations.....	99
4.6.2. Filtre de diffusion anisotrope	103
4.6.3. Caractéristiques texturales	103

4.6.4. Caractéristiques morphologiques.....	109
a) Pixels de frontière et codes de chaîne.....	109
b) Caractéristiques régionales et forme générale des masses :.....	110
c) Caractéristiques des masses basées sur les frontières :.....	111
4.6.5. Sélection de caractéristiques avec l'algorithme mRMR.....	111
4.6.6. Sélection de caractéristiques avec l'algorithme Relief.....	112
4.7. Modélisation, sélection et agrégation des caractéristiques par les systèmes de réseaux de files d'attente.....	113
4.7.1. Algorithme 1. Apprentissage de MLP à rétro propagation avec réseau à file d'attente fermée.....	116
4.7.2. Algorithme 2. Fusion de chacun des BPNN fermés avec l'autre, en utilisant un réseau de file d'attente ouvert.....	117
4.7.3. Algorithme 3. Apprentissage d'un MLP à rétro propagation avec réseau de file d'attente ouvert.....	118
4.8. Conclusion.....	122
CHAPITRE 5.....	123
Résultats, comparaison et discussion de différentes approches proposées.....	123
5.1. Introduction.....	123
5.2. Modélisation, sélection et agrégation de caractéristiques avec un système de réseau de file d'attente.....	123
5.3. Analyse des résultats expérimentaux.....	127
5.3.1. La courbe ROC.....	131
5.3.2. Résultats pour l'extraction de caractéristiques et le filtre de diffusion anisotrope..	133
5.3.3. Résultats de la modélisation, de la sélection et de l'agrégation de caractéristiques à l'aide de la mise en file d'attente et du réseau de neurones.....	136
5.4. Conclusion et recherches futures.....	139
Conclusion générale et perspectives.....	141
ANNEXE 1.....	144
A.1. Approches d'apprentissage automatique pour la détection et la localisation des tissus/maladies anatomiques.....	144
A.2. Approches d'apprentissage automatique pour la classification et prédiction images du poumon.....	146
A.3. Classification et prédication des images du foie.....	150
A.4. Classification et prédication des images du coeur.....	150
A.5. Classification et prédication des images des artères coronaires.....	151
ANNEXE2.....	152
BIBLIOGRAPHIE.....	155

LISTE DES FIGURES

Figure 1.1	Exemple des densités conditionnelles de classe pour deux classes ayant une seule variable d'entrée x (graphique de gauche) avec les probabilités a posteriori correspondantes (graphique de droite).	6
Figure 1.2	Un overview de l'algorithme EM.	14
Figure 1.3	Taxinomie des approches de réduction de dimension [23]	19
Figure 1.4	Processus de sélection de caractéristiques avec validation	21
Figure 1.5	Approches de sélection de caractéristiques [15]	24
Figure 1.6	Approche Symbiose ou Wrapper [25]	26
Figure 1.7	Approche Filtre [25]	26
Figure 1.8	Modèle de reconnaissance de formes	27
Figure 2.1	(a) Segment de la frontière. (b) Représentation comme une fonction unidimensionnelle.	47
Figure 2.2	Numéros de direction pour (a) le code de chaîne à 4 directions et (b) le code de chaîne à 8 directions.	50
Figure 2.3	(a) Frontière numérique avec grille de rééchantillonnage superposée. (b) Résultat du rééchantillonnage. (c) code de chaîne à 4 directions. (d) Code de chaîne à 8 directions. [61]	50
Figure 2.4	Différentes primitives de texture détectées par le LBP. [68]	53
Figure 3.1	Pseudo code d'application de l'évaluation de la performance du modèle dans le processus du développement du système. (Extrait de [270])	68
Figure 3.2	File simple à un serveur.	69
Figure 3.3	Un réseau mixte.	75
Figure 3.4	Plateforme préliminaire	80
Figure 3.5	Plateforme générale	81
Figure 3.6	Plateforme unifiée	81
Figure 3.7	Application de la plateforme unifiée sur les algorithmes existants.[275]	90
Figure 4.1	Block diagram of the proposed scheme for classification of mammogram using neural and queuing network.[42]	94
Figure 4.2	Construction des règles d'association à partir d'un arbre de décision pour l'extraction des régions d'intérêt. [42]	101
Figure 4.3	Recadrage des ROIs à partir d'une image mammographique faisant référence au centre de la zone anormale.[42]	102
Figure 4.4	Pixel référencé par quatre orientations	104
Figure 4.5	Calcul des matrices de cooccurrence. (a) Valeurs d'intensité de l'image d'entrée avec 5 niveaux de gris. Différentes matrices de cooccurrence (GLCM) pour la distance définie $d=1$, l'image est balayée dans quatre directions différentes telles que (b) horizontale ($\theta = 0^\circ$), (c) verticale ($\theta = 90^\circ$), (d) diagonale ($\theta = 45^\circ$), (e) antidiagonale ($\theta = 135^\circ$).	106
Figure 4.6	Matrices de cooccurrence normalisées (NGLCM) des matrices de cooccurrence correspondantes (GLCM) de la figure 5 dans quatre directions différentes telles que (a) horizontale ($\theta = 0^\circ$), (b) verticale ($\theta = 90^\circ$), (c) diagonale ($\theta = 45^\circ$), (d) antidiagonale ($\theta = 135^\circ$).	107
Figure 4.7.	Calcul des matrices de longueur de plage de niveau de gris. (a) Valeurs d'intensité de l'image d'entrée avec 5 niveaux de gris. Différentes matrices de longueur de plage (RLM), l'image est scannée dans quatre directions différentes telles que (b) horizontale ($\theta = 0^\circ$), (c) verticale ($\theta = 90^\circ$), (d) diagonale ($\theta = 45^\circ$), (e) antidiagonale ($\theta = 135^\circ$).	109
Figure 4.8	Matrices de longueur de plage de niveau de gris normalisées (NGRLM) des matrices de longueur de plage correspondantes (RLM) de la Fig. 10 dans	109

	quatre directions différentes telles que (a) horizontale ($\theta = 0^\circ$), (b) verticale ($\theta = 90^\circ$), (c) diagonale ($\theta = 45^\circ$), (d) antidiagonale ($\theta = 135^\circ$).	
Figure 4.9	(a) Code de chaîne de Freeman pour les huit directions d'un pixel limite au suivant, (b) Image échantillonnée de masse originale "mdb184", (c) sa version rééchantillonnée sur une grille plus grossière, et (d) le code de chaîne résultant.[42]	110
Figure 4.10	(a) Réseau neuronal fermé, (b) unification des deux réseaux neuronaux fermés, (c) fusion de chacun des réseaux neuronaux fermés avec l'autre	119
Figure 4.11	Un modèle générique de MLP avec q couches	119
Figure 4.12	Stockage des poids et des termes d'erreurs de la couche de sortie du NN dans la file d'attente. (a) Premier réseau fermé. (b) Deuxième réseau fermé. (c) Fusion du premier réseau fermé avec les erreurs du second. (d) Fusion du deuxième réseau fermé avec les erreurs du premier. (e) Réseau unifié ouvert.	120
Figure 4.13	Stockage des poids et des termes d'erreurs des couches cachées du NN dans la file d'attente. (a) Premier réseau fermé. (b) Deuxième réseau fermé. (c) Fusion du premier réseau fermé avec les erreurs du second. (d) Fusion du deuxième réseau fermé avec les erreurs du premier. (e) Réseau unifié ouvert.	121
Figure 5.1	Évaluation des caractéristiques pour la classification des masses normales-anormales : (a) critère de dispersion, (b) critère d'information mutuelle, (c) corrélation absolue entre chaque caractéristique et toutes les autres, (d) redondance des caractéristiques	128
Figure 5.2	Évaluation des caractéristiques pour la classification des masses bénignes-malignes : (a) critère de dispersion, (b) critère d'information mutuelle, (c) corrélation absolue entre chaque caractéristique et toutes les autres, (d) redondance des caractéristiques	129
Figure 5.3	Caractéristiques pondérées pour une classification : (a) des masses normales-anormales et (b) des masses bénignes-malignes.	129
Figure 5.4	Comparaison des courbes ROC entre le schéma de classification des mammographies utilisant la caractéristique sélectionnée optimale et les méthodes de classification SVM-RFE et CNN. (a) Avec le premier réseau fermé pour la classification ROI normale-anormale. (b) Avec deuxième réseau fermé pour la classification ROI normal-anormal. (c) Avec réseau ouvert unifié pour la classification ROI normal-anormal. (d) Avec le premier réseau fermé pour la classification bénigne-maligne de la ROI. (e) Avec deuxième réseau fermé pour la classification bénigne-maligne de la ROI. (f) Avec réseau ouvert unifié pour la classification bénigne-maligne des ROI	130
Figure 5.5	La matrice de classification canonique pour la discrimination à deux classes de l'événement anormal (A) de l'événement normal (N).	131
Figure 5.6	Comparaison de l'erreur d'apprentissage par le réseau neuro-files d'attente proposé en utilisant les algorithmes de sélection de caractéristiques MRMR ou RELIEF et les méthodes de classification SVM-RFE ou CNN. (a) Avec le premier réseau fermé pour la classification des ROI normal-anormale. (b) Avec le deuxième réseau fermé pour la classification des ROI normal-anormale. (c) Avec le réseau ouvert unifié pour la classification des ROI normal-anormale. (d) Avec le premier réseau fermé pour la classification des ROI bénin-malin. (e) Avec le deuxième réseau fermé pour la classification des ROI b bénin-malin. (f) avec le réseau ouvert unifié pour la classification des ROI bénin-malin. (g) CNN-1 pour la classification des ROI (en utilisant la demie base de données MIAS). (h) CNN-2 pour la	132

classification des ROI (en utilisant l'autre demie base de données MIAS).
(j) CNN pour la classification des ROI (en utilisant toute la base de données MIAS)

Figure 5.7 Exemples d'images de la base de données MiniMIAS utilisées pour l'évaluation expérimentale. La première rangée contient la masse originale initiale pour la mammographie (de gauche à droite, mdb184, mdb202, mdb028, mdb025, mdb015, mdb132), tandis que les rangées suivantes correspondent respectivement aux images traitées par le filtre de diffusion anisotrope, le traçage du contour de masse et les approches du cercle de masse. 134

LISTE DES TABLEAUX

Tableau 1.1	Les fonctions kernel classiques	33
Tableau 4.1	Métriques de classification de texture de Haralick et métriques de longueur d'exécutions.	106
Tableau 5.1	Différentes valeurs des mesures de performance du classificateur à l'aide de deux méthodes de sélection de caractéristiques pour la classification du retour pour la classification normal-anormal avec $n = 25$.	126
Tableau 5.2	Différentes valeurs des mesures de performance du classificateur à l'aide de deux méthodes de sélection de caractéristiques pour la classification bénigne-maligne du retour sur investissement avec $n = 25$.	126
Tableau 5.3	Comparaison des taux de précision de test optimaux et des mesures AUC entre les méthodes proposées et SVM-RFE et CNN utilisant le premier réseau fermé	133
Tableau 5.4	Comparaison des taux de précision de test optimaux et des mesures de l'AUC entre les méthodes proposées et SVM-RFE et CNN en utilisant un deuxième réseau fermé	133
Tableau 5.5	Comparaison des taux de précision de test optimaux et des mesures AUC entre les méthodes proposées et SVM-RFE et CNN en utilisant un réseau ouvert unifié	133
Tableau 5.6	Valeurs moyennes de certaines caractéristiques statistiques pour les formulaires des trois mammographies	135
Tableau 5.7	Valeurs moyennes des caractéristiques limites pour les quatre formes de mammographie	135
Tableau 5.8	Comparaison des performances de différentes méthodes avec le schéma proposé	138

Introduction générale

La principale contribution de cette thèse est l'introduction d'un nouvel paradigme d'apprentissage automatique parfait basé sur les réseaux de files d'attente.

Actuellement, les chercheurs développent un nouveau concept appelé diagnostic assisté par ordinateur (CAD) développé au cours des 20 dernières années. Leur objectif est de développer des systèmes de dépistage qui aident le radiologue à jeter un second regard sur une région suspecte de l'image. Les systèmes de CAO ont été initialement développés pour l'analyse d'images mammographiques analogiques, numérisées et analysées secondairement. La diffusion de la CAD en mammographie est aujourd'hui possible grâce à la mammographie numérique. Une fois l'image numérique obtenue, les données brutes sont immédiatement traitées à partir d'un serveur CAO où un algorithme d'apprentissage automatique permettra d'analyser différents paramètres (densité, contraste, forme et taille de l'image, spiculation, etc.) du système CAD. L'objectif du système CAD est d'augmenter la sensibilité de la mammographie sans augmenter le nombre de faux positifs.

Malgré un grand nombre de travaux de diagnostic informatique (CAO), aucun travail n'a jusqu'à présent été orienté vers les systèmes de réseaux de la mise en file d'attente, permettant de généraliser, d'unifier et de superposer plusieurs réseaux de même type, de même taille et architecture d'un algorithme d'apprentissage automatique et sélection de caractéristiques pendant l'apprentissage.

Au cours des dernières années de ma thèse, je me suis intéressée à la modélisation des méthodes d'apprentissage automatique et celles de sélection de caractéristiques par les systèmes de réseaux de files d'attentes. Notre modèle proposé est juste un modèle générique qui permet de superposer plusieurs modèles élémentaires d'algorithmes d'apprentissage automatique de même type, taille et architecture. Schématiquement, lors de l'étape d'apprentissage, tous ces modèles sont superposés à travers les systèmes de mise en file d'attente du réseau. L'idée intuitive est que tout algorithme d'apprentissage automatique (Neural Network, SVM, KNN,...) a une structure topologique de réseau en théorie des graphes. Le modèle générique vise, d'une part, à étendre la capacité de discrimination et à dilater l'espace d'apprentissage de ces réseaux et, d'autre part, à construire des algorithmes évolutifs d'apprentissage incrémental en prêtant et en absorbant l'intelligence d'un réseau à un autre. L'objectif principal de notre travail est de proposer une approche de sélection de caractéristiques pour la reconnaissance et la classification des images médicales. Cette thèse présente les étapes de la démarche que nous avons suivie pour atteindre cet objectif, les expérimentations que nous avons menées et les résultats obtenus. La reconnaissance des classes d'images est effectuée en utilisant des classificateurs de type QBPNN ayant en entrées toutes les caractéristiques sélectionnées. Nous avons opté pour la sélection des caractéristiques selon l'approche Symbiose qui est basée sur les résultats obtenus par les classificateurs QBPNN utilisés :

Le chapitre 1 est un préambule nécessaire à la compréhension des concepts de base des algorithmes d'apprentissage automatique et sélection de caractéristiques, qui nous permettent d'analyser et de discuter les résultats obtenus, afin de classer et de prédire les images médicales. On y présente de manière indépendante une taxinomie de méthodes d'un point de

vue statistique à savoir, les approches paramétriques, celles semi paramétriques(Clustering) et celles non paramétriques. Puis, on introduit une revue des algorithmes d'apprentissage automatique et de sélection de caractéristiques.

Dans **le chapitre 2**, on commence par donner les propriétés des ensembles de paramètres d'un modèle d'apprentissage automatique et sélection de caractéristiques. Ces ensembles englobent les méthodes de prétraitement, analyse, et amélioration de qualité des images médicales, à savoir les méthodes de prétraitement dans le domaine spatial et celles dans le domaine fréquentiel. On présente également les techniques de filtration des images médicales et en particulier le filtre de diffusion anisotrope. Un module d'analyse des images médicales est ensuite bénéficié pour analyser et quantifier la morphologie et la texture. Les mesures correspondantes contribuent à l'analyse de la structure et de la fonction dans les cas normaux et anormaux en abordant plusieurs aspects de données, tel que la forme, la taille, la texture,...etc.

Le chapitre 3 est consacré à la simulation, la modélisation, la sélection et à l'agrégation des caractéristiques par les systèmes de files d'attente mono classes et multi classes. On introduit les différentes plateformes de catégorisation et du choix de la méthode la plus appropriée de sélection de caractéristiques. Trois algorithmes d'apprentissage à rétro propagation via un MLP ont été présentés. Le réseau de neurones proposé est un réseau ouvert résultant d'une unification de deux réseaux fermés et fusionnés; le schéma de connexion suivant est nécessaire pour décrire les réseaux de file d'attente :

- BPNN-1 fermé et BPNN-2 fermé : Ce schéma de connexion consiste à apprendre par l'algorithme feed-forward la rétropropagation du gradient de deux réseaux de neurones (on les considère comme fermés) avec deux bases d'apprentissage différentes, en sauvegardant à chaque itération état du réseau dans la file d'attente : matrice de poids, l'erreur de sortie de chaque couche, etc.
- BPNN-1 fusionné et BPNN-2 fusionné : Ce schéma de connexion ouvre les deux réseaux fermés (BPNN-1 et BPNN-2) et fusionne l'un par l'effet de l'autre. Il est utilisé avant l'unification des poids courants des deux réseaux fermés. Ce schéma de connexion est considéré comme intermédiaire entre le premier et le dernier.
- Le résultat de la fusion est un BPNN ouvert unifié : Ce schéma de connexion consiste en l'apprentissage des deux réseaux précédents avec unification de ses couches cachée et de sortie avec une agrégation conjointe de leurs poids respectifs, qui sont disposés dans la file d'attente respective. Ce schéma additionne les nouvelles valeurs des poids générées par les deux réseaux fusionnés (BPNN-1 fusionné et BPNN-2 fusionné) etenfile (pushbacks) le résultat dans une troisième file d'attente.

Le chapitre 4 représente une approche de sélection de caractéristiques pour la reconnaissance de masses mammographiques. La méthode proposée comprend quatre modules : le premier consiste en une construction automatique d'un ensemble de règles symboliques pour la pré-classification des mammographies à travers un arbre de décision. Toutes ces images sont traitées avec un filtre de diffusion anisotrope dans le deuxième module. Le troisième module couvre deux algorithmes majeurs tels que l'extraction de caractéristiques. Le meilleur sous ensemble de caractéristique est fourni à la file d'attentes et au réseau neuronal pour la classification des mammographies comme normales, bénignes, ou malignes.

Le chapitre 5 constitue la dernière partie de cette thèse, il est dédié à la présentation des différents résultats, comparaison et discussion de différentes approches proposées. Tout au

long de ce travail, les résultats des classificateurs proposés sont comparés avec SVM-RFE et CNN pour comparer les supériorités. La fonction de coût du réseau de neurones QBPNN unifié ouvert est représentée par le produit des deux fonctions de coût des réseaux QBPNN fermés. Nous avons prouvé que le produit de deux fonctions décroissantes est une fonction plus décroissante dans l'intervalle $]0, 1[$. Lorsque la matrice de poids du réseau ouvert unifié se déplace avec une estimation de ΔW , l'erreur totale E est minimisée par l'effet de déplacement ΔW_1 du premier réseau fermé fusionné et le déplacement ΔW_2 du second. Pour valider l'efficacité du schéma proposé, la simulation a été réalisée à l'aide de la base de données MIAS. Une analyse comparative en ce qui concerne la précision et l'AUC du ROC révèle que le schéma suggéré surpasse relativement d'autres schémas. Une précision de 98,1 % et 95,2 % a été obtenue respectivement pour la classification normale-anormale et celle bénigne-maligne.

Cette thèse est axée sur le développement d'un paradigme d'apprentissage automatique basé sur des réseaux de files d'attente pour la reconnaissance d'images médicales, en mettant l'accent sur la sélection de caractéristiques et la classification. Les résultats que nous mentionnons semblent montrer de bonnes performances dans la classification des images médicales.

CHAPITRE 1

Généralités sur les algorithmes d'apprentissage automatique

1.1. Introduction

L'apprentissage automatique, également connu sous le nom de machine learning, est un sous-domaine de l'intelligence artificielle (IA) qui permet aux ordinateurs d'apprendre et de prendre des décisions sans être explicitement programmés. Au lieu de suivre des instructions spécifiques, les systèmes d'apprentissage automatique sont capables d'analyser des données et d'identifier des schémas ou des relations afin de générer des modèles prédictifs.

L'apprentissage automatique a connu un essor considérable ces dernières années grâce à l'augmentation de la puissance de calcul et à la disponibilité de grandes quantités de données. Il a été appliqué avec succès dans de nombreux domaines, y compris la reconnaissance vocale, la traduction automatique, la recommandation de produits et bien d'autres.

L'application de l'apprentissage automatique aux images médicales est un domaine en pleine expansion qui présente un grand potentiel pour améliorer les diagnostics et les traitements. Cependant, les images médicales sont souvent volumineuses et complexes, ce qui pose des défis particuliers pour leur analyse.

La sélection de caractéristiques, également appelée extraction de caractéristiques, est une étape primordiale dans le processus d'analyse des images médicales. Elle consiste à identifier et à extraire les caractéristiques les plus pertinentes des images qui permettent de distinguer différentes classes ou de prédire des variables d'intérêt. Ces caractéristiques peuvent être des propriétés géométriques, des textures, des intensités ou d'autres mesures spécifiques aux images médicales.

La sélection de caractéristiques joue un rôle essentiel dans la construction de modèles d'apprentissage automatique performants pour la classification. Une sélection judicieuse des caractéristiques peut améliorer la précision du modèle, réduire la dimensionnalité des données et accélérer le processus d'analyse. Cela peut également contribuer à une meilleure compréhension des mécanismes sous-jacents des maladies et à l'identification de biomarqueurs pertinents.

Différentes techniques de sélection de caractéristiques ont été développées, notamment les approches basées sur des mesures de corrélation, les algorithmes d'apprentissage incrémental et les méthodes d'apprentissage profond. Chaque technique a ses propres avantages et limitations, et le choix de la méthode dépendra du problème spécifique et des données disponibles.

La capacité de généralisation, qui caractérise la façon dont le résultat appris à partir d'un ensemble de données d'entraînement donné peut être appliqué à de nouvelles données invisibles, est le concept le plus central de l'apprentissage automatique. Les chercheurs ont consacré d'énormes efforts à la recherche de techniques qui pourraient conduire à un système

d'apprentissage doté d'une forte capacité de généralisation. L'un des paradigmes les plus réussis est l'apprentissage d'ensemble [1].

1.2. Estimation approximative des fonctions de densité de probabilités

1.2.1. Les distributions exponentielles de densité

L'étude des relations entre différentes catégories d'images médicales ou des régions de celles-ci doit être distinguée des autres axes de recherche en tant que lien entre les statistiques et le domaine de l'apprentissage automatique. Plusieurs modèles de distribution ont été proposés en statistique[2, 3].

Le classifieur de Bayes apparaît comme un lien entre les statistiques et le domaine de l'apprentissage automatique, où l'essentiel de son intérêt est déterminé à évaluer la probabilité a posteriori d'un objet à l'aide de la formulation de Bayes, suivie d'une prise de décision selon la règle de Bayes, qui stipule d'attribuer l'objet à la classe dans laquelle la probabilité a posteriori a la plus grande valeur. On distingue de types de variations de densités de probabilité, la densité de probabilité univariée et multivariée. Ils font référence à la modélisation de la distribution statistique des données lorsque l'on travaille respectivement avec une seule variable ou avec plusieurs variables simultanément.

La densité de probabilité univariée concerne l'estimation et la modélisation de la distribution de probabilité pour une seule variable aléatoire. Elle décrit la probabilité de chaque valeur possible que peut prendre cette variable aléatoire. La distribution de probabilité univariée est souvent représentée par une courbe, comme la distribution normale (ou gaussienne), qui est fréquemment utilisée pour modéliser des phénomènes naturels.

D'un autre côté, **la densité de probabilité multivariée** s'applique lorsque l'on travaille avec plusieurs variables aléatoires simultanément. Elle permet de modéliser les relations entre ces variables et de décrire la probabilité conjointe de plusieurs valeurs. La distribution de probabilité multivariée est généralement représentée par une fonction de densité de probabilité qui prend en compte les valeurs des différentes variables aléatoires.

Parmi les distributions multivariées couramment utilisées, on trouve la distribution normale multivariée, qui étend la distribution normale univariée au cas multivarié, la distribution Maxwell, Gamma, Beta, Poisson, Binomiale...etc. Elles sont caractérisées par un vecteur de moyenne et une matrice de covariance, qui décrivent respectivement les moyennes et les relations de variance-covariance entre les variables. [3]

1.2.2. Estimation : du maximum de vraisemblance, biais et variance

L'estimation du maximum de vraisemblance (EMV) est une technique statistique couramment utilisée pour estimer les paramètres d'un modèle probabiliste à partir d'un ensemble de données. Cela inclut également l'estimation des paramètres sur des données d'images ou médicales. Lors de l'estimation des paramètres à l'aide de l'EMV, il est important de comprendre les notions de biais et de variance associées à cette méthode.

Le problème de modélisation est considéré comme une fonction de densité de probabilité $p(x)$, étant donné un nombre fini de points de données x_n , $n = 1, \dots, N$ tiré de cette fonction de densité. Les méthodes décrites peuvent être utilisées pour construire des systèmes de classification en considérant tour à tour chacune des classes C . Les densités conditionnelles de classe correspondantes $p(x|C_k)$ est estimée en utilisant le fait que chaque point de données

est étiqueté selon sa classe. Ces densités peuvent ensuite être utilisées dans le théorème de Bayes pour trouver les probabilités a posteriori correspondant à une nouvelle mesure de x , qui peuvent à leur tour être utilisées pour faire une classification de x . [4]

L'étape d'inférence et l'étape de décision sont deux étapes clés dans la résolution d'un problème de classification:

1. **Étape d'inférence** : L'étape d'inférence consiste à utiliser un modèle de classification entraîné pour prédire les étiquettes de classe pour de nouvelles données non vues auparavant. Lors de cette étape, le modèle utilise les caractéristiques (variables) des nouvelles instances et applique les règles ou les calculs nécessaires pour estimer la classe à laquelle elles appartiennent. Selon le modèle utilisé, l'inférence peut se faire de différentes manières, telles que la comparaison avec des seuils prédéfinis, l'utilisation de fonctions de décision ou l'application de méthodes probabilistes. L'objectif principal de l'étape d'inférence est de prédire avec précision les classes des nouvelles données en se basant sur les connaissances acquises lors de l'étape d'apprentissage.
2. **Étape de décision** : L'étape de décision est l'étape où nous prenons une décision ou une action basée sur les prédictions de classe faites lors de l'étape d'inférence. Selon le problème de classification spécifique, cette décision peut varier. Par exemple, dans un système de détection de spam, la décision pourrait être de marquer un e-mail comme spam ou non-spam en fonction de la prédiction de classe. Dans un système de diagnostic médical, la décision pourrait être de recommander un traitement spécifique en fonction de la prédiction de classe qui identifie une condition médicale particulière. L'étape de décision vise à utiliser les résultats de la prédiction de classe pour guider l'action ou la décision appropriée dans le contexte du problème spécifique. [4]

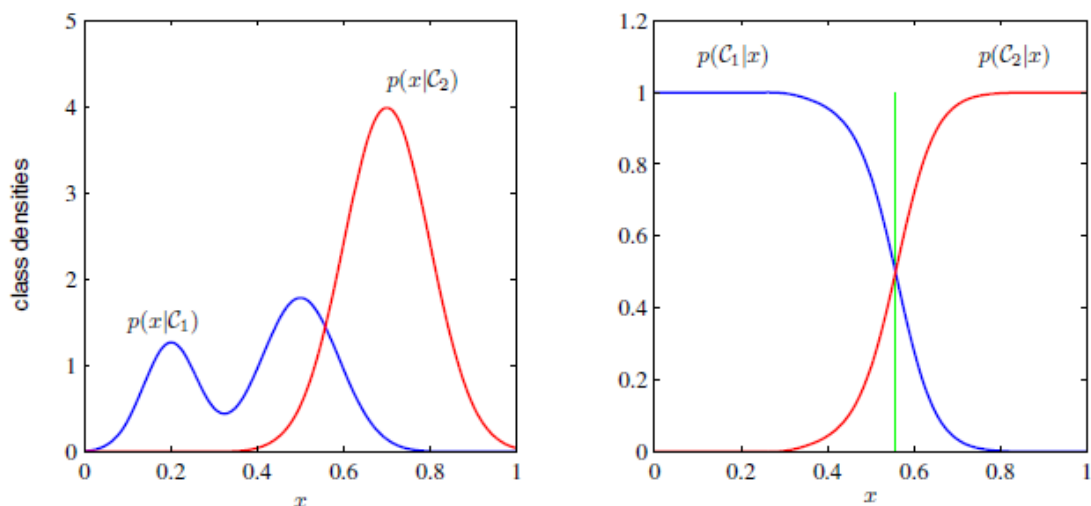


Figure 1.1. Exemple des densités conditionnelles de classe pour deux classes ayant une seule variable d'entrée x (graphique de gauche) avec les probabilités a posteriori correspondantes (graphique de droite). Notez que le mode de gauche de la densité conditionnelle de classe $p(x|C_1)$, représenté en bleu sur le graphique de gauche, n'a aucun effet sur les probabilités a posteriori. La ligne verte verticale dans le graphique de droite montre la limite de décision en x qui donne le taux d'erreur de classification minimal. [4]

L'étude présentée dans [5a] se concentre sur une base de données de patients atteints de glioblastomes, un cancer cérébral agressif, comprenant des dossiers cliniques, des images

d'IRM et de tomodensitométrie, ainsi que des données de dosimétrie 3D. L'étude utilise ces données pour développer des modèles prédictifs. La partie centrale de la thèse se penche sur le formalisme bayésien pour la modélisation des données cliniques, avec une explication détaillée du théorème de Bayes et de l'importance de la distribution a posteriori des paramètres d'intérêt. Les objectifs de l'inférence bayésienne sont abordés, notamment l'estimation de paramètres, la prédiction de nouvelles données, et la comparaison de modèles. Une partie de l'étude se concentre sur la prédiction de la zone de récurrence tumorale en utilisant des modèles linéaires généralisés et des arbres de décision appliqués aux images médicales prétraitées. Les résultats sont évalués à l'aide d'espaces ROC pour évaluer l'exactitude des prédictions.

En statistique, l'estimateur du maximum de vraisemblance est un estimateur statistique utilisé pour inférer les paramètres de la loi de probabilité d'un échantillon donné en recherchant les valeurs des paramètres maximisant la fonction de vraisemblance. En 2006, Bishop identifie trois méthodes pratiques différentes de résoudre les problèmes de décision[4] :

Les modèles génératifs: L'analyse d'images médicales en utilisant des modèles génératifs est devenue une approche prometteuse dans le domaine de l'imagerie médicale [4b,4c]. Les modèles génératifs sont des modèles statistiques capables de générer de nouvelles données à partir d'une distribution apprise à partir d'un ensemble d'apprentissage.

1. Ces approches modélisent, explicitement ou implicitement, la distribution des entrées avec *des modèles génératifs*, car il est possible de générer de nouvelles données ponctuelles synthétiques dans l'espace d'entrée. Le problème d'inférence est d'abord résolu en déterminant la densité conditionnelle de la classe $p(\mathbf{x}/C_k)$ pour chaque classe C_k séparément. Puis déduire séparément les probabilités a priori de chaque classe $p(C_k)$. Les probabilités postérieures peuvent être obtenues en modélisant et en normalisant la distribution conjointe $p(\mathbf{x}, C_k)$, la théorie de décision étant utilisée pour déterminer si chaque nouvel échantillon appartient à la classe appropriée.
2. L'utilisation de modèles génératifs dans l'analyse d'images médicales présente plusieurs avantages, notamment la capacité à générer des données synthétiques pour augmenter les ensembles de données limités, la génération de données manquantes pour l'apprentissage semi-supervisé, et la génération de variations d'images pour la validation croisée. Cependant, il est important de noter que ces modèles peuvent être sensibles à la qualité et à la quantité des données d'apprentissage, et nécessitent souvent des ensembles de données volumineux pour obtenir des résultats fiables.[5b,5c]
3. L'avantage de ces approches, est qu'elles permettent également de déterminer la densité marginale des données $p(\mathbf{x})$. Cela peut être utile pour détecter de nouveaux points de données qui ont une faible probabilité sous le modèle et pour lesquels les prédictions peuvent être de faible précision, ce que l'on appelle *la détection des valeurs aberrantes* ou *la détection de la nouveauté*.[5]

Les modèles discriminants: Ces approches sont considérées comme de complexité moyenne, où la complexité des probabilités à posteriori se fait directement au moyen *des modèles* dits *discriminants*. Cette méthode résout le problème d'inférence consistant à déterminer toutes les probabilités à posteriori $p(C_k/\mathbf{x})$ de toutes les classes, puis utilise la théorie de décision pour attribuer \mathbf{x} à l'une des classes prédéfinies.[5]

Les modèles discriminants, quant à eux, se concentrent sur la modélisation de la probabilité conditionnelle $p(Y|X)$, c'est-à-dire la probabilité d'une étiquette de classe Y étant donné une donnée d'entrée X . Ils ne modélisent pas directement la distribution conjointe $p(X, Y)$, mais plutôt la relation entre les données d'entrée et les étiquettes de classe. L'objectif principal des modèles discriminants est de trouver une frontière de décision qui permet de séparer les différentes classes de manière optimale.

Les fonctions discriminantes : Ces méthodes sont considérées comme les moins compliquées, dont une fonction distincte $f(\mathbf{x})$ est recherchée, appelée *fonction discriminante*. Les données d'apprentissage sont utilisées pour trouver une fonction discriminante $f(x)$ qui mappe chaque x directement sur une étiquette de classe, combinant ainsi les étapes d'inférence et de décision en un seul problème d'apprentissage. La fonction discriminante est une fonction mathématique qui est utilisée pour discriminer entre les différentes classes. En insérant les estimations des probabilités dans la fonction discriminante, on peut obtenir une estimation de cette fonction qui peut être utilisée pour classer de nouvelles données. Dans l'exemple de la Figure 1.1, cela correspondrait à trouver la valeur de x indiquée par la ligne verte verticale, car il s'agit de la frontière de décision donnant la probabilité minimale d'erreur de classification.

Combinaison de modèles : Pour les applications complexes, il est souhaitable de décomposer le problème en un certain nombre de sous-problèmes plus petits, chacun pouvant être traité par un module séparé. La combinaison de modèles est une technique qui vise à améliorer les performances prédictives en fusionnant les prédictions de plusieurs modèles individuels. Cette approche repose sur l'idée que les modèles individuels peuvent avoir des forces et des faiblesses différentes, et en les combinant, on peut obtenir une prédiction plus robuste et plus précise. Il existe différentes méthodes de combinaison de modèles, voici quelques-unes des plus couramment utilisées:

1. Moyenne des prédictions: Dans cette approche, les prédictions des modèles individuels sont simplement moyennées pour obtenir la prédiction finale. C'est une méthode simple mais efficace, qui fonctionne bien lorsque les modèles individuels sont diversifiés et ont des biais différents.

2. Majorité des votes: Cette méthode est souvent utilisée dans le cas des problèmes de classification. Chaque modèle donne une prédiction de classe, et la classe prédite est celle qui obtient le plus grand nombre de votes parmi les modèles.

3. Poids des modèles : Plutôt que de donner le même poids à tous les modèles, il est possible d'attribuer des poids différents en fonction de la confiance accordée à chaque modèle. Les poids peuvent être déterminés à l'aide d'une validation croisée ou d'une méthode d'apprentissage supplémentaire.

4. Méta-apprentissage : Dans cette approche, un modèle supplémentaire appelé méta-apprenant est entraîné pour apprendre à combiner les prédictions des modèles individuels. Le méta-apprenant peut être un modèle simple comme la régression logistique ou un modèle plus complexe comme un réseau de neurones. [6a]

La combinaison de modèles peut être utilisée dans divers domaines de l'apprentissage automatique, y compris la classification, la régression et même la détection d'anomalies. Cependant, il convient de noter que la combinaison de modèles n'est pas toujours garantie d'améliorer les performances et peut nécessiter une sélection judicieuse des modèles et une validation appropriée pour éviter le surajustement.

En 2019, une approche utilisant un ensemble de modèles CNN pour la classification de nodules pulmonaires est présente, en soulignant les avantages potentiels de la fusion de plusieurs modèles et en proposant différentes stratégies de fusion : [6a]

- Les auteurs soulignent les limites d'un seul modèle d'apprentissage en raison de problèmes tels que l'espace d'hypothèses limité, le risque de convergence vers des minima locaux ou une mauvaise sélection de l'espace d'hypothèses.
- Pour améliorer la précision des systèmes de prise de décision clinique, plusieurs cadres utilisant des modèles CNN ensembles ont été mis en œuvre. L'auteur propose de construire des modèles d'apprentissage en ensemble en fusionnant plusieurs modèles CNN profonds pour la classification des nodules pulmonaires. Plus précisément, huit modèles CNN profonds avec différentes architectures sont entraînés en tant que modèles primaires pour donner à chaque nodule huit valeurs de possibilité de malignité ou de bénignité.
- Ensuite, ces huit prédictions sont fusionnées en utilisant des stratégies telles que le vote majoritaire, la moyenne et plusieurs méthodes d'apprentissage automatique. L'auteur mentionne qu'une étude antérieure utilisant des images naturelles a montré que la fusion de nombreux modèles CNN profonds peut permettre d'obtenir une précision de classification plus élevée que celle d'un modèle CNN individuel.
- Les contributions et nouveautés de cette étude sont résumées en trois aspects. Premièrement, la motivation de cette recherche est l'assemblage de différents modèles CNN avec des architectures différentes, chacun ayant des perspectives d'apprentissage différentes et des caractéristiques apprises différentes des nodules pulmonaires. Deuxièmement, les relations entre les modèles CNN sont clarifiées, ce qui peut être une référence pour d'autres études similaires pour choisir les modèles CNN appropriés. Troisièmement, dix méthodes différentes, y compris le vote majoritaire, la moyenne et diverses méthodes d'apprentissage automatique, sont utilisées pour fusionner les prédictions des huit modèles CNN et générer ainsi dix modèles d'ensemble.
- Les auteurs mentionnent que la plupart des études antérieures utilisaient uniquement le vote majoritaire comme stratégie de fusion, et les relations entre les dix modèles d'ensemble sont révélées pour aider les chercheurs à identifier la stratégie de fusion optimale. [6a]

1.2.3. Estimation paramétrique de l'estimateur de Bayes

1.2.3.1. Méthodes paramétriques

L'une des approches les plus directes de l'estimation de la densité consiste à représenter la densité de probabilité $p(x)$ sous la forme d'une forme fonctionnelle spécifique qui contient un certain nombre de paramètres ajustables. Les valeurs des paramètres peuvent ensuite être optimisées pour donner le meilleur ajustement aux données. Deux principales techniques de détermination des paramètres de la distribution du modèle, appelées respectivement **maximum de vraisemblance** et **inférence bayésienne**. Ces deux approches sont couramment utilisées dans l'estimation des paramètres des modèles statistiques. Voici une brève explication de chaque concept :

1. Maximum de vraisemblance : Le maximum de vraisemblance est une méthode d'estimation des paramètres d'un modèle statistique en maximisant la vraisemblance des données observées. La vraisemblance mesure la probabilité d'observer les données

les paramètres du modèle. L'idée principale du maximum de vraisemblance est de trouver les valeurs des paramètres qui rendent les données observées les plus probables. En d'autres termes, il s'agit de trouver les paramètres qui maximisent la probabilité d'obtenir les données réelles à partir du modèle. Le maximum de vraisemblance est largement utilisé en statistique et en apprentissage automatique pour estimer les paramètres de divers modèles, tels que les modèles linéaires, les modèles de régression, les modèles de mélange, etc.

2. Inférence bayésienne : L'inférence bayésienne est une approche statistique qui combine les connaissances a priori et les données observées pour estimer les paramètres d'un modèle. Contrairement au maximum de vraisemblance qui ne se concentre que sur les données observées, l'inférence bayésienne utilise également des informations a priori sur les paramètres pour calculer une estimation postérieure de ces paramètres. Cela se fait en utilisant le théorème de Bayes, qui permet de mettre à jour les croyances a priori en fonction des nouvelles données observées. L'estimation bayésienne fournit une distribution de probabilité sur les paramètres plutôt qu'une seule valeur estimée, ce qui permet de quantifier l'incertitude associée à l'estimation des paramètres. L'inférence bayésienne est utilisée dans divers domaines, tels que l'apprentissage automatique bayésien, la modélisation statistique et l'analyse de données. [3]

Le classifieur de Bayes consiste en l'estimation de la probabilité a posteriori d'un objet décrit par x à l'aide de la formule de Bayes, suivie d'une prise de décision selon la règle de décision de Bayes: attribuer l'objet à la classe dont la probabilité a posteriori est la plus grande.

Il peut néanmoins servir de référence lorsque l'on cherche à évaluer la qualité d'un classifieur : on peut appliquer celui-ci à un problème fictif pour lequel les probabilités a priori et les vraisemblances sont connues exactement, et comparer ses performances à celles du classifieur de Bayes sur ce même problème.

$$P(C_i|x) = \frac{p(x|C_i)P(C_i)}{p(x)} = \frac{p(x|C_i)P(C_i)}{\sum_{k=1}^k p(x|C_k)P(C_k)} \quad (1.1)$$

et utiliser la fonction discriminante :

$$g_i(x) = p(x|C_i)P(C_i) \quad (1.2)$$

De manière équivalente:

$$g_i(x) = \log p(x|C_i) + \log P(C_i) \quad (1.3)$$

La distribution *normale* ou *gaussienne* est le modèle paramétrique le plus simple et le plus largement utilisé. Ce modèle est apprécié pour ses propriétés analytiques et statistiques pratiques. Si on assume que $p(x|C_i)$ sont gaussiennes alors:

$$p(x|C_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(x - \mu_i)^2}{2\sigma_i^2}\right] \quad (1.4)$$

Et

$$g_i(x) = \frac{1}{2} \log 2\pi - \log \sigma_i - \frac{(x - \mu_i)^2}{2\sigma_i^2} + \log P(C_i) \quad (1.5)$$

Le cas où on ne connaît pas les probabilités a priori des classes ($P(C_i)$) et les probabilités conditionnelles des données d'entrée étant donné les classes ($p(x|C_i)$). Dans cette situation, nous devons estimer ces probabilités à partir d'un échantillon de données. La fréquence

relative des occurrences de chaque classe dans l'échantillon est utilisée pour estimer $P(C_i)$. De même, pour estimer $p(x|C_i)$, on utilise l'échantillon pour estimer la distribution conditionnelle des données d'entrée pour chaque classe. Cela peut être fait en calculant la fréquence ou la densité des valeurs observées de x pour chaque classe dans l'échantillon. Une fois que ces probabilités sont estimées, il est possible de les insérer dans la fonction discriminante pour obtenir une estimation de la fonction discriminante elle-même (Voir l'équation 1.5).

1.2.3.2. Gestion de la complexité par le dilemme Biais/Variance

Les données de classification suivent généralement une distribution, souvent de type normale. Cependant, ces données peuvent être altérées par du bruit, ce qui rend l'apprentissage de la relation que nous cherchons à prédire difficile, voire impossible. Dans le cas de la régression polynomiale, Plusieurs ajustements polynomiaux peuvent être choisis, tels que d'ordre 1, 4 et 5. Ceci est fait dans le but de minimiser l'erreur et de trouver un équilibre entre la variance et le biais.

À mesure que l'ordre du polynôme augmente, le biais diminue, mais la variance augmente. En utilisant un modèle complexe, qui prend en compte la moyenne, nous obtenons un meilleur ajustement à la fonction sous-jacente de l'estimateur, ce qui réduit le biais. Cela crée ce que l'on appelle le dilemme biais/variance, qui est applicable à tous les systèmes d'apprentissage automatique, pas seulement à la régression polynomiale [7].

Dans le cas de la régression polynomiale, il existe plusieurs degrés de polynômes parmi lesquels choisir pour ajuster les données. Le choix du degré du polynôme est décisif car il détermine la complexité du modèle et son aptitude à capturer la relation entre les variables d'entrée et la variable cible.

Lorsque l'on ajuste un modèle polynômial, il est important de trouver un équilibre entre la variance et le biais afin d'obtenir une bonne généralisation du modèle. Le biais se réfère à l'erreur due à la simplification du modèle, tandis que la variance se réfère à l'erreur due à la sensibilité excessive du modèle aux variations aléatoires des données d'apprentissage.

Si nous choisissons un polynôme de faible degré, le modèle aura tendance à sous-ajuster les données, ce qui se traduit par un biais élevé. Cela signifie que le modèle ne parviendra pas à capturer la complexité réelle de la relation entre les variables d'entrée et la variable cible. Cependant, le modèle aura une faible variance car il est relativement insensible aux variations des données d'apprentissage.

D'autre part, si nous choisissons un polynôme de degré élevé, le modèle aura tendance à sur-ajuster les données, ce qui se traduit par une variance élevée. Cela signifie que le modèle sera très sensible aux variations des données d'apprentissage et pourra avoir du mal à généraliser correctement aux nouvelles données. Cependant, le modèle aura un biais faible car il sera capable de capturer des relations complexes dans les données. Pour trouver un équilibre entre la variance et le biais, différentes techniques peuvent être utilisées :

- 1. Validation croisée :** Une approche courante consiste à diviser les données en ensembles d'apprentissage et de validation. Différents degrés de polynômes sont ajustés sur l'ensemble d'apprentissage, et l'erreur de validation est utilisée pour évaluer les performances du modèle. Le degré du polynôme qui minimise l'erreur de validation est alors choisi.
- 2. Courbe d'apprentissage :** Une autre approche consiste à tracer la courbe d'apprentissage, qui représente l'erreur d'entraînement et l'erreur de validation en fonction du degré du polynôme. On cherche un point où les deux erreurs sont faibles et stables, ce qui indique un bon équilibre entre biais et variance.

- 3. Régularisation :** La régularisation est une technique qui ajoute une pénalité aux coefficients du polynôme pour éviter des valeurs trop grandes. Cela permet de réduire la complexité du modèle et de limiter la variance.

Pour réduire le biais, le modèle doit être flexible, mais cela augmente le risque d'une variance élevée. Si la variance reste faible, il n'est pas possible d'ajuster les données tout en maintenant un biais élevé. En augmentant progressivement l'ordre du polynôme, nous minimisons progressivement le biais, ce qui entraîne également une diminution de l'erreur de généralisation. Le modèle optimal est celui qui présente le meilleur compromis entre le biais et la variance. [4]

1.2.4. Les approches semi-paramétriques (Clustering hiérarchique)

Les approches semi-paramétriques, telles que le clustering hiérarchique, sont des méthodes d'analyse de données qui combinent des aspects paramétriques et non paramétriques. Le clustering hiérarchique est une technique de regroupement de données qui vise à créer une structure arborescente (dendrogramme) pour représenter les similarités ou dissimilarités entre les observations.

Dans le contexte du clustering hiérarchique, l'approche semi-paramétrique se réfère au fait que le choix du nombre de clusters n'est pas spécifié à priori, mais est déterminé à partir des données elles-mêmes. Cela signifie que le nombre de clusters peut être déterminé de manière adaptative en fonction de la structure des données plutôt que d'être fixé de manière rigide.

L'approche semi-paramétrique du clustering hiérarchique offre une flexibilité accrue par rapport aux méthodes entièrement paramétriques, telles que les modèles de mélange gaussien, qui supposent souvent une distribution spécifique pour les données. Au lieu de cela, le clustering hiérarchique semi-paramétrique permet aux clusters de prendre des formes arbitraires, ce qui le rend plus adapté aux données complexes ou moins bien décrites par des distributions paramétriques spécifiques. [4]

Clustering hiérarchique

Cette section offre une vue d'ensemble des différentes approches et méthodes liées au regroupement des données (clustering), en mettant en avant les aspects probabilistes, les mesures de similarité, les types d'algorithmes, et les avantages et limitations de certaines approches en fonction des caractéristiques des données.

- 1. Vue probabiliste du regroupement :** Deux perspectives probabilistes sur le regroupement des données sont évoquées. La première consiste à ajuster un modèle de mélange aux données, tandis que la seconde repose sur la recherche de mots de code minimisant l'erreur de reconstruction. Cette approche indique que le regroupement peut être considéré comme une tâche d'apprentissage probabiliste.
- 2. Méthodes de clustering basées sur les similitudes :** Il existe des méthodes de regroupement qui n'utilisent que les mesures de similarité entre les instances, sans autre contrainte sur les données. L'objectif est de former des groupes où les instances au sein d'un groupe sont plus similaires les unes aux autres qu'à celles des autres groupes. Le clustering hiérarchique est cité en tant qu'exemple de cette approche.
- 3. Mesure de similarité ou de distance :** Pour mettre en œuvre ces méthodes, une mesure de similarité (ou de distance) doit être définie entre les instances. La distance euclidienne est mentionnée comme une mesure courante, mais il est souligné qu'il est nécessaire de normaliser les attributs pour qu'ils aient la même échelle.

- 4. Types d'algorithmes de clustering** : Deux types d'algorithmes de clustering ont été présentés dans la littérature : les algorithmes agglomératifs et les algorithmes de division. Les premiers commencent avec N groupes (où N est le nombre d'instances) et fusionnent les groupes similaires pour former des groupes plus grands. Les seconds commencent avec un seul groupe et le divisent en groupes plus petits jusqu'à ce que chaque groupe ne contienne qu'une seule instance.
- 5. Critère de fusion dans les algorithmes agglomératifs** : À chaque itération des algorithmes agglomératifs, les deux groupes les plus proches sont fusionnés. Dans le cas du clustering à liaison unique, la distance utilisée est la plus petite distance entre toutes les paires possibles d'éléments des deux groupes.
- 6. Avantages de l'approche par mélange de gaussiennes** : L'approche par mélange de gaussiennes est présentée comme offrant l'avantage de mieux s'adapter à la distribution des données cibles par rapport à une approche avec un unique noyau gaussien. Cependant, elle peut être coûteuse en haute dimension, nécessitant un grand nombre de données pour estimer les paramètres des noyaux du modèle.
- 7. Utilisation de l'approche par fenêtres de Parzen** : En cas de données insuffisantes, l'approche par fenêtres de Parzen est privilégiée. Cela suggère que cette méthode peut être plus appropriée lorsque les données d'apprentissage sont limitées. [7,7a]

Choix du nombre de clusters

Comme toute méthode d'apprentissage, le clustering a également son bouton pour ajuster la complexité à travers nombre de clusters(k), Étant donné tout k, le clustering trouvera toujours k centres, qu'ils soient vraiment des groupes significatifs ou qu'ils soient imposés par la méthode utilisée. Il est possible utiliser plusieurs méthodes pour affiner k. En fonction du type de méthode de clustering utilisée, il est également possible de tracer l'erreur de reconstruction ou enregistrer la probabilité en fonction de k et rechercher le «coude». Après un k assez grand, l'algorithme commencera à diviser les groupes, auquel cas il n'y aura pas de diminution importante de l'erreur de reconstruction ou d'augmentation importante de la vraisemblance logarithmique. De même, dans le clustering hiérarchique, en regardant les différences entre les niveaux de l'arborescence, nous pouvons décider d'une bonne répartition.

1.2.5. Algorithme de maximisation des attentes

L'algorithme de maximisation des attentes [8], ou algorithme EM (Expectation Maximization), est une technique générale pour trouver des solutions de maximum de vraisemblance pour les modèles probabilistes ayant des variables latentes [8]. L'algorithme de maximisation des attentes (EM) est utilisé dans l'estimation du maximum de vraisemblance lorsque le problème implique deux ensembles de variables aléatoires dont l'un, X, est observable et l'autre, Z, est caché. Son but est de trouver le vecteur de paramètres Φ qui maximise la vraisemblance des valeurs observées de X, $L(\Phi | X)$. La distribution conjointe $p(X,Z|\theta)$ est régie par un ensemble de paramètres noté θ . L'objectif est de maximiser la fonction de vraisemblance donnée par :

$$p(X|\theta) = \sum_Z p(X,Z|\theta) \quad (1.6)$$

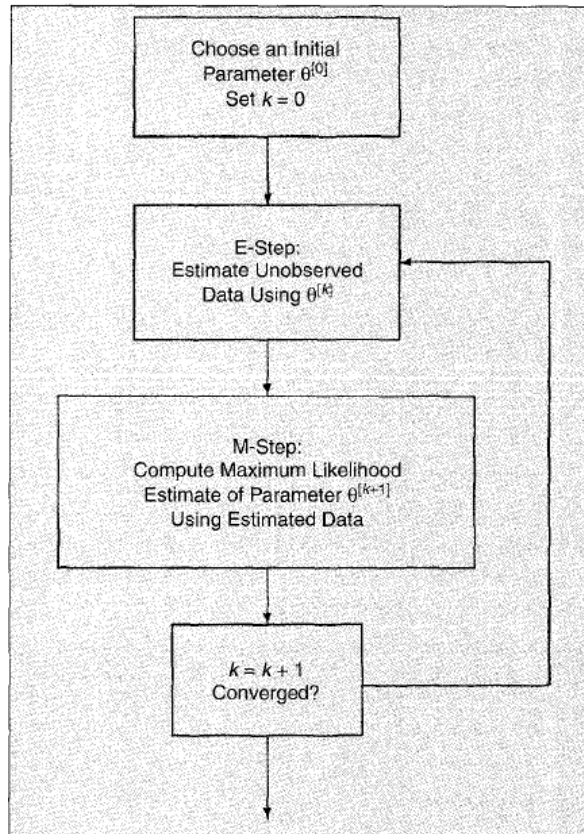


Figure 1.2. Un overview de l'algorithme EM. Après l'initialisation, l'E-step et le M-step sont alternés jusqu'à ce que l'estimation du paramètre ait convergé (plus de changement dans l'estimation).

L'idée de base derrière l'algorithme EM est de trouver θ qui maximise $\log f(x|\theta)$, mais nous n'avons pas les données x pour calculer la log-vraisemblance. Donc, à la place, nous maximisons l'espérance de $\log f(x|\theta)$ compte tenu des données y et de notre estimation actuelle de θ . Cela peut être exprimé en deux étapes.

L'algorithme EM consiste à choisir un $\theta^{[k]}$ initial, en effectuant successivement l'étape E et l'étape M jusqu'à la convergence. La convergence peut être déterminée en examinant le moment où les paramètres stoppent de changer, c'est-à-dire s'arrêtent lorsque $\|\theta^{[k]} - \theta^{[k-1]}\| < \varepsilon$ pour un ε est une mesure de distance appropriée $\|\cdot\|$. (Voir Figure 1.2)

L'importance de l'algorithme EM apparaît dans la classification avec des modèles de mélange, tout en mettant en avant ses limitations en termes de sensibilité aux valeurs initiales et de vitesse de convergence lente. Une variante de l'algorithme EM intègre une contrainte spatiale pour la segmentation d'images, ce qui peut améliorer la robustesse au bruit et la qualité des résultats de segmentation.

- 1. Utilisation de l'algorithme EM dans la classification :** L'algorithme EM est une méthode puissante pour l'estimation des paramètres dans des modèles de mélange, notamment en classification. Son utilisation dans le contexte de la classification avec des modèles de mélange est une approche bien établie et largement utilisée dans le domaine de l'apprentissage automatique et de la statistique.
- 2. Sensibilité aux valeurs initiales :** L'un des défis courants de l'algorithme EM est sa sensibilité aux valeurs initiales. Cela signifie que le choix initial des paramètres peut influencer la convergence de l'algorithme et le résultat final. Il est important de prendre

en compte cette sensibilité lors de l'application de l'algorithme EM et d'explorer différentes initialisations pour éviter de converger vers des minima locaux.

- 3. Vitesse de convergence lente** : Il est également noté que l'algorithme EM peut avoir une vitesse de convergence lente dans certaines situations. Cela signifie que le processus d'estimation des paramètres peut prendre du temps, en particulier lorsque le modèle est complexe ou que les données sont volumineuses. Des améliorations telles que l'utilisation d'algorithmes EM accélérés peuvent être envisagées pour résoudre ce problème.
- 4. Modèle de mélange gaussien avec contrainte spatiale** : Une nouvelle approche peut intégrer un modèle de mélange gaussien avec une contrainte spatiale pour la segmentation d'images. Cette méthode semble être conçue pour tenir compte de la structure spatiale des données, ce qui peut être essentiel dans le traitement d'images. L'utilisation de contraintes spatiales peut aider à améliorer la robustesse de la segmentation en réduisant l'impact du bruit et en favorisant une cohérence spatiale dans les résultats.
- 5. Robustesse au bruit grâce à la contrainte spatiale** : Cette méthode est robuste au bruit grâce à la contrainte spatiale est encourageante. Cela suggère qu'elle peut être particulièrement utile dans des environnements où les données d'image sont susceptibles de contenir du bruit, ce qui est fréquent dans la pratique. [8a]

En 2002, Chung, et al proposent un résumé d'une étude qui a comparé les performances de deux modèles de mélange différents pour le bruit de fond dans les images. Le premier modèle, le modèle de Maxwell, est une distribution simple souvent utilisée pour modéliser le bruit de fond. Le deuxième modèle, le modèle de mélange Maxwell-Gaussien, est une distribution plus complexe qui combine le modèle de Maxwell avec une distribution gaussienne. L'étude a révélé que le modèle de mélange Maxwell-Gaussien offrait un meilleur ajustement au bruit de fond que le modèle Maxwell. En effet, le modèle de mélange Maxwell-Gaussien est capable de capturer la distribution plus complexe du bruit de fond. L'étude a également révélé que l'algorithme EM modifié était un moyen efficace d'estimer les paramètres des modèles de mélange. L'algorithme EM modifié est une variante de l'algorithme EM conçu pour être plus efficace. [8b]

1.2.6. Classification et régression non paramétrique

Ces algorithmes consistent à trouver des instances similaires passées à partir de l'ensemble d'apprentissage en utilisant une mesure de distance appropriée et à les interpoler pour trouver la sortie correcte. L'estimation non paramétrique suppose une hypothèse raisonnable qui stipule que des entrées similaires doivent avoir des sorties similaires. Différentes méthodes non paramétriques diffèrent dans la façon dont elles définissent la similarité et comment elles interpolent des instances d'apprentissage similaires. Contrairement aux modèles paramétriques, dont chaque instance d'apprentissage affecte l'estimation globale et finale; les méthodes non paramétriques ne prennent aucune forme paramétrique a priori pour les densités sous-jacentes; sa complexité dépend de la taille de l'ensemble d'apprentissage, ou plutôt de la complexité du problème inhérent aux données. La vitesse de convergence de fonctions de régression est très lourde dans les zones rétrécies, le cluster de données est lisse et les fonctions changent lentement, qu'il s'agisse de densités, de discriminants ou de fonctions de régression. Des exemples similaires signifient des choses similaires.

Dans cette section, On considère les approches non paramétriques pour l'estimation de la densité, la classification et la régression. On commence par estimer une fonction de densité et on discute son utilisation dans la classification. On généralise ensuite l'approche de la régression, et voyons comment la complexité temporelle et spatiale peut être vérifiée.

1.2.6.1. Les types d'estimateurs de la densité non paramétrique

Il existe plusieurs types d'estimateurs de densité non paramétriques, qui diffèrent dans leur approche de l'estimation de la densité à partir des données sans imposer de forme paramétrique spécifique. Cette section présente différentes méthodes d'estimation de densité non paramétrique, notamment les estimateurs d'histogramme, de noyau, de Parzen et de k-plus proche voisin, en expliquant leurs principes et leurs avantages/désavantages. Ces méthodes sont couramment utilisées en statistiques pour estimer la distribution de probabilité des données. Quatre types d'estimateurs de densité non paramétriques les plus couramment utilisés :

1. Introduction aux estimateurs de densité non paramétrique : Les approches non paramétriques consistent à estimer la densité de probabilité en fixant un paramètre, la largeur d'onde (h), et en examinant combien d'échantillons tombent dans des intervalles (bacs). Ces intervalles peuvent avoir des amplitudes variables et dépendent de la densité des points autour d'un point de référence.

2. Estimateur d'histogramme :

- L'estimateur d'histogramme divise l'espace des données en intervalles de taille égale appelés "bins" et compte le nombre de points de données dans chaque intervalle.
- Le choix de la taille des bins et de l'origine affecte l'estimation de la densité. Des bins plus petits donnent une estimation plus rugueuse, tandis que des bins plus grands donnent une estimation plus lisse.
- L'estimation est obtenue en normalisant les comptages par la largeur des intervalles.
- L'estimateur d'histogramme est simple, mais le choix des paramètres (nombre et largeur des intervalles) peut influencer la qualité de l'estimation.

3. Estimateur du noyau :

- L'estimateur du noyau utilise une fonction de pondération régulière (fonction noyau) pour lisser l'estimation de la densité.
- Le noyau le plus populaire est le noyau gaussien, qui détermine la forme des influences et la largeur de la fenêtre (h) qui est le seul paramètre à estimer.
- Une petite valeur de h signifie que chaque instance d'apprentissage a un effet important sur une petite région, tandis qu'une grande valeur de h donne une estimation plus douce.
- L'estimateur de noyau est flexible mais nécessite le choix approprié du noyau et du paramètre de lissage (bande passante).

4. Estimateur de Parzen :

- L'estimateur de Parzen approxime la distribution réelle des données en combinant linéairement des densités construites autour de chaque exemple en apprentissage.
- Il utilise un noyau gaussien par défaut et estime souvent la largeur de la bande par une procédure en leave-one-out.

- Il est recommandé lorsque le nombre de données est faible, mais il est coûteux en calcul lorsque le nombre de données est élevé.
- Le paramètre h contrôlant la largeur du noyau est le seul paramètre à estimer.

5. Estimateur de k-plus proche voisin :

- Cet estimateur est basé sur la distance entre un point donné et ses voisins les plus proches pour estimer la densité en ce point.
- Il est utile lorsque les données sont regroupées ou ont des structures complexes.
- Le degré de lissage est contrôlé par k , le nombre de voisins pris en compte, et il dépend de la densité locale des données.
- Il peut utiliser une fonction noyau pour obtenir une estimation plus lisse. [7a]

1.2.6.2. Classification non paramétrique

Les méthodes non paramétriques pour l'estimation de la densité de probabilité sont des approches statistiques qui permettent d'estimer la distribution sous-jacente des données sans faire d'hypothèses spécifiques sur la forme de cette distribution. Ces méthodes sont particulièrement utiles lorsque la forme de la distribution est inconnue ou complexe.

Le principe général des méthodes non paramétriques pour l'estimation de la densité de probabilité repose sur la construction d'une fonction estimée qui se rapproche le plus possible de la véritable densité de probabilité. Parmi les méthodes les plus couramment utilisées, on trouve :

- 1. Estimation des densités conditionnelles de classe :** L'approche non paramétrique consiste à estimer les densités conditionnelles de classe, c'est-à-dire la probabilité que les instances appartiennent à une classe donnée étant donné leurs caractéristiques (représentées par la variable x). Ces densités conditionnelles sont notées $p(x | C_i)$, où C_i représente une classe spécifique.
- 2. Classification basée sur le maximum du discriminant :** Les instances d'apprentissage sont affectées à la classe pour laquelle le discriminant atteint sa valeur maximale. En d'autres termes, pour chaque instance, on cherche la classe qui maximise la probabilité conditionnelle $p(x | C_i)$.
- 3. Le vote des instances d'apprentissage :** Chaque instance d'apprentissage vote pour sa propre classe et n'a aucun effet sur les autres classes. Le poids du vote est généralement déterminé par une fonction noyau, qui attribue généralement plus de poids aux instances qui sont plus proches de l'instance en question.
- 4. Le classificateur k-NN :** L'estimateur k-NN (k plus proches voisins) est mentionné comme un cas particulier de cette approche non paramétrique. Il fonctionne en affectant l'entrée à la classe qui compte le plus d'exemples parmi les k voisins les plus proches de cette entrée. Les voisins ont un vote égal et la classe avec le plus grand nombre de votes parmi les k voisins est choisie. Si nécessaire, des méthodes sont utilisées pour rompre les égalités ou pour effectuer un vote pondéré. En général, on utilise un nombre k impair pour éviter les situations d'égalité et de confusion entre classes voisines. [4, 7]

1.2.6.3. Régression non-paramétrique et les modèles de lissage

La régression paramétrique suppose l'utilisation d'un polynôme d'un certain ordre pour modéliser la relation entre les variables et trouve les coefficients de ce polynôme qui

minimisent l'erreur quadratique sur l'ensemble d'apprentissage. En revanche, la régression non paramétrique est utilisée lorsque l'on ne peut pas supposer qu'un tel polynôme existe. Elle se base sur l'idée que les valeurs de y (la variable cible) sont proches de $g(x)$ (la fonction inconnue) pour des valeurs proches de x .

Lors de l'estimation de la régression non paramétrique, pour un certain x donné, on cherche d'abord à trouver un voisinage de x . Ensuite, on calcule la moyenne des valeurs de y dans ce voisinage pour obtenir une estimation de $g(x)$, notée $\hat{g}(x)$. Cette approche est similaire à celle utilisée dans l'estimation de densité non paramétrique.

L'estimateur de régression non paramétrique est souvent appelé *un lissage (smoother)* car il vise à fournir une estimation en lissant les données brutes. L'estimation elle-même est également appelée un lissage. L'idée sous-jacente est de réduire les variations aléatoires dans les données pour mettre en évidence les tendances sous-jacentes ou les structures plus lisses. [10]

Le choix du paramètre de lissage

Dans les méthodes non paramétriques, comme l'estimation ou la régression de la densité, le paramètre de lissage est essentiel. Il est utilisé pour déterminer la largeur de la fenêtre de lissage (dans le cas de la largeur de la bande ou l'étalement du noyau) ou le nombre de voisins utilisés (dans le cas du k plus proches voisins). L'objectif est d'obtenir une estimation moins variable que les points de données bruts. Le paramètre de lissage doit être choisi de manière à éliminer l'effet du bruit tout en préservant la variabilité de la fonction sous-jacente inconnue. Un paramètre de lissage trop grand peut entraîner un sur-lissage (une estimation trop lisse) où la variabilité de la fonction est également réduite. À l'inverse, un paramètre de lissage trop petit peut entraîner un sous-lissage (une estimation insuffisamment lisse) où l'effet du bruit n'est pas éliminé. Le choix du paramètre de lissage implique donc un compromis entre le biais et la variance de l'estimation. La nécessité de trouver un équilibre entre le biais et la variance est explicitement codée dans une fonction de coût régularisée utilisée dans *le lissage des splines*. Les splines sont des courbes lisses qui passent par certains points de données. Le paramètre de lissage intervient dans la régularisation de la fonction de coût pour contrôler la flexibilité de la spline et ajuster le niveau de lissage souhaité. [7]

Dans le contexte de l'estimation non paramétrique, la métrique de distance utilisée est un aspect critique. Lorsque les attributs sont discrets, la distance de Hamming est mentionnée comme une option simple où le nombre d'attributs non correspondants est simplement compté. Cependant, des fonctions de distance plus sophistiquées sont discutées dans d'autres références [11, 12] pour prendre en compte des aspects plus complexes de la similarité entre les instances.

Le k plus proches voisins est la méthode de classification non paramétrique la plus largement utilisée en raison de sa simplicité et de son succès dans une variété d'applications. [3]

1.3. Réduction de dimension par sélection de caractéristiques

La réduction de dimensionnalité est le processus de réduction du nombre de variables qui décrivent les données d'apprentissage. Cette technique est un sujet très important dans plusieurs domaines : Analyse exploratoire de données, Apprentissage automatique, le traitement d'image, la bio-informatique, la détection d'intrusion, l'analyse de courrier électronique et de spam Web, la classification de texte et la reconnaissance de formes. Il y a deux raisons principales pour garder la dimensionnalité de la représentation du motif (c'est-à-

dire le nombre de caractéristiques) aussi petite que possible : le coût de la mesure et la précision de la classification. Un ensemble de caractéristiques limité mais saillant simplifie à la fois la représentation du modèle et les classificateurs qui sont construits sur la représentation sélectionnée. Par conséquent, le classificateur résultant sera plus rapide et utilisera moins de mémoire [1, 13, 14]. Une différence est faite entre les techniques de réduction de la dimension par transformation (*Feature Extraction*) et celle par sélection (*Feature Selection*) de variables (voir Figure.1.3).

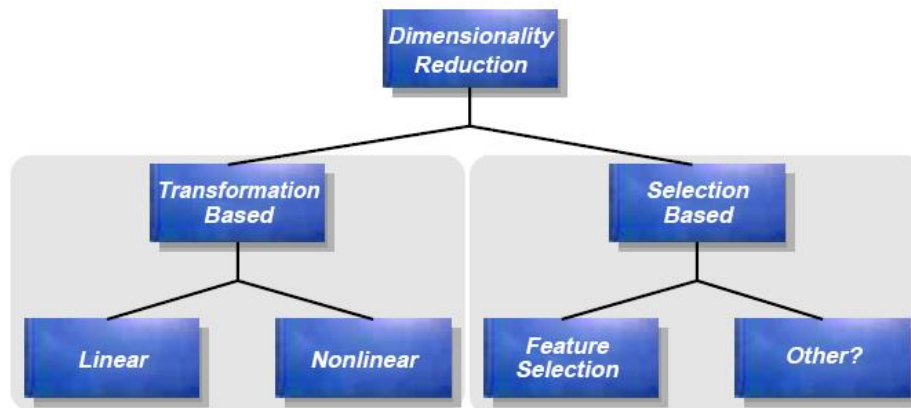


Figure 1.3: Taxinomie des approches de réduction de dimension [23]

Il est important de faire la distinction entre la sélection de caractéristiques et l'extraction de caractéristiques. Le terme sélection de caractéristiques fait référence aux algorithmes qui sélectionnent le meilleur sous-ensemble de l'ensemble de caractéristiques d'entrée, à partir d'un ensemble de départ, en utilisant divers critères et différentes méthodes, en conservant la sémantique de l'ensemble de données de départ au cours du processus de sélection.

Les méthodes qui créent de nouvelles caractéristiques basées sur des transformations ou des combinaisons de l'ensemble de caractéristiques d'origine sont appelées algorithmes d'extraction de caractéristiques. Le procédé de création utilise une transformation ou une combinaison à partir d'un espace de départ en effectuant, non seulement une réduction de la dimension, mais aussi une transformation de la sémantique de l'ensemble de données de départ. Des techniques bien connues se rattachant à cette catégorie sont, par exemple, l'analyse en composantes principales (ACP), l'analyse en composantes indépendantes (plus appropriée que l'ACP pour les distributions non-gaussiennes), l'analyse discriminante linéaire (LDA) [15].

D'après [1], Le choix entre la sélection de caractéristiques et l'extraction de caractéristiques dépend du domaine d'application et des données d'entraînement spécifiques disponibles. La sélection des caractéristiques entraîne des économies de coûts de mesure (puisque certaines caractéristiques sont supprimées) et les caractéristiques sélectionnées conservent leur interprétation physique d'origine. Les caractéristiques retenues peuvent être importantes pour comprendre le processus physique qui génère les motifs. De plus, les méthodes séquentielles classiques souffrent du problème de non-monotonie (*Nesting Effect*) [16] qui se traduit par le fait que le sous-ensemble optimal de M caractéristiques ne contient pas forcément le sous-ensemble optimal de P caractéristiques ($P < M$). L'objectif des différentes versions flottantes des méthodes séquentielles est de traiter ce type de problème. [GHE 08a]. Un petit nombre de caractéristiques peut atténuer la malédiction de la dimensionnalité lorsque le nombre d'échantillons d'apprentissage est limité. D'autre part, une réduction du nombre de

caractéristiques peut conduire à une perte du pouvoir de discrimination et ainsi diminuer la précision du système de reconnaissance résultant.

D'un autre côté, les caractéristiques transformées générées par l'extraction de caractéristiques peuvent fournir une meilleure capacité discriminante que le meilleur sous-ensemble de caractéristiques données, mais ces nouvelles caractéristiques (une combinaison linéaire ou non linéaire de caractéristiques données) peuvent ne pas avoir de signification physique claire. Dans de nombreuses situations, il est utile d'obtenir une projection bidimensionnelle ou tridimensionnelle des données multivariées données (matrice de modèle $n \times d$) pour permettre un examen visuel des données. [1]

1.3.1. Définitions de la sélection de caractéristiques

Pudil et al [17] définissent la sélection d'attributs comme suit : étant donnée une fonction permettant de mesurer la qualité d'un sous-ensemble de caractéristiques, la sélection de variables est réduite à un problème de recherche du sous-ensemble optimal par rapport à cette mesure. En effet, les principaux objectifs de la réduction de dimension par sélection sont:

- Faciliter la visualisation et la compréhension des données.
- Réduire l'espace de stockage nécessaire.
- Réduire le temps d'apprentissage et d'utilisation.
- Identifier les facteurs pertinents.
- Réduire la tâche d'extraction de primitives,
- Améliorer la précision du module de classification.
- Améliorer la fiabilité de l'estimation de la performance. [18]

Dash et Liu [19] proposent de regrouper les techniques de sélection de caractéristiques en fonction de l'objectif visé. Ils identifient alors quatre classes distinctes définies comme suit:

1. "*Idealized*" : trouver le sous-ensemble de taille minimale qui est nécessaire et suffisant pour atteindre l'objectif fixé.
2. "*Classic*" : sélectionner le sous-ensemble de m variables à partir de l'ensemble en contenant n ($m < n$), de sorte que la fonction critère choisie soit optimisée pour tous les sous-ensembles de taille m .
3. "*Improving prediction accuracy*" : choisir un sous-ensemble de caractéristiques afin d'améliorer la précision de la prédiction ou diminuer la taille de la structure sans diminution significative de la précision de prédiction du classifieur construit en utilisant seulement les variables sélectionnées.
4. "*Approximating original class distribution*" : sélectionner un sous-ensemble de variables tel que la distribution des classes résultantes soit aussi proche que possible de la distribution des classes étant donné l'ensemble complet des variables.

Webb [20] définit la sélection des caractéristiques comme étant le processus de sélection des d variables les plus discriminantes à partir d'un ensemble de n variables. Soient n le nombre total de variables et d la dimension désirée des sous-ensembles, le nombre total des sous-ensembles possibles d'attributs est égal à: $\frac{n!}{(n-d)!d!}$.

1.3.2. Processus de sélection de caractéristiques

Idéalement, les méthodes de la sélection de caractéristiques recherchent à travers les sous-ensembles de caractéristiques, et essaient de trouver le meilleur parmi les 2^N sous-ensembles

candidats selon une fonction d'évaluation. Cependant cette procédure est exhaustive, elle peut être trop coûteuse et pratiquement prohibitive, même pour une taille moyenne de l'ensemble de caractéristiques (M). Les autres méthodes basées sur la recherche heuristique ou aléatoire essaient de réduire la complexité statistique en compromettant la performance. Ces méthodes nécessitent un critère d'arrêt pour empêcher la recherche exhaustive des sous-ensembles. Dash et Liu [19] proposent qu'il y ait quatre composants de base dans une méthode typique de sélection de caractéristique (Voir Figure 1.4):

1. Une procédure de génération ou recherche pour générer le prochain sous-ensemble candidat ;
2. Une fonction d'évaluation pour évaluer le sous-ensemble considéré ;
3. Un critère d'arrêt pour décider quand s'arrêter ;
4. Une procédure de validation pour vérifier si le sous-ensemble est valide.

La procédure de génération est une procédure de recherche dont le rôle principal est de générer les sous-ensembles de caractéristiques pour l'évaluation. La procédure de recherche peut commencer par :

- (i) Un ensemble vide de caractéristiques,
- (ii) Un ensemble complet des caractéristiques,
- (iii) Un sous-ensemble aléatoire de caractéristiques.

Dans les deux premiers cas, les caractéristiques sont itérativement ajoutées ou enlevées, tandis que dans le dernier cas, les caractéristiques sont itérativement ajoutées ou enlevées ou aléatoirement générées.

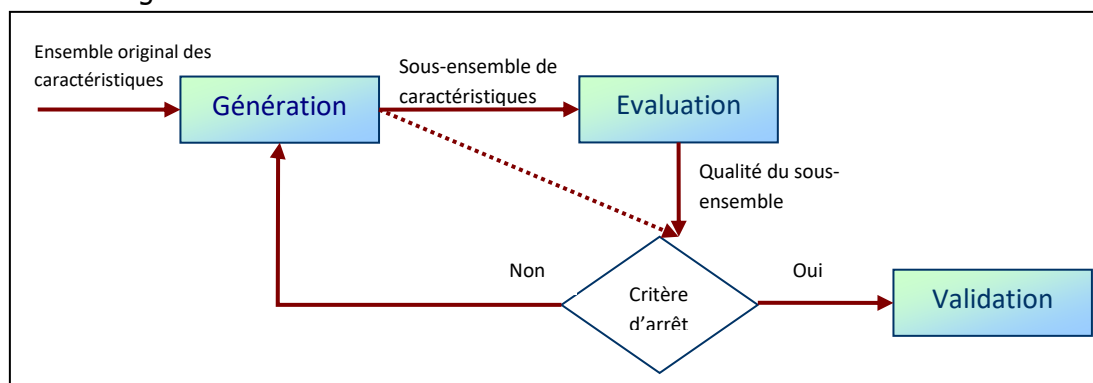


Figure 1.4: Processus de sélection de caractéristiques avec validation

Les algorithmes de sélection des caractéristiques effectuent généralement des recherches à travers l'espace des sous-ensembles de caractéristiques, et, en conséquence, doivent répondre à quatre questions de base pour orienter la direction de la recherche [20]. Ces questions sont :

- **Comment sélectionner un point de départ ?** Sélectionner un point dans l'espace de sous-ensembles de caractéristiques par lequel commencer la recherche peut affecter la direction de cette dernière. Une des options serait de commencer sans caractéristiques et successivement ajouter des attributs. Dans ce cas, la recherche est dite « parcours en avant ». Inversement, la recherche peut commencer avec toutes les caractéristiques et les éliminer une par une successivement. Dans ce cas, la recherche est dite « parcours en arrière ». Comme solution intermédiaire, il s'agit de commencer quelque part, à un point quelconque et opérer un parcours en s'éloignant de ce point (en se déplaçant vers l'extérieur de ce point).
- **Comment organiser la recherche ?** Une recherche exhaustive du sous-espace de caractéristiques est prohibitive même pour un petit nombre initial de caractéristiques. En effet,

si N représente ce nombre, il faudrait considérer les 2^N sous-ensembles possibles. Les stratégies de recherche heuristiques sont plus pratiques et donnent de bons résultats, même si elles ne garantissent pas la découverte du sous-ensemble optimal.

- **Comment opérer l'évaluation d'un sous-ensemble ?** La façon d'évaluer les sous-ensembles des caractéristiques semble être le facteur le plus discriminant des algorithmes de sélection des caractéristiques dédiés à l'apprentissage automatique. Dans l'approche orientée filtrage (Filter methods), qui est indépendante de tout algorithme d'apprentissage, on élimine les caractéristiques non désirables dès le départ par une opération de filtrage. Ces algorithmes utilisent une recherche heuristique basée sur les caractéristiques générales des données pour évaluer les sous-ensembles fructueux de caractéristiques.

D'autres méthodes prennent en considération un algorithme d'apprentissage particulier pour la sélection des caractéristiques. Dans cette approche, dite « Symbiose » (Wrapper methods), il y a utilisation d'un algorithme d'induction avec une technique de ré-échantillonnage statistique telle que la validation croisée (cross-validation) qui permet d'estimer l'exactitude des différents sous-ensembles de caractéristiques.

De manière générale, les fonctions d'évaluation effectuent des mesures sur l'ensemble des solutions candidates générées par les procédures de recherche. Ces valeurs sont comparées aux valeurs précédentes, la meilleure valeur sera retenue. Il existe plusieurs types de fonctions d'évaluation telles que la distance (distance euclidienne) qui permet mesurer les capacités de discrimination, la mesure d'information qui permet d'estimer le gain d'une caractéristique, la mesure de dépendance qui permet de quantifier la corrélation des caractéristiques, la mesure de consistance de l'ensemble des caractéristiques et la mesure du taux d'erreur lors de la classification.

- **Quel critère d'arrêt utiliser ?** Un sélectionneur de caractéristique doit décider quand arrêter la recherche dans l'espace des sous-ensembles des caractéristiques. Sans l'utilisation de ce critère, la procédure recherche peut s'exécuter indéfiniment. Selon la stratégie d'évaluation, un sélectionneur de caractéristiques peut arrêter l'ajout ou la suppression des caractéristiques quand aucune des alternatives n'améliore le sous-ensemble courant. Il est possible aussi de continuer à réviser le sous-ensemble de caractéristiques aussi longtemps que la qualité ne se dégrade pas. Une autre possibilité, serait de continuer à produire des sous-ensembles de caractéristiques jusqu'à atteindre la fin de l'espace de recherche puis de sélectionner le meilleur. Le critère d'arrêt est souvent défini comme une combinaison de la procédure de recherche et du critère d'évaluation. Une heuristique, souvent utilisée, consiste à calculer pour les différents sous-ensembles de variables sélectionnées une estimation de l'erreur de généralisation par validation croisée. Le sous-ensemble de variables sélectionnées est celui qui minimise cette erreur de généralisation.

- **Procédure de validation**

Dash et Liu [19] proposent d'ajouter une quatrième composante à un algorithme de sélection de caractéristiques : une procédure de validation. Deux alternatives sont proposées en fonction de la nature des données utilisées lors de cette procédure : artificielles ou réelles. Généralement, une base de données synthétiques est construite dans le but de tester un concept ou une application particulière. De ce fait, les variables pertinentes sont connues et identifiées. La validation d'un algorithme sera alors directe puisqu'il suffit de vérifier si le sous-ensemble retenu contient bien les variables pertinentes. Dans le cas de données réelles, les

variables pertinentes ne sont généralement pas connues. La procédure consiste alors à évaluer la précision de la classification obtenue avec le sous ensemble de variables sélectionnées par l'intermédiaire d'un classifieur. Cette précision peut alors être comparée à celles d'autres approches ou à celle obtenue par des techniques classiques.

1.3.3. Panorama des méthodes supervisées de sélection de caractéristiques

Différentes méthodes supervisées de sélection de caractéristiques, en les classant en fonction de leur exhaustivité, de leur approche heuristique ou stochastique, et de leur utilisation de la corrélation ou de l'évaluation individuelle des caractéristiques. Ces méthodes sont utilisées:

1. Méthodes optimales :

- Méthodes exhaustives : Ces méthodes évaluent toutes les solutions possibles pour trouver la solution optimale. Elles utilisent des stratégies classiques de recherche, telles que le parcours en profondeur ou en largeur (Depth first, Breath first). Elles sont généralement basées sur des critères d'évaluation et peuvent être intensives en termes de temps de calcul. Cependant, elles garantissent une solution optimale.
- Algorithme Branch and Bound (B&B) : C'est une méthode non exhaustive qui fournit une solution optimale si le critère d'évaluation des sous-ensembles est monotone. Il utilise une mesure de distance pour évaluer les variables. FOCUS est une autre méthode similaire à B&B, mais utilise un critère d'évaluation basé sur la cohérence plutôt que sur la distance.

2. Méthodes non exhaustives :

- Procédures classiques de sélection de caractéristiques : Sequential Forward Selection (SFS) et Sequential Backward Selection (SBS) sont deux algorithmes couramment utilisés. Ils ajoutent ou suppriment une seule variable à la fois lors de chaque itération du processus de recherche. Ces méthodes séquentielles peuvent souffrir du Nesting Effect, où le sous-ensemble optimal de M caractéristiques ne contient pas nécessairement le sous-ensemble optimal de P caractéristiques ($P < M$).
- Stepwise : L'algorithme Plus-l Takeaway-r (PTA (l;r)) est une variante du Stepwise. Il utilise l'algorithme SFS l fois pour ajouter l variables, puis l'algorithme SBS r fois pour en supprimer r. Ce processus est répété jusqu'à ce que le critère d'arrêt soit atteint, généralement lorsque la performance du système n'augmente plus.

3. Méthodes basées sur l'évaluation individuelle des caractéristiques :

- Algorithmes instance-based : Ces méthodes évaluent individuellement chaque variable pour construire un sous-ensemble. L'algorithme RELIEF compare une instance avec sa classe correspondante pour plusieurs sous-ensembles de données afin de calculer la pertinence de chaque attribut. Il utilise un processus aléatoire d'échantillonnage et est souvent utilisé avec l'algorithme des K plus proches voisins.

4. Méthodes heuristiques :

- Hill climbing search : Cette méthode explore les changements locaux sur le sous-ensemble courant des caractéristiques. Elle recherche des solutions en effectuant des modifications locales pour améliorer les performances.
- Best first : C'est une stratégie de recherche qui autorise le retour arrière le long du chemin de recherche. Elle explore l'espace des solutions en faisant des changements locaux sur le sous-ensemble courant, mais peut revenir en arrière si le chemin exploré semble moins prometteur.
- Beam search : Cette méthode est une version limitée de la recherche exhaustive. Elle examine seulement un nombre limité de solutions au lieu de toutes les solutions possibles.

- Tabu search : C'est une méthode méta-heuristique utilisée pour résoudre des problèmes d'optimisation combinatoire. Elle diffère des techniques de recherche locale en permettant de passer à une nouvelle solution tout en gardant la mauvaise fonction objective dans l'espoir qu'elle ne tombe pas dans les solutions optimales locales. [19, 275]

5. Sélection de caractéristiques en ligne : Il est essentiel d'adopter des méthodes de sélection de caractéristiques qui sont adaptables et dynamiques. Les approches de sélection de caractéristiques en ligne (**online feature selection**) permettent de mettre à jour régulièrement la sélection de caractéristiques à mesure que de nouvelles données arrivent. De plus, l'utilisation de techniques d'apprentissage incrémental peut aider à intégrer de nouvelles caractéristiques et à ajuster les caractéristiques existantes de manière continue. [19a]

1.3.4. Approches de sélection de caractéristiques

Tous les algorithmes de la sélection des caractéristiques peuvent correspondre soit au modèle filtre (filter) ou au modèle symbiose (wrapper). L'approche *Wrapper* consiste à utiliser la précision en termes de classification et donc l'algorithme d'apprentissage ou d'induction du système. Pour la seconde approche, *Filter*, la sélection est réalisée sans tenir compte de son influence sur la performance du système. Dans le cas d'un système de reconnaissance, elle peut être assimilée à une étape de pré-traitement. Elle est bien sûr plus rapide que l'approche *Wrapper* en termes de génération de résultats, cependant cette dernière à l'avantage de fournir généralement des résultats plus pertinents pour la classification. [19, 21, 22]

Selon Jensen [15], la littérature fait mention, généralement, de cinq sous catégories d'approches pouvant, aussi bien se trouver dans l'approche Wrapper que dans l'approche Filter. Un grand nombre d'algorithmes ont été développés en se basant sur l'une de ces cinq approches de génération de sous-ensembles (voir la Figure 1.5):

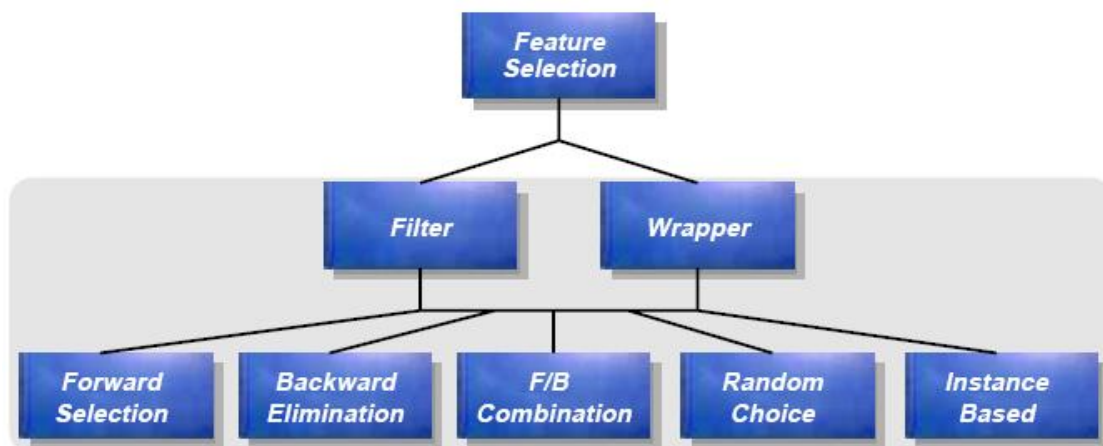


Figure1.5: Approches de sélection de caractéristiques. [15]

- *Forward* : cette approche part d'un ensemble de variables vide auquel, à chaque itération sont ajoutées une ou plusieurs variables. Elle est également appelée approche ascendante.
- *Backward* : c'est l'approche inverse; l'ensemble total de variables est considéré au départ de la procédure itérative, chaque itération permet d'en supprimer. Une autre appellation est l'approche descendante.

- *Stepwise*: cette approche est une version hybride des deux précédentes car elle consiste à ajouter ou retirer successivement des variables à l'ensemble déjà sélectionné.
- *Random choice*: Cette catégorie est la plus récente dans le domaine de la sélection d'attributs. Elle commence avec un sous ensemble aléatoire, les variables sont itérativement ajoutés ou enlevés ou aléatoirement générées.
- *Instance Based*: dans cette approche, des poids sont assignés individuellement aux variables qui sont classées en se basant sur leur pertinence par rapport au concept cible.

1.3.4.1. Approche Symbiose

L'approche symbiose (wrapper), popularisée par Kohavi et John [23] offre une manière simple et puissante pour traiter le problème de sélection de variables. Les méthodes correspondant à cette approche consistent à utiliser la précision en termes de classification et donc l'algorithme d'apprentissage ou d'induction du système. Elles se servent de l'algorithme d'induction comme une boîte noire: l'apprentissage est effectué avec les variables sélectionnées et les performances sont estimées à partir de l'erreur de généralisation (voir Figure 1.6).

Quel que soit le type de l'apprentissage automatique choisi, ce dernier est considéré comme une boîte noire parfaite et la méthode se prête à l'utilisation des progiciels (packages) disponibles de l'apprentissage automatique. Dans sa formulation la plus générale, l'approche wrapper consiste à utiliser la performance de la prédiction d'un apprentissage automatique donné pour estimer l'utilité relative de sous-ensembles de variables [24]. En pratique, on doit:

- Définir un espace de recherche des sous-ensembles de variables;
- Effectuer l'estimation de la performance de classification (ou de prédiction) d'un apprentissage automatique permettant de guider la recherche et l'arrêter;
- Définir le pouvoir discriminant (ou prédictif) utilisé.

Ces trois questions définissent les éléments essentiels d'une procédure de sélection de variables. En effet, le problème de la sélection de variables consiste à identifier les variables permettant une meilleure séparation entre les différentes classes dans le cas d'une classification et une meilleure qualité de prédiction dans le cas d'une régression. On parle alors de "pouvoir discriminant" dans le premier cas et de "pouvoir prédictif" dans le deuxième cas, pour désigner la pertinence d'une variable.

Les méthodes symbiose -dite aussi enveloppantes- sont souvent critiquées parce qu'elles exigent des quantités massives de calcul, cependant, leur utilisation ne signifie pas nécessairement sacrifier la performance de classification ou de prédiction. En effet, les stratégies de la recherche grossières peuvent limiter le problème de sur-apprentissage (overfitting) [24] défini comme l'incapacité de généralisation du modèle d'apprentissage. Schématiquement, l'erreur de généralisation de l'ensemble de validation diminue au fur et à mesure puis se met à croître bien que celle de l'apprentissage continue à diminuer.

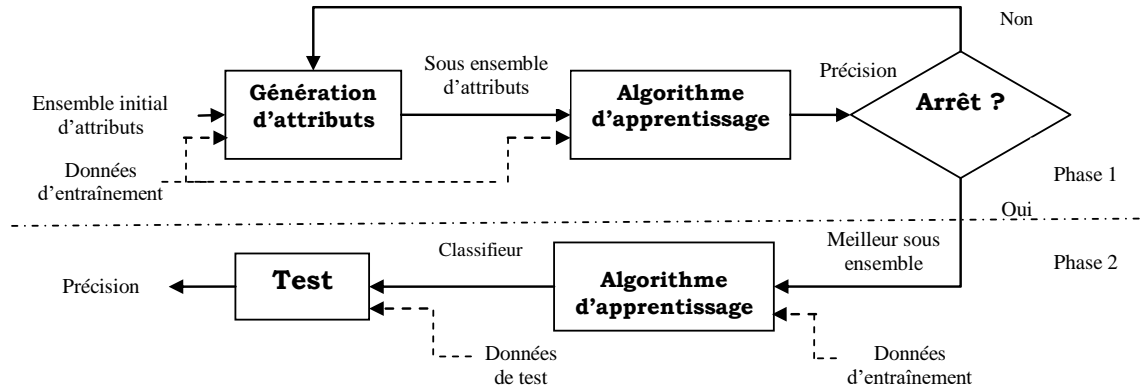


Figure 1.6 : Approche Symbiose ou Wrapper [25]

1.3.4.2. Approche Filtre

L'approche Filtre ou Filter relie généralement les attributs des données d'entraînement à travers une mesure d'évaluation sans faire intervenir les algorithmes d'apprentissage ou d'induction du système (voir Figure 1.7). Les méthodes qui se rattachent à l'approche Filtre peuvent être basées sur le poids des caractéristiques (weighting ou ranking methods), qui pondèrent les différentes caractéristiques puis les ordonnent en se basant sur leurs pertinences individuelles. Ces méthodes peuvent aussi effectuer une sélection à travers des sous-ensembles candidats de caractéristiques en utilisant une certaine mesure d'évaluation. Un sous-ensemble optimal (ou sous optimal) est sélectionné dès que le processus de recherche s'arrête.

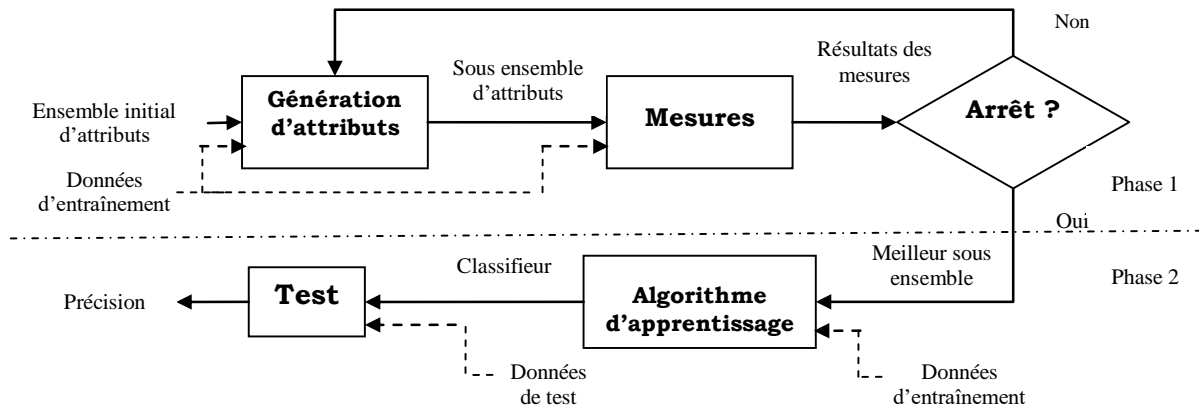


Figure 1.7: Approche Filtre [25]

Quelques mesures d'évaluation existantes sont très efficaces dans la suppression des attributs non pertinents et redondants. Les différentes stratégies de recherche, à savoir, la recherche exhaustive, heuristique, et aléatoire, sont combinées avec ces mesures d'évaluation.

Mesures de distance

Dans la littérature elles sont également appelées mesure de séparabilité, de divergence ou de discrimination. Dans le cas d'un problème à deux classes, une variable X est préférée à Y , si X introduit une plus grande différence dans la distribution de probabilité conditionnelle des classes que Y . Un exemple de cette mesure est la distance euclidienne.

Mesures d'information

Les indicateurs de cette catégorie sont basés sur le gain en termes d'information que peut apporter une variable. Il est défini, pour une variable X , comme la différence entre l'incertitude *a priori* et celle *a posteriori*, c'est-à-dire avant et après sélection de la variable X . Entre deux variables alternatives, la variable sélectionnée est, bien sûr, celle qui apporte le plus d'information. L'entropie ou l'information mutuelle sont des critères appartenant à cette catégorie.

Mesures de dépendance ou de corrélation

Elles permettent de mesurer la capacité d'une variable à prédire la valeur d'une autre. Ce critère peut être utilisé afin de mesurer la corrélation entre une variable et une classe de modélisation. Si la corrélation d'une variable X avec une classe C est plus importante que la corrélation de la variable Y avec C alors X est choisie. Une variante peut également être utilisée ; elle consiste à mesurer la dépendance entre primitives et ainsi obtenir un indicateur de redondance entre les variables.

Mesures de consistance

Ce sont des mesures apparues plus récemment dans la littérature, elles sont différentes des autres dans le sens où elles se basent fortement sur le corpus d'apprentissage et qu'elles utilisent le *Min-Features Bias*. Cette procédure permet de sélectionner le plus petit sous-ensemble de variables basé sur des hypothèses de consistance forte. Un algorithme utilisant cette approche est *Focus*, La mesure de consistance essaie de trouver un nombre minimal de caractéristiques séparant les classes aussi consistantes que l'ensemble complet de caractéristiques. L'inconsistance qualifie les cas où deux exemples qui ont les mêmes valeurs de caractéristique sont étiquetés par deux classes différentes.[19, 20, 16, 26]

1.4. Algorithmes de classification et de reconnaissance de formes

1.4.1. La reconnaissance de formes : définition et approches

La reconnaissance des formes est un sujet interdisciplinaire qui s'est considérablement développée dans les années 1960, couvrant les développements dans les domaines de diverses disciplines techniques et scientifiques telles que des statistiques, de l'ingénierie, de l'intelligence artificielle, de l'informatique, de la psychologie et de la physiologie, la biologie, la médecine, le marketing, la vision par ordinateur [12]. Le problème de la reconnaissance des formes consiste à classer ou catégoriser des entités en fonction de leur similitude avec des modèles préétablis. On distingue quatre approches principales dans ce domaine (Voir Figure 1.8):

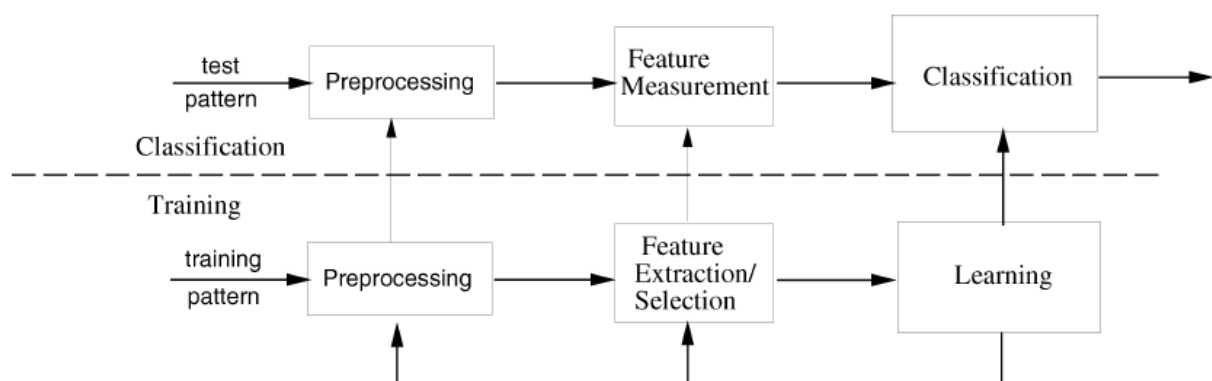


Figure 1.8. Modèle de reconnaissance de formes

Correspondance de modèle (template matching)

Cette approche consiste à comparer un modèle préétabli à une entité à reconnaître en prenant en compte les changements de pose (translation et rotation) et d'échelle. Cette méthode est exigeante en termes de calcul, mais elle peut être utilisée pour la reconnaissance de formes rigides. [1]

Approche statistique

Dans cette approche, chaque modèle est représenté par des caractéristiques ou mesures, et l'objectif est de choisir des caractéristiques qui permettent de séparer efficacement les modèles de différentes catégories dans un espace de caractéristiques à plusieurs dimensions. Les limites de décision sont établies en fonction des distributions de probabilité des motifs appartenant à chaque classe [1].

Approche syntaxique

Dans cette perspective hiérarchique, un modèle complexe est vu comme une combinaison de sous-modèles simples, qui sont à leur tour constitués de sous-modèles élémentaires. Les sous-modèles élémentaires sont appelés primitives, et le modèle complexe est représenté en termes de relations entre ces primitives. Une analogie est établie entre cette structure hiérarchique et la syntaxe d'un langage, où les motifs sont considérés comme des phrases générées par une grammaire [1].

La correspondance syntaxique ou structurelle

Cette approche vise à non seulement classifier les motifs, mais aussi à décrire comment un modèle donné est construit à partir de primitives. Elle est particulièrement adaptée lorsque les motifs présentent une structure définie, telle que les formes d'onde ECG, les images texturées et l'analyse de la forme des contours. Cependant, la mise en œuvre de cette approche peut être complexe, car elle nécessite la segmentation des motifs (pour détecter les primitives) et l'inférence de la grammaire à partir des données d'apprentissage [1].

1.4.2. La classification : définition et algorithmes

La description, la classification et le regroupement de modèles sont des problèmes importants de la reconnaissance de forme, d'après Watanabe dans [1], le modèle est défini comme l'opposé d'un chaos; c'est une entité, vaguement définie, qui pourrait recevoir un nom. Par exemple, un modèle pourrait être une image d'empreinte digitale, un mot cursif manuscrit, un visage humain ou un signal vocal. La reconnaissance/classification d'un modèle consiste en l'une des deux tâches suivantes [1] :

1) Classification supervisée : le problème est de reconnaître le modèle, qui est identifié comme un membre d'une classe prédéfinie, un expert (ou oracle) est employé pour étiqueter correctement des exemples. L'apprenant doit alors trouver ou approximer la fonction qui permet d'affecter la bonne étiquette à ces exemples. L'analyse discriminante linéaire ou les SVM sont des exemples typiques.

2) Classification non supervisée: dans lequel le modèle est assigné à une classe jusqu'ici inconnue. Aucun expert n'est disponible. L'algorithme doit découvrir par lui-même la structure des données. Le clustering et les modèles de mélanges de gaussiennes, C-means, des classificateurs bayésiens non supervisés, des classificateurs hiérarchiques de regroupement et de division, des classificateurs théoriques des graphes sont des algorithmes d'apprentissage non supervisés. [27]

1.4.2.1. Apprentissage basée sur les instances : K Plus proche Voisin

L'algorithme NN (Nearest Neighbor) a été introduit pour la première fois par J.G. Skellam [28]. la classification k-plus proche voisin (kNN) est une approche plus sophistiquée, trouve un groupe de k objets dans l'ensemble d'apprentissage qui sont les plus proches de l'objet de test, et fonde l'attribution d'une étiquette sur la prédominance d'une classe particulière dans ce quartier. Il existe plusieurs éléments clés de cette approche : (i) l'ensemble d'objets étiquetés à utiliser pour évaluer la classe d'un objet de test, (ii) une métrique de distance ou de similarité qui peut être utilisée pour calculer la proximité des objets, (iii) la valeur de k, le nombre de voisins les plus proches, et (iv) la méthode utilisée pour déterminer la classe de l'objet cible en fonction des classes et des distances des k voisins les plus proches. Plus généralement, kNN est un cas particulier d'apprentissage par instance. Cela inclut le raisonnement à base de cas, qui traite des données symboliques.

1.4.2.2. Approches arborescentes de classification

- L'apprentissage par arbre de décision est l'une des approches les plus largement utilisées et pratiques pour l'inférence inductive. Toutes les méthodes d'induction d'arbre commencent par un nœud racine qui représente l'ensemble de données et divisent récursivement les données en sous-ensembles plus petits en testant un attribut donné à chaque nœud. Les sous-arbres désignent les partitions de l'ensemble de données d'origine qui satisfont aux tests de valeur d'attribut spécifiés. Ce processus se poursuit généralement jusqu'à ce que les sous-ensembles soient « purs », c'est-à-dire que toutes les instances du sous-ensemble appartiennent à la même classe, moment auquel la croissance de l'arbre est terminée.
- La plupart des algorithmes qui ont été développés pour apprendre les arbres de décision sont des variantes d'un algorithme de base qui utilise une recherche descendante et gourmande dans l'espace des arbres de décision possibles. Cette approche est illustrée par l'algorithme ID3 [29] et son successeur C4.5. Dans le domaine de sélection de caractéristiques, l'arbre de décision intègre la sélection de caractéristiques lors de la construction de modèle. Il existe divers apprenants d'arbres de décision descendants tels que
- ID3 "Decision Tree Induction" est la désignation d'un ancien algorithme de construction d'un arbre d'induction de graphes [19, 29], dans lequel le critère de sélection de variables est *le gain informationnel*. ID3, apprend les arbres de décision en les construisant de haut en bas, en commençant par la question « quel attribut doit être testé à la racine de l'arbre ? » Pour répondre à cette question, chaque attribut d'instance est évalué à l'aide d'un test statistique pour déterminer dans quelle mesure elle seule classe les exemples d'apprentissage.
- Le meilleur attribut est sélectionné et utilisé comme test au nœud racine de l'arbre. Un descendant du nœud racine est ensuite créé pour chaque valeur possible de cet attribut, et les exemples d'apprentissage sont triés vers le nœud descendant approprié (c'est-à-dire vers le bas de la branche correspondant à la valeur de l'exemple pour cet attribut). [31]
- C4.5 a été développé par Quinlan, et est le descendant de ID3. L'algorithme C4.5 utilise un critère inspiré de la théorie de l'information tels que le gain (réduction de l'entropie de la distribution de classe due à l'application d'un test), et plus élaboré basé sur l'entropie $Entropy(t) = -\sum_{i=0}^{c-1} p(i|t) \log_2 p(i|t)$ qui est appelé *le rapport de gain d'information*, le nœud dans l'arbre est créé les sous-arbres sont générés

progressivement en fractionnant les valeurs des attributs dans chaque nœud de manière répétée et récursive. C4.5 permet d'élaguer les arbres et de simplifier les règles pour gérer les données manquantes, etc...[30].

- **CART** : CART est un algorithme de régression qui estime les fonctions par un mappage des variables d'entrée sur des sorties continues¹. Ces *approches sont basées sur la régression* [32] en estimant des fonctions qui mappent les variables d'entrée sur des sorties continues. CART est l'un des algorithmes qui entre dans cette catégorie. est proposé par un cadre méthodologique général pour générer une induction binaire sérieuse. Il utilise *l'index de Gini* pour choisir l'attribut qui partitionne les données. Chaque nœud de l'arborescence est optimisé et conçu pour fractionner les données afin de former des grappes où une prédiction précise qui peut être effectuée avec des modèles plus simples. Ceci permet d'*élaguer la complexité de coût* où une série d'arbres sont implantés, chacun obtenu à partir du précédent en remplaçant un ou plusieurs sous-arbres par une feuille. Le dernier arbre de la série ne comprend qu'une seule feuille qui prédit une classe spécifique.[30, 32]
- **Forêt aléatoire (Random forest)**. RF est une méthode de classification qui affecte certains objets dans l'une des catégories données et fournit l'importance des caractéristiques. Une forêt aléatoire est un classificateur consistant en une collection de classificateurs arborescents, chacun d'eux étant construit par des instances avec des entités échantillonnées de manière aléatoire. Les prévisions sont faites à la majorité des voix des arbres. Le classificateur de la Forêt aléatoire est composé à son tour d'une collection de sous-classificateurs arborescents. Chacun de ceux-ci est construit par des instances avec des caractéristiques échantillonnées au hasard. Les prédictions sont faites par un vote majoritaire des arbres. Une étude sur les RFs est présentée par Breiman dans son article [33].
- **Le PBT (Probabilistic Boosting Tree)** est un cadre d'apprentissage basé sur un arbre binaire dans lequel chaque nœud est constitué d'un certain nombre de classificateurs faibles combinés dans un seul classificateur puissant. Dans la phase d'apprentissage, l'arbre est construit de manière récursive. L'ensemble des données est divisé en deux ensembles, basés sur la décision du nœud racine. Chacun de ces ensembles est ensuite utilisé pour former les sous-arbres gauche et droit de manière récursive. Dans la phase de test, la probabilité conditionnelle, $p(y|x)$, qu'un échantillon donné x appartient à la classe positive ou négative, est calculée à chaque nœud de l'arbre sur la base du classifieur appris, qui guide la propagation de probabilité dans ses sous-arbres. La profondeur maximale de l'arbre est définie par l'utilisateur en fonction de la complexité de l'ensemble de données. Pour plus de détails sur le PBT, nous renvoyons les lecteurs à [34].

1.4.2.3. Approches basées sur les règles d'association

Cette section se concentre sur les approches basées sur les règles d'association dans le contexte de l'apprentissage automatique, en mettant ses avantages en termes de compréhension humaine et en abordant les défis liés à l'adaptation à des ensembles de données complexes:

¹Contrairement à la *classification* dont les variables prédites sont discrètes.

- 1. L'induction de règles** : L'induction de règles est présentée comme une technique d'apprentissage automatique couramment utilisée. L'objectif principal de cette technique est d'extraire un ensemble de règles à partir de données. Ces règles doivent capturer les connaissances générales contenues dans les données tout en restant aussi simples et concises que possible.
- 2. Phase d'apprentissage** : Pendant la phase d'apprentissage, des règles sont générées à partir d'un ensemble d'apprentissage, en se basant sur les caractéristiques et les étiquettes de classe des échantillons d'apprentissage. Cela signifie que les règles sont déduites en fonction des informations contenues dans les données d'apprentissage.
- 3. Application des règles** : Une fois que ces règles sont générées, elles peuvent être appliquées pendant la phase de classification pour étiqueter de nouvelles données de test non vues. Cela permet de généraliser les connaissances extraites des données d'apprentissage pour prendre des décisions sur de nouvelles données.
- 4. Avantages de l'induction de règles** : Plusieurs avantages de l'induction de règles. Tout d'abord, les règles sont faciles à comprendre pour les êtres humains car elles sont généralement formulées sous la forme de règles simples "si-alors". Deuxièmement, il est mentionné que les systèmes basés sur l'apprentissage de règles peuvent surpasser les arbres de décision sur de nombreux problèmes.
- 5. Inconvénients de l'induction de règles** : Néanmoins, l'induction de règles présente des limites. Il est noté que cette méthode peut ne pas bien s'adapter aux ensembles de données volumineux, en particulier lorsqu'ils sont bruités. Cela peut entraîner une complexité excessive dans la génération de règles.
- 6. Induire des ensembles de règles** : Une caractéristique distinctive de l'algorithme C4.5 est la capacité à élaguer des règles à partir de l'arbre de décision induit. Cela signifie que l'arbre peut être modélisé comme une combinaison de règles conditionnelles, où chaque règle correspond à un chemin de la racine à une feuille de l'arbre.
- 7. Élagage des règles** : L'élagage des règles est effectué pour simplifier les ensembles de règles. Il est possible d'utiliser des méthodes d'élagage pessimistes de C4.5. Ces méthodes permettent de sélectionner un sous-ensemble de règles simplifiées pour chaque classe en évaluant le coût de la théorie impliquée dans l'encodage des règles.[35-42]

1.4.2.4. Approches bayésiennes

Ces approches sont incluses dans le cadre de raisonnement inférentiel de l'intelligence artificielle, et modélisent le problème d'apprentissage à l'aide d'un cadre bayésien probabiliste. En apprentissage automatique, les approches de Bayes sont des modèles probabilistes qui s'imposent comme un lien entre les statistiques et le domaine de l'apprentissage automatique. Les méthodes bayésiennes calculent la probabilité a posteriori $P[C_i|(x_1, \dots, x_n)]$ d'un objet (x_1, \dots, x_n) à l'aide de la formulation de Bayes, suivie d'une prise de décision selon la règle de Bayes, qui indique que l'objet est inclus dans la classe C_i dont la probabilité a posteriori a la plus grande valeur. Le cœur de la méthode naïve de Bayes réside dans l'estimation de la probabilité à priori $p((x_1, x_2, \dots, x_n)|C_i)$. La méthode naïve de Bayes suppose que les composants de x sont indépendants au sein de chaque classe, c à d chacune des distributions marginales est univariées, et la probabilité $p(x_j|C_i)$ est estimée séparément de sorte que

$p(x_i|C_i) = \prod_{j=1}^p p(x_i|C_i)$ [3, 43]. C'est une hypothèse forte et naïve dite la propriété de *l'Indépendance conditionnelle des attributs* [4], ce qui simplifie le calcul de la probabilité a posteriori.

1.4.2.5. Approches à Noyaux pour la classification : SVM

L'astuce du noyau est une autre technique couramment utilisée pour résoudre des problèmes linéairement inséparables. Le problème est de définir une fonction de noyau appropriée basée sur le produit interne entre les données, comme une transformation non linéaire des données de l'espace d'entrée vers un espace de caractéristiques avec une dimension plus élevée (voire infinie) afin de rendre les problèmes linéairement séparables.

La machine à vecteurs de support (SVM), initialement proposée par Vladimir Vapnik et Corinna Cortes [44], est une technique d'apprentissage supervisé basée sur la théorie de l'apprentissage statistique qui peut être appliquée à la classification, à la régression et à la reconnaissance de formes. Un SVM est une sorte de classificateur binaire qui prend un ensemble de données d'entrée, puis classe chaque entrée en deux classes ou catégories possibles. L'idée est de mapper l'espace d'entrée à n dimensions dans un espace de caractéristiques de dimension supérieure, puis le nouvel espace de caractéristiques est classé en construisant un classificateur linéaire. Dans SVM, un point de données est considéré comme un vecteur de dimension p et SVM les séparera par un hyperplan de dimension $(p-1)$.

Pour une tâche d'apprentissage linéairement séparable à deux classes, le but de SVC est de trouver un hyperplan qui peut séparer deux classes d'échantillons donnés avec une marge maximale qui s'est avérée capable d'offrir la meilleure capacité de généralisation. La capacité de généralisation fait référence au fait qu'un classificateur a non seulement de bonnes performances de classification (par exemple, la précision) sur les données d'entraînement, mais garantit également une précision prédictive élevée pour les données futures à partir de la même distribution que les données d'entraînement.

Intuitivement, une marge peut être définie comme la quantité d'espace, ou de séparation, entre les deux classes telles que définies par un hyperplan. Géométriquement, la marge correspond à la distance la plus courte entre les points de données les plus proches de n'importe quel point de l'hyperplan.

Dans de nombreux cas, les points de données ne sont pas séparables linéairement, dans ce cas, les données d'entrée peuvent être transformées à l'aide d'un mappage non linéaire (ϕ) dans un autre espace de dimension [45]. Dans cette nouvelle cartographie, une frontière linéaire peut être trouvée.

Le choix du type de fonction noyau et de son degré de complexité diffère d'un problème de classification/prédiction à un autre, une déviation de la fonction noyau affecte les résultats générés par les classificateurs. Son paramètre qui ne reflète pas et ne correspond pas à la solution optimale entraîne une baisse relative des performances, et est choisi a priori par l'utilisateur, qui détermine le type de classifieur. D'autres paramètres (nombre d'unités cachées, poids, seuil) sont retrouvés lors de l'apprentissage. En revanche, le degré de pertinence de la caractéristique diffère d'une caractéristique à l'autre ; certaines caractéristiques sont plus fortes et plus descriptives que d'autres. La sélection de caractéristiques est conforme au classement de ces caractéristiques en fonction du degré de ces points forts et de sa pertinence dans le système de classification.[31]

1) Noyaux SVM

SVM peut effectuer efficacement une classification non linéaire à l'aide de l'astuce du noyau en mappant leur entrée dans des espaces de caractéristiques de grande dimension. Les fonctions à noyau les plus fréquemment utilisées sont la fonction de base radiale (RBF) linéaire, sigmoïde et polynomiale. (Voir le Tableau 1.1)

Tableau 1. Les fonctions kernel classiques.

Classification: $f(x) = \text{sgn}\left(\sum_i \lambda_i k(x, x_i) + b\right)$
 Weights: $\lambda_1, \lambda_2, \lambda_3, \lambda_4$

Comparaison		
<i>Linear function</i>	$k(x, x_i) = x \cdot x_i + c$	Le noyau linéaire n'a qu'un seul paramètre accordable qui est c. Le noyau linéaire fonctionne très bien et très rapidement sur des ensembles de données linéairement séparables, malheureusement la plupart des problèmes du monde réel ne sont pas linéairement séparables.
Classical architectures: <i>polynomial classifiers of degree d</i>	$k(x, x_i) = (x \cdot x_i)^d$	Comparé aux autres noyaux SVM, le noyau polynomial a plus de paramètres qui doivent être optimisés. Outre C et γ (gamma), il a au moins 2 paramètres plus importants : le degré polynomial d et le coefficient de degré r. Le paramètre d doit être défini avec soin, si la valeur de d est trop grande, les valeurs du noyau peuvent aller à l'infini ou à zéro.
<i>Radial basis function classifiers with Gaussian kernel of width c > 0,</i>	$k(x, x_i) = \exp(-\ x - x_i\ ^2 / c)$	Le RBF est généralement le plus fréquemment appliqué, car il peut classer des données non linéaires séparables, contrairement à une fonction de noyau linéaire. De plus, le RBF a moins de paramètres à définir qu'un noyau polynomial. Le paramètre ajustable joue un rôle majeur dans les performances du noyau.
<i>Neural networks with tanh activation function,</i>	$k(x, x_i) = \tanh(k(x, x_i) + c)$	Une fonction de noyau sigmoïde est équivalente à un réseau de neurones de perception à deux couches. Le noyau sigmoïde provient du domaine des réseaux de neurones où la fonction sigmoïde est souvent utilisée comme fonction d'activation pour les neurones

		artificiels. Il y a deux paramètres ajustables dans le noyau sigmoïde : c et γ
Support vectors:	x_1, \dots, x_4	-
Input vector:	x	-

2) La modélisation (Classificateur/prédicteur)

Le résultat de l'apprentissage est aussi nommé modèle d'identification de classifieur, reconnaissance de formes, prédicteur, etc. La pertinence d'un classifieur peut être appréciée de différentes manières. On s'accorde pourtant à reconnaître l'importance de certains critères globaux que nous précisons dans ce paragraphe ; certaines ont été prises de [31] :

- La précision :

C'est certainement le critère le plus important, en tous cas le plus souvent cité de la formation encadrée. Il montre la capacité intrinsèque du classifieur à reconnaître la variable à prédire dans la population, lorsque la précision est parfaite, c.-à-d. Zéro erreur, on peut penser qu'on a trouvé une expression du concept cible à apprendre. Notons qu'une expression mise au point sur un fichier d'apprentissage ne reflète pas nécessairement une bonne qualité de prédiction.

- La simplicité :

Elle est souvent liée au nombre de paramètres qui interviennent dans le modèle. Le modèle est d'autant plus intelligible qu'il a moins de paramètres.

- La compréhensibilité :

Lui dans [31] affirme que les connaissances extraites doivent être sémantiquement et structurellement similaires à celles que peut produire un expert humain. Même si ce point de vue nous semble trop excessif, ce n'en est pas moins à défaut que nous avons préféré, dans la mesure du possible, garder et utiliser des prédicateurs intelligibles.

- Intelligibilité :

La solution d'ingénierie apportée par le KDD (Knowledge Discovery in Data) est liée à l'intelligibilité du modèle construit automatiquement. Un système de fouille de données produit des connaissances pouvant se présenter sous des formes variées selon la méthode d'apprentissage utilisée : des arbres de décision, des règles d'associations, etc.

1.4.2.6. Approches d'ensemble

Cette section offre une introduction complète aux approches d'ensemble en apprentissage automatique, en mettant en évidence les principes de base, les avantages de la combinaison de classifieurs, et en décrivant en détail les méthodes de Bagging et de Boosting, notamment AdaBoost. Ces techniques sont largement utilisées dans le domaine de l'apprentissage automatique pour améliorer la performance des modèles. Un classificateur d'ensemble, est une méthode qui combine plusieurs classificateurs pour améliorer la performance de classification globale. L'idée est de tirer parti des forces de modèles de base plus simples pour résoudre un problème complexe.

1. Principe de "diviser pour régner" : Le principe sous-jacent à l'apprentissage d'ensemble consiste à décomposer un problème complexe en plusieurs sous-problèmes plus simples. Cette approche permet de mieux comprendre et résoudre chaque sous-problème individuellement, puis de combiner les résultats pour obtenir une meilleure performance globale.

2. Approches de classification composées d'un classifieur unique : Les approches de classification composées d'un seul classifieur ont été explorées précédemment, mais il suggère que la combinaison de plusieurs classifieurs peut offrir des améliorations en termes de flexibilité, de performances et de stabilité. Il fait référence à des études qui ont montré les avantages de cette approche.

3. Différentes approches de combinaison : Il existe plusieurs principes de combinaison, notamment les approches de combinaison séquentielle et parallèle. Il explique que dans la combinaison séquentielle, l'ordre des classifieurs est essentiel, tandis que le boosting est un exemple de cette approche. Le boosting est particulièrement intéressant car il transforme des classifieurs faibles en un classifieur fort. [46]

4. Bagging : Le concept de Bagging (Bootstrap Aggregating), consiste à construire plusieurs jeux de données bootstrap et à apprendre un ensemble de classifieurs individuels, puis à combiner leurs prédictions. Le bagging est principalement utilisé avec des classifieurs tels que les arbres de décision et est efficace pour stabiliser les classifieurs instables. [46]

5. Boosting : Boosting vise à augmenter les performances d'un ensemble de classificateurs faibles en un classificateur fort. Chaque classifieur est appris à partir des ensembles d'apprentissage basés sur les performances du précédent [47]. AdaBoost est cité comme une méthode de boosting influente et efficace. [9, 10]

6. Applications d'AdaBoost : Parmi les méthodes de Boosting, AdaBoost, a été appliqué avec succès à divers domaines en raison de leur solide base théorique, de leurs prédictions précises et de leur simplicité. [9, 10]

1.4.2.7. Algorithmes évolutionnaires

Cette section fournit une introduction complète aux algorithmes évolutionnaires, en mettant les principes fondamentaux, les différentes approches, en se concentrant sur les algorithmes génétiques et en expliquant leur utilisation dans une variété de domaines. Les EA sont présentés comme des méthodes puissantes pour résoudre des problèmes complexes d'optimisation en explorant efficacement les espaces de recherche. Les algorithmes évolutionnaires (EA) sont couramment utilisés pour résoudre des problèmes difficiles, mais qu'ils diffèrent des approches de recherche brute ou par grille. Ils utilisent des mécanismes inspirés de l'évolution biologique pour chercher la meilleure solution dans un espace de solutions.

1. Principes de base des EA : Les EA utilisent des mécanismes tels que la reproduction, la mutation, la recombinaison et la sélection, inspirés de l'évolution biologique. Chaque individu représente une solution potentielle au problème et est évalué par une fonction de fitness. Les meilleurs individus ont une probabilité plus élevée d'être sélectionnés pour la reproduction, ce qui conduit à l'évolution de la population vers de meilleures solutions. [48, 49]

2. Différentes approches des EA : Trois approches principales des EA : les algorithmes génétiques (GA), la programmation évolutive (EP) et les stratégies évolutives (ES). Chacune de ces approches est complètement différentes en termes de représentation des solutions et d'opérateurs génétiques. [49]

3. Applications des EA : Les EA sont utilisés pour résoudre une gamme variée de problèmes d'optimisation dans différents domaines, tels que l'ingénierie, la finance, la biologie et

l'intelligence artificielle. Leur flexibilité et leur capacité à explorer efficacement des espaces de recherche complexes sont mises en avant. [50, 51]

- 4. Représentation des solutions** : Les solutions candidates sont codées dans des chromosomes contenant des gènes. Il mentionne l'importance de choisir une représentation de génotype appropriée pour réduire le temps de traitement des EA.
- 5. Détails sur les algorithmes génétiques (AG)** : Les algorithmes génétiques sont basés sur le concept de l'évolution darwinienne. La description du processus d'AG y compris la sélection, le croisement et la mutation, utilisés en divers domaines.
- 6. Comparaison avec les méthodes d'optimisation standard** : Deux principales différences entre les algorithmes génétiques et les méthodes d'optimisation standard : la génération d'une population à chaque itération et l'utilisation de générateurs de nombres aléatoires pour la sélection.

1.4.2.8. Optimisation de l'essaim de particules (PSO)

PSO comme une technique de calcul évolutive développée par Kennedy et Eberhart en 1995, et s'inspire du comportement des oiseaux en vol pour résoudre des problèmes où la destination n'est pas entièrement connue. Cette section fournit une explication complète des principes de base de l'optimisation par essaim de particules (PSO). PSO utilise un essaim de particules pour explorer l'espace des solutions, les particules sont mises à jour à chaque itération en fonction de leurs meilleures expériences personnelles et globales, PSO peut se terminer selon différents critères prédéfinis. Cette méthode est souvent utilisée pour résoudre des problèmes d'optimisation dans divers domaines.

- 1. Principes de base de PSO** : PSO fonctionne en utilisant une population d'individus appelée un essaim, composé de particules. Ces particules sont mises à jour itération après itération dans le but de découvrir la solution optimale. Chaque particule ajuste sa direction de recherche en fonction de deux facteurs : sa propre meilleure expérience précédente (pbest) et la meilleure expérience globale de tout l'essaim (gbest).
- 2. Représentation des solutions** : Chaque particule dans l'essaim représente une solution potentielle au problème. Au départ, les particules ont des positions et des vitesses initiales aléatoires. Les équations sont données pour montrer comment la position et la vitesse de chaque particule sont mises à jour à chaque itération. Des facteurs tels que le poids inertiel (w), les facteurs d'apprentissage ($c1$ et $c2$), et des paramètres de pondération ($r1$ et $r2$) influencent ces mises à jour.
- 3. Mise à jour de la fitness** : Pour chaque itération, la fitness de chaque particule est calculée. Ensuite, les meilleures expériences personnelles et globales sont mises à jour en utilisant des équations spécifiques. Ces mises à jour permettent aux particules de converger vers de meilleures solutions au fil du temps.
- 4. Critère de terminaison** : PSO peut s'arrêter lorsque certains critères sont atteints, tels qu'une fitness est suffisamment bonne, un nombre prédéfini de générations ou un facteur de convergence basé sur un seuil de changement de population minimum. [52-56]

1.5. Conclusion

En résumé, l'étape d'inférence consiste à utiliser un modèle de classification pour prédire les étiquettes de classe des nouvelles données, tandis que l'étape de décision consiste à prendre une décision ou une action en se basant sur ces prédictions de classe. Ensemble, ces deux

étapes sont essentielles pour résoudre efficacement les problèmes de classification et utiliser les résultats de la modélisation pour prendre des décisions éclairées dans divers domaines d'application.

La conclusion de ce chapitre réside dans la présentation de plusieurs concepts et approches liés à l'apprentissage automatique, à l'optimisation, à l'estimation de densité, à la sélection de caractéristiques à la classification, et à d'autres domaines connexes. Ces approches et concepts couvrent un large éventail de sujets en statistiques, en traitement de données et en apprentissage automatique. En combinant ces approches, il est possible de développer des modèles précis et fiables pour la classification et la prédiction, ce qui peut conduire à des améliorations significatives dans les soins de santé et la prise de décision clinique. Ce chapitre souligne l'importance de choisir la méthode appropriée en fonction du problème spécifique et des données disponibles, ainsi que la nécessité de comprendre les principes sous-jacents pour prendre des décisions éclairées dans le domaine de l'apprentissage automatique et de l'analyse de données.

CHAPITRE 2

Reconnaissance Des images médicales et Sélection de Caractéristiques

2.1. Introduction

Les images médicales sont souvent détériorées par le bruit en raison de diverses sources d'interférences et d'autres phénomènes qui affectent les processus de mesure dans les systèmes d'imagerie et d'acquisition de données. La nature du système physiologique étudié et les procédures utilisées en imagerie diminuent également le contraste et la visibilité des détails.

Le bruit est une quantité de données qui empêche la vision claire de l'image médicale et qui interdit à la fois, la compréhension des moindres détails par le radiologue, et l'obtention de mesures correctes qui affectent les systèmes d'imagerie. Le bruit est émis par diverses modalités de dépistage (méthodes d'examen) de l'imagerie médicale, et est modélisé à travers plusieurs formules purement mathématiques et vise à améliorer la qualité de ces images, autrement dit, à les réduire ou les supprimer complètement.

Les techniques d'amélioration d'image sont des techniques mathématiques qui visent à améliorer la qualité d'une image donnée. Le résultat est une autre image qui montre certaines fonctionnalités d'une manière qui est meilleure dans un certain sens par rapport à leur apparence dans l'image d'origine. On peut également dériver ou calculer plusieurs versions traitées de l'image d'origine, chacune présentant une caractéristique sélectionnée dans une apparence améliorée. Des techniques simples d'amélioration de l'image sont développées et appliquées de manière ad hoc. Des techniques avancées optimisées en fonction de certaines exigences spécifiques et de critères objectifs sont également disponibles.

2.2. Prétraitement, analyse, et amélioration de qualité des images médicales

Dans cette section, nous présentons des méthodes d'amélioration d'image qui dépendent uniquement du niveau de gris des pixels et ne prennent pas en compte le voisinage des pixels ou les caractéristiques de l'image entière. Ces méthodes sont basées sur le changement intrinsèque de chaque pixel séparément. Elles consistent à ajuster la valeur de l'intensité lumineuse de chaque pixel qui a perdu ou reçu des quantités excessives.

2.2.1. Manipulation individuelle des pixels des images médicales

Trois principes sont essentiels en matière de représentation de l'image : récupération des caractéristiques non linéaires des supports d'affichage, la mise à l'échelle de l'intensité, et

l'égalisation de l'histogramme. Les approches associées à ces principes sont examinées en détail.

- **Compensation des caractéristiques non linéaires des supports d'affichage**

L'opération de compensation des caractéristiques non linéaires des supports d'affichage est une étape essentielle pour obtenir une image nette et stable. Les supports d'affichage englobent les méthodes qui ajustent les caractéristiques du système d'affichage pour produire une image sans distorsion. Les images numériques affichées sur des dispositifs tels que les tubes cathodiques (CRT) peuvent présenter un faible contraste et des détails peu visibles en raison de la non-linéarité du système d'affichage. Inverser cette non-linéarité à l'aide d'une transformation peut améliorer la visibilité des détails structurels. [57, 58].

- **Mise à l'échelle de l'intensité (Rétrécir ou élargir la plage dynamique)**

La mise à l'échelle de l'intensité est une technique d'amélioration d'image qui ajuste la plage dynamique des données d'image par rapport aux caractéristiques de l'affichage. Lorsque la plage dynamique des données acquises dépasse celle du système d'affichage, la mise à l'échelle permet de concentrer l'intérêt sur des bandes spécifiques d'intensité dans l'image. Cette mise à l'échelle améliore la visibilité des détails dans ces bandes en ajustant l'image de manière à ce que l'intervalle d'intensité d'intérêt corresponde à la plage dynamique de l'affichage. [57, 58]

- **Égalisation de l'histogramme**

L'égalisation de l'histogramme est une technique qui répartit les informations d'intensité dans l'image de manière uniforme sur la plage d'intensité disponible. Elle est utile lorsque les bandes d'intensité pertinentes ne sont pas connues à l'avance. L'histogramme normalisé de l'image est utilisé pour interpréter la distribution des niveaux de gris comme une fonction de densité de probabilité. En utilisant l'histogramme cumulatif normalisé, la transformation des niveaux de gris est effectuée pour étirer l'histogramme de l'image entrante, rendant les niveaux de gris plus uniformément répartis. [59, 60]

- **Égalisation locale de l'histogramme**

L'égalisation d'histogramme de zone locale est une méthode puissante qui applique l'égalisation d'histogramme à des zones locales qui se chevauchent dans l'image. Cela amplifie les caractéristiques subtiles et améliore l'observabilité, mais la complexité de calcul augmente avec la taille du noyau. Bien que cette méthode puisse améliorer les détails, elle peut également introduire une non-linéarité et une saturation. [61, 63]

- **Amélioration des Bords et des Contours**

L'amélioration des contours est essentielle, car les contours sont essentiels pour la compréhension visuelle des images. Les contours améliorés peuvent être combinés avec l'image originale pour préserver le contexte. Les opérateurs de dérivation, tels que Sobel, calculent les gradients et améliorent les caractéristiques de contour en fonction des changements de niveaux de gris des pixels voisins. Cette amélioration peut être obtenue par convolution avec des noyaux de dérivation. L'évaluation des contours peut être réalisée dans diverses directions pour extraire des informations directionnelles spécifiques [59, 61, 62].

Un opérateur dérivé du premier ordre, tel que Sobel, calcule les informations de gradient dans une direction spécifique. L'opérateur dérivé peut être codé dans un masque de pondération. Le gradient dans les directions x et y (défini par $\frac{\partial f(x,y)}{\partial x}$ et $\frac{\partial f(x,y)}{\partial y}$)

Ces masques de poids de 3 x 3 pixels chacun sont utilisés pour la convolution afin de calculer les images de gradient respectives. Pour l'amélioration de l'image spatiale sur la base des informations de gradient de premier ordre, l'image de gradient résultante peut être simplement ajoutée à l'image d'origine et redimensionnée en utilisant la gamme dynamique complète des valeurs de gris. Quatre masques de pondération calculent le gradient respectivement dans les directions horizontales, verticale à 45 degrés et à 135 degrés. Les informations de gradient ainsi obtenues peuvent être davantage utilisées pour l'amélioration et l'extraction de caractéristiques directionnelles spécifiques pour la segmentation d'images. Un opérateur dérivé du second ordre, connu sous le nom de Laplacien, peut être défini comme :

$$\begin{aligned} \nabla^2 f(x,y) &= \frac{\partial^2 f(x,y)}{\partial x^2} + \frac{\partial^2 f(x,y)}{\partial y^2} \\ &= [f(x+1,y) + f(x-1,y) + f(x,y+1) + f(x,y-1) + 4f(x,y)] \end{aligned} \quad (2.1)$$

Où $\nabla^2 f(x,y)$ représente la dérivée du second ordre ou laplacienne de l'image $f(x,y)$.

L'ajout du Laplacien (informations sur les contours) à l'image fournit une méthode simple d'amélioration de l'image basée sur les contours.

2.2.2. Représentation et filtration des images médicales dans le domaine spatial

La validation et l'évaluation des méthodes de quantification et d'analyse des textures et des morphologies dans les images médicales se trouvent au cœur de cette section. Ces aspects sont essentiels pour assurer la fiabilité et l'applicabilité clinique de ces méthodes. Cette démarche implique généralement des comparaisons avec des annotations manuelles ou des diagnostics de référence, l'examen de la reproductibilité des méthodes, ainsi que l'analyse de l'impact clinique sur les performances diagnostiques et pronostiques.

Les méthodes du domaine spatial manipulent les valeurs des pixels de l'image dans le domaine spatial en fonction des statistiques de distribution de l'image globale ou des régions locales. Dans cette catégorie visant à améliorer la qualité des images et leurs caractéristiques, on retrouve des exemples tels que la transformation d'histogramme, le filtrage spatial, la croissance de régions, le traitement morphologique des images, ainsi que les méthodes d'estimation d'images basées sur des modèles.

L'histogramme d'une image, un outil fréquemment utilisé dans l'amélioration et la caractérisation d'images, joue un rôle central dans cette section. Il offre une vue de la distribution des valeurs de niveaux de gris dans une image et est employé pour mettre en évidence les tendances et les caractéristiques de l'image. Ce dernier point est important dans le cadre de l'amélioration des images médicales et de leur analyse. L'histogramme, $h(i)$, peut être défini comme :

$$h(i) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \delta(f(m, n) - i), \quad i = 0, 1, \dots, P - 1 \quad (2.2)$$

Où

$$\delta(w) = \begin{cases} 1 & w = 0 \\ 0 & \text{sinon} \end{cases} \quad (2.3)$$

Les méthodes du domaine spatial traitent une image avec une transformation pixel par pixel basée sur les statistiques de l'histogramme ou les opérations de voisinage. Ces méthodes sont généralement plus rapides dans la mise en œuvre informatique par rapport aux méthodes de filtrage fréquentiel qui nécessitent le calcul de la transformée de Fourier pour la représentation du domaine fréquentiel.

2.2.2.1. Filtration des images médicales dans le domaine spatial

Dans cette section, l'attention se porte sur la filtration des images médicales en utilisant des opérations de voisinage, où les valeurs des pixels d'une image sont manipulées en fonction des valeurs correspondantes dans une sous-image du même format. Ces méthodes visent à améliorer les caractéristiques de l'image en ciblant spécifiquement les variations locales. [61]

- **Opérateurs Locaux** Lorsqu'on effectue des transformations sur une image à l'aide d'opérateurs, ces derniers prennent une image en entrée et produisent une nouvelle image en sortie. Les opérateurs locaux jouent un rôle clé dans l'amélioration des images en attribuant une nouvelle valeur à chaque pixel, en fonction de lui-même et de ceux dans son voisinage. Contrairement aux méthodes qui dépendent du changement individuel des valeurs des pixels, les opérateurs locaux dépendent du changement qui prend en compte les points voisins de chaque pixel dans l'image originale. Cette dépendance locale peut englober des opérateurs linéaires, des opérateurs non linéaires, et même des techniques d'égalisation d'histogramme dans une fenêtre donnée. Cette section présente divers filtres standard couramment employés pour l'amélioration des images.

- **Suppression du Bruit par Filtre Moyen** Le filtrage moyen se réalise en convoluant l'image avec un noyau de dimension $(2K+1) \times (2L+1)$, où chaque coefficient équivaut à l'inverse du nombre total de coefficients dans le noyau. Le filtrage moyen implique de remplacer la valeur de chaque pixel par la moyenne des valeurs de pixels dans un voisinage local (généralement une fenêtre $N \times N$, avec $N = 3, 5, 7$, etc.). Par opposition, le filtrage médian remplace chaque pixel par la valeur médiane dans ce voisinage local. Contrairement au filtrage moyen, le filtrage médian préserve les contours tout en éliminant les variations de texture mineures. Cependant, une conséquence est la perte de netteté des contours, même si le bruit est réduit dans l'image. [61, 62, 64]

- **Suppression du Bruit par Filtre Médian** Le filtrage médian, méthode non linéaire courante pour éliminer le bruit, se distingue par son approche unique. Contrairement à la convolution classique, le filtrage médian ne manipule pas les valeurs par convolution avec des coefficients de noyau. Au lieu de cela, pour chaque position de la matrice de noyau, un pixel de l'image d'entrée à cet emplacement devient le pixel de sortie au centre du noyau. La valeur médiane des pixels dans la matrice de noyau est calculée, et le pixel de sortie prend cette valeur. Cette opération non linéaire du filtrage médian permet de réduire significativement certains types de bruit tout en préservant les discontinuités telles que les bords et les

changements importants d'intensité. [61,62, 64,65]. Dans [89], les auteurs ont présenté plusieurs traitements d'image, parmi lesquels le filtrage médian adaptatif pour la réduction du bruit et une égalisation d'histogramme pour l'amélioration du contraste.

2.2.3. Représentation et filtration des images médicales dans le domaine fréquentiel

Les méthodes de représentation et de filtrage des images médicales dans le domaine fréquentiel traitent une image acquise dans le domaine de Fourier, en exploitant les caractéristiques fréquentielles inhérentes à l'image. Le filtrage fréquentiel, le filtrage homomorphe et les approches basées sur les ondelettes figurent parmi les exemples illustrant cette catégorie d'amélioration d'images et de caractéristiques basée sur la représentation fréquentielle. [61]

2.2.3.1. Transformation de Fourier Rapide (FFT)

L'application directe ou inverse de la transformation de Fourier à une image $N \times N$, en utilisant les définitions conventionnelles, requiert un nombre significatif de calculs, proportionnel à N^2 . L'algorithme de Transformation de Fourier Rapide (FFT) permet de réduire drastiquement le nombre d'opérations nécessaires, avec une complexité en temps de l'ordre de $N \log_2 N$. L'efficacité accrue de la FFT devient plus apparente avec des valeurs plus élevées de N , offrant ainsi des avantages significatifs en termes de calcul. La transformée de Fourier $F(u, v)$ d'une image $f(m, n)$ est définie comme :

$$F(u, v) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) e^{-2\pi j \left(\frac{um}{M} + \frac{vn}{N} \right)}, \quad (2.4)$$

$$u = 0, 1, 2, \dots, M-1 \quad v = 0, 1, 2, \dots, N-1$$

Où u et v sont les paramètres de fréquence spatiale. La transformée de Fourier fournit la représentation spectrale d'une image, qui peut être modifiée pour améliorer les propriétés souhaitées. Une image du domaine spatial peut être obtenue à partir d'une image du domaine spectral avec la transformée de Fourier inverse donnée par

$$f(m, n) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u, v) e^{-2\pi j \left(\frac{um}{M} + \frac{vn}{N} \right)}, \quad (2.5)$$

$$m = 0, 1, 2, \dots, M-1 \quad n = 0, 1, 2, \dots, N-1$$

2.2.3.2. Filtration des images médicales dans le domaine fréquentiel

Les méthodes de filtrage dans le domaine fréquentiel se concentrent sur le traitement d'une image transformée dans le domaine de Fourier, permettant d'accentuer ou d'atténuer des composantes fréquentielles spécifiques. De manière générale, ces composantes peuvent être différenciées en gammes basses et hautes. Les composants à basse fréquence sont souvent associés à des formes et structures floues dans l'image, tandis que les hautes fréquences englobent les détails nets, les contours et le bruit. [61]

- **Filtrage Passe-Bas** Le filtrage passe-bas idéal, utilisé pour réduire le bruit et les composantes à haute fréquence, produit un effet de lissage sur l'image. Il consiste en la multiplication d'une fonction de filtre passe-bas 2D, notée $H(u, v)$, par la transformation de Fourier $F(u, v)$ de l'image. Cette opération aboutit à une image lissée [61]. Sun et al. [152] ont utilisé un filtre passe-bas directionnel pour moyenner les intensités de l'image dans la direction parallèle au bord du vaisseau.
- **Filtrage Passe-Haut** Le filtrage passe-haut est employé pour accentuer la netteté de l'image et extraire des informations à haute fréquence, telles que les contours. Les informations à basse fréquence sont atténuées ou bloquées en fonction de la conception du filtre. Un filtre passe-haut idéal se caractérise par une fonction de fenêtre rectangulaire pour la bande passante des hautes fréquences. [61]
- **Filtrage de Wiener** Le filtrage de Wiener constitue une méthode bien établie et efficace pour la restauration d'images. Bien que l'approche soit conceptuellement simple, sa mise en œuvre pratique peut poser des défis. Parmi ces défis, la détermination de la Fonction de Réponse d'Impulsion (PSF) et les valeurs faibles ou nulles dans la fonction de filtre $H(u, v)$ peuvent engendrer des problèmes de calcul. Pour pallier ces problèmes, des stratégies telles que la déconvolution contrainte et le filtrage pondéré sont employées pour éviter la division par zéro dans les calculs. [5].

2.2.3. Filtre de diffusion anisotropique

Dans le contexte de l'apprentissage automatique, lorsque l'algorithme cherche à ajuster les paramètres pour un modèle en utilisant un ensemble de données limité, il peut modéliser non seulement les formes générales dans les données, mais également le bruit présent dans les données qui peut entraîner une dégradation des performances du modèle lorsqu'il est appliqué à des données de test. Cela signifie que le modèle peut ne pas généraliser correctement s'il prend en compte le bruit. Les auteurs, Perona et Malik, ont introduit en 1990 un filtre adaptatif appelé diffusion anisotrope pour résoudre ce problème. [67] Ce filtre vise à réduire le bruit tout en renforçant les bords (contours) dans les images. Le filtre de diffusion anisotropique fonctionne en considérant les valeurs d'intensité des pixels comme semi-absorbantes. Les intensités de chaque pixel sont mises à jour numériquement à chaque itération de l'algorithme. L'idée clé est que l'intensité se propage progressivement vers les pixels voisins, en suivant le gradient négatif. En d'autres termes, les pixels voisins du pixel central absorbent une petite quantité d'intensité du pixel central, mais uniquement s'ils ont des intensités plus faibles que le pixel central. Le processus de diffusion s'arrête une fois que tout le contraste a été éliminé. En d'autres termes, le filtre continue à ajuster les intensités des pixels jusqu'à ce que le bruit soit réduit, tout en conservant les bords nets. Le filtrage anisotrope de diffusion est suivi mathématiquement, l'équation différentielle partielle :

$$\frac{\partial I(x,y,t)}{\partial t} = \text{div}(c(x,y,t)\nabla I(x,y,t)) = c(x,y,t)\Delta I(x,y,t) + \nabla c(x,y,t)\nabla I(x,y,t) \quad (2.6)$$

Où $I(x, y, t)$ est l'intensité d'image au temps t , $c(x, y, t)$ est une constante, ∇ et Δ indiquent respectivement le gradient et l'opérateur de Laplacien, div indique l'opérateur de divergence. Les auteurs ont décrit le comportement de la diffusivité du filtre à travers une fonction monotoniquement décroissante et non négative dans la limite du module de gradient, appelée la fonction $g(\nabla I)$ de diffusion. Les fonctions g proposées par Perona et Malik sont comme suit :

$$g(\nabla I(x, y)) = e^{-\left(\frac{\nabla I}{K}\right)^2} \quad (2.7)$$

$$g(\nabla I(x, y)) = \frac{1}{1 + |\nabla I/K|^2} \quad (2.8)$$

$$\text{Avec } g(0) = 1. \quad (2.9)$$

Le paramètre K sert à placer la limite entre les gradients forts correspondant aux régions non homogènes de la masse (en particulier bords), et les petits gradients correspondant aux régions homogènes.

La représentation discrète du gradient d'intensité soustrait l'intensité de chaque Pixel à ses quatre voisins est comme suit :

$$\begin{aligned} (\nabla I)_N(x, y) &= I(x, y - 1) - I(x, y) \\ (\nabla I)_S(x, y) &= I(x, y + 1) - I(x, y) \\ (\nabla I)_E(x, y) &= I(x + 1, y) - I(x, y) \\ (\nabla I)_W(x, y) &= I(x - 1, y) - I(x, y) \end{aligned} \quad (2.10)$$

Le choix le plus simple de la constante de diffusion peut être calculé en rapprochant la norme du gradient à chaque endroit de l'arc avec la valeur absolue de sa projection le long de la direction de l'arc :

$$\begin{aligned} c_N(x, y) &= g((\nabla I)_N(x, y)) \\ c_S(x, y) &= g((\nabla I)_S(x, y)) \\ c_E(x, y) &= g((\nabla I)_E(x, y)) \\ c_W(x, y) &= g((\nabla I)_W(x, y)) \end{aligned} \quad (2.11)$$

Le processus itératif de la diffusion répète plusieurs fois et approches progressivement par un moment Δt , la représentation discrète de l'Eq. (1) devient :

$$I(t + \Delta t) = I(t) + \Delta t [c_N(\nabla I)_N + c_S(\nabla I)_S + c_W(\nabla I)_W + c_E(\nabla I)_E] |_{\forall x, y} \quad (2.12)$$

2.3. Analyse et quantification bidirectionnelle morphologique et texturale des images médicales

L'analyse et la quantification des textures et des morphologies dans les images médicales offrent une multitude d'applications cliniques couvrant divers aspects des données. Par exemple, dans le contexte du cancer, l'analyse des textures peut servir à caractériser les tumeurs et à anticiper leur agressivité. La quantification des morphologies permet d'évaluer l'évolution des pathologies, de mesurer la taille des lésions ou encore de détecter des anomalies anatomiques. Ces informations trouvent des applications dans le diagnostic, le suivi des traitements et la préparation des interventions chirurgicales. Cette section explore des techniques bien établies de quantification de forme et de texture en deux dimensions (2D), qui contribuent significativement aux domaines cliniques et de recherche.

2.3.1. Quantification de la Forme

Les méthodes de quantification de forme, opérant sur l'image segmentée de trois manières distinctes. Les approches basées sur la compacité et les moments spatiaux génèrent des mesures quantitatives de forme en effectuant des calculs géométriques et statistiques sur tous les pixels d'une région segmentée. Les mesures de distance radiale, les codes de chaîne et les descripteurs de Fourier agissent exclusivement sur les pixels limites, employant des calculs géométriques, statistiques et spectraux pour encoder et représenter un contour fermé. Lorsque les structures sont étirées ou ramifiées, les informations de forme essentielles résident dans les lignes médianes, accessibles via des algorithmes d'amincissement. Les mesures de forme quantitatives telles que longueur, angle, courbure ou orientation peuvent ensuite être calculées sur cette représentation squelettisée.

2.3.1.1. Quantification par des mesures globales de forme

Dans [68], plusieurs mesures de forme ont été décrites, mais il ne s'agit pas d'une liste exhaustive. Il existe de nombreuses mesures alternatives pour ces formes, ainsi que des mesures pour d'autres formes spécialisées. Il existe également de nombreuses mesures de forme globale standard dans la littérature, par exemple : moments, descripteurs de Fourier, modèles auto-régressifs. Selon Chen et Wang [68] : « Ces types d'approches se sont avérés efficaces, bien qu'un inconvénient soit que leurs valeurs ne sont souvent pas facilement compréhensibles. Dans de nombreuses applications, il est préférable que les descripteurs de région puissent être analysés par les experts du domaine, car cela facilite la validation et le développement du système de vision par ordinateur. » Néanmoins, étant donné que les données réelles ont tendance à être bruyantes et déformées, il est nécessaire de pouvoir faire face aux erreurs, et donc des mesures de forme continues sont nécessaires plutôt que des tests de forme binaires. Différentes mesures de forme dans le domaine de la vision par ordinateur sont traitées et sont utilisées pour caractériser les propriétés géométriques des régions d'intérêt dans des images numériques. [68]

- 1. Diversité des mesures:** Plusieurs mesures de forme sont discutées et peuvent être plus adaptées à certaines formes spécialisées. Les mesures sont généralement basées sur le calcul de l'écart entre une région donnée et une forme de référence, mais en raison du bruit et des déformations dans les données réelles, des mesures continues sont préférées aux tests binaires.
- 2. Mesures de forme globale:** Dans [68], différentes mesures globales de forme, telles que l'excentricité, l'allongement, la circularité, etc. sont décrites. Ces mesures permettent de caractériser des aspects tels que la compacité, l'aspect elliptique, la rectilinéarité, etc., des régions. Il mentionne également que certaines de ces mesures sont basées sur des descripteurs de Fourier et des moments centrés.
- 3. Spécificités des mesures et méthodes spécifiques de calcul:** Certaines considérations spécifiques peuvent être prise en compte à certaines mesures, comme la sensibilité aux bruits, la nécessité de calculer les valeurs de manière robuste, et comment certaines mesures peuvent réagir différemment à des variations de forme. Chaque mesure est accompagnée d'une méthode de calcul ainsi sa complexité, certaines d'entre elles dépendent de la méthode d'approximation polygonale utilisée et que leur complexité varie. Par exemple, pour la circularité, la mesure est basée sur le rapport entre le périmètre et l'aire de la région, et comment cette mesure peut être

affectée par le bruit dans les données. Pour l'ellipticité, une méthode basée sur les moments normalisés est mentionnée et des approches alternatives pour l'ajustement des ellipses sont décrites. Il note également que certaines mesures sont plus coûteuses en termes de calcul que d'autres en raison de leur méthode de calcul plus complexe.

2.3.1.2. Les moments spatiaux

La forme des segments limites (et des formes d'onde de signature) peut être décrite quantitativement en utilisant des moments statistiques simples. Ces moments sont liés aux variations d'intensité au sein de la région, ils doivent être utilisés avec prudence car ils ne sont pas toujours bien corrélés avec l'interprétation visuelle de la texture. Néanmoins, ils peuvent être utilisés comme caractéristiques extraites dans des applications de reconnaissance de formes où leur valeur potentielle est explorée avec une analyse quantitative.

Puisqu'une région lisse contient des pixels avec des valeurs proches les unes des autres et qu'une région rugueuse a une grande variabilité dans les valeurs de pixels, les moments statistiques de l'histogramme de région peuvent être utilisés comme métriques de texture.

Considérons la Figure 2.1 (a), qui montre le segment d'une frontière, et la Figure 2.2(b), qui montre le segment représenté comme une fonction unidimensionnelle $g(r)$ d'un nombre arbitraire en terme de la variable r . Cette fonction est obtenue en reliant les deux extrémités du segment et en faisant tourner le segment de droite jusqu'à ce qu'il soit horizontal. Les coordonnées des points sont tournées du même angle. [61]

L'utilisation des moments statistiques pour quantifier la forme des segments limites et des formes d'onde de signature peut être quantitativement décrite. Elle explore comment ces moments peuvent être utilisés pour caractériser la texture des régions, mais avertit que ces mesures doivent être utilisées avec prudence, car elles ne sont pas toujours directement corrélées avec l'interprétation visuelle de la texture:

- 1. Moments spatiaux et description quantitative:** Les moments statistiques pour décrire quantitativement la forme des segments limites et des formes d'onde de signature, sont des mesures mathématiques qui fournissent des informations sur la distribution des valeurs dans une région donnée.
- 2. Corrélation avec l'interprétation visuelle:** Les moments statistiques doivent être utilisés avec précaution, car ils ne sont pas toujours directement liés à la manière dont les humains interprètent visuellement la texture. Cela signifie que bien que ces mesures puissent être utiles dans certaines applications, elles ne peuvent pas toujours capturer pleinement les caractéristiques visuelles de la texture.
- 3. Utilisation des moments statistiques pour caractériser la texture:** Les moments statistiques de l'histogramme de la région peuvent être utilisés comme métriques de texture. Ces moments peuvent être exploités et suggérés comme des caractéristiques extraites dans des applications de reconnaissance de formes. La texture est caractérisée en examinant la variation des valeurs de pixels à l'intérieur de la région.

Les Figures 2.1 et 2.2 aident à expliquer ces concepts. Ils montrent un segment de frontière et comment il peut être représenté sous forme d'une fonction unidimensionnelle. Cela permet de visualiser comment les coordonnées des points sont ajustées en fonction de la rotation du segment, ce qui est pertinent pour la description des moments spatiaux.

L'amplitude de g est traitée comme une variable aléatoire discrète v et l'histogramme est formé d'amplitude $p(v_i)$, $i = 0, 1, 2, \dots, A-1$, où A est le nombre d'incrément d'amplitude discrets dont plusieurs échelles d'amplitude sont divisées. Ensuite, en gardant à l'esprit que $p(v_i)$ est une estimation de la probabilité que la valeur v_i se produise, le $i^{\text{ème}}$ moment de v autour de sa moyenne est :

$$\mu_n(v) = \sum_{i=0}^{A-1} (v_i - m)^n p(v_i) \quad (2.13)$$

Où

$$m = \sum_{i=0}^{A-1} v_i p(v_i) \quad (2.14)$$

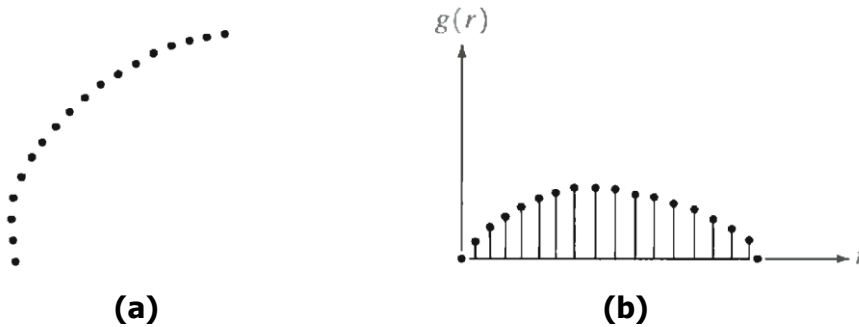


Figure 2.1. (a) Segment de la frontière. (b) Représentation comme une fonction unidimensionnelle.

La quantité m est reconnue comme la valeur moyenne ou moyenne de v et μ_2 comme sa variance. Généralement, seuls les premiers instants sont nécessaires pour différencier des signatures de formes clairement distinctes.

Une autre approche consiste à normaliser $g(r)$ par unité de surface et à le traiter comme un histogramme. En d'autres termes, $g(r_i)$ est maintenant traité comme la probabilité que la valeur r_i se produise. Dans ce cas, r est traité comme la variable aléatoire et les moments sont :

$$m = \sum_{i=0}^{A-1} v_i p(v_i) \quad (2.15)$$

Dans cette notation, K est le nombre de points sur la frontière, et $\mu_n(r)$ est directement lié à la forme de $g(r)$. Par exemple, le deuxième moment $\mu_2(r)$ mesure l'étalement de la courbe autour de la valeur moyenne de r et le troisième moment $\mu_3(r)$ mesure sa symétrie par rapport à la moyenne.

Fondamentalement, ce procédé consiste à réduire la tâche de description à celle de décrire des fonctions 1-D. Bien que les moments soient de loin la méthode la plus populaire, ils ne sont pas les seuls descripteurs pouvant être utilisés à cette fin.

Par exemple, une autre méthode consiste à calculer la transformée de Fourier discrète 1-D, à obtenir son spectre et à utiliser les q premières composantes du spectre pour décrire $g(r)$. L'avantage des moments par rapport aux autres techniques est que la mise en œuvre des

moments est simple et ils portent également une interprétation "physique" de la forme des limites. La normalisation de la taille, si on le souhaite, peut être obtenue en mettant à l'échelle la plage de valeurs de g et r . [61]

Le concept des moments sont utilisés pour analyser les distributions statistiques et de représenter les distributions spatiales des valeurs des pixels.

Les moments peuvent être calculés aussi bien pour les images binaires que pour les images en niveaux de gris. Dans les images binaires, les moments quantifient strictement la forme de la région segmentée; en revanche, les moments appliqués aux images en niveaux de gris incluent des informations concernant la distribution d'intensité en plus de la forme. Les moments constituent un ensemble infini de coefficients de transformée à partir desquels $f(i, j)$ peut être récupéré de manière unique. Le nombre fini de moments utilisés en pratique ne conserve pas toutes les informations d'image, mais ils peuvent fournir un ensemble efficace de descripteurs de forme et peuvent contribuer à la classification.

Dans les applications où la forme d'une région segmentée doit être quantifiée d'une manière qui est insensible à son orientation, des métriques invariantes de rotation sont utilisées. Les moments centraux invariants d'échelle donnés peuvent être combinés davantage pour obtenir des descripteurs invariants de translation, d'échelle et de rotation [5].

Les moments statistiques ont été utilisés pour des applications telles que la quantification de la texture dans les images ultrasonores du tissu hépatique [61]. Les moments statistiques ne contiennent pas d'informations spatiales car ils sont dérivés de l'histogramme de l'image. Étant donné que le concept de rugosité est souvent associé à la taille du grain dans l'objet, les mesures de texture qui transmettent des informations spatiales peuvent être utiles.

2.3.1.3. Mesures de Distance Radiale, code de chaînes et les descripteurs de Fourier

Plusieurs techniques utilisées sont traitées dans la littérature pour l'analyse de formes dans des images numériques en utilisant des méthodes basées sur des codes de chaîne et des descripteurs de Fourier. Typiquement, la représentation en séquence connectée de la frontière est basée sur la connectivité quatre ou huit des segments. La direction de chaque segment est codée à l'aide d'un schéma de numérotation tel que ceux illustrés à la Figure 2.2(a) ou (b). La manière dont les caractéristiques d'une région segmentée sont représentées et analysées en termes de caractéristiques externes (frontière) et internes (pixels).

La distance radiale est présentée comme une technique commune pour quantifier la forme d'une région en mesurant la distance de chaque pixel sur la frontière par rapport à un point central (généralement le centroïde).

L'invariance d'échelle est obtenue en normalisant la séquence de distance radiale. Différentes métriques de forme, telles que l'entropie et la compacité, peuvent être extraites de cette séquence [69, 81]. La représentation en séquence connectée de la frontière est basée sur la connectivité 4 ou 8 des segments. (Voir la Figure 2.2 (a) ou (b)). La séquence des points de la frontière peuvent apparaître dans une image digitale dans un plan xy , les paires $(x_0, y_0), (x_1, y_1), \dots, (x_{fin}, y_{fin})$ représentent les coordonnées des points du contour contourné qui traverse la frontière, disant dans le sens inverse des aiguilles d'une montre. Chaque paire de coordonnées peut être traité comme un nombre complexe :

$$s(k) = x(k) + jy(k) \quad (2.16)$$

Il est également possible de quantifier la forme de la frontière avec les moments de zernike [82], l'ondelette discrète [77-79]. Par ailleurs, l'utilisation des codes de chaîne pour représenter des formes analyse les transitions entre les points de la frontière. Les codes de chaîne sont adaptés en rééchantillonnant la frontière pour réduire la longueur de la chaîne et minimiser les perturbations dues au bruit, et sont normalisés pour l'invariance à la rotation en utilisant la première différence du code.

Un espacement de grille plus grand est sélectionné dont la frontière est traversée par un point de frontière attribué à chaque nœud de la grande grille, en fonction de la proximité de la frontière d'origine à ce nœud, comme illustré à la Figure 2.3(a) et la Figure 2.3(b) respectivement. La frontière rééchantillonnée obtenue de cette manière peut alors être représentée par un code 4 ou 8, comme indiqué sur les Figures 2.3(c) et (d), respectivement. [61]

Il est également noté que les codes de chaîne différentiels sont invariants à la rotation et à la translation mais ne quantifient pas directement les attributs de forme. C'est une représentation qui diffère pour différentes limites ou contours d'objets et peut être utilisée pour extraire des informations pertinentes à partir des contours d'objets. Il n'est pas une mesure directe d'un attribut de forme particulier ni un outil de comparaison d'échelle entre formes.

1. Invariance à la Rotation et à la Translation :

- Le code de chaîne différentiel est invariant à la rotation et à la translation de la frontière, ce qui signifie que la manière dont l'objet est orienté ou déplacé n'affecte pas ce code.

2. Extraction de Métriques de Forme :

- Malgré le fait que le code de chaîne différentiel ne quantifie pas directement les attributs de forme, il contient toutes les informations essentielles pour extraire des métriques de forme.

- Par exemple, le lissage de la frontière est lié à la courbure locale de la frontière, et cette courbure peut être quantifiée à partir du code de chaîne différentiel.

- Les frontières lisses se traduisent par des transitions douces entre les points adjacents, ce qui se reflète dans de petits écarts entre les chiffres consécutifs du code de chaîne différentiel.

3. Symétrie et Sections Concaves :

- La symétrie des frontières peut être détectée dans le code de chaîne différentiel. Si le point de départ est sur l'axe de symétrie, le code est également symétrique autour de son milieu.

- En cas de point de départ arbitraire, le code de chaîne différentiel présente généralement deux parties symétriques autour de leurs propres centres.

- La présence de sections concaves dans la frontière peut également être identifiée en analysant les signes des chiffres du code de chaîne différentiel. Les concavités produisent des chiffres négatifs.

4. Représentation sous forme de Nombre Complexe :

- Il est possible de représenter chaque pixel du contour d'une région binaire sous forme de nombre complexe, où les parties réelle et imaginaire correspondent aux coordonnées x et y du pixel. Cela permet d'exprimer le contour comme une séquence complexe 1D. [61, 90]

5. Applications des Codes de Chaîne :

- Plusieurs exemples d'applications des codes de chaîne, notamment la classification de noyaux cellulaires dans des images de "Pap smear" et dans la segmentation des limites ventriculaires gauches dans des images échographiques cardiaques.

- Dans ces applications, les codes de chaîne sont utilisés pour extraire diverses caractéristiques de forme, telles que la longueur de la limite, le diamètre minimum, le diamètre maximum, l'ellipticité, le lissage, la symétrie, la détection de sections concaves, etc. [87, 89]

2.3.1.4. Codes de chaîne et les Descripteurs de Fourier

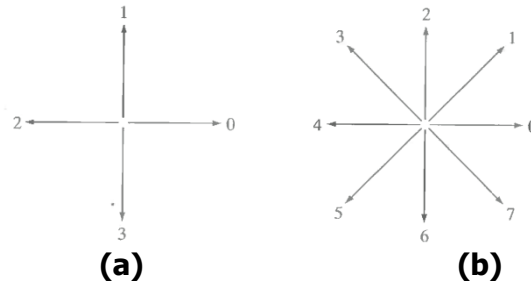


Figure 2.2. Numéros de direction pour (a) le code de chaîne à 4 directions et (b) le code de chaîne à 8 directions.

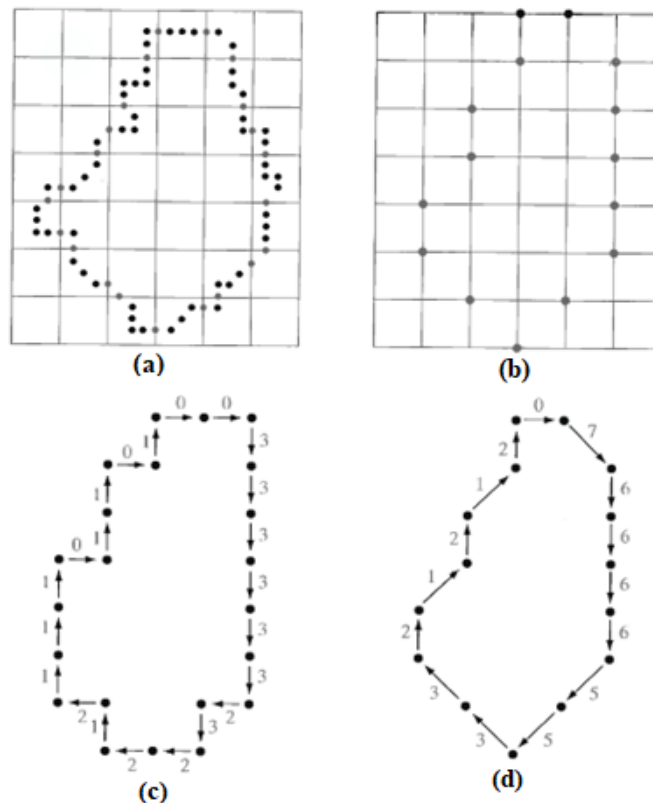


Figure 2.3. (a) Frontière numérique avec grille de rééchantillonnage superposée. (b) Résultat du rééchantillonnage. (c) code de chaîne à 4 directions. (d) Code de chaîne à 8 directions. [61]

Ces normalisations ne sont exactes que si les frontières elles-mêmes sont invariantes à la rotation et au changement d'échelle, ce qui, en pratique, est rarement le cas. Une séquence de code invariante à la rotation peut également être obtenue en utilisant la première différence du code en chaîne ; la différence entre deux chiffres consécutifs est définie comme le nombre de directions entre eux, qui est considéré comme positif dans le sens antihoraire. [61]

2.3.1.5. Amincissement (Thinning)

L'aminçissement est une technique utilisée dans le traitement d'images pour extraire les informations essentielles de la forme d'une structure en représentant cette structure sous forme de son squelette, qui consiste en des lignes médianes le long des principaux composants de la structure. L'objectif est d'obtenir une représentation simplifiée tout en préservant les caractéristiques clés de la forme. Un algorithme spécifique appelé l'algorithme de Zhang et Suen, qui est utilisé pour effectuer l'aminçissement de manière itérative :

- 1. Objectif de l'aminçissement :** L'objectif principal de l'aminçissement est de représenter une structure sous forme de son squelette en utilisant les lignes médianes des principaux composants.
- 2. Transformée de l'axe médian (MAT) :** La transformée de l'axe médian (MAT) comme une technique pour déterminer les lignes médianes. Elle consiste à calculer la distance entre chaque pixel intérieur de la structure et chaque pixel limite. La distance minimale est utilisée pour déterminer la ligne médiane.
- 3. Complexité de calcul du MAT :** Le calcul de la transformée de l'axe médian est coûteux en termes de calcul, car il nécessite le calcul de nombreuses distances, en particulier si la distance euclidienne est utilisée. [96-98]

Algorithmes itératifs : En raison de la complexité de calcul du MAT, de nombreux algorithmes itératifs ont été développés pour déterminer les lignes médianes avec moins de calculs. [97, 98]. Dans [98a], une étude présentant un aperçu des différentes techniques d'aminçissement utilisées dans le domaine de l'imagerie médicale.

- 4. Algorithme de Zhang et Suen :** Cet algorithme est mentionné comme particulièrement efficace. Il fonctionne de manière itérative et comporte deux étapes :
 - À l'étape 1, il identifie les pixels de bordure (ceux qui ont une valeur de 1) qui ont au moins un pixel connecté à 8 voisins égal à 0. Ensuite, il supprime les pixels de bordure qui satisfont quatre conditions spécifiques pour éviter de diviser la région à éclaircir.
 - À l'étape 2, il marque les pixels pour suppression en fonction de ces mêmes quatre conditions. Après avoir visité chaque pixel de l'image, tous les pixels marqués sont supprimés. Ensuite, l'algorithme recommence l'étape 1 avec le résultat de l'étape 2. Ce processus itératif se poursuit jusqu'à ce qu'aucun pixel ne soit marqué à l'étape 1 et à l'étape 2.
 - Résultat : L'algorithme de Zhang et Suen produit un squelette constitué de brins de pixels connectés en 8 directions. [99]

2.3.2. Quantification des textures

Dans l'analyse d'images médicales, la perception des tissus repose souvent sur des termes tels que lissé, granuleux, régulier ou homogène. Cette caractéristique se rapporte aux variations locales d'intensité et peut être quantifiée à l'aide de métriques de texture. Les moments statistiques proviennent directement de l'histogramme d'intensité de l'image. Les mesures de matrice de cooccurrence s'établissent à partir d'un histogramme 2D, préservant ainsi les informations spatiales. Les mesures spectrales issues de la transformation de Fourier permettent la quantification de la texture, particulièrement lorsque des motifs répétitifs sont présents. Les statistiques de longueur de série (run-length statistics) quantifient la texture en analysant les brins de pixels linéaires ayant la même valeur dans l'image.

2.3.2.1. Moments statistiques

Diverses méthodes de quantification des textures dans des images mettent l'accent sur les caractéristiques extraites pour décrire la texture. Les moments statistiques sont utilisés pour quantifier les variations d'intensité dans une région d'une image. Ils ne contiennent pas d'informations spatiales car ils sont dérivés de l'histogramme de l'image et peuvent être utilisés dans des applications de reconnaissance de formes, la quantification de la texture dans les images ultrasonores du tissu hépatique [61], mais ils ne sont pas toujours bien corrélés avec l'interprétation visuelle de la texture. Les moments statistiques incluent des mesures telles que la moyenne, la variance, skewness et kurtosis, qui fournissent des informations sur l'asymétrie et l'uniformité de l'histogramme. [61]

2.3.2.2. Mesures de la matrice de cooccurrence

Les métriques de texture basées sur la matrice de cooccurrence contiennent des informations spatiales. Cette matrice compile des informations spatiales et statistiques en utilisant la position relative de paires de pixels. Les valeurs d'orientation (θ) et de distance (d) sont utilisées pour former plusieurs matrices de cooccurrence, mettant en évidence la structure dans différentes directions. Ces mesures sont utiles pour caractériser la texture, en particulier dans les images où la taille du grain est importante. Typiquement, quatre valeurs de θ , à savoir 0° , 45° , 90° et 135° , couvrent les orientations, et le choix de distance le plus courant est $d = 1$ lorsque θ vaut 0° ou 90° , et $d = 2$ lorsque θ vaut 45° ou 135° .

2.3.2.3. Matrice de cooccurrence des niveaux de gris (GLCM)

Plusieurs métriques de texture qui contiennent des informations spatiales sont basées sur la matrice de cooccurrence [117, 118], La taille de cette matrice dépend du nombre de niveaux de gris dans l'image, ce qui peut la rendre assez grande. Les valeurs de pixel sont souvent mappées pour quantifier l'homogénéité et mettre en évidence les transitions d'intensité. Cette méthode est sensible aux transitions d'intensité mais insensible à l'amplitude de ces transitions. La matrice de cooccurrence est constituée d'éléments, à savoir la probabilité d'avoir des paires de pixels avec des valeurs i et j dans des emplacements qui présentent la position relative sélectionnée.

2.3.2.4. Matrice des longueurs de plage de niveaux de gris (Gray level run lengths matrix GRLM)

Les pixels consécutifs peuvent être utilisés pour quantifier la texture, en se concentrant sur la longueur des séquences de pixels ayant la même valeur. Les histogrammes de longueur de course sont calculés pour différentes orientations, ce qui permet de caractériser la texture en fonction de la longueur des séquences. Une métrique courante est le pourcentage d'exécution, qui est calculé en fonction du nombre de pixels, du nombre d'échelles de gris et de la longueur maximale d'une séquence. Ces mesures peuvent être utilisées pour caractériser le tissu dans une image. Les informations de longueur de plage sont collectées généralement en utilisant les orientations de θ (à savoir 0° , 45° , 90° , et 135°). Pour chaque orientation θ , le nombre d'exécutions d'une longueur de m pixels à l'échelle de gris k est calculé pour former l'histogramme de longueur d'exécution [127,128].

- Mesures spectrales

La transformée de Fourier peut être utilisée pour quantifier les textures avec une structure périodique ou presque périodique. La DFT (Transformée de Fourier Discrète) fournit des informations sur l'orientation, la taille du grain et le contraste de la texture dans une image. Elle est particulièrement utile pour quantifier les textures avec des motifs globaux répétitifs.

2.3.2.5. Analyse de texture avec des motifs binaires locaux (LBP : Local Binary Patterns)

Les méthodes d'analyse de texture ont été traditionnellement divisées en deux catégories, à savoir l'approche statistique/stochastique et l'approche structurelle. Il explique comment ces approches sont utilisées pour décrire et caractériser les textures dans les images.

- 1. Approche statistique/stochastique :** Cette approche traite les textures comme des phénomènes statistiques. Elle repose sur l'analyse des propriétés statistiques des intensités et des positions des pixels dans une texture. Les mesures statistiques de texture, telles que les statistiques de cooccurrence et les histogrammes de différence, sont utilisées pour représenter et caractériser la formation de la texture. Ces modèles fonctionnent généralement bien avec les microtextures stochastiques, c'est-à-dire les textures ayant des variations aléatoires et imprévisibles dans les intensités des pixels.
- 2. Approche structurelle :** Cette catégorie d'approche introduit le concept de primitives de texture, également appelées texels ou textons. L'idée est de décrire une texture en utilisant un ensemble de primitives de texture et en décrivant leurs relations dans la texture. Le but est de représenter des structures complexes en utilisant des primitives plus simples. Cette approche est particulièrement efficace pour les macrotextures qui ont des structures claires et bien définies. Les modèles basés sur les primitives de texture ont été largement utilisés pour expliquer la perception humaine des textures.

La méthode LBP (Local Binary Pattern) peut être considérée comme une approche fédératrice entre ces deux catégories. Contrairement à l'approche statistique qui se concentre sur les statistiques des pixels individuels, LBP se concentre sur la formation de motifs locaux dans la texture. Chaque pixel est étiqueté avec un code de primitive de texture qui correspond le mieux à son voisinage local. Quelques exemples sont montrés dans la Figure 2.4 avec l'opérateur LBP. Dans la figure, les uns sont représentés par des cercles blancs et les zéros sont noirs. Ces codes LBP peuvent être considérés comme des micro-textons, c'est-à-dire des primitives locales détectées par la méthode LBP. Ces primitives peuvent représenter diverses caractéristiques locales telles que les taches, les zones plates, les bords, les extrémités des bords, les courbes, etc. [68]

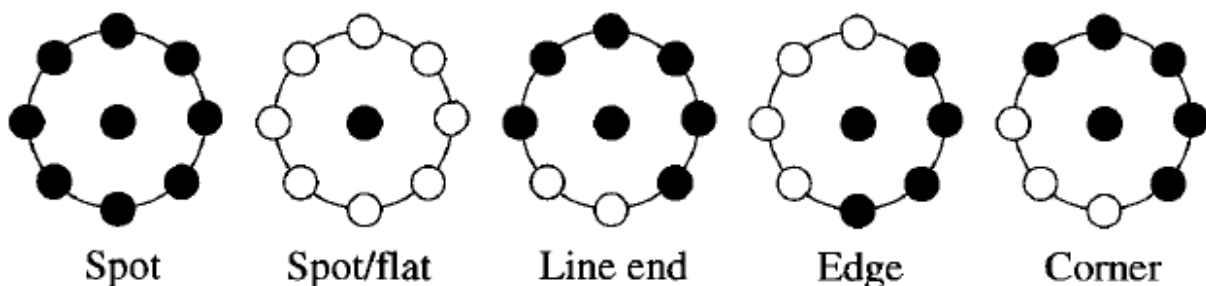


Figure 2.4. Différentes primitives de texture détectées par le LBP. [68]

L'importance et les avantages de l'opérateur Local Binary Pattern (LBP) dans l'analyse de texture, en particulier pour les applications en imagerie médicale, ses propriétés s'éclairent à travers l'étude de plusieurs aspects :

- 1. Rapidité d'extraction des caractéristiques :** L'opérateur LBP est considéré comme très simple et efficace pour extraire rapidement les caractéristiques de texture. En raison de sa simplicité et de ses performances, il a été largement appliqué dans diverses applications.
- 2. Classification supervisée avec des caractéristiques de texture :** L'opérateur LBP a été utilisé dans la classification supervisée pour prendre en compte la texture de l'emphysème et d'autres anomalies dans les images CT pulmonaires. D'autres caractéristiques, telles que les moments d'histogrammes de réponse de filtre, les matrices de co-occurrence de niveaux de gris (GLCM), les matrices de longueur de niveau de gris (GLRLM), et les moments d'histogramme d'atténuation, ont également été utilisées pour la classification. [129,134]
- 3. Performances de classification excellentes avec des opérateurs locaux de petite taille :** Les opérateurs locaux de petite taille, tels que les modèles binaires locaux (LBP) et les approches basées sur texton, ont montré d'excellentes performances de classification de texture sur des bases de données standard. Ces opérateurs sont particulièrement adaptés aux situations où la région d'intérêt (ROI) est petite, ce qui est souvent le cas en imagerie médicale pour les pathologies localisées.[129a, 129b]
- 4. Unification des approches statistiques et structurelles :** L'opérateur LBP est considéré comme unificateur des modèles statistiques et structurels traditionnels d'analyse de texture. Il décrit la texture en termes de micro-primitives (textons) et de leurs règles statistiques de placement. Cette approche combine ainsi des informations structurelles et statistiques pour une meilleure représentation de la texture.
- 5. Tolérance aux changements d'éclairage :** L'opérateur LBP est relativement invariant par rapport aux changements d'éclairage dans l'image. Cela signifie qu'il peut maintenir sa performance malgré les variations d'intensité lumineuse, ce qui le rend adapté à des conditions d'imagerie réelles où l'éclairage peut varier.
- 6. Simplicité de calcul et implémentations disponibles :** L'opérateur LBP est simple à calculer, ce qui permet son utilisation en temps réel pour analyser des images dans des conditions difficiles. De plus, des implémentations C++ et Matlab de l'opérateur sont disponibles gratuitement, facilitant ainsi le travail des chercheurs et développeurs d'applications. [68]

2.4. Segmentation d'images médicales dans le domaine spatial

2.4.1. Objectifs, et techniques de segmentation d'images médicales

Cette section fournit une vue d'ensemble complète de la segmentation d'images médicales, en mettant en évidence son importance, ses nombreuses applications et la variabilité des techniques utilisées en fonction des besoins spécifiques de chaque application. Il offre également un aperçu des différentes catégories de techniques de segmentation et de leurs caractéristiques.

- 1. Objectif de la segmentation :** L'objectif principal de la segmentation d'images médicales est de diviser une image en régions ou classes homogènes en fonction de

caractéristiques spécifiques. Ces caractéristiques peuvent varier en fonction de l'application médicale, telles que la détection de la bordure coronaire, la quantification des lésions de sclérose en plaques, etc. [82, 87, 90, 94, 5, 105, 107, 113, 118, 120]

- 2. Applications de la segmentation :** Plusieurs applications de la segmentation d'images médicales, notamment la détection de différentes structures anatomiques, la mesure de volumes tumoraux, la classification automatisée des cellules sanguines, la détection de tumeurs, etc. Ces applications couvrent un large éventail de domaines médicaux. [81,85,86,95,96,98,102,106,110,120,121,124]
- 3. Variabilité des techniques de segmentation :** Il n'existe pas de technique de segmentation standard qui fonctionne pour toutes les applications d'imagerie médicale en raison de la variabilité des objectifs et des types de données. Différentes hypothèses sur la nature des images analysées conduisent à l'utilisation d'algorithmes différents.
- 4. Classification des techniques de segmentation :** Les techniques de segmentation peuvent être classées en plusieurs catégories en fonction de différents critères, tels que manuel, semi-automatique et automatique, basé sur les pixels ou sur la région, délimitation manuelle, seuillage, segmentation basée sur un modèle, techniques classiques, statistiques, floues et de réseaux de neurones. [165, 172, 176].
- 5. Deux principales catégories de techniques :** Les techniques de segmentation peuvent être classées en deux grandes catégories : celles basées sur les régions, qui recherchent des régions homogènes en fonction de critères prédéfinis, et celles basées sur les bords, qui identifient les frontières d'objets et les régions de segments entourées par les frontières. [61,143,163,169,174]
- 6. Méthode de seuillage :** Le seuillage est une méthode courante de segmentation de région, où un seuil est utilisé pour diviser l'image en groupes de pixels en fonction de leurs valeurs. Plusieurs méthodes de seuillage sont présentées dans la littérature, notamment les méthodes globales basées sur des histogrammes et les méthodes locales basées sur des propriétés locales. [148,164,170,174,179]
- 7. Méthodes basées sur les bords :** Les méthodes de segmentation basées sur les bords sont conçues pour trouver les frontières d'objets et les régions entourées par ces frontières. Elles utilisent généralement des images d'amplitude et/ou de phase de bord produites par des opérateurs de bord adaptés aux caractéristiques de l'image. [141, 143, 61, 153,173, 62]
- 8. Combinaison d'images de différentes sources :** L'amélioration des performances de la segmentation peut être obtenue en combinant des images de plusieurs sources (segmentation multispectrale) [149,167,178] ou en intégrant des images dans le temps (segmentation dynamique ou temporelle) [162,168,175].

2.4.2. Prétraitement d'image et seuillage

Cette section offre un aperçu complet des techniques de segmentation d'images médicales, en se concentrant sur l'importance du prétraitement d'image, du seuillage, de la croissance de région et des techniques basées sur les bords. Il explique comment ces méthodes peuvent être utilisées en imagerie médicale pour extraire des informations utiles à partir d'images complexes [156,160,62,171].

- 1. Importance du prétraitement d'image et seuillage :** L'importance du prétraitement d'image est pour faciliter la segmentation, en particulier lorsque les

images médicales présentent des contours flous ou un faible contraste. Il mentionne des techniques telles que le lissage d'image à l'aide de filtres moyens ou médians, ainsi que le lissage gaussien. Le prétraitement vise à rendre l'histogramme de l'image plus fortement bimodal, ce qui simplifie la segmentation par seuillage. Plusieurs techniques de seuillage ont été développées [143,148,61,153,155,62,57,70,174,179]. Certains d'entre eux sont basés sur l'histogramme de l'image ; d'autres sont basés sur des propriétés locales, telles que la valeur moyenne locale et l'écart type, ou le gradient local.

2. **Seuillage global** : Le seuillage global est une technique où un seul seuil est sélectionné pour l'ensemble de l'image en se basant sur l'histogramme de l'image. Cette technique est efficace lorsque l'histogramme de l'image est bimodal, et elle génère une image binaire où les pixels sont classés comme objets ou arrière-plan en fonction du seuil choisi [148,163,174]. Certaines des techniques de sélection de seuil sont décrites dans les références [143,148,62,179].
3. **Seuillage local (adaptatif)** : Dans plusieurs applications, le seuillage global ne fonctionne pas bien, en particulier lorsque l'arrière-plan varie sur l'image ou que le contraste des objets varie. Une autre solution consiste à appliquer un seuillage local, (également appelé seuillage adaptatif) [144,148,153,179]. C'est une approche qui divise l'image en sous-images ou examine les intensités locales des pixels pour déterminer des seuils locaux. Cela permet d'obtenir de meilleurs résultats de segmentation dans des zones d'image présentant des variations locales.

2.4.3. Régions en croissance (Region Growing)

La méthode de régions en croissance vise à regrouper des pixels ayant des intensités similaires. Cette méthode commence par un ou plusieurs pixels initiaux (graines) appartenant à la structure d'intérêt. Les pixels voisins sont ensuite ajoutés à la région de croissance en fonction d'un critère d'uniformité. La croissance se poursuit jusqu'à ce qu'aucun autre pixel ne puisse être ajouté. Cette méthode est utile pour segmenter des régions de même intensité qui sont spatialement séparées [186, 187, 61, 163, 189, 190, 191, 192, 62, 174].

Plusieurs applications médicales de La croissance des régions (Region growing), telles que la segmentation des ventricules sur les images cardiaques [195], l'extraction des vaisseaux sanguins sur les données d'angiographie ou l'extraction de la surface du cerveau [196]. Il est noté également que la segmentation morphologique peut être trouvées dans les références [197-199].

2.4.4. Techniques de segmentation basées sur les bords

Un bord ou une limite sur une image est défini par le gradient d'intensité de pixel local. Un gradient est une approximation de la dérivée de premier ordre de la fonction image.

L'amplitude et la direction du gradient peuvent être affichées sous forme d'images. L'image d'amplitude aura des niveaux de gris qui sont proportionnels à l'amplitude des changements d'intensité locaux, tandis que l'image de direction aura des niveaux de gris représentant la direction du gradient local maximal dans l'image d'origine.

1. **Techniques de segmentation basées sur les bords** : Les techniques de segmentation basées sur les bords, se concentrent sur la détection des bords définis par le gradient d'intensité local. Les opérateurs de gradient courants tels que Roberts, Prewitt, Robinson, Kirsch, et Frei-Chen. Ces méthodes sont rapides en termes de calcul,

mais elles nécessitent souvent une étape de post-traitement pour relier les bords et former des frontières fermées autour des régions. [61,153,156, 62, 57]

- 2. Lissage avant la détection des bords :** Les méthodes de détection de bord basées sur le gradient sont sensibles au bruit. L'image peut être lissée avec un filtre Gaussien avant d'appliquer un opérateur de détection des bords, tel que le Laplacien de Gaussienne (LoG). Cette approche réduit les effets du bruit tout en préservant les bords importants. [139, 143, 148, 61, 153, 207,62, 1,74,191]

Une technique de recherche de graphes pour la détection des frontières a été utilisée dans de nombreuses applications médicales [187,181,203,204,173,205,206].

2.4.5. Fuzzy C-Means et son application en imagerie médicale

FCM est un algorithme de clustering qui peut être utilisé pour la segmentation en regroupant des pixels similaires dans un espace de caractéristiques. Il offre une alternative aux méthodes de seuillage et de détection de bord, en mettant particulièrement en évidence l'algorithme Fuzzy C-Means (FCM) et ses applications en imagerie médicale.

- 1. Espace paramétrique de la segmentation d'image:** L'opération de la division d'image se déroule dans un espace paramétrique où les pixels similaires sont regroupés. Les méthodologies traditionnelles de segmentation incluent le seuillage (définir des seuils pour séparer les pixels), la détection de contours (détection des bordures entre différentes régions) et la détection de régions (identification de régions homogènes dans une image). Cependant, ces méthodes classiques peuvent être moins efficaces en présence de structures voisines ou de bruit.
- 2. Clustering pour la segmentation d'image:** Les algorithmes de clustering sont introduits comme des méthodes pour segmenter des régions en partitionnant l'image en ensembles de pixels similaires dans l'espace des caractéristiques. L'idée est d'examiner chaque pixel et de l'assigner à un groupe qui représente le mieux la valeur de son vecteur de caractéristiques. Une autre catégorie d'algorithmes de segmentation appelée "croissance de région" est mentionnée dans la section 2.4.3. Ces algorithmes attribuent des pixels ou des régions adjacentes au même segment si leurs valeurs d'image sont proches selon un critère prédéfini de proximité.
- 3. Fuzzy C-Means (FCM):** L'attention se tourne vers l'algorithme Fuzzy C-Means (FCM). Contrairement aux méthodes de partition dure qui forcent chaque élément à appartenir exclusivement à une classe, FCM permet à chaque élément d'appartenir partiellement à différentes classes. Cela signifie qu'un pixel peut avoir une appartenance partielle à plusieurs classes, reflétant ainsi une certaine incertitude ou flou. Cela s'avère utile pour conserver davantage d'informations de l'image d'origine dans le processus de segmentation. Cependant, FCM peut être plus sensible aux bruits que d'autres méthodes basées sur le niveau de gris.
- 4. Améliorations de FCM:** Différentes méthodes ont été proposées pour améliorer FCM et le rendre plus robuste face au bruit. De nombreuses méthodes ajoutent un terme de pénalité basé sur les informations spatiales des pixels voisins à la fonction objective de FCM standard. Cela permet de rendre les fonctions objectifs FCM modifiées plus lisses sur le plan spatial, ce qui peut aider à extraire des objets complets à partir d'images [6, 13, 49-54].

2.5. État de l'art de classification et de prédiction des images médicales

Les techniques de sélection de caractéristiques du domaine de l'apprentissage automatique correspondent à des algorithmes de sélection d'un sous-ensemble de caractéristiques à partir d'un ensemble de départ, à l'aide de divers critères et méthodes. Tous les algorithmes de sélection de caractéristiques peuvent correspondre soit au modèle de filtre (filter) soit au modèle de symbiose (wrapper). L'approche Wrapper consiste à utiliser la précision en termes de classification et donc l'algorithme d'apprentissage ou d'induction du système. Pour la deuxième approche, Filtre, la sélection se fait sans tenir compte de son influence sur les performances du système. Dans le cas d'un système de reconnaissance, l'approche filtre est plus rapide que celle Wrapper en termes de génération de résultats, cependant cette dernière a l'avantage de fournir généralement des résultats plus pertinents pour la classification.) [19] [211]. Le but de la sélection des caractéristiques est la visualisation des caractéristiques sélectionnées les plus pertinentes pour la tâche de classification des mammographies, ainsi son chemin d'attribution de la caractéristique la plus puissante à la plus faible. Ce sont des techniques qui ne créent pas de nouvelles caractéristiques en utilisant une transformation ou une combinaison de l'espace de données de départ, comme par exemple l'analyse en composantes principales (ACP), l'analyse en composantes indépendantes (plus appropriée que l'ACP pour les distributions non gaussiennes), l'analyse discriminante linéaire (LDA). De plus, la disponibilité préalable de la connaissance du domaine est un facteur facultatif de sélection des caractéristiques qui peut être utilisé pour aider le radiologue, certaines caractéristiques peuvent être plus indicatives que d'autres.

2.5.1. Classification et prédiction des images des seins

L'approche générale proposée dans cette thèse fait partie d'un cadre de sélection de caractéristiques wrapper (voir la Figure 1.8 du premier chapitre) qui utilise la précision de la classification et donc l'algorithme d'apprentissage ou d'induction du système. Contrairement aux méthodes basées sur la transformation, la sélection de caractéristiques préserve la sémantique du domaine sans transformer ni détruire l'espace d'entrée. Plusieurs investigations dans la littérature ont été proposées pour le diagnostic informatisé pour la classification des images de mammographie en appliquant différentes méthodes d'apprentissage automatique et de sélection de caractéristiques :

- En 2001, Liu et al ont proposé un nouveau schéma pour la détection des lésions spiculées. Ils ont utilisé une analyse multi-résolution basée sur la transformée en ondelettes discrètes bidimensionnelles (2D-DWT) pour calculer les caractéristiques statistiques et ont trouvé un taux de précision de $Az = 84,2\%$ en utilisant un arbre binaire comme classificateur dans la classification des mammographies.[223]
- En 2007, Rashed, Ismail, & Zaki, ont proposé un système utilisant la transformée discrète en ondelettes en décomposition multiniveaux, dans leurs travaux ; trois ondelettes sont utilisées et sont décomposées en quatre niveaux de processus de décomposition : les Daubechies-4, Daubechies-8 et Daubechies-16. Les résultats obtenus ont montré que la classification utilisant les ondelettes Daubechies-8 surpasse les autres.[224]
- En 2009, Yoon & Kim ont présenté quatre cadres de machines à vecteurs de support basées sur la sélection de caractéristiques, la machine à vecteurs de support (SVM), l'élimination de caractéristiques récursives SVM (SVM-RFE), la pertinence maximale de redondance minimale

(mRMR) et SVM-RFE avec RMRm. Ils ont proposé un nouveau critère de redondance de mRMR, la redondance des caractéristiques est minimisée en utilisant le coefficient de corrélation absolu de Pearson au lieu de l'information mutuelle. La précision Az obtenue de leur schéma est relativement plus performante que les autres schémas en utilisant le SVM avec la fonction de noyau RBF comme classificateur dans la classification des mammographies.[225]

- En 2009, Lladó et al ont proposé l'utilisation de modèles binaires locaux (LBP) pour représenter les propriétés texturales des masses. Leur travail utilise un classificateur de machine à vecteurs de support (SVM) pour séparer les vraies masses de celles qui sont du parenchyme normal. L'approche a été évaluée à l'aide de 1792 ROI extraites de la base de données DDSM.[227]

- En 2009, le concept d'indice de Moran et les mesures de coefficient de Geary sont proposés. Tout d'abord, une étape d'amélioration du tissu mammaire avec égalisation d'histogramme est appliquée, les caractéristiques de texture définies sont ensuite extraites des tissus mammaires (tissu anormal et normal), de quatre directions, égales à 0° , 45° , 90° et 135° . À l'aide d'une machine à vecteurs de support avec fonction de base radiale comme classificateur, la précision obtenue est de 96,04 % et Az ROC de 0,946 avec le coefficient de Geary et une précision de 99,39 % et Az ROC de 1 avec l'indice de Moran pour discriminer les tissus dans les mammographies comme normaux ou anormaux. Ils ont également obtenu une précision de 88,31 % et Az ROC de 0,804 avec le coefficient de Geary et une précision de 87,80 % et Az ROC de 0,89 avec l'indice de Moran pour discriminer les tissus dans les mammographies comme bénins et malins.[228]

- En 2011, Tahmasbi et al ont utilisé les moments de Zernike comme des descripteurs de caractéristiques de forme et de marge des masses. Un classificateur Perceptron multicouche (MLP), utilisant à la fois des algorithmes génériques de rétropropagation (BP) et d'apprentissage basé sur l'opposition (OBL), est choisi et appliqué pour la classification des mammographies. Les systèmes conçus donnent Az = 0,976, ce qui représente une sensibilité passable, et Az = 0,975, démontrant une spécificité passable. Les meilleurs FNR et FPR obtenus sont respectivement de 0,0 et 5,5 %.[229]

- En 2012, Ramirez-Villegas & Ramirez-Moreno ont présenté deux schémas de classification de modèles tels que les machines à vecteurs de support (SVM) avec polynôme, fonction de base radiale (RBF) et noyaux quadratiques et réseau de neurones (NN) avec fonction d'activation sigmoïde. Les performances obtenues ont atteint un taux de précision de 93,75 % en utilisant le MLP avec fonction d'activation sigmoïde comme classificateur dans la classification de la microcalcification. La sélection des caractéristiques a été effectuée en appliquant un schéma hybride intégrant un test KS non paramétrique, une analyse de corrélation, un modèle de régression logistique (LR) et une sélection séquentielle en avant (SFS).[230]

- En 2012, Wu, Lin et Moon proposent un schéma de classification des mammographies pour classer les tissus des tumeurs mammaires comme bénignes ou malignes, ils ont utilisé une combinaison de caractéristiques texturales et morphologiques. Les performances obtenues ont atteint un taux de précision de 95,24 % en utilisant le SVM avec la fonction de noyau RBF comme classificateur dans la classification des mammographies.[231]

- En 2015, Wajid & Hussain ont proposé un nouvel histogramme de forme basé sur l'énergie comme ensemble de caractéristiques pour la reconnaissance des anomalies dans les mammographies, appelé histogramme de forme basé sur l'énergie locale (LESH). Les performances remarquables ont été obtenues par le classificateur linéaire SVM, qui a fonctionné à un taux stable de 99,73 % sur tous les ensembles de sélection de caractéristiques

multiples et avec une valeur A_z de $0,9975 \pm 0,0010$ lors de l'utilisation de l'ensemble de données INbreast. Les résultats ont été vérifiés pour l'ensemble de données MIAS lorsqu'une précision de 100 % a été obtenue par le classificateur linéaire SVM et qu'une valeur A_z de 1,0000 a été obtenue pour 70 entités LESH sélectionnées. Les deuxièmes classificateurs les plus dominants étaient le polynôme et le RBF ; MLP a montré une précision de performance globale faible.[232]

- En 2015, Vaidehi and Subashini proposent une technique de segmentation basée sur le clustering en utilisant une approche floue K-means. Dans la zone délimitée par le contour de la tumeur, deux ensembles de caractéristiques basées sur l'intensité sont extraits. Tout d'abord, les caractéristiques de texture sont extraites de la matrice de cooccurrence de niveau de gris (GLCM) et sont calculées à distance $d=1$ et $\theta=0^\circ, 45^\circ, 90^\circ, 135^\circ$. Deuxièmement, l'extraction des éléments de moment de Zernike encerclant la masse. En plus des mesures de performance standard, les auteurs ont proposé un autre critère d'évaluation de la précision appelé Mathews Correlation Coefficient (MCC). Les résultats expérimentaux utilisant le SRC (sparse representation classifier) comme Le but de la sélection des caractéristiques est la visualisation des plus pertinentes pour la tâche de classification des mammographies, ainsi son chemin d'attribution de la caractéristique la plus puissante à la plus faible. Il s'agit de techniques qui ne créent pas de nouvelles caractéristiques en utilisant une transformation ou une combinaison à partir de l'espace de données de départ, telles que, par exemple, l'analyse en composantes principales (ACP), l'analyse en composantes indépendantes (plus appropriée que l'ACP pour les distributions non gaussiennes), l'analyse discriminante linéaire (LDA). De plus, le classificateur de disponibilité précédent surpasse les autres schémas. Les performances obtenues avec ce classificateur en termes de précision, de sensibilité, de spécificité et de MCC sont respectivement de 93,75 %, 90,62 %, 99,99 % et 87,35 %.[233]

- Pour améliorer les performances de la classification des mammographies, Zyout et al introduisent plusieurs aspects basés à la fois sur les niveaux de gris de l'image et les coefficients d'ondelettes. Premièrement, les statistiques du premier ordre des coefficients d'ondelettes (FOSWC) et les statistiques du second ordre des coefficients d'ondelettes (SOSWC) basées sur les caractéristiques d'énergie et d'entropie, les descripteurs de cooccurrence (ou Haralick). Deuxièmement, l'aspect GLCM multi-échelles (à savoir MDGLCM), obtenu en concaténant des vecteurs de caractéristiques GLCM correspondant à quatre distances de pixels d , $d \in \{1, 3, 6, 9\}$ au lieu d'autres ensembles de caractéristiques texturales GLCM à distance unique. Troisièmement, un aspect de sélection de modèle qui consiste en une procédure de recherche heuristique basée sur l'optimisation de l'essaim de particules (PSO) dans tout le système CAO. Il est démontré que l'utilisation des caractéristiques SOSWC en termes de validation croisée moyenne et de performances de test à partir de différents ensembles de caractéristiques SOSWC utilisant tous les filtres à ondelettes était supérieure à celles fournies par les caractéristiques texturales FOSWC et GLCM.[234]

- En 2015, les concepts méthodologiques présentés par Oliveira et al sont inspirés de la Phylogénie. Les régions d'intérêt sont divisées en couches (masque interne et externe) et soumises à la sous-étape d'extraction de caractéristiques de texture à l'aide d'un arbre phylogénétique pour décrire les relations évolutives entre les espèces (intensités de pixels). L'indice de diversité taxonomique (Δ) et la distinction taxonomique (Δ^*), initialement utilisés en écologie, ont été utilisés pour décrire la texture des régions d'intérêt. Ces indices ont été calculés sur la base d'arbres phylogénétiques pour générer une matrice de caractéristiques.

Ses valeurs sont utilisées par le classificateur SVM avec la fonction de base radiale (RBF) pour classer les régions en masse et non masse. La méthodologie proposée a réussi à classer les masses et les non-masse, avec une précision moyenne de 98,88 %.[235]

- En 2009, Subashini et al ont comparé deux schémas de classification de modèles tels que les machines à vecteurs de support (SVM) avec noyau polynomial et réseau de neurones (RBFNN) avec fonction de base radiale. Les mesures de performance sont calculées en termes de précision, d'indice de Youden et de pouvoir discriminant. Pour SVM avec fonction de noyau polynomial, ces mesures de performance sont respectivement de 92,1316 %, 0,823468 et 2,907. Alors que pour RBFNN ces mesures sont respectivement de 96,5664%, 0,924942 et 3,6304. [236]

- En 2015, les mesures de performances sont calculées par rapport à la normale vs. anormal et bénin vs malin. Pour la base de données MIAS, les valeurs AUC sont respectivement de 0,9899 et 0,9504. Alors que pour la base de données DDSM, ces mesures AUC ont atteint 0,9945 et 0,9945. Pour la base de données MIAS, les mesures de précision sont respectivement de 98,0 % et 94,2 %, alors que pour la base de données DDSM, elles sont de 98,8 % et 97,4 %. Ils ont utilisé une analyse multirésolution basée sur la transformation en ondelettes discrètes bidimensionnelles (2D-DWT) ainsi qu'une matrice de cooccurrence de niveaux de gris (GLCM) pour calculer les caractéristiques de texture, avec deux algorithmes de sélection de caractéristiques pour la réduction de la dimensionnalité par sélection, à l'aide d'un réseau de neurones comme classificateur.[237]

- En 2016, Shi et al ont proposé un nouvel algorithme d'apprentissage en profondeur appelé réseau polynomial profond (DPN) pour améliorer les performances de représentation de la caractéristique de texture extraite pour un petit ensemble de données ultrasonores ; l'algorithme Stacked DPN est proposé pour améliorer encore la capacité d'apprentissage en profondeur du DPN d'origine. Les résultats expérimentaux montrent que le S-DPN atteint les meilleures performances avec des précisions de classification de $92,40 \pm 1,1$ % et $90,28 \pm 2,78$ sur les ensembles de données d'échographie du sein et de la prostate, respectivement.[238]

- En 2016, Kong et al ont proposé une nouvelle méthode de sélection de caractéristiques et d'analyse discriminante appelée Jointly Sparse Discriminant Analysis (JSDA). Des expériences sur trois ensembles de données sur le cancer du sein (Breast cancer Wisconsin (BCW), Wisconsin Prognostic Breast Cancer (WPBC) et cancer du sein) montrent que la JSDA surpasse plusieurs méthodes d'apprentissage sur les données du cancer pour le diagnostic.[239]

- En 2016, Un nouvel cadre qui utilise ELM (Extreme Learning Machine) comme classificateur est proposé par Xie et al, le modèle de segmentation de la masse est basé sur la méthode des ensembles de niveaux, la méthode de sélection des caractéristiques basée sur la combinaison de SVM et ELM est ensuite prise en compte. Un sous-ensemble optimal de 30 caractéristiques sélectionnées à partir de 300 images mammographiques dans DDSM et 60 mammographies dans MIAS sont considérées comme un ensemble de données à classer par ELM. Les performances obtenues à l'aide de ce classificateur en termes de précision ont atteint 96,02 %.[240]

- En 2017, Khan et al ont proposé un réseau de neurones à ondelettes (WNN) comme classificateur, ce dernier pouvant former des paramètres différentiables et non différentiables en utilisant le concept de programmation génétique cartésienne (CGP). Leur approche a été évaluée sur la tâche à deux spirales et sur des ensembles de données du monde réel pour la

détection du cancer du sein et de la maladie de Parkinson. Les performances de CGPWNN sur le jeu de données DDSM sont comparées à l'aide de plusieurs structures d'ondelons et les précisions d'apprentissage augmentent pour toutes les stratégies lorsque le nombre d'ondelons dans la couche cachée, et ils ont atteint une précision maximale de 98,48 % avec une répartition des données 50/50 et 97,76 % de précision avec validation croisée 10 fois. [241]

- Un ensemble de 1070 mammographies de la base de données DDSM ont été utilisé dans les travaux de Soriano et al, cette base est divisée en deux catégories : les mammographies bénignes et malignes. MSE et CCE (Entropie croisée catégorielle) sont utilisées comme fonctions de coût. Ils ont implémenté Adadelta, Rmsprop et SGD de la bibliothèque TensorFlow comme méthodes d'optimisation. La meilleure précision et la MSE sont obtenues à l'aide de l'optimiseur SGD. Ils sont respectivement de 85,00 % et 15 % en utilisant un classificateur CNN dans la classification mammaire maligne/bénigne.[242]

- En 2017, Zhang et al. ont proposé une technique de classification en trois classes (cas normaux, malins et bénins) pour un grand ensemble de données de mammographies en utilisant une méthode d'apprentissage en profondeur. Ils ont présenté deux méthodes, à savoir l'ondelette discrète de réseau de neurones convolutif (CNN-DW) et la transformée de réseau de neurones convolutif-curvelet (CNN-CT). L'ensemble de données Image Retrieval in Medical Applications (IRMA) est filtré par l'égalisation d'histogramme adaptatif à contraste limité (CLAHE) pour améliorer le contraste des images de mammographie. Dans la méthode CNN-DW, les images de mammographie améliorées sont décomposées en quatre sous-bandes au moyen d'une transformée en ondelettes discrète bidimensionnelle (2D-DWT), tandis que dans la deuxième méthode, la transformée en courbes discrètes (DCT) est utilisée. Dans les deux méthodes, la caractéristique invariante d'échelle dense (DSIFT) pour toutes les sous-bandes est extraite. Une matrice de données d'entrée contenant ces caractéristiques de sous-bande de tous les patches de mammographie est créée et traitée comme entrée dans le réseau neuronal convolutif (CNN). La couche Softmax et la couche de machine à vecteurs de support (SVM) sont utilisées pour former CNN à la classification. En utilisant l'ensemble de données IRMA, CNN-DW et CNN-CT ont atteint un taux de précision de 81,83 % et 83,74 %, respectivement.[243]

2.6. Conclusion

En conclusion, ce chapitre offre une perspective détaillée sur les méthodes de prétraitement, d'analyse et d'amélioration d'images médicales, ainsi que sur l'analyse et la quantification des textures et des morphologies. Il souligne l'importance de ces étapes dans le domaine de l'imagerie médicale pour obtenir des images de haute qualité, qui va faciliter l'interprétation clinique, et d'améliorer la précision des diagnostics. Les différentes techniques présentées, telles que la matrice de cooccurrence, la segmentation d'images, la transformée de Fourier, et bien d'autres, démontrent la diversité des outils à disposition des professionnels pour extraire des informations pertinentes à partir d'images médicales.

De plus, ce chapitre met en lumière la nécessité d'adapter ces méthodes en fonction des caractéristiques spécifiques des images médicales et des besoins de chaque application, tout en reconnaissant les limites et les complexités associées à certaines de ces techniques.

Enfin, il appuie l'importance de la représentation, du filtrage et de l'analyse fréquentielle des images médicales, ainsi que de l'utilisation de filtres de diffusion anisotropique pour améliorer la qualité des données tout en préservant les détails essentiels.

CHAPITRE 3

Modélisation, Sélection et Agrégation des Caractéristiques par les systèmes de files d'attente

3.1. Introduction

Les concepteurs de systèmes de traitement de l'information ont besoin de méthodes pour quantifier les facteurs de conception tels que la performance et la fiabilité. Les systèmes modernes de traitement de l'information, de communication et de chaîne de production doivent faire face à des charges de travail complexes et des demandes de service aléatoires. Pour évaluer la performance et la fiabilité de ces systèmes, on utilise couramment des méthodes probabilistes et statistiques.

Ce chapitre vise à explorer les principales techniques de modélisation probabiliste utilisées dans l'amélioration de performance des systèmes de traitement de l'information. Une approche présentée combine des réseaux neuronaux artificiels avec des réseaux de files d'attente pour la sélection de caractéristiques. L'utilisation de files d'attente permet de stocker l'erreur et de pondérer les matrices de poids pour chaque couche du réseau. Cette approche peut être utilisée pour des systèmes avec un grand nombre de ressources et des caractéristiques complexes. La sélection de caractéristiques permet d'identifier les variables pertinentes qui améliorent les performances du modèle d'apprentissage. En outre, les concepts de réseaux de file d'attente ouverts et fermés sont introduits pour décrire différents types de réseaux. [243, 244]

Les mesures de performances qui revêtent un intérêt commun incluent le débit, l'utilisation des ressources, la probabilité de perte et le délai (ou temps de réponse). Ces mesures fournissent des informations importantes pour évaluer la capacité des systèmes à répondre aux demandes et à fonctionner efficacement. En plus, ce chapitre fournira une exploration approfondie des techniques de modélisation des algorithmes de sélection de caractéristiques utilisées pour analyser les performances des systèmes d'apprentissage automatique. Nous soulignerons également l'importance des méthodes statistiques et examinerons diverses mesures de performances pertinentes pour évaluer les systèmes.

3.2. Modélisation et Abstractions pour l'Évaluation des Performances : Vers une Compréhension du Système Réel

Pendant le processus de formalisation, les abstractions et le processus de modélisation dans le cadre d'une évaluation de performance basée sur un modèle sont appliquées par rapport au système réel :

Modélisation et conceptualisation dans l'évaluation des performances

- 1. Le processus de modélisation :** La première étape d'une évaluation de performance basée sur un modèle consiste en un processus de formalisation. Pendant cette étape, le modélisateur génère une description formelle du système du monde réel. L'objectif est de passer d'une description informelle du système (par exemple, en langage naturel) à un modèle formel du système réel en utilisant une conceptualisation. [270]
- 2. Conceptualisation :** Une conceptualisation est un résumé simplifié de la réalité de référence, représentée dans un certain but. Deux types de conceptualisations sont présentés pour évaluer les performances :
 - a. Réseau de file d'attente :** Dans cette conceptualisation, le système est représenté comme un réseau de stations-service reliées par des bords par lesquels des entités circulent. Ces entités séjournent dans les files d'attente et les serveurs des stations-service. Ce modèle est davantage orienté vers la structure du système réel.
 - b. Chaîne de Markov :** Dans cette autre conceptualisation, le système est modélisé à l'aide d'un graphe représentant les transitions possibles entre les états du système. Ce modèle met l'accent sur la description du comportement du système au niveau de l'espace d'état sous-jacent et "imite" le comportement du système du monde réel. [270]
- 3. Abstractions appliquées au cours de la formalisation :** Pendant le processus de formalisation, certaines abstractions sont appliquées par rapport au système du monde réel :
 - a. Discrétisation du comportement du système :** Même si le système du monde réel peut avoir des composants à comportement continu, le modèle les considère comme évoluant de manière discrète.
 - b. Abstraction des mécanismes de synchronisation :** Les mécanismes de synchronisation présents dans le système du monde réel sont abstraits dans les deux modèles. Si ces mécanismes sont décisifs pour obtenir des résultats utiles de l'évaluation, des variantes de réseaux de Petri stochastiques peuvent être utilisées comme description alternative.
 - c. Utilisation de variables aléatoires :** Les durées d'activité du système sont associées à des variables aléatoires, et des probabilités de branchement sont utilisées pour représenter des évolutions alternatives du système.
 - d. Résolution du non-déterminisme :** Les deux modèles résolvent le non-déterminisme inhérent au système du monde réel en transformant le réseau de file d'attente formel ou le prototype de chaîne de Markov en une spécification "exécutable". Cela signifie qu'à tout moment de leur fonctionnement, chaque évolution future possible a une probabilité bien définie d'occurrence.
 - e. Interprétation en temps discret et continu :** Selon le type de variables aléatoires utilisées pour représenter les durées des activités du système, on peut obtenir une interprétation en temps discret (DTMC) ou en temps continu (CTMC) du comportement du système. L'interprétation en temps continu est plus appropriée pour les systèmes avec des composants évoluant de manière asynchrone, tandis que l'interprétation en temps discret est généralement utilisée pour les systèmes dont les composants évoluent de manière séquentielle déclenchée par une horloge globale. [270]

3.2.1. Évaluation de la performance de modélisation

La deuxième étape de l'évaluation de la performance d'un système basée sur un modèle se concentre sur les différentes méthodes de résolution afin de déduire des mesures de performance qui peuvent être utilisées selon la conceptualisation choisie lors du processus de formalisation.

- 1. Méthodes de résolution disponibles :** Selon la conceptualisation choisie lors du processus de formalisation, on distingue trois principales catégories des méthodes de résolution:
 - a. Solutions analytiques :** Les méthodes analytiques consistent à représenter la description formelle du système sous forme d'une équation unique ou d'un ensemble d'équations du système. À partir de ces équations, les mesures de performance intéressantes peuvent être obtenues sous forme de solutions fermées ou approximatives en utilisant des algorithmes appropriés issus des mathématiques numériques. Les mesures peuvent être calculées soit par programmation ad-hoc, soit à l'aide de progiciels de calcul formel tels que Mathematica [247].
 - b. Solutions numériques :** Pour certaines descriptions de système, il n'est pas possible d'obtenir des solutions de forme fermée. Dans de tels cas, des solutions approchées peuvent être obtenues en utilisant des algorithmes numériques implémentés dans des packages de calcul formel ou des outils d'analyse de performance. Ces méthodes sont intégrés dans des outils d'analyse de performance tels que SHARPE [248], SPNP [249] ou TimeNET [250], et sont appliquées aux systèmes modélisés sous forme de réseaux de files d'attente, de réseaux de Petri stochastiques ou d'autres formalismes de modélisation de haut niveau.
 - c. Solutions de simulation :** Lorsqu'aucune méthode de solution analytique n'est réalisable pour des modèles complexes, les solutions peuvent être obtenues par la simulation à événements discrets (DES). Cette méthode "exécute" le modèle en collectant des informations sur le comportement observé pour dériver ultérieurement les mesures de performance. Les sorties de simulation de multiples exécutions du modèle sont collectées et utilisées pour calculer les mesures d'intérêt à l'aide de méthodes statistiques. Les réseaux de files d'attente, les réseaux de Petri stochastiques ou les chaînes de Markov peuvent être utilisés comme entrée pour la simulation à événements discrets, ce qui en fait la méthode de résolution la plus flexible et généralement applicable.
- 2. Solutions hybrides :** l'utilisation de solutions hybrides consiste à combiner différents formalismes de modélisation et méthodes de résolution pour exploiter leurs forces complémentaires. Par exemple, des approches mixtes de simulation et analytiques/numériques peuvent être utilisées, ou la combinaison d'arbres de défaillances, de schémas fonctionnels de fiabilité, de graphes de fiabilité et de modèles de Markov. Cette approche de combinaison de méthodes basées sur l'espace d'états et non basées sur l'espace d'états permet de tirer parti des avantages de chaque méthode. [251-253]
- 3. Tolérance de grandeur et évitement de la largeur :** Les techniques de spécification de haut niveau, telles que les systèmes de files d'attente, les réseaux de Petri stochastiques généralisés (GSPN) et les réseaux de récompense stochastiques (SRN), peuvent être utilisées pour automatiser la génération de modèles. Une autre

approche pour traiter les grands modèles consiste à éviter leur création dès le début en utilisant des techniques d'évitement de la largeur, telles que les réseaux de mise en file d'attente de forme de produit. Ces approches permettent de gérer efficacement la complexité des modèles et de réduire les exigences de calcul.

3.2.2. Phases et activités du processus d'évaluation des performances de modélisation

La Figure 3.1 présente un pseudo-code qui résume les différentes phases et activités du processus d'évaluation des performances basé sur un modèle. Deux scénarios principaux sont envisagés : le premier concerne l'évaluation des performances basée sur un modèle appliquée aux premières phases du développement du système afin de prédire les propriétés de performance ou de fiabilité du produit final. Si les propriétés prévues ne répondent pas aux exigences données, des modifications doivent être apportées à la conception proposée afin d'éviter les problèmes de performances attendus. Dans le deuxième scénario, le produit final est déjà disponible, et une évaluation des performances basée sur un modèle est utilisée pour déterminer les paramètres de configuration système optimaux, résoudre des problèmes de planification de capacité ou vérifier si le système existant fonctionnerait de manière satisfaisante après une modification de son environnement.

Dans les deux scénarios, la première activité du processus d'évaluation consiste à recueillir des informations sur la structure et le comportement fonctionnel du système existant ou prévu. La participation d'un expert du domaine est souvent essentielle, en particulier pour les applications complexes. Les informations collectées sont généralement exprimées de manière informelle et enregistrées dans un document sous forme textuelle ou combinée textuelle/graphique. Selon l'objectif de l'étude d'évaluation, la description informelle inclut également les exigences de performance, de fiabilité, de disponibilité, de sûreté de fonctionnement ou de performabilité que le système doit satisfaire.

À partir de la description informelle, le modélisateur procède à la formalisation du système. Si une conceptualisation de chaîne de Markov est choisie, une représentation formelle au niveau de l'espace d'état est créée. En revanche, si un réseau de file d'attente ou un concept de réseau de Petri stochastique est choisi, un modèle formel de haut niveau est dérivé. Ce modèle formel représente le système et son interaction avec l'environnement à un niveau conceptuel, en négligeant les détails considérés comme non pertinents pour l'évaluation.

Dans le cas d'un système existant, une validation conceptuelle de l'exactitude du modèle de haut niveau peut être effectuée par un processus itératif de raffinement par étapes. Cette validation conceptuelle peut être affinée davantage par une "validation apparente", où les experts du domaine tentent de parvenir à un consensus sur la pertinence du modèle par le biais de dialogues, et par une "validation des hypothèses du modèle", qui consiste à vérifier les hypothèses implicites ou explicites. [259, 260]

Certaines propriétés importantes du modèle sont vérifiées en répondant à des questions telles que sa logique, son exhaustivité, son niveau de détail, la justification des hypothèses de distribution, la sensibilité des résultats aux simplifications des hypothèses de distribution, la validité des autres propriétés stochastiques telles que l'indépendance, et si le modèle est représenté au niveau approprié pour le contexte de l'application. [261, 262]

La déduction des mesures de performance peut être réalisée en appliquant des méthodes de résolution fermée/simulation basées sur une description de haut niveau, ou par une analyse numérique/simulation d'une chaîne de Markov. Le lien entre le modèle de haut niveau et le

modèle sémantique est établi par l'activité de génération d'espace d'état, qui est souvent automatisée à l'aide d'outils d'analyse de performance appropriés. L'utilisation de tels outils permet de générer un modèle sémantique qui représente une implémentation de niveau inférieur reflétant toutes les propriétés du système telles que décrites dans le modèle de haut niveau.

Une fois les mesures de performance dérivées, elles peuvent être validées en les comparant aux données recueillies par des mesures, si un système existant est disponible. Les résultats de la validation sont utilisés pour modifier le modèle de haut niveau. Dans le scénario de développement du système, les mesures de performance calculées sont utilisées pour vérifier si le produit final satisfait les exigences de performance.

Si la réponse à l'étape de vérification est "NON", la description formelle du système doit être revue et le processus d'évaluation des performances doit recommencer. Si la réponse est "OUI", le processus de développement du système peut passer aux phases ultérieures de conception détaillée et de mise en œuvre.

Dans les scénarios de configuration du système et de planification de la capacité, il existe une forte interdépendance entre les mesures et la modélisation. Les données de mesure doivent être utilisées pour valider le modèle, tandis que la paramétrisation du modèle repose largement sur les données d'entrée des résultats de mesure. Les paramètres du modèle sont souvent dérivés à partir de mesures effectuées lors d'études antérieures sur des systèmes similaires, y compris dans le scénario de développement du système. En retour, les études de mesure peuvent être mieux planifiées et exécutées si elles sont complétées et guidées par une évaluation basée sur un modèle.

```
// Phase 1: Collecte des informations et description informelle du système
```

```
informalDescription = collectInformalDescription()
```

```
// Phase 2: Formalisation du modèle
```

```
modelType = selectModelType()
```

```
if modelType == "Chaîne de Markov":
```

```
    formalModel = createMarkovChainModel(informalDescription)
```

```
else if modelType == "Réseau de file d'attente":
```

```
    formalModel = createQueueingNetworkModel(informalDescription)
```

```
else if modelType == "Réseau de Petri stochastique":
```

```
formalModel = createStochasticPetriNetModel(informalDescription)
```

```
// Phase 3: Validation conceptuelle du modèle
```

```
conceptualValidation = iterativeRefinement(formalModel)
```

```
apparentValidation = validateApparent(conceptualValidation)
```

```
assumptionsValidation = validateModelAssumptions(conceptualValidation)
```

```

// Phase 4: Déduction des mesures de performance
performanceMeasures = derivePerformanceMeasures(formalModel)

// Phase 5: Validation des mesures de performance
validationResults = null
if existingSystemAvailable:
    validationResults = validateMeasuresWithExistingSystem(performanceMeasures)

// Phase 6: Modification du modèle de haut niveau (si nécessaire)
if validationResults == "NON":
    reviseFormalModel(formalModel)
go to Phase 1

// Phase 7: Utilisation des mesures de performance pour la vérification des exigences
if validationResults == "OUI":
    verifyPerformanceRequirements(performanceMeasures)
// Phase 8: Phases ultérieures du développement du système (conception détaillée et mise en
œuvre)
if validationResults == "OUI":
    detailedDesign = createDetailedDesign(performanceMeasures)
implementation = implementSystem(detailedDesign)

```

Figure. 3.1. Pseudo code d'application de l'évaluation de la performance du modèle dans le processus du développement du système. (Extrait de [270])

3.3. Définitions et notation des systèmes de réseaux de files d'attente

Les réseaux de file d'attente, également connus sous le nom de systèmes de file d'attente, sont des modèles mathématiques utilisés pour analyser et étudier les phénomènes de files d'attente et de congestion dans les systèmes de traitement de demandes. Ils sont largement utilisés dans divers domaines tels que les télécommunications, les réseaux informatiques, les systèmes de transport, les centres d'appels, les systèmes de distribution, etc.

3.3.1. Réseaux mono classe

Un **réseau de files d'attente** est un ensemble de files interconnectées dans lequel les clients peuvent être dirigés après avoir été servis d'une file vers une autre. Une file est composée d'une file d'attente et d'un ou plusieurs serveurs. Les clients qui arrivent dans la file d'attente attendent d'être servis par un serveur avant de quitter la file. (Voir Figure 3.2)

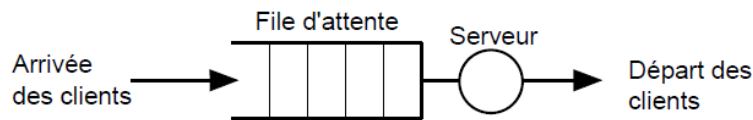


Figure 3.2.File simple à un serveur.

On distingue deux types de réseaux, mono classes c'est-à-dire les réseaux où tous les clients appartiennent à une seule classe, et multi-classes dont le modèle mono-classe peut être étendu en incluant plusieurs catégories d'emplois dans le réseau. Différents aspects liés à la notion des réseaux de files d'attente peuvent être spécifiés :

1. **Capacité et discipline de service** : Chaque file d'attente a une capacité maximale qui détermine le nombre maximal de clients qu'elle peut contenir. Si une file atteint sa capacité maximale, tout nouveau client arrivant est rejeté. De plus, une discipline de service est utilisée pour déterminer l'ordre dans lequel les clients reçoivent le service. Par exemple, la discipline FIFO (First In First Out) indique que les clients sont servis dans l'ordre d'arrivée.
2. **Modélisation stochastique** : L'arrivée des clients dans les files d'attente est décrite par un processus stochastique, c'est-à-dire un processus probabiliste. De même, le temps de service est également modélisé par une loi de probabilité. La loi exponentielle est mentionnée comme étant la plus couramment utilisée pour modéliser les temps d'arrivée et de service, ce qui induit un processus de Poisson, qui est un processus stochastique markovien (noté M).
3. **Notation de Kendall** : La notation de Kendall est utilisée pour normaliser la description d'une file d'attente. La notation de Kendall comprend les symboles T, X, S, C et Z, qui représentent respectivement la distribution des arrivées, la distribution de service, le nombre de serveurs, la capacité de la file et la discipline de service.
4. **File M/M/1** : M/M/1 est un exemple donné d'une file qui indique que cette file a des temps d'inter-arrivée et de service suivant une distribution exponentielle et indépendante (M/M), et qu'elle est servie par un seul serveur ($S = 1$). Lorsque certaines variables ne sont pas spécifiées, elles prennent des valeurs par défaut, par exemple, pour C non renseignée, cela signifie que la file a une capacité infinie ($C = \infty$) et pour Z non renseigné, cela signifie que la discipline de service est FIFO.
5. **Réseau ouvert et fermé** : Un réseau est dit ouvert si les clients peuvent arriver de l'extérieur du réseau et en ressortir. En revanche, dans un réseau fermé, les clients ne sont ni autorisés à quitter le réseau ni à y entrer.
6. **Réseaux monoclasses et multiclassés** : Dans un réseau multiclassés, les clients se partagent le réseau, mais chaque client appartient à une classe spécifique et est servi puis dirigé en fonction de sa classe. En revanche, dans un réseau monoclasse, il n'y a pas de distinction entre les clients et ils sont traités de manière identique. [263-265]

Les symboles suivants sont utilisés dans la description des réseaux de file d'attente [265a]:

N	Nombre de nœuds
K	Le nombre constant d'emplois dans un réseau fermé
(k_1, k_2, \dots, k_N)	L'état du réseau

k_i	Le nombre de travaux au ième nœud ; pour les réseaux fermés $\sum_{i=1}^K k_i = K$
m_i	Le nombre de serveurs parallèles au ième nœud $m_i \geq 1$
μ_i	Taux de service des travaux au ième nœud
$1/\mu_i$	Le temps de service moyen des travaux au ième nœud
P_{ij}	Probabilité de routage, la probabilité qu'un travail soit transféré au jème nœud après l'achèvement du service au ième nœud. (Dans les réseaux ouverts, le nœud avec l'index 0 représente le monde extérieur au réseau.)
P_{0j}	La probabilité qu'un travail entrant dans le réseau depuis l'extérieur entre en premier dans le jème nœud
P_{i0}	La probabilité qu'un travail quitte le réseau juste après avoir terminé le service au nœud i

$$p_{i0} = 1 - \sum_{j=1}^N p_{ij}$$

λ_{0i}	Le taux d'arrivée des emplois de l'extérieur au ième nœud
λ	Le taux d'arrivée global de l'extérieur vers un réseau ouvert $\lambda = \sum_{i=1}^N \lambda_{0i}$
λ_i	le taux d'arrivée global des emplois au ième nœud

3.3.2. Réseaux multi classes

Le type de modèle décrit dans la section précédente peut être étendu en incluant plusieurs catégories d'emplois dans le réseau. Les classes de travaux peuvent différer dans leurs temps de service et dans leurs probabilités de routage. La distinction entre les classes de travaux dans les réseaux de files d'attente.

- 1. Différence entre les classes de travaux :** Dans un réseau de files d'attente, il peut exister plusieurs classes de travaux. Chaque classe de travail peut avoir des caractéristiques différentes, telles que des temps de service distincts et des probabilités de routage différentes. Cela signifie que les travaux appartenant à différentes classes peuvent être servis à des vitesses différentes et suivre des chemins de routage distincts dans le réseau.
- 2. Changement de classe de travail :** Le travail peut changer de classe lorsqu'il passe d'un nœud à un autre dans le réseau. Cela signifie qu'un travail peut être attribué à une classe spécifique lors de son arrivée dans le réseau, mais ensuite, en fonction des règles de routage du réseau, il peut changer de classe lorsqu'il est dirigé vers un autre nœud ou service.
- 3. Classes ouvertes et fermées :** Une classe de travail est dite fermée si le nombre de travaux dans cette classe est constant, c'est-à-dire qu'aucun travail de cette classe n'entre ou ne sort du réseau. En d'autres termes, le nombre de travaux dans la classe reste fixe tout au long de la simulation ou de l'observation. En revanche, une classe de

travail qui n'est pas fermée est dite ouverte, ce qui signifie que de nouveaux travaux peuvent entrer dans la classe ou que des travaux peuvent quitter la classe.

- 4. Réseaux mixtes :** Si un réseau de files d'attente contient à la fois des classes ouvertes et fermées, alors on dit qu'il s'agit d'un réseau mixte. Cela signifie qu'il peut y avoir des classes de travaux dont le nombre de travaux reste constant et d'autres classes de travaux où de nouveaux travaux peuvent entrer ou sortir. [270]

Les symboles supplémentaires suivants sont nécessaires pour décrire les réseaux de mise en file d'attente contenant plusieurs classes de travaux, à savoir[265a] :

R Le nombre de catégories d'emplois dans un réseau

k_{ir} Le nombre de travaux de la $r^{\text{ème}}$ classe au $i^{\text{ème}}$ nœud ; pour un réseau fermé :

$$\sum_{i=1}^N \sum_{r=1}^R k_{ir} = K$$

K_r Le nombre d'emplois de la $r^{\text{ème}}$ classe dans le réseau ; pas nécessairement constant, même en réseau fermé :

$$\sum_{i=1}^N k_{ir} = K_r$$

K Le nombre d'emplois dans les différentes classes, appelé la population

vecteur ($K = (K_1, \dots, K_R)$)

S_i L'état du $i^{\text{ème}}$ nœud : ($S_i = (k_{i1}, \dots, k_{iR})$) :

$$\sum_{i=1}^N S_i = K$$

S L'état global du réseau avec plusieurs classes ($S = (S_1, \dots, S_N)$)

μ_{ir} Le taux de service du $i^{\text{ème}}$ nœud pour les travaux de la $r^{\text{ème}}$ classe

$p_{ir,js}$ La probabilité qu'un emploi de la $r^{\text{ème}}$ classe au $i^{\text{ème}}$ nœud soit transféré vers la $s^{\text{ème}}$ classe et le $j^{\text{ème}}$ nœud (probabilité de routage)

$p_{0,js}$ La probabilité dans un réseau ouvert qu'un travail de l'extérieur du réseau entre dans le $j^{\text{ème}}$ nœud en tant que travail de la $s^{\text{ème}}$ classe

$p_{ir,0}$ La probabilité dans un réseau ouvert qu'un travail de la $r^{\text{ème}}$ classe quitte le réseau après avoir été servi au $i^{\text{ème}}$ nœud, donc

$$p_{ir,0} = 1 - \sum_{j=1}^N \sum_{s=1}^R p_{ir,js}$$

X Le taux d'arrivée global de l'extérieur vers un réseau ouvert

$\lambda_{0,ir}$ Le taux d'arrivée de l'extérieur au nœud i pour les travaux de la classe r $\lambda_{0,ir} = \lambda p_{0,ir}$

Xi Le taux d'arrivée des emplois de la r-ème classe au i-ème nœud :

$$\lambda_{ir} = \lambda \cdot p_{0,ir} + \sum_{j=1}^N \sum_{s=1}^R \lambda_{js} \cdot p_{js,ir}$$

pour les réseaux fermés, et on obtient, $p_{0,ir} = 0 (1 < i < N, 1 < r < R)$ et on obtient :

$$\lambda_{ir} = \sum_{j=1}^N \sum_{s=1}^R \lambda_{js} \cdot p_{js,ir}$$

3.3.3. Mesures de performance

3.3.3.1. Réseaux mono classe

Dans les réseaux mono classes, les méthodes analytiques sont utilisées pour calculer les probabilités d'état et d'autres mesures de performance du réseau. Le problème central de la théorie des files d'attente consiste à déterminer *les probabilités d'état stationnaires* $\pi(k_1, k_2, \dots, k_N)$ pour tous les états possibles du réseau. Ces probabilités permettent ensuite de calculer les valeurs moyennes d'autres mesures de performance importantes du réseau :

• **Probabilités marginales** $\pi_i(k)$: pour les réseaux à file d'attente fermée, les probabilités marginales $\pi_i(k)$ que le i-ème nœud contienne exactement $k_i = k$ emplois sont calculées comme suit :

$$\pi_i(k) = \sum_{\substack{k_1, \dots, k_N \\ \sum_{j=1}^N k_j = K \\ k_i = k}} \pi(k_1, \dots, k_N). \quad (3.1)$$

$$\sum_{j=1}^N k_j = K \quad (3.2)$$

Ainsi, $\pi_i(k)$ est la somme des probabilités de tous les états possibles (k_1, \dots, k_N) , $0 \leq k_i \leq K$, qui satisfont la condition $\sum_{j=1}^N k_j = K$ où un nombre fixe d'emplois, K , est spécifié pour le i-ème nœud. La condition de normalisation selon laquelle la somme des probabilités de tous les états possibles (k_1, \dots, k_N) qui satisfont à la condition, $\sum_{j=1}^N k_j = K$ avec $(0 \leq k_i \leq K)$ doit être 1, c'est-à-dire

$$\sum_{\substack{k_1, \dots, k_N \\ \sum_{j=1}^N k_j = K}} \pi(k_1, \dots, k_N) = 1. \quad (3.3)$$

$$\sum_{j=1}^N k_j = K \quad (3.4)$$

De même, pour les réseaux ouverts, nous avons

$$\pi_i(k) = \sum_{k_i=k} \pi(k_1, \dots, k_N). \quad (3.5)$$

avec la condition de normalisation

$$\sum \pi(k_1, \dots, k_N) = 1. \quad (3.6)$$

Nous pouvons maintenant utiliser les probabilités marginales pour obtenir d'autres mesures de performance intéressantes pour les réseaux ouverts et fermés. Pour les réseaux fermés, nous devons tenir compte du fait que $\pi_i(k) = 0 \quad \forall k > K$

• **Utilisation** ρ_i : L'utilisation ρ_i d'un nœud serveur unique avec l'indice i est donnée par

$$\rho_i = \sum_{k=1}^{\infty} \pi_i(k), \quad (3.7)$$

où ρ_i est la probabilité que le $i^{\text{ème}}$ nœud soit occupé, c'est-à-dire

$$\rho_i = 1 - \pi_i(0). \quad (3.8)$$

Pour les nœuds avec plusieurs serveurs, nous avons

$$\rho_i = \frac{1}{m_i} \sum_{k=0}^{\infty} \min(m_i, k) \pi_i(k) = 1 - \sum_{k=0}^{m_i-1} \frac{m_i - k}{m_i} \cdot \pi_i(k), \quad (3.9)$$

et si le taux de service est indépendant de la charge que nous obtenons

$$\rho_i = \frac{\lambda_i}{m_i \mu_i}. \quad (3.10)$$

• **Le débit (Throughput)** λ_i : Le débit λ_i d'un nœud individuel d'indice i représente en général la vitesse à laquelle les travaux quittent le nœud :

$$\lambda_i = \sum_{k=1}^{\infty} \pi_i(k) \mu_i(k), \quad (3.11)$$

où le taux de service $\pi_i(k)$ est, en général, dépendant de la charge, c'est-à-dire du nombre d'emplois au nœud. Par exemple, un nœud avec plusieurs serveurs ($m_i > 1$) peut être considéré comme un serveur unique dont le taux de service dépend de la charge $\mu_i(k) = \min(k, m_i) \cdot \mu_i$, où μ_i est le taux de service d'un serveur individuel. Cela est également vrai pour les taux de service indépendants :

$$\lambda_i = m_i \cdot \rho_i \cdot \mu_i. \quad (3.12)$$

Nous notons que pour un nœud en équilibre, le taux d'arrivée et le débit sont égaux. Notez également que lorsque nous considérons des nœuds avec des tampons finis, les clients qui arrivent peuvent être perdus lorsque le tampon est plein. Dans ce cas, le débit du nœud sera inférieur au taux d'arrivée au nœud.

• **Débit global** λ : Le *débit global* λ d'un réseau ouvert est défini comme la vitesse à laquelle les travaux quittent le réseau. Pour un réseau à l'équilibre, ce taux de départ est égal au taux d'entrée des emplois dans le réseau, c'est-à-dire

$$\lambda = \sum_{i=1}^N \lambda_{0i}. \quad (3.13)$$

Le débit global d'un réseau fermé est défini comme le débit d'un nœud particulier d'indice i pour lequel $e_i = 1$. Alors le débit global des travaux dans les réseaux fermés est

$$\lambda = \frac{\lambda_i}{e_i}. \quad (3.14)$$

• **Nombre moyen d'emplois \bar{K}_i** : le nombre moyen d'emplois au i ème nœud est donné par

$$\bar{K}_i = \sum_{k=1}^{\infty} k \cdot \pi_i(k). \quad (3.15)$$

Du théorème de Little, il s'ensuit que

$$\bar{K}_i = \lambda_i \cdot \bar{T}_i, \quad (3.16)$$

Où, \bar{T}_i dénotes le temps de réponse moyen.

• **Longueur moyenne de la file d'attente \bar{Q}_i** : La longueur moyenne de la file d'attente au i ème nœud est déterminée par

$$\bar{Q}_i = \sum_{k=m_i}^{\infty} (k - m_i) \cdot \pi_i(k), \quad (3.17)$$

ou, en utilisant le théorème de Little,

$$\bar{Q}_i = \lambda_i \bar{W}_i, \quad (3.18)$$

Où \bar{W}_i est le temps d'attente moyen.

• **Temps de réponse moyen \bar{T}_i** : Le temps de réponse moyen des travaux au i ème nœud peut être calculé à l'aide du théorème de Little pour un nombre moyen de travaux donné \bar{K}_i

$$\bar{T}_i = \frac{\bar{K}_i}{\lambda_i}. \quad (3.19)$$

• **Temps d'attente moyen \bar{W}_i** : Si les taux de service sont indépendants de la charge, alors le temps d'attente moyen au nœud i est

$$\bar{W}_i = \bar{T}_i - \frac{1}{\mu_i}. \quad (3.20)$$

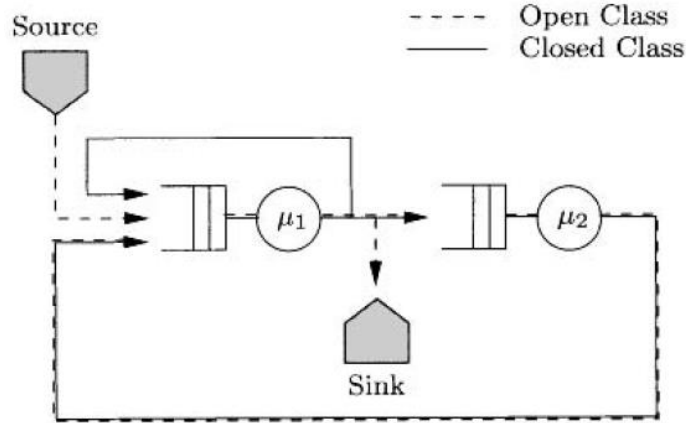


Figure 3.3. Un réseau mixte.

3.3.3.2. Réseaux multiclassés

Lorsqu'on étend les réseaux de files d'attente pour inclure plusieurs catégories d'emplois (réseaux multiclassés), cela entraîne une extension des mesures de performance. Les probabilités d'état d'un réseau multiclassés sont représentées par $\pi(S_1, \dots, S_N)$. La condition de normalisation stipule que la somme des probabilités de tous les états possibles doit être égale à 1.

• **Probabilité marginale $\pi_i(k)$** : pour les réseaux fermés, la *probabilité marginale*, c'est-à-dire la probabilité que le i ème nœud soit dans l'état $S_i = k$, est donnée par

$$\pi_i(k) = \sum_{\substack{\sum_{j=1}^N S_j = K \\ S_i = k}} \pi(S_1, \dots, S_N) \quad (3.21)$$

Et pour les réseaux ouverts, nous avons

$$\pi_i(k) = \sum_{s_i = k} \pi(S_1, \dots, S_N). \quad (3.22)$$

Les formules suivantes pour calculer les mesures de performance peuvent être appliquées aux réseaux ouverts et fermés.

• **Utilisation ρ_{ir}** : L'*utilisation* du i ème nœud par rapport aux emplois de la r ème classe est

$$\rho_{ir} = \frac{1}{m_i} \sum_{\substack{\text{all states } k \\ \text{with } k_r > 0}} \pi_i(k) \frac{k_{ir}}{k_i} \min(m_i, k_i), \quad k_i = \sum_{r=1}^R k_{ir}, \quad (3.23)$$

et si les taux de service sont indépendants de la charge, on a

$$\rho_{ir} = \frac{\lambda_{ir}}{m_i \mu_{ir}} \quad (3.24)$$

• **Débit (Throughput) λ_{ir}** : Le débit λ_{ir} , est le taux auquel les travaux de la r ème classe sont servis et quittent le i ème nœud [266]:

$$\lambda_{ir} = \sum_{\substack{\text{all states } k \\ \text{with } k_r > 0}} \pi_i(k) \frac{k_{ir}}{k_i} \mu_i(k_i); \quad (3.25)$$

ou si les taux de service sont indépendants de la charge, on a :

$$\lambda_{ir} = m_i \cdot \rho_{ir} \cdot \mu_{ir} \quad (3.26)$$

• **Débit global (Overall Throughput) λ_r** :: Le *débit global* des travaux (jobs) de la rème classe dans les réseaux fermés avec multiclassés est

$$\lambda_r = \frac{\lambda_{ir}}{e_{ir}}, \quad (3.27)$$

et pour les réseaux ouverts on obtient

$$\lambda_r = \sum_{i=1}^N \lambda_{0,ir} \quad (3.28)$$

• **Nombre moyen d'emplois (Jobs) \bar{K}_{ir}** : le nombre moyen *d'emplois (jobs)* de la rème classe au ième nœud est

$$\bar{K}_{ir} = \sum_{\substack{\text{all states } k \\ \text{with } k_r > 0}} k_r \cdot \pi_i(k). \quad (3.29)$$

Le théorème de Little peut également être utilisé ici :

$$\bar{K}_{ir} = \lambda_{ir} \cdot \bar{T}_{ir}. \quad (3.30)$$

• **Longueur moyenne de la file d'attente \bar{Q}_{ir}** : la longueur moyenne de la file d'attente des travaux de classe r au ième nœud peut être calculée à l'aide du théorème de Little comme suit :

$$\bar{Q}_{ir} = \lambda_{ir} \cdot \bar{W}_{ir}. \quad (3.31)$$

• **Temps de réponse moyen \bar{T}_{ir}** : le temps de réponse moyen des travaux de la r e classe au ième nœud peut également être déterminé à l'aide du théorème de Little:

$$\bar{T}_{ir} = \frac{\bar{K}_{ir}}{\lambda_{ir}} \quad (3.32)$$

• **Temps d'attente moyen \bar{W}_{ir}** : Si les taux de service sont indépendants de la charge, alors le temps d'attente moyen est donné par

$$\bar{W}_{ir} = \bar{T}_{ir} - \frac{1}{\mu_{ir}}. \quad (3.33)$$

3.3.3.3. Réseaux de file d'attente de forme de produit

Dans cette section, les réseaux de files d'attente qui ont une structure spéciale appelée réseaux en forme de produit ou réseaux séparables. Ces réseaux ont la particularité de permettre des solutions sans générer leur espace d'état sous-jacent.

• Équilibre global

Le comportement de nombreux modèles de système de file d'attente peut être décrit à l'aide de CTMC (Continuous-Time Markov Chain). Un CTMC est caractérisé par les taux de transition entre les états du modèle correspondant. Si le CTMC est ergodique, alors un unique vecteur

de probabilité en régime permanent indépendant du vecteur de probabilité initial existe. Le système d'équations pour déterminer le vecteur de probabilité en régime permanent π est donné par $\pi Q = 0$, où Q est la matrice génératrice infinitésimale du CTMC. Cette équation indique que pour chaque état d'un réseau de file d'attente en équilibre, le flux sortant d'un état est égal au flux entrant dans cet état. [270]

Cette conservation du débit en régime permanent peut s'écrire

$$\sum_{j \in S} \pi_j q_{ji} = \pi_i \sum_{j \in S} q_{ij}, \quad \forall i \in S, \quad (3.34)$$

Où q_{ij} est le taux de transition de l'état i à l'état j . Après soustraction $\pi_i \cdot q_{ii}$ des deux côtés de l'équation. (3.34) et en notant que $q_{ii} = -\sum_{j \neq i} q_{ij}$, on obtient l'équation d'équilibre global (voir Eq. (2.61)) :

$$\forall i \in S: \sum_{j \neq i} \pi_j q_{ji} - \pi_i \sum_{j \neq i} q_{ij} = 0, \quad (3.35)$$

Ce qui correspond à l'équation matricielle :

$$TQ = 0$$

(3.36)

Avec la condition de normalisation

$$\sum_{i \in S} \pi_i = 1. \quad (3.37)$$

Équilibre local

L'équilibre local dans les réseaux de files d'attente, en présentant des techniques de résolution alternatives aux équations d'équilibre global :

- a) Décomposition des équations d'équilibre global en équations d'équilibre local :** Chandy a remarqué que dans certains cas, les équations d'équilibre global peuvent être décomposées en équations plus simples, appelées équations d'équilibre local. Ces équations d'équilibre local permettent de décrire le comportement du système de manière non ambiguë en évitant la génération et la résolution des équations d'équilibre global, qui peuvent être très coûteuses, surtout pour les grands réseaux. [267]
- b) Algorithmes de résolution efficaces :** Il existe plusieurs algorithmes de résolution efficaces et exacts pour une large classe de réseaux de files d'attente. Ces algorithmes évitent l'utilisation des équations d'équilibre global en tirant profit de certaines hypothèses sur les distributions des temps d'arrivée, des temps de service et de la discipline de file d'attente dans tous les nœuds du réseau. En se basant sur ces hypothèses, des équations d'équilibre locales peuvent être déduites pour chaque nœud individuel du réseau. [267]
- c) Importance du concept d'équilibre local :** Le concept d'équilibre local est la base théorique de l'applicabilité des méthodes d'analyse pour les réseaux de files d'attente. En utilisant l'équilibre local, on peut simplifier considérablement l'analyse des grands réseaux, ce qui est essentiel dans les cas où le nombre d'équations d'équilibre global serait extrêmement grand. [267]

d) La propriété d'équilibre local pour un nœud : signifie le taux de départ d'un état du réseau de file d'attente dû au départ d'un travail du nœud i est égal au taux d'arrivée à cet état dû à l'arrivée d'un travail à ce nœud. Cela peut également être étendu aux réseaux de mise en file d'attente avec plusieurs classes de travaux, c à d le taux de départ d'un état du réseau de file d'attente dû au départ d'un travail de classe r du nœud i est égal au taux d'arrivée à cet état dû à l'arrivée d'un travail de classe r à ce nœud. [267]

• Réseaux de Forme produit

Les réseaux de files d'attente qui ont une solution non ambiguë des équations d'équilibre locales sont appelés réseaux de forme produit. La solution en régime permanent aux probabilités d'état de ces réseaux consiste en des facteurs multiplicatifs, chaque facteur se rapportant à un seul nœud. Cette propriété de réseaux de forme produit permet une simplification significative par rapport aux équations d'équilibre global, car les équations peuvent être divisées en un certain nombre d'équations uniques pour chaque nœud individuel du réseau. Le concept de "forme produit" explique comment cette propriété permet de simplifier le calcul des probabilités d'état stationnaire et des mesures de performance [267] :

Forme produit : Le terme "forme produit" a été introduit pour décrire des réseaux de files d'attente ouverts et fermés, avec des temps d'arrivée et de service distribués de manière exponentielle, et une discipline de file d'attente de type FCFS (First-Come-First-Served) dans toutes les stations. Le résultat le plus important de la théorie des files d'attente est que pour ces réseaux, la solution pour les probabilités d'état stationnaire peut être exprimée comme un produit de facteurs décrivant l'état de chaque nœud du réseau. Cette solution est appelée "solution produit-forme". [268- 271]

- 1. Équilibre local :** La "propriété d'équilibre local" est une condition nécessaire et suffisante pour l'existence de solutions de forme-produit. Elle implique que les probabilités d'état stationnaire peuvent être obtenues en résolvant des équations d'équilibre d'état stable, qui équilibrent la vitesse à laquelle le processus de Markov quitte un état avec la vitesse à laquelle il y entre. L'équilibre local permet de réduire l'effort de calcul, car le nombre d'équations à résoudre est réduit par rapport aux équations d'équilibre global.
- 2. Propriété $M \Rightarrow M$ (Markov implique Markov) :** Une station-service a la propriété $M \Rightarrow M$ si et seulement si elle transforme un processus d'arrivée de Poisson en un processus de départ de Poisson. Si tous les nœuds du réseau ont cette propriété, le réseau a une solution sous forme de produit. Cela signifie que chaque station-service du réseau est caractérisée par un comportement Markovien qui facilite le calcul de la solution produit-forme.
- 3. Propriété d'équilibre de poste :** Une discipline de service a la propriété d'équilibre de poste si les taux de service auxquels les travaux dans une position de la file d'attente sont servis sont proportionnels à la probabilité qu'un travail entre dans cette position. Si le réseau a cette propriété, alors il a une solution sous forme de produit. Cela permet une représentation plus simple des équations d'équilibre et facilite le calcul des probabilités d'état stationnaire.

3.4. Pourquoi les réseaux de files d'attente

Lorsqu'on parle de conceptualisations telles que les files d'attente ou les chaînes de Markov dans le contexte de modélisation, il est important de noter que ces concepts ne se limitent pas uniquement aux systèmes d'évaluation de performance, mais peuvent également être utilisés dans d'autres domaines, y compris le domaine des modèles de machine learning inspirés de la biologie.

Modélisation via les files d'attente

Les modèles de machine learning, tels que les réseaux de neurones, les algorithmes génétiques et les CNN (Convolutional Neural Networks), sont largement inspirés du fonctionnement de certains processus biologiques ou de phénomènes du monde réel. Ces modèles tentent de reproduire certaines caractéristiques ou comportements observés dans la nature afin de résoudre des problèmes complexes. Par exemple :

- 1. Réseaux de neurones :** Les réseaux de neurones artificiels sont inspirés du fonctionnement du cerveau humain. Ils sont composés de "neurones" artificiels interconnectés, qui traitent les informations et apprennent à partir des données fournies. Le concept de neurones interconnectés est une abstraction du fonctionnement des neurones biologiques dans le cerveau. Les réseaux de neurones sont largement utilisés dans l'apprentissage automatique et le deep learning.
- 2. Algorithmes génétiques :** Les algorithmes génétiques sont inspirés du processus de sélection naturelle et de l'évolution biologique. Ils utilisent des mécanismes d'évolution pour trouver des solutions à des problèmes d'optimisation ou de recherche. Les individus au sein de populations d'individus (solutions candidates) subissent des mutations et des croisements pour produire de nouvelles générations, et les individus les mieux adaptés ont plus de chances de survivre et de transmettre leurs caractéristiques aux générations suivantes.
- 3. CNN (Convolutional Neural Networks) :** Les CNN sont un type de réseau de neurones conçu spécifiquement pour le traitement des données en grille, tels que les images. Ils sont inspirés par la manière dont le système visuel humain traite l'information visuelle. Les CNN utilisent des opérations de convolution pour extraire automatiquement des caractéristiques pertinentes des images et sont largement utilisés pour des tâches de reconnaissance d'images et de vision par ordinateur.

Ces exemples montrent comment les concepts de modélisation, tels que les files d'attente et les chaînes de Markov, peuvent être étendus au-delà des systèmes d'évaluation de performance et appliqués à des modèles de machine learning inspirés de la biologie. Cette approche permet de créer des modèles plus complexes et sophistiqués, qui imitent certains aspects du monde réel et peuvent fournir des solutions efficaces à des problèmes difficiles.

3.4.1. Plateformes de catégorisation des méthodes de sélection des caractéristiques

Les systèmes courants de sélection de caractéristique sont très inefficaces à utiliser parce qu'ils impliquent l'utilisateur dans trop de détails de bas niveau. Il serait souhaitable que la méthode de sélection de caractéristiques la plus convenable puisse être recommandée automatiquement. Afin d'automatiser le choix d'une méthode de sélection, plusieurs plateformes ont été proposées [274].

3.4.1.1. Une plateforme préliminaire

Dans [273,274], une plateforme préliminaire (voir Figure 3.4) a été développée pour choisir automatiquement un algorithme de sélection des caractéristiques convenable dans un champ limité. La recherche s'opère uniquement par des algorithmes de sélection de caractéristiques basés sur des mesures de consistance. On distingue quatre algorithmes représentatifs: FOCUS forward exhaustive search, ABB - backward complete search, QBB - hybrid search (random plus ABB), and SetCover-heuristic search. L'idée a été validée en utilisant un nombre estimé (M') de caractéristiques à choisir. L'algorithme utilise M' qui est une évaluation de M par SetCover. M'' est une évaluation de M par QBB, et M est le nombre des caractéristiques optimales qui devraient être sélectionnées. FOCUS et ABB sont utilisés autant de fois que possible. Ils optimisent le sous-ensemble des caractéristiques sélectionnées sous la condition que M ou $N-M$ soient petits. Le facteur temps est impliqué dans l'utilisation de l'estimation de M .

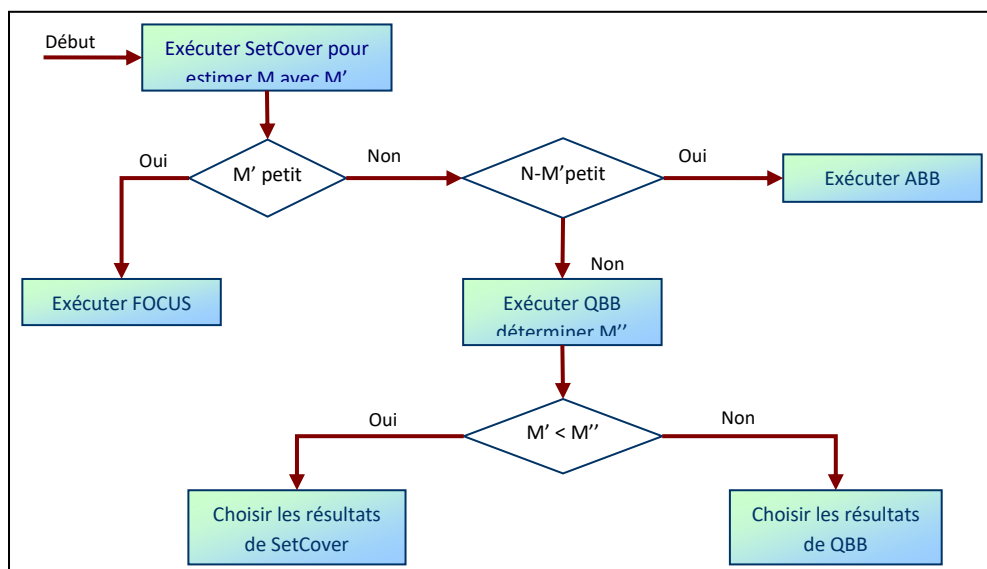


Figure 3.4. Plateforme préliminaire

3.4.1.2. Une plateforme générale

Une plate-forme plus générale a été proposée dans [274], elle inclut la plateforme préliminaire et permet aussi de choisir automatiquement un algorithme de sélection des caractéristiques. Ce choix est déterminé par trois facteurs majeurs: *l'algorithme* utilisé pour sélectionner des caractéristiques selon les mesures d'évaluation et les stratégies de recherche, le type *des données* disponibles, et celui de *la performance* attendue. Chacun des trois facteurs a sa propre dimension comme illustré dans la Figure 3.5.

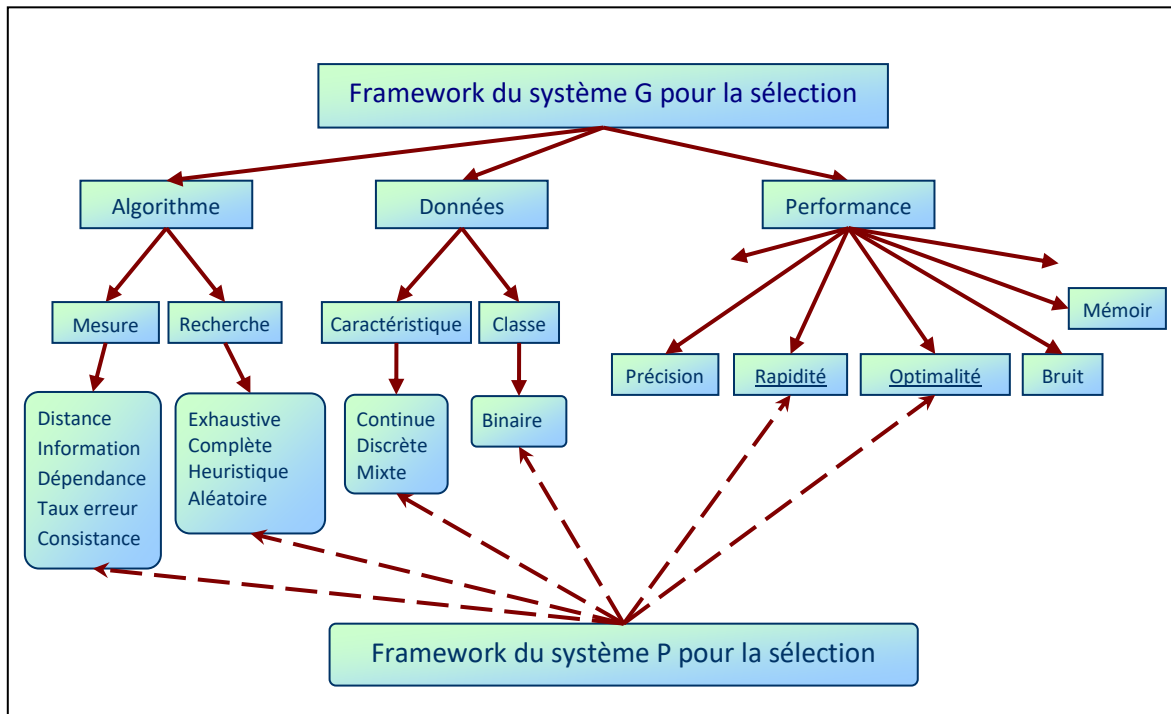


Figure 3.5. Plateforme générale

3.4.1.3. Plateforme unifiée

Dans [274,275], une plate-forme unifiée est proposée (Figure 3.6). Elle est basée sur les deux plateformes précitées et le développement récent des algorithmes de sélection de caractéristiques existants.

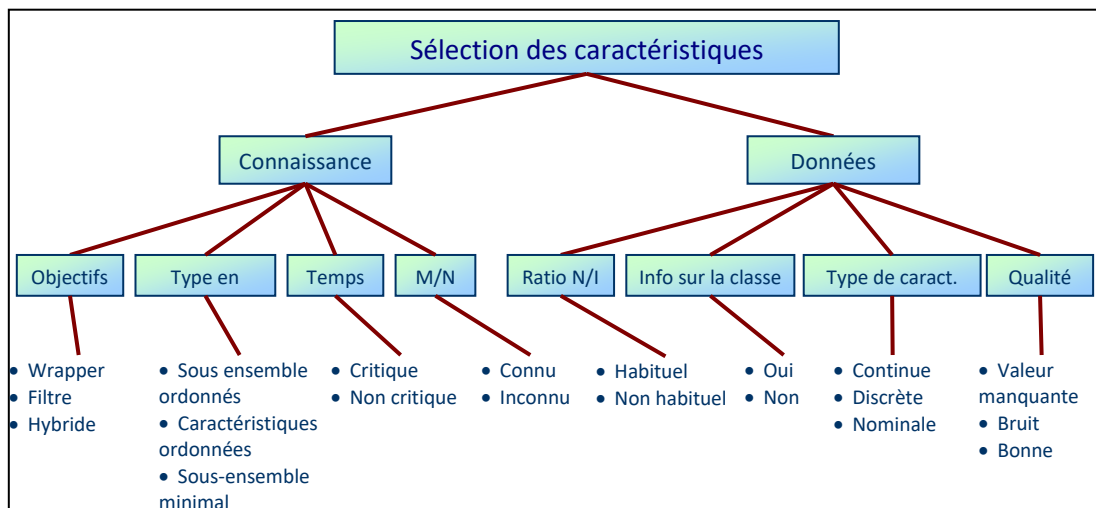


Figure 3.6. Plateforme unifiée

Au sommet de la plateforme, les deux facteurs déterminants sont ceux relatifs aux connaissances et aux données. Le facteur de la connaissance répond à la question de Pourquoi la sélection des caractéristiques est effectuée (Objectifs). L'objectif de la sélection de caractéristiques peut être la visualisation, la compréhension ou le nettoyage des données, l'amélioration des performances, la suppression de la redondance et/ou l'impertinence. Aucune

méthode de sélection des caractéristiques ne peut atteindre tous ces objectifs à la fois, dans [274], ces différents buts sont résumés à Filtre (Filter) ou Symbiose (Wrapper) ou une combinaison des deux.

Si on analyse les méthodes de sélection de caractéristiques selon leur dépendance vis à vis des traitements en aval, on peut distinguer deux approches :

- L'approche Symbiose contenant les méthodes dites « Wrapper methods » qui utilisent la performance en aval pour sélectionner les attributs.
- L'approche par filtre « Filter methods » qui englobe les méthodes de sélection de caractéristiques indépendamment des traitements aval.

Chen et Liu dans [276] énoncent que si l'objectif désiré de la sélection de caractéristiques est l'obtention d'une grande précision alors l'utilisation d'un classifieur est conseillée pour l'évaluation des variables, autrement dit, l'approche *Wrapper* est préférable. Par contre, si le but de la procédure de sélection est la réduction de la dimension et la suppression de variables redondantes, il vaut mieux utiliser une mesure moins coûteuse en termes de temps de calcul et donc utiliser une approche *Filter*. On a parfois besoin de combiner les deux approches pour atteindre des objectifs plus complexes.

L'information sur M/N est très utile pour déterminer la stratégie de la recherche. Si le M est petit, une stratégie de recherche complète en avant (a forward complete search) peut être appliquée; si le N-M est petit, une stratégie de la recherche complète moins évoluée (a backward complete search strategy) peut être adoptée même en situation de temps critique [274].

Le facteur des *données* est relatif aux points suivants:

- Le ratio N/I entre le nombre des caractéristiques N et le nombre d'instances I est devenu récemment un paramètre intéressant surtout lorsque la sélection des caractéristiques est utilisée dans des applications telles que l'analyse exploratoire de textes (text mining), la bio-informatique. En général, un grand nombre de caractéristiques pose beaucoup plus de problèmes qu'un grand nombre d'instances. Si $I \gg N$ ($N > 20$), la recherche est refusée car elle exige des quantités massives de calculs inutiles. Par contre si $N \gg I$ alors il faut s'orienter vers les algorithmes qui travaillent au long de la dimension.
- Peu des études de comparaison des méthodes de sélection de caractéristiques ont été effectuées dans la littérature, on peut citer celles de [277-281]. Selon celle effectuée par [281] la dimensionnalité de l'ensemble de donnée est considérée comme un petit, moyen, une grande échelle si elle appartient aux intervalles $[0, 19]$, $[20, 49]$, $[50, \infty]$ respectivement.
- Information sur la classe : répond à la question, si un algorithme de sélection de caractéristiques est supervisé ou non, autrement dit si les données contiennent une information sur leurs classes d'appartenance.
- Le type de caractéristiques : les caractéristiques continues et discrètes peuvent être ordonnées selon leurs valeurs contrairement aux caractéristiques nominales.
- La qualité des données : bonne, contenant du bruit ou ayant des valeurs manquantes.

3.4.2. Aspects proposés pour améliorer la catégorisation

Dans [275], les auteurs proposent de tenir compte de quelques aspects supplémentaires permettant aux chercheurs dans ce domaine de mieux catégoriser les méthodes de sélection des caractéristiques.

a) Sur apprentissage (Overfitting)

Lorsque la dimension des données est de grande échelle, le risque de sur-apprentissage (Overfitting) est amplifié. Le sur-apprentissage se produit lorsque le modèle d'apprentissage capture non seulement les structures générales des données, mais aussi le bruit spécifique à l'ensemble de données d'entraînement. Cela conduit à une mauvaise performance du modèle sur de nouvelles données ou données de test.

Le sur-apprentissage est un problème sérieux dans les ensembles de données de grande dimension, car plus le nombre de caractéristiques augmente, plus le modèle a de possibilités de trouver des relations artificielles et de mémoriser des exemples spécifiques au lieu de généraliser correctement. Cela peut entraîner une dégradation significative des performances du modèle sur des données inconnues.

La réduction de dimension par sélection de caractéristiques peut contribuer à atténuer le phénomène de sur-apprentissage. En éliminant certaines caractéristiques moins pertinentes ou redondantes, la sélection de caractéristiques réduit la complexité du modèle et limite ainsi les risques de sur-apprentissage. En se concentrant uniquement sur les caractéristiques les plus informatives et discriminantes, la sélection de caractéristiques permet de créer un modèle plus général qui peut mieux généraliser sur de nouvelles données.

En sélectionnant judicieusement les caractéristiques les plus significatives, la réduction de dimension par sélection de caractéristiques peut améliorer la capacité de généralisation du modèle en évitant de mémoriser le bruit et en se concentrant sur les modèles sous-jacents importants dans les données. Cela permet de créer des modèles plus simples, plus faciles à interpréter et plus performants sur de nouvelles données, tout en évitant le sur-apprentissage. Parmi les algorithmes qui souffrent de ce phénomène, on peut citer FOCUS [276], un biais fort vers sa consistance peut être statistiquement injustifié et mener au sur-apprentissage des données d'entraînement. Pour remédier à ce problème, il est possible d'avoir recours à la validation croisée.

b) Non-monotonie

La non-monotonie est un problème qui se produit dans le contexte de la sélection séquentielle de caractéristiques [280] lorsqu'il n'y a pas de garantie que le sous-ensemble optimal de M caractéristiques (où M est le nombre de caractéristiques que l'on souhaite sélectionner) contient le sous-ensemble optimal de P caractéristiques (où P est le nombre total de caractéristiques disponibles). En d'autres termes, lorsque l'on sélectionne un certain nombre de caractéristiques, il est possible que certaines caractéristiques importantes soient omises, même si elles faisaient partie du meilleur sous-ensemble de caractéristiques.

Ce problème de non-monotonie peut entraîner des résultats suboptimaux lors de la sélection de caractéristiques. Si une méthode de sélection de caractéristiques souffre de non-monotonie, cela signifie qu'elle peut sélectionner un sous-ensemble de caractéristiques qui n'est pas le meilleur en termes de performances du modèle, simplement parce que certaines caractéristiques ont été omises dans le processus de sélection.

Pour remédier à ce problème, différentes versions flottantes des méthodes de sélection de caractéristiques ont été développées. Les méthodes flottantes permettent de réévaluer et d'ajuster la sélection de caractéristiques après que certaines caractéristiques ont été choisies, afin de s'assurer que le meilleur sous-ensemble de caractéristiques est effectivement inclus dans la sélection finale.

En utilisant des méthodes flottantes, on peut s'assurer que la sélection de caractéristiques est plus complète et optimale, en prenant en compte toutes les interactions entre les caractéristiques et en ajustant la sélection en fonction des performances du modèle. Cela permet de surmonter le problème de non-monotonie et d'améliorer la qualité de la sélection de caractéristiques, ce qui conduit à des modèles plus performants et plus fiables.

c) Sélection complètement/partiellement linéaire

La sélection complètement ou partiellement linéaire des caractéristiques se réfère à la manière dont la réduction de dimension est appliquée sur l'ensemble des données.

Dans la sélection complètement linéaire, la réduction de dimension par sélection de caractéristiques est appliquée sur l'ensemble complet de données, c'est-à-dire que toutes les instances ou observations de l'ensemble de données sont prises en compte lors du processus de sélection. Cela signifie que la réduction de dimension est réalisée en utilisant l'ensemble des caractéristiques disponibles dans l'ensemble de données pour sélectionner un sous-ensemble optimal de caractéristiques qui sont les plus informatives et pertinentes pour la tâche d'apprentissage.

En revanche, dans la sélection partiellement linéaire, la réduction de dimension est appliquée sur un sous-ensemble limité de données plutôt que sur l'ensemble complet. Cela signifie que seules certaines instances ou observations de l'ensemble de données sont utilisées pour effectuer la sélection des caractéristiques. Ce sous-ensemble peut être choisi de manière aléatoire ou selon certains critères spécifiques.

La sélection complètement linéaire est généralement plus exhaustive et prend en compte l'ensemble de l'information disponible dans les données, ce qui peut conduire à une sélection plus précise et pertinente des caractéristiques. Cependant, elle peut être coûteuse en termes de temps de calcul, surtout pour les ensembles de données de grande taille.

La sélection partiellement linéaire, quant à elle, peut-être plus rapide car elle utilise un sous-ensemble plus restreint de données pour effectuer la sélection des caractéristiques. Cependant, cela peut entraîner une perte d'information et une sélection moins précise des caractéristiques.

Le choix entre la sélection complètement linéaire et la sélection partiellement linéaire dépend des contraintes de calcul, de la taille de l'ensemble de données et des performances souhaitées du modèle. Dans certains cas, la sélection complètement linéaire peut être préférée pour obtenir les meilleures performances du modèle, tandis que dans d'autres cas, la sélection partiellement linéaire peut être suffisante et plus efficace en termes de temps de calcul.

d) Interaction avec l'utilisateur et connaissances à priori:

L'interaction avec l'utilisateur et les connaissances à priori peuvent grandement aider à réduire l'espace de recherche lors du processus de sélection de caractéristiques. Lorsque l'utilisateur a une connaissance spécifique du domaine ou des données, il peut fournir des informations supplémentaires sur la pertinence des caractéristiques, leur corrélation, le type et le niveau de bruit présent dans les données, etc. Ces connaissances peuvent être utilisées pour guider le processus de sélection des caractéristiques de manière plus ciblée et efficace.

En effet, l'utilisateur peut identifier des caractéristiques qui sont a priori considérées comme étant importantes ou non pertinentes pour la tâche d'apprentissage. Ces informations peuvent être utilisées pour partitionner l'ensemble complet de données en plusieurs sous-ensembles de données, chacun contenant des caractéristiques spécifiques. En utilisant ces sous-

ensembles, l'algorithme de sélection de caractéristiques peut se concentrer uniquement sur les caractéristiques jugées pertinentes et ignorer celles qui ne le sont pas, ce qui réduit considérablement l'espace de recherche.

En fournissant une telle interaction avec l'utilisateur et des connaissances à priori, on peut également éviter de sélectionner des caractéristiques qui peuvent être redondantes ou inutiles pour la tâche d'apprentissage. Cela permet de réduire le risque de surapprentissage (overfitting) en éliminant les caractéristiques qui ne contribuent pas de manière significative à la performance du modèle.

e) **Augmentation de la dimension des données**

L'augmentation de la dimension de l'ensemble des données risque d'augmenter explosivement, de manière combinatoire, la dimension de l'espace de la recherche pour le modèle d'induction. De plus, elle augmente les chances qu'un algorithme d'analyse exploratoire des données trouve des modèles non valides.

Lorsqu'on fait face à un ensemble de données de grande dimension, c'est-à-dire avec un grand nombre de caractéristiques, cela peut entraîner ce que l'on appelle "la malédiction de la dimension" (curse of dimensionality). Cette malédiction se réfère aux problèmes et défis qui surviennent lorsque la dimension de l'espace de recherche devient très élevée par rapport à la taille des données disponibles.

Une augmentation de la dimension de l'espace de recherche peut avoir les conséquences suivantes :

- **Explosion combinatoire** : Plus le nombre de caractéristiques augmente, plus le nombre de combinaisons possibles de ces caractéristiques augmente de manière exponentielle. Cela signifie que l'espace de recherche devient immense, rendant difficile pour les algorithmes d'apprentissage de trouver les modèles pertinents et significatifs dans les données.
- **Surapprentissage (Overfitting)** : Avec un grand nombre de caractéristiques, les algorithmes d'apprentissage peuvent être plus susceptibles de surapprendre les spécificités et le bruit présents dans les données d'apprentissage. Cela peut conduire à un modèle qui performe très bien sur les données d'apprentissage, mais qui généralise mal sur de nouvelles données (sur lesquelles le modèle n'a pas été entraîné).

La sélection de caractéristiques vise à remédier à ces problèmes en identifiant les caractéristiques les plus informatives et en éliminant les caractéristiques redondantes, bruitées ou non pertinentes. L'objectif est de réduire la dimension de l'espace de recherche tout en conservant ou en améliorant la précision du modèle d'apprentissage. En sélectionnant les caractéristiques les plus informatives, on peut :

- **Réduire la complexité du modèle** : En réduisant le nombre de caractéristiques, on diminue l'espace de recherche, ce qui permet aux algorithmes d'apprentissage de trouver plus facilement les modèles pertinents et significatifs dans les données.
- **Réduire le risque de surapprentissage** : En éliminant les caractéristiques redondantes ou bruitées, on limite la probabilité que le modèle surapprenne les spécificités des données d'apprentissage.
- **Améliorer la généralisation** : Un modèle plus simple et moins complexe a généralement de meilleures chances de généraliser correctement sur de nouvelles données.

- **Réduire le coût computationnel** : En réduisant la dimension de l'espace de recherche, le temps et les ressources nécessaires pour entraîner le modèle sont également réduits.

f) **Données changeantes et connaissance** :

Un changement rapide (non stationnaire) des données peut rendre invalides les données réduites précédemment sélectionnées. De plus, les variables mesurées peuvent être modifiées, effacées, ou ajoutées avec le temps en intégrant de nouvelles dimensions.

Lorsque les données sont sujettes à des changements rapides et non stationnaires, cela signifie que les caractéristiques informatives peuvent évoluer dans le temps. Les caractéristiques qui étaient pertinentes et utiles à un moment donné peuvent devenir obsolètes ou moins informatives à un autre moment en raison des variations dans les données. De plus, de nouvelles caractéristiques peuvent être introduites dans les données au fil du temps, tandis que certaines caractéristiques existantes peuvent être modifiées ou supprimées.

L'influence de ces changements dans les données sur la sélection de caractéristiques est importante car cela peut affecter la performance du modèle d'apprentissage. Voici comment ces changements peuvent impacter la sélection de caractéristiques :

- **Obsolescence des caractéristiques** : Les caractéristiques qui étaient utiles dans le passé peuvent ne plus être pertinentes dans le présent en raison de l'évolution des données. Lorsque la sélection de caractéristiques est effectuée sur un ensemble de données obsolète, le modèle peut perdre en précision car il se base sur des caractéristiques qui ne reflètent plus le comportement actuel des données.
- **Introduction de nouvelles caractéristiques** : L'apparition de nouvelles caractéristiques dans les données peut apporter de nouvelles informations et nuances qui n'étaient pas présentes auparavant. Si la sélection de caractéristiques n'est pas régulièrement mise à jour pour tenir compte de ces nouvelles caractéristiques, le modèle peut manquer des informations importantes pour bien généraliser sur les données actuelles.
- **Impact des caractéristiques externes** : Lorsqu'on ajoute de nouvelles caractéristiques qui sont en dehors de l'ensemble initial de données, leur relation avec les caractéristiques existantes peut être complexe. Cela peut entraîner des interactions imprévues entre les différentes caractéristiques et rendre la sélection de caractéristiques plus difficile. Il peut être nécessaire d'ajuster les critères de sélection pour prendre en compte ces nouvelles interactions.
- **Coût de mise à jour** : La mise à jour régulière de la sélection de caractéristiques peut être coûteuse en termes de temps et de ressources, surtout si les données sont volumineuses et que les changements sont fréquents. Il est donc important de trouver un équilibre entre la fréquence de mise à jour nécessaire pour garder le modèle à jour et le coût associé à cette mise à jour.

3.4.3. Critères d'analyse d'évaluation individuelles des algorithmes de sélection des caractéristiques

1. Critère de dispersion (Scatter Criterion)

Critère de dispersion, - également connu sous le nom d'analyse discriminante linéaire (LDA) - considérez une matrice de dispersion intra-classe pour chaque classe, mesurant la dispersion des échantillons autour de la moyenne de classe respective, et la matrice de dispersion inter-classes, mesurant la dispersion des moyennes de classe autour de la moyenne du mélange, et trouve une transformation qui maximise la dispersion inter-classes et minimise la dispersion intra-classe, de sorte que la séparabilité des classes est maximisée dans l'espace dimensionnel réduit. [211]

La dispersion intra-classe de la kème caractéristique est

$$SC_{w,k} = \sum_{j=1}^C P_j SC_{jk} , \quad (3.38)$$

Où SC_{jk} est la variance de la classe ω_j , et P_j est la probabilité a priori de cette classe et trouvée par :

$$SC_{jk} = \frac{1}{n_j} \sum_{i=s_j}^{s_j+n_j} (x_{ik} - \bar{x}_{jk})^2 , \quad (3.39)$$

$$P_j = \frac{n_j}{N} . \quad (3.40)$$

Où n_j est le nombre d'échantillons dans la classe ω_j , sur un total de N échantillons. La dispersion inter-classes est la variance des centres de classe par rapport au centre global et est trouvée par

$$SC_{b,k} = \sum_{j=1}^C P_j (\bar{x}_{jk} - \bar{x}_k)^2 . \quad (3.41)$$

Et la dispersion du mélange est la somme des dispersions intra et inter-classes, et est égale à la variance de toutes les valeurs par rapport au centre global.

$$SC_{m,k} = SC_{w,k} + SC_{b,k} = \frac{1}{N} \sum_{i=1}^N (x_{ik} - \bar{x}_k)^2 \quad (3.42)$$

Le critère de dispersion d'une certaine caractéristique est défini comme un rapport de la dispersion du mélange à la dispersion intra-classe. Le critère de dispersion J_k de la kème caractéristique est

$$J_k = \frac{SC_{m,k}}{SC_{w,k}} . \quad (3.43)$$

Plus ce ratio est élevé, plus le pouvoir prédictif de la caractéristique est élevé. Notez également que le critère de dispersion ne dépend pas du fait que les données soient normalisées ou non.

2. Informations mutuelles

Les informations mutuelles sont considérées comme des informations partagées au sein de deux entités de la même classe, et entre une entité et une classe différente. Cette approche est utilisée pour sélectionner les caractéristiques entre chaque caractéristique et les classes cibles [3, 211,296]. Cependant, les caractéristiques avec de grandes valeurs peuvent avoir une plus grande influence sur la fonction de coût que les caractéristiques avec de petites valeurs ; le problème est surmonté en normalisant les caractéristiques afin que leurs valeurs se situent dans des plages similaires. La technique de centrage et de réduction des caractéristiques est la normalisation via les estimations respectives de la moyenne et de la variance. Pour N données disponibles de la k^{ème} caractéristique, nous avons

$$\hat{x}_{ik} = \frac{x_{ik} - \bar{x}_k}{\sigma_k}, \quad \sigma_k^2 = \frac{1}{N} \sum_{i=1}^N (x_{ik} - \bar{x}_k)^2 \quad (3.44)$$

Ensuite, les valeurs normalisées sont discrétisées en L niveaux finis pour faciliter la recherche de probabilités. Les valeurs discrètes correspondantes sont $\tilde{x}_{lk} = 1, 2, \dots, N$ et $k = 1, 2, \dots, M$. L'information mutuelle de la ke caractéristique est

$$MI(\mathbf{x}_k, \omega) = \sum_{l=1}^L \sum_{j=1}^C P(\tilde{x}_{lk}, \omega_j) \log_2 \frac{P(\tilde{x}_{lk}, \omega_j)}{P(\tilde{x}_{lk})P(\omega_j)} \quad (3.45)$$

Où $P(\omega_j)$ est la probabilité a priori de la classe ω_j , $P(\tilde{x}_{lk})$ est la distribution de la kème caractéristique et $P(\tilde{x}_{lk}, \omega_j)$ est la probabilité conjointe. Cet indicateur mesure dans quelle mesure les distributions des valeurs de caractéristiques et des classes cibles diffèrent de l'indépendance statistique.

3. Coefficient de corrélation

La corrélation mesure le degré de dépendance de deux caractéristiques. L'ajustement linéaire -appelé coefficient de corrélation de Pearson- [211] revient à calculer le coefficient de corrélation entre deux variables en recherchant le lien qui existe entre les variables à l'aide d'une droite. La relation linéaire permet de supprimer les entités avec une corrélation linéaire proche de zéro de la classe. Nous utilisons le coefficient de corrélation entre deux caractéristiques k et o par

$$\rho_{ko} = \frac{\sum_{i=1}^N \hat{x}_{ik} \hat{x}_{io}}{\sqrt{\sum_{i=1}^N \hat{x}_{ik}^2 \sum_{i=1}^N \hat{x}_{io}^2}} \quad (3.46)$$

La moyenne de la corrélation absolue entre la caractéristique k et toutes les autres caractéristiques calculées par :

$$\rho_k = \frac{1}{M-1} \sum_{\substack{o=1 \\ o \neq k}}^M |\rho_{ko}| \quad (3.47)$$

C'est une mesure de la redondance de la caractéristique k et de toutes les autres caractéristiques. Une caractéristique avec une ρ_k relativement faible est « unique ».

4. Redondance

Le procédé des méthodes d'évaluation individuelle des caractéristiques (sélection de caractéristiques de classement ou de pondération) est simple d'un point de vue de calcul. Cependant, de telles approches ne prennent pas en compte les caractéristiques redondantes existantes. Les notions de redondance de caractéristiques sont en termes de corrélation de caractéristiques. Il est largement admis que deux caractéristiques sont redondantes si leurs valeurs sont complètement corrélées. Il existe deux types de mesures largement utilisées pour la corrélation entre deux variables : linéaire et non linéaire. Cependant, il n'est pas souhaitable d'utiliser la mesure de corrélation linéaire car la recherche médicale génère une énorme difficulté à interpréter les données de la variabilité naturelle de l'objet d'étude. De plus, les mesures de relation linéaire ne peuvent pas capturer les corrélations d'attributs qui ne sont pas de nature linéaire. Les mesures non linéaires sont basées sur le concept d'entropie, ou la mesure de l'incertitude d'une variable aléatoire. La redondance moyenne de la caractéristique k est définie comme suit :

$$R(\mathbf{x}_k) = \frac{1}{M-1} \sum_{\substack{o=1 \\ o \neq k}}^M I(\mathbf{x}_k, \mathbf{x}_o) \quad (3.48)$$

$$\text{Où } MI(\mathbf{x}_k, \mathbf{x}_o) = \sum_{l=1}^L \sum_{p=1}^L P(\tilde{x}_{lk}, \tilde{x}_{po}) \log_2 \frac{P(\tilde{x}_{lk}, \tilde{x}_{po})}{P(\tilde{x}_{lk})P(\tilde{x}_{po})} \quad (3.49)$$

3.4.4. Application de la plateforme unifiée sur les algorithmes existants

La Figure 3.7 donne un exemple d'utilisation de la plateforme unifiée pour catégoriser des algorithmes existants et, éventuellement, futurs. Nous nous sommes limités aux algorithmes représentatifs de la sélection supervisée de caractéristiques sont à catégoriser selon les aspects suivants : information sur la classe, but, type en sortie, la stratégie de recherche qui est influencée par le temps, et l'information sur M/N.

		Type en sortie			Stratégie de recherche	temps
		Sous-ensemble minimal	Sous-ensembles ordonnés	Caractéristiques ordonnées		
Objectif	Wrapper	B&B, BreathS, DepthS		Ranking Feature	Complète	
		SFS, SBS, BestFirst, BeamS, HillClimbing			Heuristique	
		SFFS, SBFS, ASFFS, ASBFS			Flottante	
				MRMR	Corrélation	
		Algo. génétiques	LVW		Non déterministe	
	Filter	B&B, BreathS, DepthS	ABB, Focus	Ranking Feature	Complète	
		SFS, SBS, BestFirst, BeamS, HillClimbing		Relief	Heuristique	
		SFFS, SBFS, ASFFS, ASBFS			Flottante	
			CFS	FCBF, MRMR	Corrélation	
		Algo. génétiques	LVF		Non déterministe	
		Supervisée			Ratio M/N	
		Information de la classe				

Figure 3.7. Application de la plateforme unifiée sur les algorithmes existants.[275]

L'aspect « Information sur la classe » est relatif au fait que les données contiennent (ou pas) l'information sur la classe. Avec cette information, les algorithmes de sélection de caractéristique supervisés sont applicables, tandis que sans cette information de classe, les algorithmes de la sélection de caractéristique non supervisés sont applicables.

L'aspect « But » est fait pour répondre à la question : Pourquoi la sélection de caractéristique est effectuée ? Le « Type en sortie » de la sélection de caractéristique peut être parfois connu a priori. Cet aspect divise les algorithmes de la sélection de caractéristique en trois groupes: une liste ordonnée des caractéristiques, une liste classifiée des sous-ensembles et les sous-ensembles minimums. La différence entre ces trois groupes est relative à l'ordre entre les caractéristiques sélectionnées. Si un sous-ensemble de caractéristiques est sélectionné, alors il n'y a aucun ordre parmi les caractéristiques. On ne peut pas enlever facilement des caractéristiques du sous-ensemble, mais on peut le faire pour une liste ordonnée en enlevant les caractéristiques les moins importantes.

La stratégie de la recherche est un autre aspect important qui est déterminé par la contrainte du temps et la connaissance de M : le nombre prévu de caractéristiques sélectionnés et N : le nombre total de caractéristiques originales. Le M peut ne pas être connu et quelquefois pourrait être inféré des autres sources de connaissance telle que le but de la sélection de caractéristique, etc. Cinq stratégies de recherche sont identifiées : Complète, Heuristiques, flottante, corrélation et non déterministe.

3.5. Conclusion et perspectives

En conclusion, ce chapitre nous a exposé l'importance des réseaux de files d'attente dans la gestion de l'ordre d'exécution des tâches dans un système informatique. Il a également mis en lumière l'application de ces concepts aux algorithmes de sélection de caractéristiques, des techniques primordiales pour identifier les attributs les plus pertinents dans un ensemble de données.

Nous avons exploré différentes méthodes de résolution pour évaluer la performance des systèmes basés sur des modèles, notamment des approches analytiques, numériques, de simulation, et hybrides, ainsi que des techniques de tolérance de grandeur et d'évitement de la largeur pour gérer des modèles complexes. Chacune de ces méthodes offre ses avantages et est adaptée à des situations d'évaluation spécifiques.

La première étape de l'évaluation des performances basée sur un modèle est la formalisation du système réel à travers une conceptualisation. Deux approches courantes sont la modélisation en réseau de files d'attente et la modélisation en chaîne de Markov, chacune mettant en évidence des aspects différents du système.

En fin de compte, l'application des réseaux de files d'attente aux algorithmes de sélection de caractéristiques peut prendre diverses formes, selon les besoins spécifiques de chaque problème. Il s'agit d'une approche puissante pour gérer l'ordre d'exécution des caractéristiques et maximiser la précision des modèles d'apprentissage, en contribuant ainsi à l'amélioration des performances et à la pertinence des résultats obtenus. En résumé, la compréhension des réseaux de files d'attente est un atout précieux dans le domaine de l'analyse de données et de l'optimisation des systèmes informatiques.

Comme perspective, il est possible de modéliser dans le cadre des chaînes de Markov les processus biologiques tels que les réseaux de neurones comme des systèmes dynamiques stochastiques, où l'état du réseau à un moment donné dépend de son état précédent et est également influencé par des variables aléatoires. La taille de l'espace d'état et le nombre de transitions possibles peuvent cependant devenir rapidement énormes, ce qui rend la modélisation et l'analyse est très complexe et laborieuses, en particulier pour les réseaux de neurones profonds et massivement connectés.

CHAPITRE 4

Approche de Sélection de Caractéristiques pour la Reconnaissance de Masses Mammographiques

4.1. Introduction

Le cancer du sein est un problème de santé majeur dans le monde avec 1,7 million de nouveaux cas de cancer du sein diagnostiqués et 522 000 décès en 2012, la prévalence est de plus de 4,4 millions de femmes [285]. Les taux les plus élevés sont observés dans les pays industrialisés (Amérique du Nord, Australie et Europe de l'Ouest) et sont cinq fois plus élevés que dans les pays en développement (Asie, Afrique). C'est le cancer le plus courant dans les pays en développement, mais les taux augmentent dans les pays à faible incidence, en particulier ceux qui adoptent un mode de vie « occidental » [286]. Les taux de mortalité par cancer du sein chez la femme ont chuté de 36 % entre 1989 et 2015. On pense que ces baisses du taux de mortalité sont le résultat des progrès du traitement, d'une détection plus précoce grâce au dépistage, se traduisant par l'évitement d'environ 249 000 décès par cancer du sein. En 2017, environ 252.710 nouveaux cas de cancer du sein invasif seront diagnostiqués chez les femmes, ainsi qu'environ 63.410 cas supplémentaires de cancer du sein in situ. En outre, environ 40 610 femmes devraient mourir du cancer du sein. [287]

La mammographie est l'examen radiologique de référence pour le dépistage du cancer du sein, première cause de décès chez la femme. Pour obtenir une bonne qualité d'image et compte tenu des caractéristiques anatomiques et radiologiques du sein, il est nécessaire d'utiliser un mammographe qui est un appareil de radiographie aux caractéristiques particulières et aux performances adaptées. Les performances des détecteurs de radiographie numérique peuvent être caractérisées à partir de mesures souvent réalisées dans des conditions de laboratoire. Ils sont exprimés à partir de la fonction de transfert de modulation (MTF), le spectre de puissance de bruit (NPS) et l'efficacité de détection quantique (QDE). [58]

La fonction de transfert de modulation (MTF) caractérise la résolution spatiale et la nécessité des conditions de contraste élevé et sans bruit très éloignées des conditions cliniques.

Le spectre de puissance de bruit (NPS) présenté dans les images médicales rend plus complexe la tâche de détection des lésions à faible contraste. La propriété de bruit d'un système d'imagerie est décrite par le spectre de puissance de bruit (NPS). Ce dernier représente l'écart type de l'intensité du signal pour les différentes fréquences spatiales contenue dans l'image.

L'efficacité de détection quantique (QDE) est une grandeur utile pour caractériser les performances d'un détecteur. Le QDE décrit l'efficacité avec laquelle le système diminuera les

signaux de rapport qui lui sont adressés, entre l'entrée et la sortie du détecteur en fonction de la fréquence spatiale.

Les modalités d'imagerie mammaire jouent un rôle majeur dans la détection précoce, le diagnostic et la prise en charge clinique du cancer du sein. Généralement, il existe trois modalités de mammographie : la mammographie sur film d'écran (SFM), la mammographie numérique plein champ (FFDM) et la tomosynthèse mammaire numérique (DBT). Leurs performances se différencient en fonction de leur profondeur de résolution spatiale [287].

Pour la SFM, l'acquisition des images est obtenue à l'aide d'un couple écran-film contenu dans une cassette de taille standard, choisie en fonction du volume mammaire. Après la génération du faisceau de rayons X, les photons X transmis à la sortie du sein sont convertis en photons lumineux par les écrans intensificateurs des récepteurs. SFM décrit les détails structurels et anatomiques des tissus mammaires (résolution spatiale 30 μm); il est très efficace pour détecter les MC (Micros-Calcifications), les masses mammaires et les distorsions architecturales, cependant, il échoue à détecter les masses denses.

Pour la FFDM. Le processus complet est plus complexe et utilise un détecteur numérique. Un tel détecteur doit avoir une large gamme dynamique et une haute efficacité d'absorption des rayons X afin de produire des images de haute qualité. Les photons de rayon X sont transformés avec précision et efficacité en signaux numériques (résolution spatiale 45 μm). Cependant, le diagnostic du radiologue ne se fait pas à partir des données numériques mais à partir d'une image analogique imprimée sur une reprographie laser ou directement affichée sur une console d'interprétation. La détection par mammographie numérique améliore la précision du diagnostic avec un dépistage mammaire dense en raison d'une résolution de contraste plus élevée.

La DBT est un autre développement de la mammographie numérique qui offre la possibilité d'images tomographiques à différentes profondeurs de sein. L'image mammographique conventionnelle est une projection bidimensionnelle de l'ensemble volume mammaire composé de structures anatomiques normales superposées et de toute lésion bénigne ou maligne. Plusieurs études en dépistage ont montré l'intérêt de la tomosynthèse pour la détection de lésions malignes chez seins denses [289].

En plus des techniques ci-dessus, les chercheurs développent actuellement un nouveau concept appelé Computed Assisted Diagnosis (CAD) développé au cours des vingt dernières années. Leur objectif est de développer des systèmes de dépistage qui aident le radiologue à jeter un second regard sur une région suspecte de l'image. Les systèmes de CAO ont été initialement développés pour l'analyse d'images mammographiques analogiques, numérisées et analysées secondairement, mais leur utilisation était donc peu pratique. La diffusion de la CAD en mammographie est aujourd'hui possible grâce à la mammographie numérique. [42]

Une fois l'image numérique obtenue, les données brutes sont immédiatement transférées vers la console et le serveur CAO où un algorithme de machine learning permettra d'analyser différents paramètres (densité, contraste, forme et taille de l'image, spiculation, etc.) et de les comparer avec une base de données interne d'images mammographiques du système CAD. L'objectif du système CAD est d'augmenter la sensibilité de la mammographie sans augmenter le nombre de faux positifs. Plusieurs études ont montré que ces systèmes de CAO peuvent augmenter le taux de cancers détectés en situation de dépistage de 15 à 20 %. [290].

Les systèmes assistés par ordinateur sont généralement divisés en deux catégories : les schémas CADe et CADx. Les systèmes CADe suivent deux étapes du paradigme de classification des modèles, impliquant la détection (segmentation) des régions de masse candidates et la classification des masses candidates en tissus normaux ou anormaux. Concernant les CADx dont l'initialisation nécessite l'identification d'une lésion ROI fournie soit par un radiologue soit par la sortie d'un schéma CADe, et formule la ROI mammographique ou la caractérisation de la lésion dans le cadre de l'extraction quantitative de caractéristiques d'image et de la classification supervisée des motifs de caractéristiques en deux classes (bénignité ou malignité). [288]

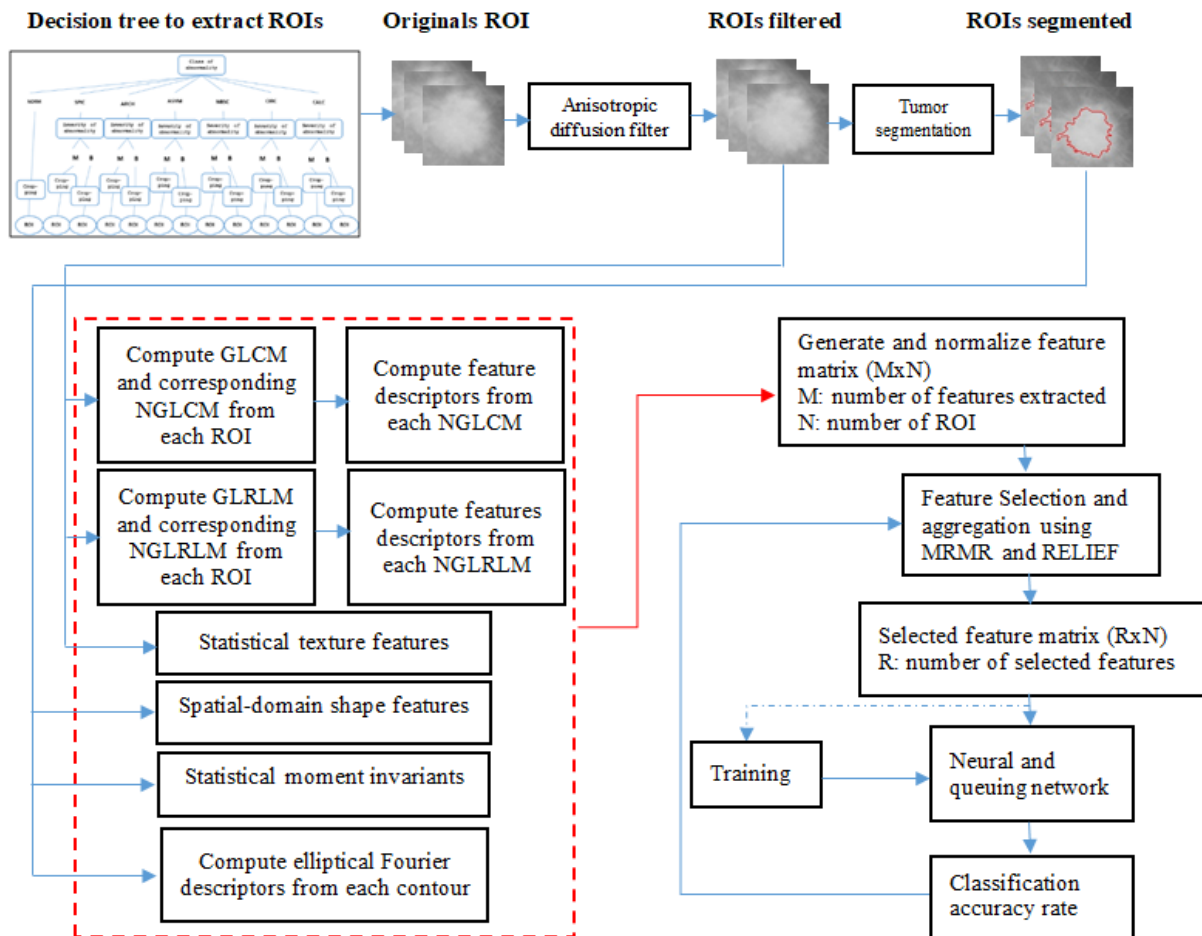


Figure 4.1. Block diagram of the proposed scheme for classification of mammogram using neural and queuing network.[42]

4.2. Pourquoi les réseaux de file d'attente sont importants :

Les réseaux de file d'attente sont essentiels pour comprendre et optimiser les performances des systèmes où la congestion et les files d'attente peuvent entraîner des inefficacités ou des temps d'attente excessifs. Ils permettent de prédire le comportement du système, d'identifier les goulots d'étranglement, de déterminer la capacité optimale du système, d'évaluer les niveaux de service et de proposer des améliorations pour réduire les temps d'attente et les temps de réponse. Malgré un grand nombre de travaux de diagnostic informatique (CAO) (Voir la section 2.5 du deuxième chapitre), aucun travail n'a jusqu'à présent été orienté vers les systèmes de réseaux en file d'attente, permettant de généraliser et de superposer plusieurs

réseaux de même type, de même taille et architecture d'un algorithme d'apprentissage automatique /sélection de caractéristiques lors de l'apprentissage. Le modèle proposé dans [31] est juste un modèle générique qui permet de superposer plusieurs modèles élémentaires d'algorithmes d'apprentissage automatique de même type, taille et architecture. Schématiquement, lors de l'étape d'apprentissage, tous ces modèles sont superposés à travers les systèmes de mise en file d'attente du réseau. L'idée intuitive est que tout algorithme d'apprentissage automatique (réseau de neurones, SVM, KNN, ...) possède une structure topologique de réseau en théorie des graphes. Le modèle générique vise, d'une part, à étendre la capacité de discrimination, et à dilater l'espace d'apprentissage de ces réseaux et, d'autre part, de construire des algorithmes évolutifs d'apprentissage incrémental en prêtant et en absorbant l'intelligence d'un réseau à un autre. [31]

En utilisant les modèles mathématiques des réseaux de file d'attente, les concepteurs de systèmes peuvent prendre des décisions éclairées concernant la conception, la gestion et l'optimisation des processus pour offrir une meilleure expérience à leurs utilisateurs ou clients tout en minimisant les coûts opérationnels.

4.3. Intégration des algorithmes de sélection de caractéristiques dans les réseaux de files d'attente.

Intégrer les algorithmes de sélection de caractéristiques dans les réseaux de files d'attente peut être bénéfique pour améliorer l'efficacité et la performance globale du système. Dans ce contexte, les réseaux de files d'attente sont utilisés pour modéliser et optimiser les processus de flux de travail, de traitement de requêtes ou de services dans un environnement avec plusieurs tâches en attente d'exécution. Voici une approche générale pour intégrer les algorithmes de sélection de caractéristiques dans les réseaux de files d'attente :

- 1. Formulation du problème :** Avant d'intégrer les algorithmes de sélection de caractéristiques, il est essentiel de bien définir le problème du réseau de files d'attente. Cela implique de spécifier les tâches à accomplir, les caractéristiques disponibles pour chaque tâche, les mesures de performance à optimiser (par exemple, le temps de réponse, l'efficacité, etc.), et les contraintes du système.
- 2. Collecte et prétraitement des données :** Le prétraitement des données est une étape importante. Les données utilisées pour les algorithmes de sélection de caractéristiques doivent être nettoyées, normalisées et préparées de manière appropriée. Les caractéristiques pertinentes et les étiquettes associées (le cas échéant) doivent être extraites à partir des données brutes.
- 3. Choix de l'algorithme de sélection de caractéristiques :** En fonction du type de données, du problème et des contraintes spécifiques du réseau de files d'attente, choisissez un algorithme de sélection de caractéristiques approprié. Vous pouvez opter pour des méthodes du filtre, de l'emballage ou des méthodes intégrées, selon les besoins du système.
 - a) Méthode du filtre (Filter Methods) :** Les méthodes du filtre évaluent les caractéristiques indépendamment du modèle d'apprentissage utilisé. Elles se basent sur des mesures statistiques ou des scores pour classer les caractéristiques selon leur capacité à être pertinentes pour la tâche d'apprentissage. Les caractéristiques sont sélectionnées en fonction de leur score, indépendamment de l'algorithme de prédiction. Exemples de méthodes

du filtre : corrélation de Pearson, information mutuelle, coefficient de corrélation, test du chi-carré, etc.

b) Méthode de l'emballage (Wrapper Methods) : Les méthodes de l'emballage utilisent le modèle d'apprentissage pour évaluer la performance des caractéristiques. Elles évaluent différentes combinaisons de caractéristiques et sélectionnent celles qui optimisent les performances du modèle. Ces méthodes sont généralement plus coûteuses en termes de calcul, car elles nécessitent l'entraînement et l'évaluation du modèle pour chaque sous-ensemble de caractéristiques. Exemples de méthodes de l'emballage : recherche exhaustive, recherche par sélection avant-arrière (forward-backward), algorithme génétique, etc.

c) Méthode intégrée (Embedded Methods) : Les méthodes intégrées effectuent la sélection des caractéristiques pendant le processus d'apprentissage du modèle lui-même. Elles optimisent simultanément les caractéristiques et les paramètres du modèle, ce qui les rend plus efficaces que les méthodes du filtre et de l'emballage. Les méthodes intégrées sont généralement spécifiques à certains algorithmes d'apprentissage. Exemples de méthodes intégrées : LASSO (Least Absolute Shrinkage and Selection Operator) pour la régression linéaire, RFE (Recursive Feature Elimination) pour les modèles à noyaux, etc.

d) Méthode d'importance des caractéristiques (Feature Importance Methods) : Certaines techniques d'apprentissage automatique, comme les arbres de décision et les forêts aléatoires, peuvent fournir des mesures d'importance des caractéristiques. Ces mesures permettent de sélectionner les caractéristiques les plus influentes pour la prédiction. Par exemple, dans un arbre de décision, les caractéristiques utilisées en haut de l'arbre pour séparer les classes auront généralement une plus grande importance.

4. Intégration des algorithmes de sélection de caractéristiques : L'intégration des algorithmes de sélection de caractéristiques dans un réseau de files d'attente peut être effectuée de différentes manières en fonction de la nature du système :

a. Prétraitement en amont : Avant d'entrer dans le réseau de files d'attente, les tâches peuvent être filtrées et leurs caractéristiques sélectionnées en utilisant les algorithmes de sélection appropriés. Cela permet de réduire la dimensionnalité des données et d'améliorer l'efficacité du système en aval.

b. Intégration en ligne : Certains algorithmes de sélection de caractéristiques peuvent être intégrés directement dans les nœuds du réseau de files d'attente. À chaque nœud, les caractéristiques pertinentes sont sélectionnées en fonction de la nature de la tâche et des données disponibles.

c. Optimisation globale : Dans certains cas, l'optimisation de la sélection des caractéristiques peut être effectuée de manière globale pour l'ensemble du réseau de files d'attente. Cela peut nécessiter des algorithmes d'optimisation spécifiques pour sélectionner les caractéristiques les plus appropriées pour l'ensemble du système.

5. Évaluation et ajustement : Après l'intégration des algorithmes de sélection de caractéristiques, il est important d'évaluer les performances du réseau de files d'attente en termes des mesures de performance spécifiées au début. Si nécessaire, ajustez les

paramètres des algorithmes de sélection ou du réseau de files d'attente pour obtenir les meilleurs résultats.

L'intégration réussie des algorithmes de sélection de caractéristiques dans les réseaux de files d'attente peut permettre de réduire le temps de traitement, d'améliorer la précision des prédictions et d'optimiser les performances globales du système. Cependant, il est essentiel de comprendre le contexte et les besoins spécifiques du problème pour choisir et appliquer les bonnes techniques de sélection de caractéristiques.

4.4. Modélisation et optimisation du BPNN par un réseau de file d'attente

La modélisation du BPNN (Back Propagation Neural Network) par un réseau de file d'attente représente les propriétés texturales et morphologiques des masses. Le réseau de neurones proposé est un réseau ouvert issu d'une fusion (hybridation) de deux réseaux fermés. L'apprentissage est effectué par une rétropropagation en aval de deux réseaux de neurones multicouches (NN). La méthode de résolution la plus appropriée pour modéliser un algorithme de machine learning qui utilise la fonction coût MSE dépend du type de modèle spécifique utilisé dans l'algorithme. Pour les problèmes d'optimisation simples, des méthodes analytiques comme les moindres carrés peuvent être utilisées. Pour les modèles plus complexes, des méthodes d'optimisation numérique basées sur la descente de gradient ou d'autres approches évolutives peuvent être appliquées. Il est important de choisir la méthode de résolution en fonction de la complexité du modèle et de la nature du problème d'optimisation que l'algorithme de machine learning cherche à résoudre.

La base de données standard MIAS est utilisée pour la validation du schéma proposé. L'apprentissage est utilisé à travers deux BPNN (nous considérons ces réseaux comme fermé) avec deux bases de données d'apprentissage différentes (ici, nous divisons la base de données MIAS en deux), en sauvegardant à chaque itération l'état du réseau sur la structure de la file d'attente. L'apprentissage est alors effectué à travers l'un des deux réseaux, (nous le considérons dans ce cas comme ouvert) avec mise à jour de ses couches cachées avec une agrégation conjointe des états de l'autre réseau, qui sont disposés sur sa structure de file d'attente.

Lorsque l'on souhaite modéliser un algorithme de machine learning qui utilise la fonction coût MSE (Mean Squared Error) comme fonction objective d'un problème d'optimisation sans contraintes, la méthode de résolution la plus appropriée dépend du type de modèle spécifique utilisé dans l'algorithme. Voici quelques exemples de méthodes de résolution couramment utilisées pour différents types d'algorithmes de machine learning :

- 1. Régression linéaire :** Pour un algorithme de régression linéaire, où l'objectif est de trouver une ligne droite qui minimise l'erreur quadratique moyenne (MSE) entre les prédictions du modèle et les valeurs réelles, la méthode la plus courante est la méthode des moindres carrés (Ordinary Least Squares). Cette méthode permet de trouver analytiquement les coefficients optimaux de la ligne droite qui minimise l'erreur.
- 2. Réseaux de neurones :** Pour les réseaux de neurones, l'optimisation de la fonction coût MSE peut être réalisée à l'aide de techniques d'optimisation basées sur la descente de gradient. La descente de gradient stochastique (SGD) et ses variantes, comme l'Adam ou RMSprop, sont des méthodes couramment utilisées pour optimiser les paramètres du réseau afin de minimiser la fonction coût MSE.

- 3. Méthodes d'optimisation numérique :** Pour les problèmes d'optimisation plus complexes qui ne peuvent pas être résolus analytiquement, des méthodes d'optimisation numérique peuvent être utilisées. Ces méthodes incluent la recherche aléatoire, les algorithmes génétiques, la recherche par essaim de particules, et d'autres approches évolutives. Ces méthodes peuvent être appliquées à différents types de modèles de machine learning et sont souvent utilisées lorsque la fonction coût n'a pas de forme analytique simple.
- 4. SVM (Support Vector Machines) :** Pour les algorithmes de type SVM, l'objectif est de trouver un hyperplan qui maximise la marge entre les classes tout en minimisant les erreurs de classification. La méthode de résolution couramment utilisée est la programmation quadratique (QP) pour résoudre le problème d'optimisation quadratique associé au SVM.
- 5. Arbres de décision :** Pour les arbres de décision, la fonction coût MSE peut être utilisée pour déterminer les meilleures divisions des nœuds de l'arbre. La méthode de résolution consiste généralement à utiliser des algorithmes de type "Diviser pour régner" pour construire l'arbre en trouvant récursivement les divisions qui minimisent l'erreur MSE.

4.5. Une Approche Complète pour la Sélection de Caractéristiques en Reconnaissance de la Masse Mammaire

L'objectif principal de ce travail est de proposer une approche de sélection de caractéristiques pour la reconnaissance de la masse mammaire. Les étapes présentées décrivent les étapes de la démarche que nous avons suivie pour atteindre cet objectif (voir Figure 4.1). Nous effectuons la reconnaissance de ces classes en utilisant des classificateurs de type QBPNN ayant en entrées toutes les caractéristiques sélectionnées. Nous avons opté pour la sélection des caractéristiques selon l'approche Symbiose (Wrapper, voir Figure 1.6 du premier chapitre) qui est basée sur les résultats obtenus par les classificateurs QBPNN utilisés. Pour chacune des catégories de ROI, notre approche de sélection des caractéristiques fonctionne comme suit.

1. Construction d'une base de données ROI en utilisant une approche basée sur des règles et une représentation graphique d'un arbre de décision (voir la section 4.6.1; les Figures 4.2, 4.3).
2. Filtration des ROI à l'aide d'une technique appelée diffusion anisotrope (voir section 4.6.2).
3. Pour chaque ROI de catégorie filtrée:

- Calcul des matrices de cooccurrence, l'image est balayée en quatre directions différentes : horizontale, verticale, diagonale, antidiagonale (Voir les Figure 4.4, 4.5 et 4.6, et le Tableau 4.1).
- Calcul des matrices de longueur de plage de niveau de gris. (Voir la Figure 4.7 et 4.8, et le Tableau 4.1)
- Caractéristiques morphologiques issues du tracé des contours de masse (Voir la section 4.6.4; et la Figure 4.9).

A chaque agrégation de caractéristiques, l'apprentissage et le test de deux classificateurs fermés QBPNN correspondant à la catégorie de masse considérée sont effectués et leurs performances évaluées.

4. Le processus d'apprentissage itératif consiste à :

- Fusionner chaque réseau QBPNN fermé en mettant en file d'attente (Enqueuedans la file d'attente respective) à la fois les poids de l'un et les erreurs de l'autre, (voir Algorithmes 1 et 2 de la section 4.7.1 et 4.7.2).

- Unifier les deux réseaux QBPNN en sortant (popFront de la file d'attente) en même temps le réseau fusionné poids de chacun. (Voir Algorithme 3 de la section 4.7.3)

Dans la section suivante, nous présentons sept sous-sections, en commençant par la pré-classification basée sur des règles à partir d'un arbre de décision pour extraire les régions d'intérêt (ROI). Ensuite, la filtration de la diffusion anisotrope est présentée dans la sous-section suivante. Les troisième et quatrième sous-sections décrivent l'étape d'extraction des caractéristiques texturales et morphologiques des mammographies déjà filtrées. La dernière sous-section est la sous-section la plus fondamentale de ce travail. Il décrit la modélisation, la sélection et l'agrégation de fonctionnalités avec des systèmes de réseau de file d'attente. Enfin, nous présentons nos expériences ; discuter des résultats, suivis d'une conclusion et des perspectives d'extensions futures.

4.6. La méthode proposée

La méthode proposée de classification des images mammographies comprend quatre modules : le premier consiste en une construction automatique d'un ensemble de règles d'associations pour la pré-classification des mammographies à travers un arbre de décision. Toutes ces images sont traitées avec un filtre de diffusion anisotrope dans le deuxième module. Le troisième module couvre deux algorithmes majeurs tels que l'extraction et la sélection de caractéristiques. L'algorithme d'extraction de caractéristiques consiste à générer deux types de caractéristiques : texturaux à partir de la matrice de co-occurrence et de la matrice de longueur de plage et, morphologiques à partir des différents tracés de contour de masse. Dans la sélection des caractéristiques, les algorithmes MRMR et RELIEF sont utilisés pour découvrir les caractéristiques efficaces et significatives et sont fournis à la file d'attente et au réseau neuronal pour la classification des mammographies comme normales, bénignes ou malignes dans le dernier module. Le schéma fonctionnel général du schéma est présenté à la Figure 4.1.

4.6.1. Pré-classification des mammographies basée sur des règles d'associations

De nombreux problèmes de classification CAD manquent généralement de règles systématiques et génériques formulées par les radiologues. Cette représentation permet l'appariement des caractéristiques relatives aux formes et leur affectation aux sous-classes correspondantes[291,292]. Pour accélérer le processus d'induction des règles et réduire sa complexité, une étape de sélection est nécessaire ; il réduit la dimensionnalité des grands ensembles de caractéristiques, minimisant l'espace de recherche et la perte d'informations nécessaires à l'induction des règles. La sélection supprime la redondance des données fondamentales et simplifie la conception et la mise en œuvre du classificateur ; en déterminant quelles caractéristiques doivent être fournies au système. De plus, la réduction de la dimensionnalité d'entrée augmente la vitesse de fonctionnement du classifieur, tout en maintenant un temps de réponse acceptable. Pour simplifier le processus de classification des seins, nous proposons des règles symboliques de base qui reflètent une attribution préalable des seins à partir de particularités tissulaires (le fond, la classe et la sévérité de l'anomalie) prenant en compte différentes formes de sein selon leur classe d'anomalie et leur type de sévérité maligne ou bénigne (Cf. fig.3). Ces règles sont de type : **IF** Caractéristiques **THEN** Attribution. L'arbre de décision représentant la pré-classification proposée est décrit dans la Figure 4.2. Le premier nœud est appelé la "racine" de l'arbre. Il est situé au premier niveau.

Nous observons la distribution de fréquence de la variable pour prédire le "caractère du tissu de fond". Nous constatons qu'il existe 272 observations correspondant à 13x3 les classes de seins, dont 89 "D : Dense-glandulaire", 101 "F : Gras" et 103 "G : Gras-glandulaire". Nous ne détaillons que la collection d'images F :

La variable « Classe d'anomalie » est la première variable utilisée ; comme elle est composée de 6 modalités {NORM (Tissus normaux), SPIC (lésions spéculées), ARCH (Distorsion architecturale), ASYM (Asymétrie), MISC (Miscumscribe), CIRC (masses circonscrites)}, cette variable produit 6 nœuds enfant. La première arête (le premier chemin), à gauche, au deuxième niveau, est produite à partir de la modalité NORM, produit 66 observations. Du deuxième bord jusqu'au septième, on nous dirige, à son tour, vers une autre même variable, "Sévérité de l'anomalie". Chacune de ces dernières est composée de deux modalités « M : malin, B : bénin ». La deuxième arête est produite à partir de la modalité SPIC on se dirige vers la variable "Sévérité de l'anomalie" et est produite à partir des modalités "M" et "B" qui produisent respectivement 3 et 2 observations. La troisième arête est produite à partir de la modalité ARCH nous amène vers la même variable de segmentation et est produite à partir des modalités « M » et « B » qui couvrent chacune respectivement 2 et 4 observations. Le quatrième front est produit à partir de la modalité ASYM évoluant vers la même variable de segmentation "Sévérité de l'anomalie" et est produit à partir des modalités "M" et "B" qui produisent respectivement 3 et 1 observations. La cinquième arête est produite à partir de la modalité MISC qui nous amène vers la même variable de segmentation "Sévérité de l'anomalie" et est produite à partir des modalités "M" et "B" résultant respectivement de 5 et 3 observations. La dernière arête est produite à partir de la modalité CIRC qui nous oriente vers la variable "Sévérité de l'anomalie" et est produite à partir des modalités "M" et "B" qui produisent respectivement 2 et 10 observations. Chaque mammographie obtenue a été soumise à une étape de recadrage pour extraire les régions d'intérêt (ROI) des centres des coordonnées d'anomalie (x, y) de l'image du sein et la valeur approchée du rayon (en pixels) du cercle fermé.

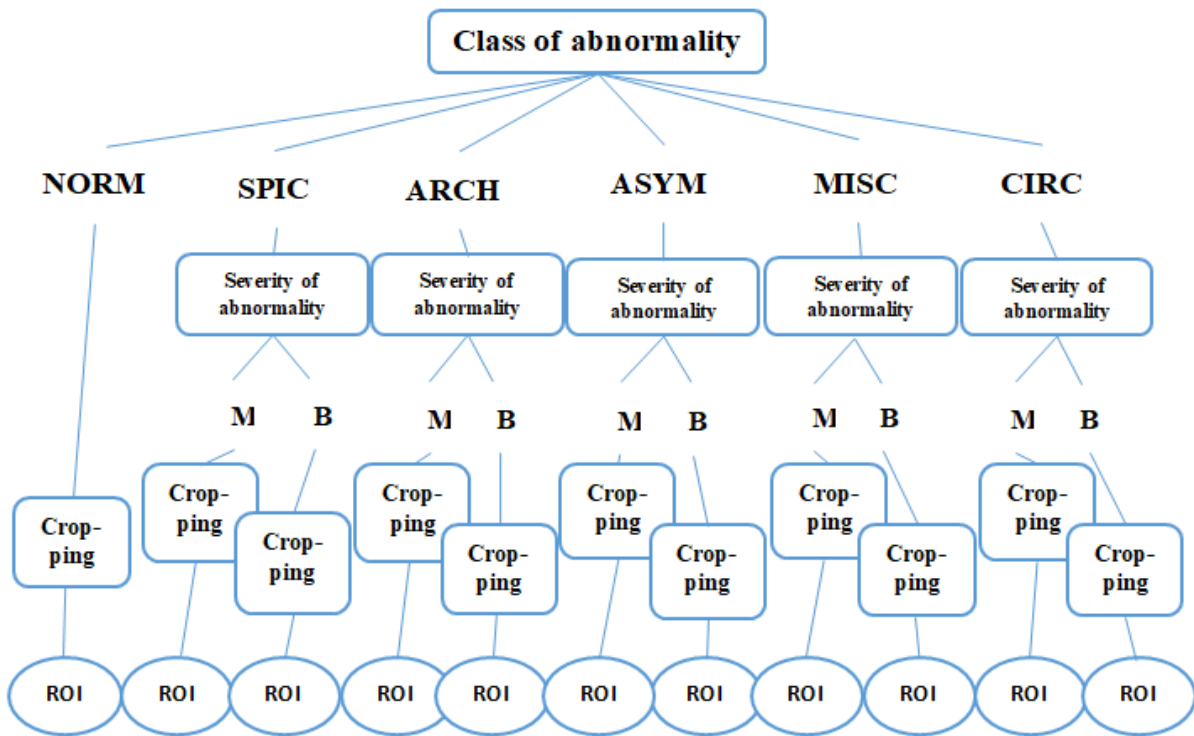


Figure 4.2. Construction des règles d'association à partir d'un arbre de décision pour l'extraction des régions d'intérêt. [42]

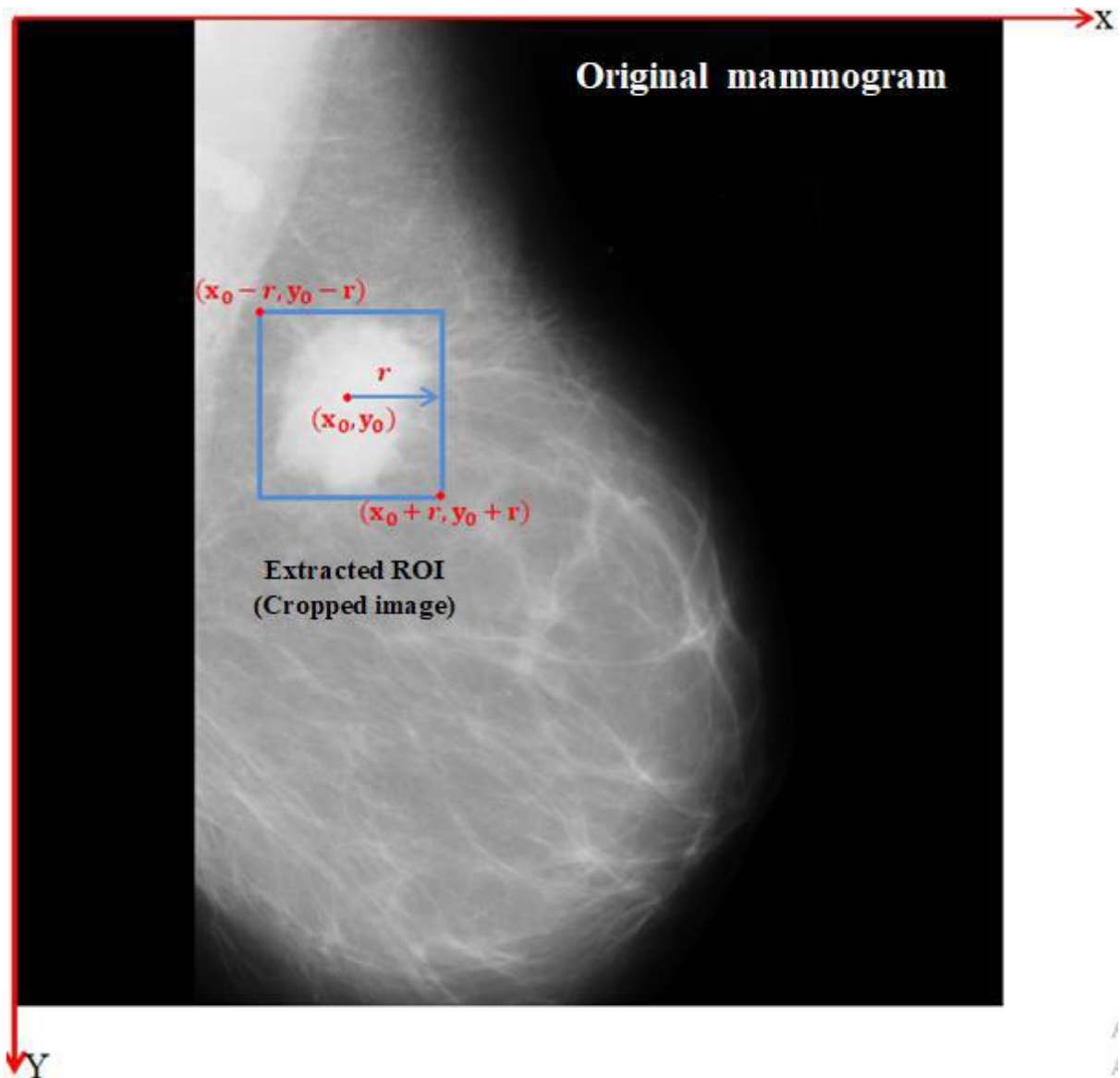


Figure 4.3. Recadrage des ROIs à partir d'une image mammographique faisant référence au centre de la zone anormale.[42]

Le sein est une structure volumique et hautement déformable en trois dimensions et la mammographie crée une projection en deux dimensions. La forme du sein, sa taille et la dimension du parenchyme glandulaire peuvent être variables. La glande mammaire prend place devant la paroi thoracique entre la clavicule, et la 6^{ème} ou 8^{ème} côte, latéralement elle s'étend du sternum à la ligne axillaire moyenne. En mammographie, les images sont généralement assez complexes, elles contiennent du bruit sous forme à la fois de rayonnement diffusé et de texture de fond. Plus l'image est bruitée, plus son interprétation est difficile. La détection et la reconnaissance des lésions dans ces images peuvent être difficiles car les lésions, en particulier les nodules ou les masses, sont généralement intégrées dans un fond de bruit anatomique. L'arrière-plan dans cette définition peut inclure d'autres objets. Pour distinguer l'objet d'intérêt de l'arrière-plan, une opération de recadrage a été appliquée sur les images de mammographie pour extraire les régions d'intérêt (ROI) qui contiennent les anomalies, éliminant les parties indésirables de l'image (voir la Figure 4.3). Ce processus est

exécuté en faisant référence au centre de la zone anormale comme centre de la ROI et en prenant le rayon approximatif (en pixels) d'un cercle entourant l'anormal. Pour l'extraction du ROI normal, la même procédure de recadrage est effectuée sur des images mammographiques normales avec une sélection aléatoire de l'emplacement (comme 60 x 60). [41]

4.6.2. Filtre de diffusion anisotrope

Lorsque l'algorithme d'apprentissage automatique recherche les meilleurs paramètres pour une forme particulière à l'aide d'un ensemble limité de données, il peut modéliser non seulement les formes générales dans les données, mais également tout bruit spécifique à l'ensemble de données, ce qui génère une mauvaise performance du modèle d'apprentissage sur les données de test. En 1990 Perona & Malik [67]; introduit un filtre adaptatif très puissant appelé diffusion anisotrope, réduisant le bruit dans la zone et renforçant les bords d'une masse. Ils ont supposé que les valeurs d'intensité des pixels étaient semi-absorbantes. Les intensités de chaque pixel changent numériquement à chaque itération de l'algorithme. Ils diffusent dans le temps dans les pixels voisins suivant le gradient négatif c'est-à-dire que les pixels voisins de chaque pixel central absorbent une petite quantité d'intensité du pixel central, et ce uniquement pour les voisins ayant des concentrations plus faibles par rapport au pixel central. Le processus de diffusion se termine dès que tout contraste est supprimé. Le filtrage par diffusion anisotrope suit mathématiquement l'équation aux dérivées partielles :

$$\frac{\partial I(x,y,t)}{\partial t} = \text{div}(c(x,y,t)\nabla I(x,y,t)) = c(x,y,t)\Delta I(x,y,t) + \nabla c(x,y,t)\nabla I(x,y,t) \quad (4.1)$$

Où $I(x,y,t)$ est l'intensité de l'image au temps t , $c(x,y,t)$ est une constante, ∇ et Δ indiquent respectivement le gradient et les opérateurs laplaciens et div indique l'opérateur de divergence. (Perona & Malik, 1990) ont décrit le comportement de la diffusivité du filtre à travers une fonction monotone non négative décroissante en terme de module de gradient, appelée fonction de diffusion $g(\nabla I)$. Les fonctions g initialement proposées par Perona et Malik sont les suivantes :

$$g(\nabla I(x,y)) = e^{-\left(\frac{|\nabla I|}{K}\right)^2} \quad (4.2)$$

$$g(\nabla I(x,y)) = \frac{1}{1+|\nabla I/K|^2} \quad (4.3)$$

$$\text{With } g(0) = 1 \quad (4.4)$$

Le paramètre K sert à fixer la limite entre les forts gradients correspondant aux régions inhomogènes de la masse (notamment les arêtes), et les faibles gradients correspondant aux régions homogènes.

4.6.3. Caractéristiques texturales

- Caractéristiques texturales du premier ordre

Ces caractéristiques étudient la distribution des intensités de pixels, extraites de l'histogramme de l'image de la masse, indiquent en quelque sorte leur variation, leurs contrastes, leur asymétrie, leur prédominance, leur étroitesse et leur uniformité. Ces caractéristiques comprennent : la moyenne, la variance, l'asymétrie, l'aplatissement, l'énergie et l'entropie. [68]

- Caractéristiques texturales de second ordre : Matrice de cooccurrence de niveaux de gris (GLCM)

Plusieurs métriques de texture contenant des informations spatiales sont basées sur la matrice de co-occurrence [117, 118], -également connue sous le nom de matrice de dépendance spatiale des niveaux de gris-. Tout comme la construction de l'histogramme est une étape de prétraitement qui prépare les données pour les moments statistiques, la formation des matrices de cooccurrence est une étape initiale qui compile les informations spatiales et statistiques pour calculer les métriques de texture. $C_d^\theta(i, j)$ est un histogramme bidimensionnel des valeurs de niveau de gris i et j où $C_d^\theta(i, j)$ spécifie la probabilité qu'un pixel de niveau de gris j soit situé à une distance d et une direction θ à partir d'un pixel de niveau de gris i . La position relative des paires de pixels, définie par la distance d en nombre de pixels ($d = 1$ pour les pixels voisins, etc.) et l'orientation θ constituent cette information spatiale. L'orientation θ est quantifiée dans quatre directions : horizontale, diagonale, verticale et antidiagonale (0° , 45° , 90° , 135°) comme le montre la Figure 4.4.

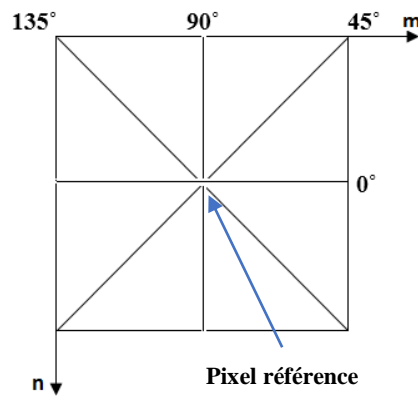


Figure 4.4. Pixel référencé par quatre orientations

Dans [118], les auteurs ont défini une métrique de caractéristique globale pour classer la texture en décrivant quantitativement la forme de la matrice de cooccurrence (voir tableau 4.1). Soit $P(i, j, d, \theta)$ décrit la probabilité composée d'une paire de pixels (i, j) au déplacement d et, direction de déplacement θ (0° , 45° , 90° , 135°) :

$$P(I_1, I_2, d, \theta) = \frac{\text{number of pairs of pixels at distance } d \text{ with values } (I_1, I_2)}{\text{total number of possible pairs}} \quad (4.5)$$

$$\text{for } \theta = 0^\circ : P(I_1, I_2, d, \theta) = P(I(m, n) = I_1, I(m \pm d, n) = I_2, d, 0^\circ) \quad (4.6)$$

$$\text{for } \theta = 45^\circ : P(I_1, I_2, d, \theta) = P(I(m, n) = I_1, I(m \pm d, n \pm d) = I_2, d, 45^\circ) \quad (4.7)$$

$$\text{for } \theta = 90^\circ : P(I_1, I_2, d, \theta) = P(I(m, n) = I_1, I(m, n \pm d) = I_2, d, 90^\circ) \quad (4.8)$$

$$\text{for } \theta = 135^\circ : P(I_1, I_2, d, \theta) = P(I(m, n) = I_1, I(m \pm d, n \pm d) = I_2, d, 135^\circ) \quad (4.9)$$

Pour chacun de ces histogrammes, un tableau est défini, appelé matrice de cooccurrence ou de dépendance spatiale. La matrice de cooccurrence pour une paire (d, θ) est définie comme la matrice $N_g \times N_g$:

$$C_d^\theta = \frac{1}{R} \begin{bmatrix} \eta(0,0) & \eta(0,1) & \dots & \eta(0, N_g - 1) \\ \eta(1,0) & \eta(1,1) & \dots & \eta(1, N_g - 1) \\ \dots & \dots & \dots & \dots \\ \eta(N_g - 1, 0) & \eta(N_g - 1, 1) & \dots & \eta(N_g - 1, N_g - 1) \end{bmatrix} \quad (4.10)$$

Où $\eta(I_1, I_2)$ est le nombre de paires de pixels, à la position relative (d, θ) , qui ont des valeurs de niveau de gris I_1, I_2 , respectivement, et g est le nombre de niveaux de gris distincts dans la ROI quantifiée. R est le nombre total de paires de pixels possibles.

Le processus de calcul de GLCM d'une matrice d'intensité d'image de test donnée considère chaque pixel de niveau de gris de la matrice d'image comme pixel de référence. La matrice de cooccurrence en niveaux de gris (GLCM) est générée pour chaque ROI de mammographie. Pour chaque pixel, la distance d et les valeurs de direction θ sont prises dans quatre directions de pixel voisines possibles ($0^\circ, 45^\circ, 90^\circ, 135^\circ$). Les matrices de cooccurrence de niveau de gris normalisées (NGLCM) considèrent chaque cellule dans les matrices qui contient une valeur de probabilité en divisant R (le nombre total de paires de pixels possibles).

Le processus de calcul du GLCM d'une matrice d'intensité d'image de test donnée est décrit sur la Figure 4.5. Chaque pixel de niveau de gris de la matrice d'image est considéré comme un pixel de référence. La matrice de cooccurrence en niveaux de gris (GLCM) est générée pour chaque ROI de mammographie. Ici, le nombre de niveaux de gris est de cinq, nous supposons que pour chaque pixel la distance $d=1$ et la direction θ sont prises dans quatre directions de pixel voisines possibles ($0^\circ, 45^\circ, 90^\circ, 135^\circ$)

La Figure 4.6 montre les matrices de cooccurrence normalisées en niveaux de gris (NGLCM) où chaque cellule des matrices contient une valeur de probabilité. Chaque élément de NGLCM est calculé en divisant $R=40$ en cas de direction horizontale, $R=38$ en cas de direction verticale, et $R=30$ en cas de direction diagonale gauche et de direction diagonale droite, à chaque élément de GLCM symétrique correspondant.

$$I(m, n) = \begin{bmatrix} 0 & 1 & 1 & 4 & 3 \\ 1 & 2 & 2 & 2 & 4 \\ 3 & 3 & 3 & 3 & 3 \\ 1 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 1 & 3 \end{bmatrix} \quad (a)$$

$$C_{d=1}^{0^\circ} = \frac{1}{40} \begin{bmatrix} 4 & 2 & 1 & 0 & 1 \\ 2 & 2 & 2 & 1 & 1 \\ 1 & 2 & 4 & 0 & 1 \\ 0 & 1 & 0 & 8 & 1 \\ 1 & 1 & 1 & 1 & 2 \end{bmatrix} \quad (b)$$

$$C_{d=1}^{90^\circ} = \frac{1}{38} \begin{bmatrix} 2 & 2 & 1 & 1 & 0 \\ 2 & 0 & 2 & 2 & 1 \\ 1 & 2 & 0 & 4 & 1 \\ 1 & 2 & 4 & 0 & 4 \\ 0 & 1 & 1 & 4 & 0 \end{bmatrix} \quad (c)$$

$$C_{d=1}^{45^\circ} = \frac{1}{30} \begin{bmatrix} 2 & 0 & 1 & 1 & 1 \\ 0 & 2 & 1 & 1 & 1 \\ 1 & 1 & 0 & 4 & 1 \\ 1 & 1 & 4 & 0 & 2 \\ 1 & 1 & 1 & 2 & 0 \end{bmatrix}$$

(d)

$$C_{d=1}^{135^\circ} = \frac{1}{30} \begin{bmatrix} 0 & 2 & 2 & 1 & 0 \\ 2 & 0 & 2 & 1 & 0 \\ 2 & 2 & 0 & 4 & 0 \\ 1 & 1 & 4 & 0 & 2 \\ 0 & 0 & 0 & 2 & 2 \end{bmatrix}$$

(e)

Figure 4.5. Calcul des matrices de cooccurrence. (a) Valeurs d'intensité de l'image d'entrée avec 5 niveaux de gris. Différentes matrices de cooccurrence (GLCM) pour la distance définie $d=1$, l'image est balayée dans quatre directions différentes telles que (b) horizontale ($\theta = 0^\circ$), (c) verticale ($\theta = 90^\circ$), (d) diagonale ($\theta = 45^\circ$), (e) antidiagonale ($\theta = 135^\circ$).

$$C_{d=1}^{0^\circ} = \frac{1}{50} \begin{bmatrix} 0,08 & 0,04 & 0,02 & 0 & 0,02 \\ 0,04 & 0,04 & 0,04 & 0,02 & 0,02 \\ 0,02 & 0,04 & 0,08 & 0 & 0,02 \\ 0 & 0,02 & 0 & 0,16 & 0,02 \\ 0,02 & 0,02 & 0,02 & 0,02 & 0,04 \end{bmatrix}$$

(a)

$$C_{d=1}^{90^\circ} = \frac{1}{38} \begin{bmatrix} 0,0526 & 0,0526 & 0,0263 & 0,0263 & 0 \\ 0,0526 & 0 & 0,0526 & 0,0526 & 0,0263 \\ 0,0263 & 0,0526 & 0 & 0,1052 & 0,0263 \\ 0,0263 & 0,0526 & 0,1052 & 0 & 0,1052 \\ 0 & 0,0263 & 0,0263 & 0,1052 & 0 \end{bmatrix}$$

(b)

$$C_{d=1}^{45^\circ} = \frac{1}{29} \begin{bmatrix} 0,0689 & 0 & 0,0345 & 0,0345 & 0,0345 \\ 0 & 0,0689 & 0,0345 & 0,0345 & 0,0345 \\ 0,0345 & 0,0345 & 0 & 0,138 & 0,0345 \\ 0,0345 & 0,0345 & 0,138 & 0 & 2 \\ 0,0345 & 0,0345 & 0,0345 & 0,0689 & 0 \end{bmatrix}$$

(c)

$$C_{d=1}^{135^\circ} = \frac{1}{30} \begin{bmatrix} 0 & 0,0667 & 0,0667 & 0,0333 & 0 \\ 0,0667 & 0 & 0,0667 & 0,0333 & 0 \\ 0,0667 & 0,0667 & 0 & 0,1333 & 0 \\ 0,0333 & 0,0333 & 0,1333 & 0 & 0,0667 \\ 0 & 0 & 0 & 0,0667 & 0,0667 \end{bmatrix}$$

(d)

Figure 4.6. Matrices de cooccurrence normalisées (NGLCM) des matrices de cooccurrence correspondantes (GLCM) de la figure 5 dans quatre directions différentes telles que (a) horizontale ($\theta = 0^\circ$), (b) verticale ($\theta = 90^\circ$), (c) diagonale ($\theta = 45^\circ$), (d) antidiagonale ($\theta = 135^\circ$).

Tableau 4.1. Métriques de classification de texture de Haralick et métriques de longueur d'exécutions. [42]

Texture Feature	Formula	Run-length Feature	Formula
energy	$\sum_i \sum_j P^2(i, j, d, \theta^\circ)$	Short-run emphasis	$SRE = \sum_{g=0}^{G-1} \sum_{l=1}^L \frac{Q_{RL}(g, l)}{l^2}$
Entropy	$-\sum_i \sum_j P(i, j, d, \theta^\circ) \log_2 P(i, j, d, \theta^\circ)$	Long-run emphasis	$LRE = \sum_{g=0}^{G-1} \sum_{l=1}^L Q_{RL}(g, l) l^2$

Contrast	$\sum_i \sum_j i - j P(i, j, d, \theta^\circ)$	Low gray-level emphasis	LGRE = $\sum_{g=0}^{G-1} \sum_{l=1}^L \frac{Q_{RL}(g, l)}{(g + 1)^2}$
Inertia	$\sum_i \sum_j (i - j)^2 P(i, j, d, \theta^\circ)$	High gray-level emphasis	HGRE = $\sum_{g=0}^{G-1} \sum_{l=1}^L Q_{RL}(g, l)(g + 1)^2$
correlation	$\sum_i \sum_j \frac{ijP(i, j, d, \theta^\circ) - \mu_i \mu_j}{\sigma_i \sigma_j}$	Gray-level non-uniformity	GLNU = $\sum_{l=1}^L \left[\sum_{g=0}^{G-1} Q_{RL}(g, l) \right]^2$
Texture homogeneity	$\sum_i \sum_j \frac{P(i, j, d, \theta^\circ)}{1 + i - j }$	Run-length non-uniformity	RLNU = $\sum_{g=0}^{G-1} \left[\sum_{l=1}^L Q_{RL}(g, l) \right]^2$
Inverse difference	$\sum_i \sum_j \frac{P(i, j, d, \theta^\circ)}{ i - j }$	Run percentage	RPC = $\sum_{g=0}^{G-1} \sum_{l=1}^L \frac{1}{Q_{RL}(g, l)}$
Inverse difference moment	$\sum_i \sum_j \frac{P(i, j, d, \theta^\circ)}{1 + (i - j)^2}$	Short run, low gray-level emphasis	SRLGE = $\sum_{g=0}^{G-1} \sum_{l=1}^L \frac{Q_{RL}(g, l)}{l^2 (g + 1)^2}$
Cluster tendency	$\sum_i \sum_j (i - \mu_i + j - \mu_j)^2 P(i, j, d, \theta^\circ)$	Long run, high gray-level emphasis	LRHGE = $\sum_{g=0}^{G-1} \sum_{l=1}^L Q_{RL}(g, l) l^2 (g + 1)^2$
-	-	Short-run, high gray-level emphasis	SRHGE = $\sum_{g=0}^{G-1} \sum_{l=1}^L \frac{Q_{RL}(g, l)(g + 1)^2}{l^2}$
-	-	Long-run, low gray-level emphasis	LRLGE = $\sum_{g=0}^{G-1} \sum_{l=1}^L \frac{Q_{RL}(g, l) l^2}{(g + 1)^2}$

- Matrice des longueurs de plage de niveaux de gris (GLRLM : Gray level run lengths matrix)

La méthode run-length est une technique appliquée sur des images en niveaux de gris pour la classification des textures, utilisée pour améliorer la représentation d'une texture spécifique par un prétraitement de texture ; les paramètres de cette technique sont le niveau de gris, sa longueur et la direction de balayage. Une série de niveaux de gris est une séquence successive de pixels ayant la même valeur de niveau de gris, la longueur de la série est le nombre de pixels dans la direction de balayage horizontale ou verticale, ou le nombre de pixels multiplié par $\sqrt{2}$ pour une direction diagonale [127, 128]. Généralement, l'image est scannée dans quatre directions : 0° , 45° , 90° et 135° . Les quatre directions restantes ont des histogrammes identiques, en raison de la symétrie. La probabilité d'une longueur de plage spécifique est $Q_{RL}(g, l)$, où g est la case de valeur de gris et l est la longueur. Si l'histogramme a été normalisé, alors

$$\sum_{g=0}^{G-1} \sum_{l=1}^L Q_{RL}(g, l) = 1 \quad (4.11)$$

Le processus de calcul de GLRLM d'une matrice d'intensité d'image de test donnée consiste à définir la matrice de longueur de plage correspondante Q_{RL} pour chacune des quatre directions de balayage (0° , 45° , 90° , 135°). Son élément (g, l) donne le nombre de fois qu'un niveau de gris $g - 1$ ($g = 1 \dots G$) apparaît dans l'image avec une longueur de plage l , ($l = 1, 2 \dots L$), où G est la plus grande valeur d'intensité de pixel, et L est la plus grande longueur de plage possible dans l'image. Le tableau 1 décrit les caractéristiques extraites à l'aide de la méthode de longueur de plage.

La Figure 4.7 et la Figure 4.8 décrivent respectivement le processus de calcul du GRLM (Gray-level Run Length Matrix) et du GRLM normalisé correspondant d'une matrice d'intensité d'image de test. Prenons par exemple l'image I avec cinq niveaux de gris possibles ($G = 5$). Pour chacune des quatre directions (0° , 45° , 90° , 135°). Nous définissons la matrice de longueur de plage correspondante Q_{RL} . Son élément (g, l) donne le nombre de fois qu'un niveau de gris $g-1, g=1 \dots G$ apparaît dans l'image avec une longueur de plage $l; l = 1, 2 \dots L$. Il s'agit d'un tableau $G \times L$, où L est la plus grande longueur de plage possible dans l'image.

$$I = \begin{bmatrix} 0 & 1 & 1 & 4 & 3 \\ 1 & 2 & 2 & 2 & 4 \\ 3 & 3 & 3 & 3 & 3 \\ 1 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 1 & 3 \end{bmatrix}$$

(a)

$$Q_{RL}(0^\circ) = \begin{bmatrix} 2 & 0 & 1 & 0 & 0 \\ 3 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 2 & 0 & 0 & 0 & 1 \\ 2 & 1 & 0 & 0 & 0 \end{bmatrix}$$

(b)

$$Q_{RL}(90^\circ) = \begin{bmatrix} 3 & 1 & 0 & 0 & 0 \\ 5 & 0 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 0 \\ 7 & 0 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(c)

$$Q_{RL}(45^\circ) = \begin{bmatrix} 3 & 1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 0 \\ 7 & 0 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(d)

$$Q_{RL}(135^\circ) = \begin{bmatrix} 5 & 0 & 0 & 0 & 0 \\ 5 & 0 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 0 \\ 7 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 \end{bmatrix}$$

(e)

Figure 4.7. Calcul des matrices de longueur de plage de niveau de gris. (a) Valeurs d'intensité de l'image d'entrée avec 5 niveaux de gris. Différentes matrices de longueur de plage (RLM), l'image est scannée dans quatre directions différentes telles que (b) horizontale ($\theta = 0^\circ$), (c) verticale ($\theta = 90^\circ$), (d) diagonale ($\theta = 45^\circ$), (e) antidiagonale ($\theta = 135^\circ$).

$$Q_{RL}(0^\circ) = \begin{bmatrix} 0.133 & 0 & 0.067 & 0 & 0 \\ 0.2 & 0.067 & 0 & 0 & 0 \\ 0.067 & 0 & 0.067 & 0 & 0 \\ 0.133 & 0 & 0 & 0 & 0.067 \\ 0.133 & 0.067 & 0 & 0 & 0 \end{bmatrix}$$

(a)

$$Q_{RL}(90^\circ) = \begin{bmatrix} 0.125 & 0.0416 & 0 & 0 & 0 \\ 0.208 & 0 & 0 & 0 & 0 \\ 0.166 & 0 & 0 & 0 & 0 \\ 0.291 & 0 & 0 & 0 & 0 \\ 0.166 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(b)

$$Q_{RL}(45^\circ) = \begin{bmatrix} 0.13 & 0.043 & 0 & 0 & 0 \\ 0.13 & 0.043 & 0 & 0 & 0 \\ 0.173 & 0 & 0 & 0 & 0 \\ 0.304 & 0 & 0 & 0 & 0 \\ 0.173 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(c)

$$Q_{RL}(135^\circ) = \begin{bmatrix} 0.208 & 0 & 0 & 0 & 0 \\ 0.208 & 0 & 0 & 0 & 0 \\ 0.166 & 0 & 0 & 0 & 0 \\ 0.291 & 0 & 0 & 0 & 0 \\ 0.083 & 0.041 & 0 & 0 & 0 \end{bmatrix}$$

(d)

Figure 4.8. Matrices de longueur de plage de niveau de gris normalisées (NGRLM) des matrices de longueur de plage correspondantes (RLM) de la Fig. 10 dans quatre directions différentes telles que (a) horizontale ($\theta = 0^\circ$), (b) verticale ($\theta = 90^\circ$), (c) diagonale ($\theta = 45^\circ$), (d) antidiagonale ($\theta = 135^\circ$).

4.6.4. Caractéristiques morphologiques

Les caractéristiques morphologiques décrivent la variation de la structure extérieure de la forme. Le but de l'analyse de forme est d'obtenir plusieurs métriques quantitatives par des images binaires, c'est-à-dire une image partitionnée en pixels à l'intérieur de la masse et en pixels à l'extérieur de la masse. Les caractéristiques morphologiques utilisées dans ce travail sont la longueur du périmètre ; périmètre à diagonale, rapport d'aspect, compacité, irrégularité, énergie de flexion, descripteur de Fourier, caractéristiques basées sur la longueur radiale normalisée (NRL) et orientation du gradient relatif (RGO). Toutes ces métriques utilisent les points frontières et ont été adoptées dans les travaux antérieurs pour la classification de masse [293,294].

a) Pixels de frontière et codes de chaîne

Le codage en chaîne est l'une des techniques les plus largement utilisées pour la description de la forme des frontières. Dans [295], la courbe limite est approximée via une séquence de segments de ligne droite connectés de direction et de longueur présélectionnées. Chaque

segment de ligne est codé avec un numéro de codage spécifique en fonction de sa direction. Le nombre de pixels frontières m est directement trouvé en comptant les pixels frontières $(x_i, x_j), i = 1, 2, \dots, m$. Ensuite, le code de chaîne de Freeman est utilisé pour coder de manière compacte les pixels limites. La direction de chaque pixel limite au pixel limite suivant est mise dans la chaîne. La direction du dernier pixel au premier pixel est le dernier code de la chaîne. Les codes de direction utilisés et le code de chaîne résultant de la masse sont illustrés à la Figure 4.9.

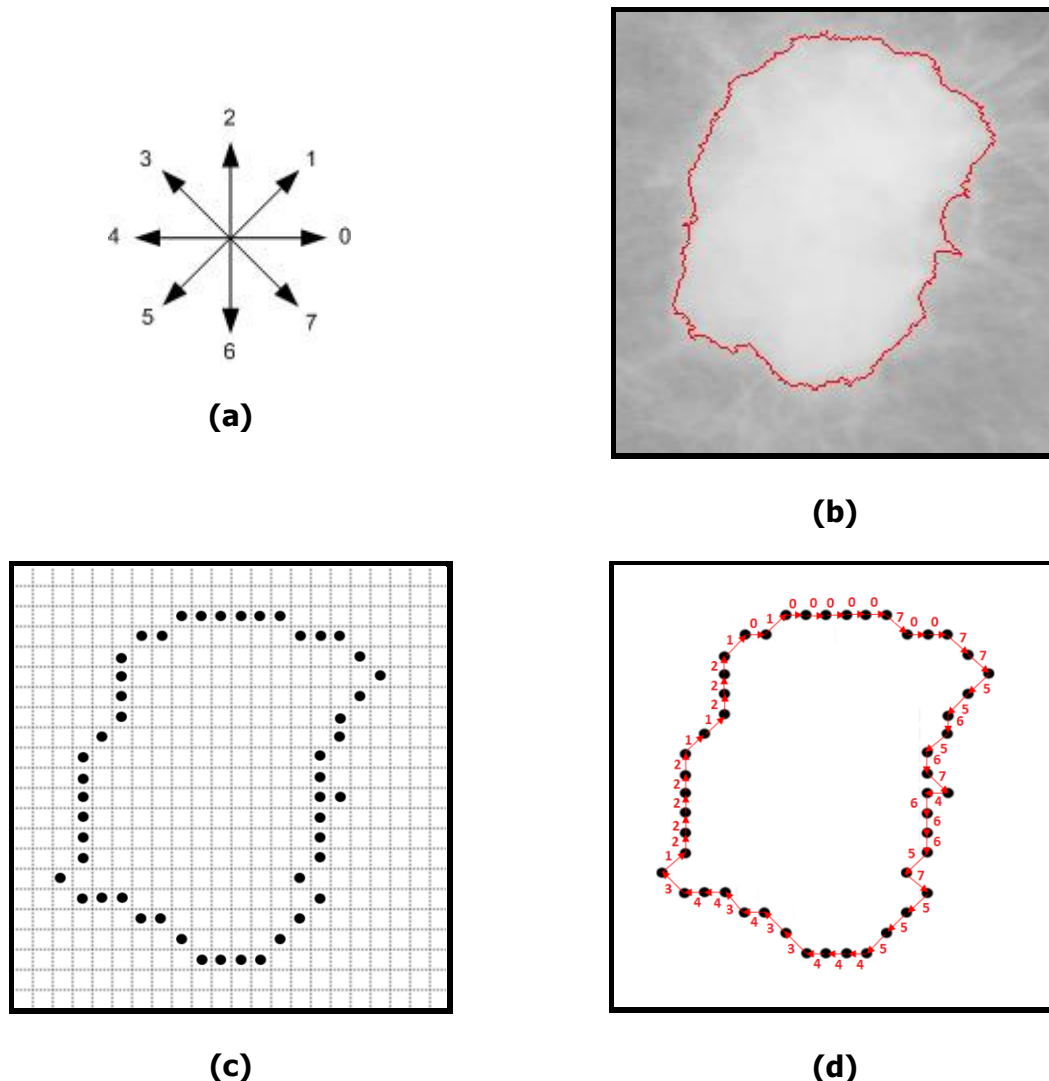


Figure 4.9. (a) Code de chaîne de Freeman pour les huit directions d'un pixel limite au suivant, (b) Image échantillonnée de masse originale "mdb184", (c) sa version rééchantillonnée sur une grille plus grossière, et (d) le code de chaîne résultant.[42]

b) Caractéristiques régionales et forme générale des masses :

- **Répartition des quarts et des moitiés de masse sur la surface totale :** La surface A est le nombre de pixels dans la masse et sur le bord de la masse. Les quatre quarts résultants relatifs à A sont : UR/ A supérieur droit, LR/ A inférieur droit, LL/ A inférieur gauche et UL/ A supérieur gauche, et les quatre moitiés résultantes relatives à A sont : U supérieur / A , droite R/ A , bas Lo/ A et gauche Lt/ A . [42, 68]

- **Moments invariants** : les moments invariants sont un ensemble de grandeurs mesurables pour décrire les masses qui sont insensibles à des déformations particulières et qui fournissent un pouvoir de discrimination suffisant pour distinguer des objets appartenant à des classes malignes ou bénignes. Ces caractéristiques comprennent : les moments de Hu, les moments centraux normalisés, l'allongement, l'orientation et la rondeur de la mammographie. [42, 68]

c) Caractéristiques des masses basées sur les frontières :

- **Longueur du périmètre** : La longueur du périmètre T de la masse est trouvée en additionnant les distances d'un pixel au suivant. C'est le nombre de pixels dans la direction de balayage horizontale ou verticale, ou le nombre de pixels multiplié par $\sqrt{2}$ pour une direction diagonale. [42, 68]
- **Rapport périmètre/diagonale** : cette caractéristique est le rapport de la moitié de la longueur du périmètre à la diagonale du rectangle de masse principal coupé. [42, 68]
- **Rapport d'aspect** : (Excentricité et Allongement) : Le rapport d'aspect A est défini comme la fraction des rayons du cercle circonscrit au cercle inscrit, qui correspond au rapport de la plus grande à la plus petite des longueurs des axes de l'ellipse d'image dans le tissu. [42, 68]
- **Rapport de compacité** : Il s'agit d'une mesure de l'écart par rapport à la complexité d'un contour de forme circulaire par rapport à la zone fermée [42, 68]. Ce rapport est de 1 pour un cercle rempli de masse normale et est supérieur à 1 pour les formes complexes réparties comme masse bénigne et maligne.
- **Irrégularité** : L'irrégularité du contour peut être mieux quantifiée par le coefficient de variation ; la variation est la distance entre les points les plus éloignés des masses à tous les angles par rapport à l'axe vertical. La variation est obtenue en faisant tourner la masse extraite à différents angles. [42, 68]
- **Énergie de flexion** : L'énergie de flexion E est une mesure de la courbure de la courbe limite de masse. Elle peut être trouvée à partir du code de chaîne en additionnant les carrés des changements de direction d'un pixel limite à l'autre. [42, 68]

7. Descripteurs de Fourier elliptiques (EFD)

EFD [68] est utile pour discriminer les formes de mammographie avec des contours lisses et irréguliers, et décrire la manière dont le comportement de la limite change. La courbe passe par tous les pixels de la frontière. Les points $(x(t), y(t)), t=1, 2, \dots, m$ peuvent être approximés en utilisant les descripteurs de Fourier elliptiques.

4.6.5. Sélection de caractéristiques avec l'algorithme mRMR

La méthode de sélection des caractéristiques Minimal-Redundancy-Maximal-Relevance (MRMR) a été utilisée pour supprimer les caractéristiques inutiles [296]. La méthode vise à trouver un sous-ensemble S de caractéristiques, en maximisant le critère d'information mutuelle, tout en essayant d'éviter la redondance entre les caractéristiques elles-mêmes. Si $MI(M, N)$ désigne l'information mutuelle entre deux variables aléatoires quelconques M et N . Alors MRMR est défini comme suit :

$$\begin{aligned} \max_S \Phi = D - R, \\ \text{where} \quad D = \frac{1}{q} \sum_{x_i \in S} MI(x_i, \omega) \quad R = \frac{1}{q^2} \sum_{x_i, x_j \in S} MI(x_i, x_j) \end{aligned} \quad (4.12)$$

La quantité Φ est maximisée sur l'ensemble S contenant les indices des entités sélectionnées, et la variable x_i dénote la i -ème entité et ω désigne les étiquettes de classe. L'optimisation utilisant la méthode MRMR peut être efficacement effectuée de manière incrémentale en étendant l'ensemble des caractéristiques sélectionnées d'un candidat à la fois, tout en garantissant une optimalité de premier ordre par rapport au critère de dépendance maximale. Le critère suivant est utilisé pour ajouter la caractéristique x_j à l'ensemble précédent de caractéristiques $q-1$:

$$\max_{x_j \in X - S_{q-1}} \left[MI(x_j, \omega) - \frac{1}{q-1} \sum_{x_i \in S} MI(x_i, x_j) \right] \quad (4.13)$$

Pour l'ensemble complet de caractéristiques X , le sous-ensemble S de q caractéristiques qui a le critère de pertinence maximal est l'ensemble qui satisfait la valeur moyenne maximale de toutes les valeurs d'information mutuelle entre les caractéristiques individuelles de x_i et la classe ω .

$$\max D(S, \omega), \quad D = \frac{1}{q} \sum_{x_i \in S} MI(x_i, \omega) \quad (4.14)$$

Le sous-ensemble S de q caractéristiques qui a le critère de redondance minimale est l'ensemble qui satisfait la valeur moyenne minimale de toutes les valeurs d'informations mutuelles entre toutes les paires de caractéristiques de x_i and x_j .

$$\min R(S), \quad R = \frac{1}{q^2} \sum_{x_i, x_j \in S} MI(x_i, x_j) \quad (4.15)$$

4.6.6. Sélection de caractéristiques avec l'algorithme Relief

Relief utilise une méthode statistique pour sélectionner les éléments pertinents. Il s'agit d'un algorithme basé sur le poids de la caractéristique qui s'inspire des algorithmes d'apprentissage basés sur les instances [19]. L'idée principale de cet algorithme est d'estimer la qualité des caractéristiques selon le fait suivant : "dans quelle mesure leurs valeurs distinguent les exemples qui sont proches les uns des autres ?". Pour cela, étant donné un exemple X choisi aléatoirement dans un jeu de données S avec k caractéristiques, Relief recherche dans le jeu de données ses deux plus proches voisins : l'un de la même classe, appelé plus proche hit H , et l'autre d'une classe différente, appelée plus proche miss M . Il met à jour l'estimation de la qualité $W[A_i]$ pour toutes les caractéristiques A_i en fonction des valeurs de la fonction différence $\text{diff}()$ pour X , H et M . m fois, où m est un paramètre défini par l'utilisateur. Pour les exemples, X_1, X_2 , la fonction $\text{diff}(A_i, X_1, X_2)$ calcule la différence entre les valeurs (X_{1i}, X_{2i}) pour la caractéristique A_i :

$$\text{diff}(A_i, x_1, x_2) = \begin{cases} |x_{1i} - x_{2i}| & \text{if } A_i \text{ is numerical} \\ 0 & \text{if } A_i \text{ is nominal and } x_{1i} = x_{2i} \\ 1 & \text{if } A_i \text{ is nominal and } x_{1i} \neq x_{2i} \end{cases} \quad (4.16)$$

The quality $W [A_i]$ is updated as follows:

$$W[A_i] := W[A_i] - \frac{\text{diff}(A_i, X, H)}{m} + \frac{\text{diff}(A_i, X, H)}{m};$$

(4.17)

4.7. Modélisation, sélection et agrégation des caractéristiques par les systèmes de réseaux de files d'attente

Les réseaux neuronaux artificiels sont une tentative d'imiter la capacité de traitement des systèmes neuronaux biologiques.

L'idée de base est de réaliser des systèmes capables d'effectuer des tâches de traitement complexes en interconnectant un nombre élevé d'éléments de traitement agencés qui peuvent même fonctionner en parallèle [283]. La mise à jour des couches cachées des réseaux de neurones (couche par couche et motif par motif) a une relation étroite avec le pouvoir descriptif de l'entité dans la couche d'entrée. Une variable pertinente est une variable telle que son élimination entraîne une dégradation des performances du système d'apprentissage : le pouvoir discriminant en classification ou la qualité de prédiction en régression. Les algorithmes de sélection de caractéristiques tentent de rechercher le sous-ensemble de l'espace des caractéristiques en insérant ou en supprimant progressivement les caractéristiques de leur ensemble actuel [284]. Cette notion de séquence nous permet, d'une part, d'organiser les caractéristiques selon leur poids à partir de la caractéristique la plus descriptive jusqu'à la caractéristique la plus faible ; et, d'autre part, d'ordonner l'état du réseau - vecteur de poids, terme d'erreur, temps de calcul, etc. - pendant le processus d'apprentissage.

Parmi les différents algorithmes d'apprentissage supervisé, un réseau de neurones à propagation vers l'arrière (BPNN) est un algorithme d'apprentissage supervisé qui utilise une architecture de perceptron à plusieurs couches (MLP) et un schéma de connexion à anticipation.

Le BPNN est basé sur la règle d'apprentissage par correction d'erreurs et est basé sur la technique de descente de gradient pour résoudre un problème d'optimisation [283].

Dans cette thèse, nous avons associé deux files d'attente pour stocker l'erreur δ à chaque nœud du réseau de neurones et pondérer les matrices $w_{ij}(l)$ pour chaque couche cachée et de sortie. Les réseaux de file d'attente constitués de plusieurs stations-service sont plus adaptés pour représenter la structure de nombreux systèmes avec un grand nombre de ressources que les modèles constitués d'une seule station-service. Dans un réseau de file d'attente, au moins deux stations-service sont connectées l'une à l'autre. Une station, c'est-à-dire un nœud, dans le réseau représente une ressource dans le système réel. Les travaux peuvent en principe être transférés entre deux nœuds quelconques du réseau ; en particulier, un emploi peut être directement renvoyé au nœud qu'il vient de quitter [270].

Nous avons distingué deux types de réseaux, un « ouvert » et un « fermé ». Un réseau de file d'attente est dit ouvert lorsque des travaux peuvent entrer dans le réseau depuis l'extérieur et que des travaux peuvent également quitter le réseau. Les emplois peuvent arriver de l'extérieur du réseau à chaque nœud et quitter le réseau à partir de n'importe quel nœud [270].

Un réseau de file d'attente est dit fermé lorsque les travaux ne peuvent ni entrer ni sortir du réseau. Le nombre d'emplois dans un réseau fermé est constant. Un réseau dans lequel un nouvel emploi entre chaque fois qu'un emploi quitte le système peut être considéré comme

un réseau fermé [270]. Le réseau de neurones que nous avons proposé est un réseau ouvert (voir Figure 4.6 b) résultant d'une unification de deux poids fusionnés des deux réseaux fermés (voir Figure 4.6a, c) ; le schéma de connexion suivant est nécessaire pour décrire les réseaux de file d'attente, à savoir.

- **BPNN-1 fermé et BPNN-2 fermé** : Ce schéma de connexion consiste à apprendre par l'algorithme feed-forward la rétropropagation du gradient de deux réseaux de neurones (on les considère comme fermés) avec deux bases d'apprentissage différentes, en sauvegardant à chaque itération état du réseau dans la file d'attente : matrice de poids, l'erreur de sortie de chaque couche, etc. (Voir Algorithme 1)

- **BPNN-1 fusionné et BPNN-2 fusionné** : Ce schéma de connexion ouvre les deux réseaux fermés (BPNN-1 et BPNN-2) et fusionne l'un par l'effet de l'autre. Il est utilisé avant l'unification des poids des courants des deux réseaux fermés. (Voir Algorithme 2, la Figure 4.12 et la Figure 4.13). Ce schéma de connexion est considéré comme intermédiaire entre le premier et le dernier.

- **Le résultat de la fusion est un BPNN ouvert unifié** : Ce schéma de connexion consiste en l'apprentissage des deux réseaux précédents avec unification de ses couches cachée et de sortie avec une agrégation conjointe de leurs poids respectifs, qui sont disposés dans la file d'attente respective. Ce schéma additionne les nouveaux poids générés par les deux réseaux fusionnés (Merged BPNN-1 and Merged BPNN-2) et enfile le résultat dans une troisième file d'attente (Voir Algorithme 3 et la Figure 4.12 et la Figure 4.13).

La règle d'apprentissage de correction d'erreur est la différence entre la sortie cible et la sortie réelle. L'algorithme BPNN est conçu de manière à mettre à jour les poids dans le sens de la descente du gradient de l'erreur cumulée. Cela se fait de manière itérative, ce qui implique la minimisation de l'erreur cumulée du réseau [281]. Ci-dessous, nous donnons un bref plan de réseau de file d'attente appliqué sur l'un des modèles NN les plus populaires, le perceptron multicouche (MLP).

Considérons un problème de reconnaissance de formes à n dimensions avec c classes. Un réseau de neurones obtient un vecteur de caractéristiques $x = [x_1, \dots, x_n]^T \in R^n$ à son entrée, et produit des valeurs pour les c fonctions discriminantes $g_1(x), \dots, g_c(x)$ à sa sortie. Généralement, les NN sont entraînés pour minimiser l'erreur quadratique sur un ensemble d'apprentissage étiqueté $Z = \{z_1, \dots, z_n\}, z_j \in R^n$ [283]. Dans le cas d'un réseau à q couches (voir Figure 4.7), une notation simplifiée est utilisée pour simplifier la formulation. Notant que $E(t)$ est la somme des carrés des erreurs à la t -ième présentation :

$$E(t) = \frac{1}{2} \sum_{i=1}^{n_q} (z_i^{(q)} - s_i(t))^2 \quad (4.18)$$

Le problème d'optimisation peut être formulé comme suit :

$$\min_W E(t) = \min_W \left(\frac{1}{2} \sum_{i=1}^{n_q} (z_i^{(q)} - s_i(t))^2 \right) \quad (4.19)$$

$$z_i^{(1)}(t) = f(y_i^{(1)}) \quad (4.20)$$

$$i = 1, 2, \dots, n_1$$

Avec :

$$y_i^{(l)}(t) = \sum_{j=1}^{n_{l-1}} w_{ij}^{(l)}(t) z_j^{(l-1)} \quad (4.21)$$

$$i = 1, 2, \dots, n_l$$

La fonction objective des algorithmes d'apprentissage entre dans le cadre des problèmes d'optimisation sans contraintes dont la fonction est quadratique. La méthode de propagation du gradient met à jour progressivement les poids de réseau par une série d'optimisations partielles, couche par couche et exemple par exemple. Les poids sont ajustés à chaque itération en propageant les erreurs vers l'arrière. Ceci est poursuivi jusqu'à ce que l'erreur quadratique moyenne soit minimisée à un niveau acceptable.

Dans la suite, W représente l'ensemble des poids d'interconnexion du réseau de neurones et $W^{(l)}(t)$ la valeur à l'instant t des poids reliant les neurones des couches $(l-1)$ à ceux de la couche (l) . L'application de la méthode du gradient pour les poids de la couche (l) conduit à l'algorithme :

$$W^{(l)}(t+1) = W^{(l)}(t) - \eta(t) \frac{\partial E(t)}{\partial W^{(l)}} \quad (4.22)$$

Considérons le réseau neuronal de la Figure 4.7 dans lequel $\text{sum}_i^l(t)$ est le total de tous les signaux atteignant le i ème neurone de la couche (l) . L'optimisation globale par la méthode du gradient de rétropropagation conduit à un algorithme de la forme (en utilisant la règle de chain) :

$$w_{ij}^{(l)}(t+1) = w_{ij}^{(l)}(t) - \eta \left(\frac{\partial E(t)}{\partial z_i^{(l)}(t)} \right) \left(\frac{\partial z_i^{(l)}(t)}{\partial \text{sum}_i^l(t)} \right) \left(\frac{\partial \text{sum}_i^l(t)}{\partial w_{ij}^{(l)}} \right) \quad (4.23)$$

Soit pour la couche de sortie ($l = q$):

$$w_{ij}^{(q-1)}(t+1) = w_{ij}^{(q-1)}(t) - \eta(t) n_q \bar{\Delta}_1 E_2(t) \delta_{i,1}^{(q)}(t) z_{i,1}^{(q-1)} + w_{ij}^{(q-1)}(t) - \eta(t) n_q \bar{\Delta}_2 E_1(t) \delta_{i,2}^{(q)}(t) z_{i,2}^{(q-1)}, \quad (4.24)$$

$$w_{ij}^{(q-1)}(t+1) = w_{ij,1}^{(q-1)}(t+1) + w_{ij,2}^{(q-1)}(t+1) \quad (4.25)$$

Avec:

$$\begin{cases} w_{ij,1}^{(q-1)}(t+1) = w_{ij,1}^{(q-1)}(t) - \eta(t) n_q \bar{\Delta}_1 E_2(t) \delta_{i,1}^{(q)}(t) z_{i,1}^{(q-1)} \\ w_{ij,2}^{(q-1)}(t+1) = w_{ij,2}^{(q-1)}(t) - \eta(t) n_q \bar{\Delta}_2 E_1(t) \delta_{i,2}^{(q)}(t) z_{i,2}^{(q-1)} \end{cases} \quad (4.26a-b)$$

Et :

$$\begin{cases} \delta_{i_1}^{(q)} = \left(z_{j_1}^{(q)}(t) - s_i(t) \right) \frac{\partial f(\text{sum}_{i_1}^{(q)}(t))}{\partial \text{sum}_{i_1}^{(q)}(t)} \\ \delta_{i_2}^{(q)} = \left(z_{j_2}^{(q)}(t) - s_i(t) \right) \frac{\partial f(\text{sum}_{i_2}^{(q)}(t))}{\partial \text{sum}_{i_2}^{(q)}(t)} \end{cases} \quad (4.27a-b)$$

Pour les couches cachées:

$$w_{ij}^{(l-1)}(t+1) = w_{ij_1}^{(l-1)}(t) - \eta(t)n_q \overline{\Delta_1} E_2(t) \delta_{i_1}^{(l)}(t) z_{j_1}^{(q-1)} + w_{ij_2}^{(l-1)}(t) - \eta(t)n_q \overline{\Delta_2} E_1(t) \delta_{i_2}^{(l)}(t) z_{i_2}^{(l-1)}, \quad (4.28)$$

$$w_{ij}^{(l-1)}(t+1) = w_{ij_1}^{(l-1)}(t+1) + w_{ij_2}^{(l-1)}(t+1) \quad (4.29)$$

Avec:

$$\begin{cases} w_{ij_1}^{(l-1)}(t+1) = w_{ij_1}^{(l-1)}(t) - \eta(t)n_q \overline{\Delta_1} E_2(t) \delta_{i_1}^{(l)}(t) z_{j_1}^{(l-1)} \\ w_{ij_2}^{(l-1)}(t+1) = w_{ij_2}^{(l-1)}(t) - \eta(t)n_q \overline{\Delta_1} E_1(t) \delta_{i_2}^{(l)}(t) z_{j_2}^{(l-1)} \end{cases} \quad (4.30a-b)$$

Et :

$$\begin{cases} \delta_{i_1}^{(l)} = \frac{\partial f(\text{sum}_{i_1}^{(l)}(t))}{\partial \text{sum}_{i_1}^{(l)}(t)} \sum_{m=1}^{n_{l+1}} \delta_{m_1}^{(l+1)} w_{mi_1}^{(l+1)} \\ \delta_{i_2}^{(q)} = \frac{\partial f(\text{sum}_{i_2}^{(l)}(t))}{\partial \text{sum}_{i_2}^{(l)}(t)} \sum_{m=1}^{n_{l+1}} \delta_{m_2}^{(l+1)} w_{mi_2}^{(l+1)} \end{cases}, \quad (4.31a-b)$$

Et :

$$\begin{cases} \overline{\Delta_1} = \overline{z_{i_1}^{(q)} - s_{i_1}} \\ \overline{\Delta_2} = \overline{z_{i_2}^{(q)} - s_{i_2}} \end{cases} \quad (4.32a-b)$$

Tel que $\overline{\Delta_1}$ et $\overline{\Delta_2}$ sont respectivement les moyennes des différences entre la valeur prédite mesurée et celle réelle, et $n_q = n_{q_1} = n_{q_2}$ correspond au nombre des nœuds de la couche de sortie des deux réseaux.

Au début, tous les poids sont initialement affectés par des nombres aléatoires de petites valeurs, ainsi la valeur du taux d'apprentissage η . Toutes les valeurs de u_j dans le réseau sont ensuite calculées en utilisant les valeurs courantes des poids et le vecteur initial de caractéristiques de la première image mammographique (propagation directe). Pour le δ , cependant, nous procédons vers l'arrière, c'est-à-dire de la couche sortie, vers la couche d'entrée (Back propagation).

4.7.1. Algorithme 1. Apprentissage de MLP à rétro propagation avec réseau à file d'attente fermée

Supposons que la topologie du réseau de neurones soit choisie et fixée dès le début (le nombre de couches cachées et le nombre de nœuds à chaque couche cachée), et que la fonction d'activation est différentiable. Le processus d'apprentissage d'un réseau de neurones fermé consiste, à *enfiler* conjointement les différents états de mise à jour de la matrice de poids en file d'attente jusqu'à atteindre la convergence de sa représentation optimale et, à fermer le

réseau de neurones pour que la tâche de mise à jour ne puisse ni entrer ni quitter le réseau depuis l'extérieur d'un autre réseau de neurones. Un modèle de réseau neuronal fermé et de mise en file d'attente est illustré à la Figure 4.10(a), (Voir la Figure 4.12 et la Figure 4.13). [42]

1. *Choisissez une structure MLP* Choisissez le nombre de couches cachées, le nombre de nœuds à chaque couche et les fonctions d'activation.
2. *Initialiser la procédure d'apprentissage* Choisissez de petites valeurs aléatoires pour tous les poids (y compris les biais) du NN. Choisissez le taux d'apprentissage $\eta > 0$, le nombre maximal d'époques T et l'objectif d'erreur $\varepsilon > 0$.
3. Mettez $E > \infty$, le compteur d'époques $t = 1$ et le compteur d'objets $j = 1$, génération aléatoire de w_{prec} .
4. Tant que ($E > \varepsilon$ et $t \leq T$) faire
 - (a) Soumettez z_j comme exemple d'apprentissage suivant.
 - (b) Calculez la sortie de chaque nœud du NN avec les poids actuels (propagation directe), par Eq. (4.20)
 - (c) Calculez le terme d'erreur δ à chaque nœud de la couche de sortie par Eq. (4.26a).
 - (d) Calculez le terme de mise à jour $-\eta m_q \bar{\Delta} \delta u$ à chaque nœud de la couche de sortie en utilisant les valeurs respectives de δ et u .
 - (e) Mettre en file d'attente (*enfiler*) et enregistrer dans la file d'attente respective tous les termes de mise à jour aux nœuds de la couche de sortie.
 - (f) Calculez de manière récursive tous les termes d'erreur aux nœuds des couches cachées à l'aide de l'équation. (4.31a) (propagation vers l'arrière).
 - (g) Calculez le terme de mise à jour $-\eta m_q \bar{\Delta} \delta u$ à chaque nœud des couches cachées en utilisant les valeurs respectives de δ et u .
 - (h) Mettre en file d'attente (*enfiler*) et enregistrer dans la file d'attente respective tous les termes de mise à jour aux nœuds des couches cachées.
 - (i) Pour chaque couche cachée et chaque couche de sortie, mettez à jour les poids en utilisant :

$$w_{curr} = w_{prec} - \eta m_q \bar{\Delta} \delta u, (4.33)$$
 en utilisant les valeurs respectives de δ et u .
 - (j) Mettre en file d'attente (*enfiler*) et enregistrer séparément chaque poids caché et chaque poids de sortie dans les files d'attente respectives.
 - (k) Calculez E en utilisant les poids actuels et Eq. (4.18).
 - (l) Mettre en file d'attente (*enfiler*) et enregistrer E dans une file d'attente séparément.
 - (m) Si $j = N_j$ (un passage complet à travers Z (époque) est terminé), alors mettez $t = t + 1$ et mettez $j = 0$. Sino ; mettez $j = j + 1$.
5. Fin % (Tant que).

4.7.2. Algorithme 2. Fusion de chacun des BPNN fermés avec l'autre, en utilisant un réseau de file d'attente ouvert

Le processus de fusion des deux réseaux de neurones fermés consiste, à ouvrir chacun de sorte que nous fusionnions respectivement les termes des mises à jour $-\eta m_q \bar{\Delta} \delta u$ d'un réseau fermé avec les autres erreurs E de l'autre. (Voir la Figure 4.10)

L'Algorithme 2 *défile* progressivement en même temps, le terme de mise à jour $-\eta m_q \bar{\Delta} \delta u$ et le poids respectif pour le réseau fermé *courant*, et le terme d'erreur E de *l'autre* réseau fermé.

Ensuite, il calcule les *poids fusionnés* des deux réseaux fermés fusionnés : un modèle de réseau fermé de neurones et de files d'attente est illustré à la Figure 4.10 (c), (Voir la Figure 4.12 et la Figure 4.13). [42]

1. *Choisissez une structure MLP* Choisissez le nombre de couches cachées, le nombre de nœuds à chaque couche et les fonctions d'activation.
2. *Initialisez la procédure d'apprentissage*, Utilisez un tableau de file d'attente qui stocke les valeurs de tous les poids fusionnés de chaque couche séparément, y compris les biais. Prendre le même taux d'apprentissage η , et le même nombre d'époques obtenues t_1 et t_2 ; ceux utilisés des deux réseaux de file d'attente fermés. Et *mettre à jour* les files d'attente, les files d'attente d'erreurs et les files d'attente de matrices de *poids* régulées à chaque itération, utilisées par les deux réseaux fermés.
3. Définissez les compteurs d'époque $t_{curr} = t_1$, $t_{other} = t_2$ des deux réseaux de neurones fermés.
4. Tant que ($t > 0$ et $t \leq \min(t_{curr}, t_{other})$) faire :
 - (a) Retirer (*défiler*) de la file d'attente "d'erreur", le terme d'erreur E d'*autre réseau* de neurones fermé.
 - (b) Retirer (*défiler*) de la file d'attente "des mises à jour", le terme de mise à jour $-\eta m_q \bar{\Delta} \delta u$ du réseau de neurones fermé *courant*.
 - (c) Retirer de la file d'attente (*défiler*) de la file d'attente des "poids", pondère les termes du réseau de neurones fermé actuel.
 - (d) Pour chaque poids caché et chaque poids de sortie, mettez à jour les poids en utilisant :

$$w'_{curr} = w_{prec} - \eta m_q \bar{\Delta} E \delta u, \quad (4.34)$$
 en utilisant les poids, les erreurs et les mises à jour respectifs stockés dans les files d'attente respectives.
 - (e) Mettre en file d'attente (*enfiler*), les "poids fusionnés" dans les files d'attente respective séparément.
 - (f) Mettez $t = t - 1$.
5. Fin % (Tant que).

4.7.3. Algorithme 3. Apprentissage d'un MLP à rétro propagation avec réseau de file d'attente ouvert

Le processus d'apprentissage d'un réseau de neurones ouvert consiste, à ouvrir les deux réseaux de neurones fermés afin d'unifier les états de sa matrice de poids. Algorithme 3 *défile(popFronts)* progressivement les poids fusionnés des deux réseaux de neurones fusionnés jusqu'à vider leurs files d'attente et *enfile* la somme dans un troisième réseau. Sur la Figure 4.10 (c), une unification du modèle de réseau de neurones et de file d'attente fusionné est illustrée. (Voir la Figure 4.12 et la Figure 4.13)

1. *Choisissez une structure MLP* Choisissez le nombre de couches cachées, le nombre de nœuds à chaque couche et les fonctions d'activation.
2. *Initialiser la procédure d'apprentissage* Utiliser un tableau de file d'attente qui stocke les valeurs de tous les nouveaux poids fusionnés dans la file d'attente séparément, y compris les biais. Choisissez le même taux d'apprentissage η , et le même nombre d'époques t_1 , t_2 obtenues des deux réseaux fermés, et les matrices files d'attente *pondère* régulées à chaque itération, utilisées par les deux réseaux fermés.

3. Mettre les nouveaux poids fusionnés $w'_{curr_1} = w'_{0_1}, w'_{curr_2} = w'_{0_2}$ (w'_{0_1}, w'_{0_2} sont les poids initiaux des deux réseaux fermés et fusionnés générés à l'aide de l'algorithme 2), et $w'_{curr} = w'_{curr_1} + w'_{curr_2}$.

4. Tant que (not **isEmpty**(w'_{curr_1})) faire (Nous avons $\text{length}(w'_{curr_1}) = \text{length}(w'_{curr_2})$)

(a) Retirer de la file d'attente (*popFront*) les vecteurs de poids stockés dans les files d'attente respectives des deux réseaux de neurones fermés et mettez à jour les poids par :

$$w'_{curr} = w'_{curr_1} + w'_{curr_2} \quad (4.35)$$

(b) Mettez en file d'attente (*Pushback*) et enregistrez séparément chacun des nouveaux poids fusionnés, cachés et ceux en sortie dans les files d'attente.

(c) Calculez E en utilisant les poids actuels et l'équation (4.18).

5. Fin % (Tant que).

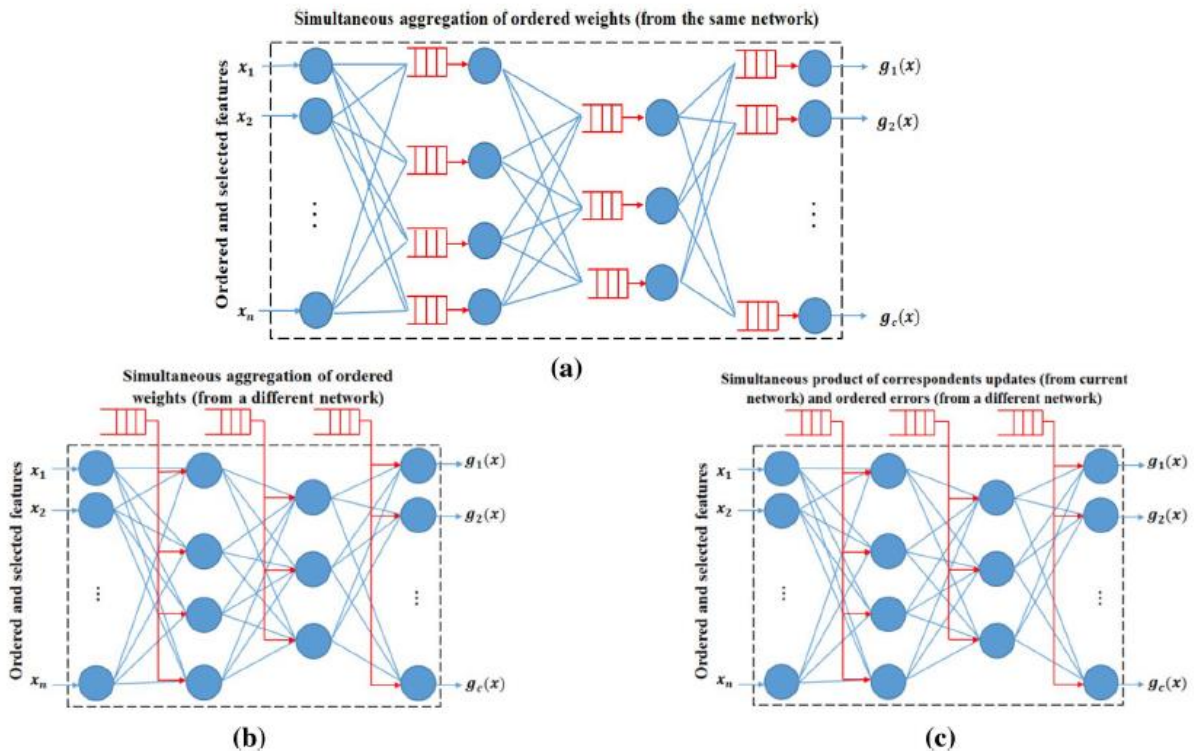


Figure 4.10. (a) Réseau neuronal fermé, (b) unification des deux réseaux neuronaux fermés, (c) fusion de chacun des réseaux neuronaux fermés avec l'autre

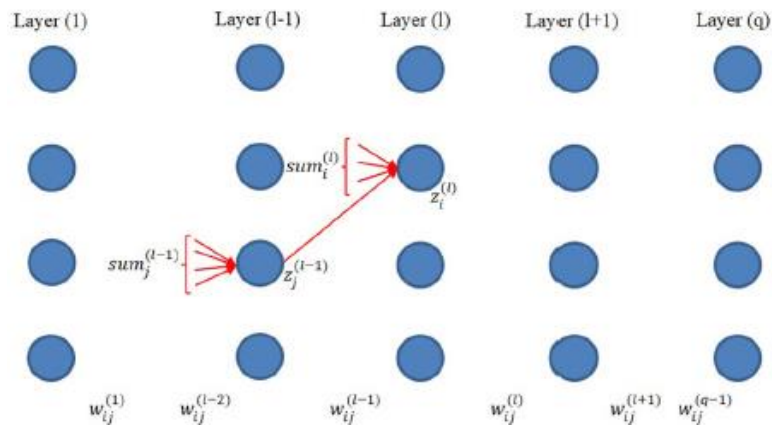


Figure 4.11. Un modèle générique de MLP avec q couches.

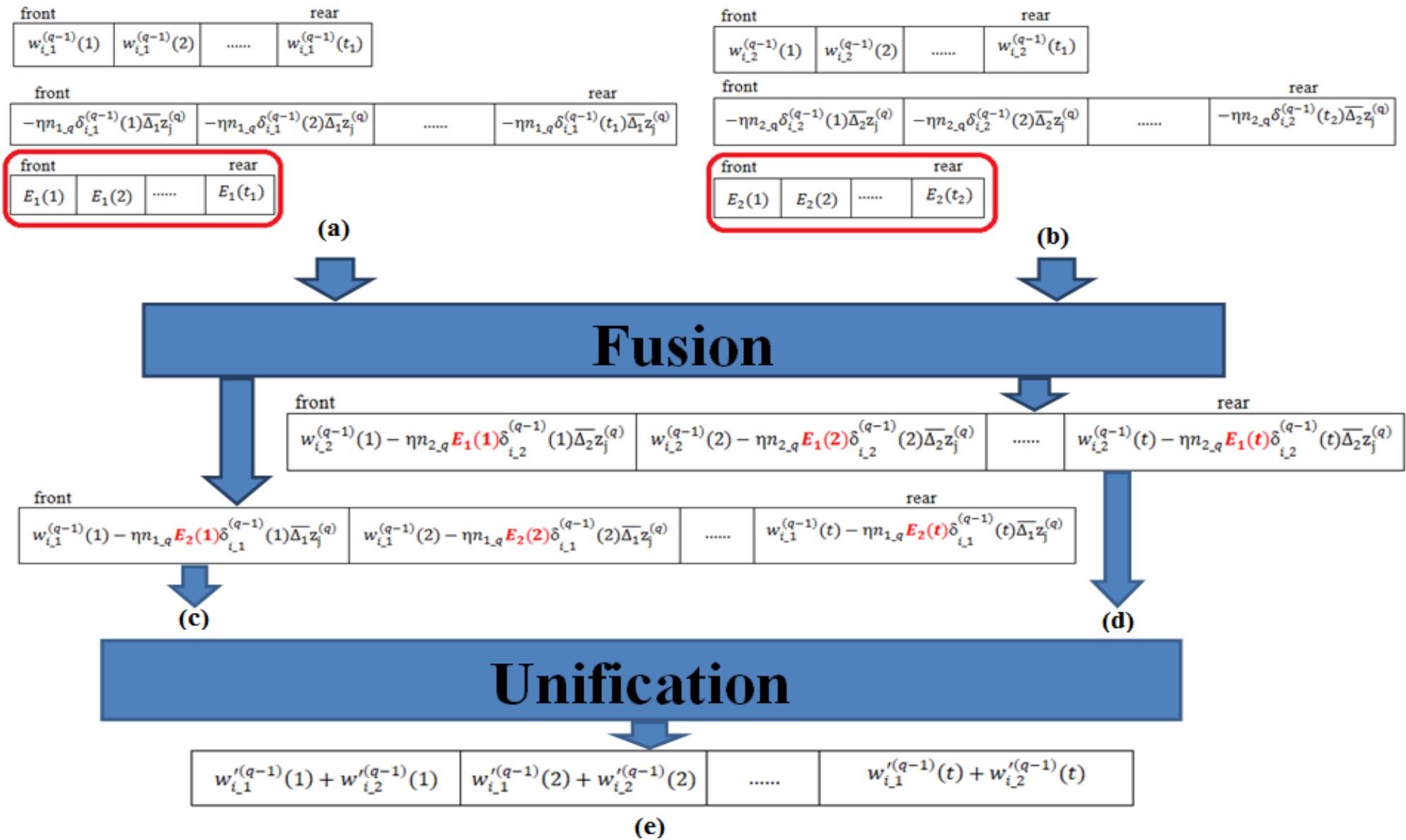


Figure 4.12. Stockage des poids et des termes d'erreurs de la couche de sortie du NN dans la file d'attente. (a) Premier réseau fermé. (b) Deuxième réseau fermé. (c) Fusion du premier réseau fermé avec les erreurs du second. (d) Fusion du deuxième réseau fermé avec les erreurs du premier. (e) Réseau unifié ouvert.

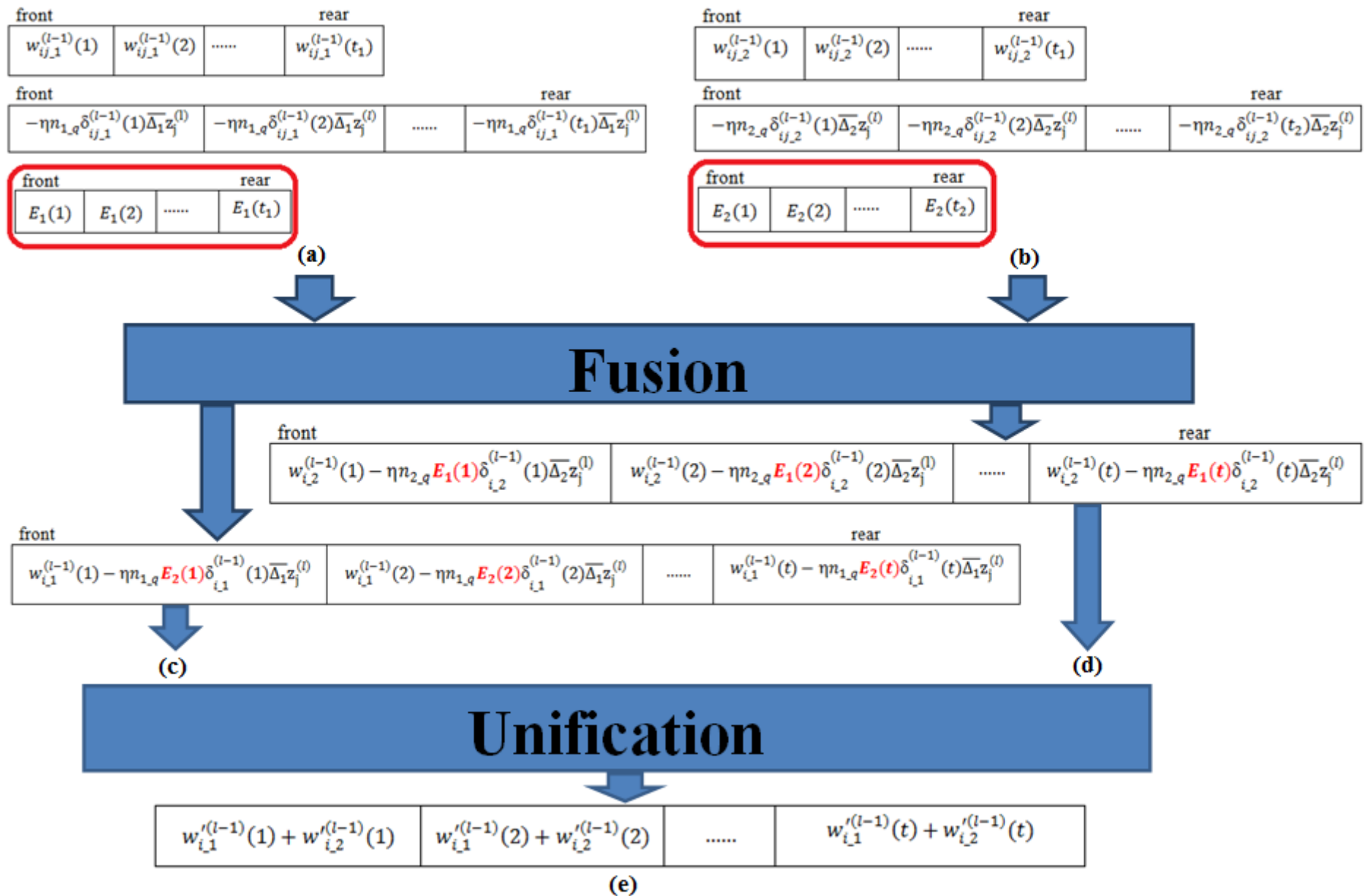


Figure 4.13. Stockage des poids et des termes d'erreurs des couches cachées du NN dans la file d'attente. (a) Premier réseau fermé. (b) Deuxième réseau fermé. (c) Fusion du premier réseau fermé avec les erreurs du second. (d) Fusion du deuxième réseau fermé avec les erreurs du premier. (e) Réseau unifié ouvert

4.8. Conclusion

Ce chapitre décrit une méthode de classification des images de mammographies en utilisant quatre modules différents. La méthode proposée utilise des règles symboliques, un filtre de diffusion anisotrope et des caractéristiques texturales et morphologiques pour pré-classifier et classifier les images mammographies. La sélection de caractéristiques est effectuée à l'aide des algorithmes MRMR et Relief pour identifier les caractéristiques les plus informatives et pertinentes :

1. Méthode de classification : La méthode de classification comprend quatre modules.
 - Module 1 : Construction automatique d'un ensemble de règles symboliques pour la pré-classification des mammographies à travers un arbre de décision.
 - Module 2 : Traitement des images avec un filtre de diffusion anisotrope pour réduire le bruit spécifique aux ensembles de données.
 - Module 3 : Extraction et sélection de caractéristiques à partir des images traitées, en utilisant des caractéristiques texturales et morphologiques.
 - Module 4: Classification des mammographies en utilisant les caractéristiques sélectionnées avec les algorithmes MRMR et RELIEF.
2. Pré-classification des mammographies basée sur des règles d'associations : Cette étape utilise des règles symboliques de base pour pré-classifier les seins en fonction de particularités tissulaires.
3. Filtre de diffusion anisotrope : Utilisation d'un filtre adaptatif pour réduire le bruit dans la zone et renforcer les bords d'une masse.
4. Caractéristiques texturales : Utilisation de caractéristiques du premier et du second ordre pour étudier la distribution des intensités de pixels dans les images.
5. Caractéristiques morphologiques : Utilisation de différentes mesures pour décrire la variation de la structure extérieure de la forme des masses dans les images.
6. Critères d'analyse d'évaluation individuelles des algorithmes de sélection des caractéristiques: Utilisation de critères tels que la dispersion, l'information mutuelle, le coefficient de corrélation et la détection de redondance pour sélectionner les caractéristiques les plus pertinentes.
7. Sélection de caractéristiques avec l'algorithme MRMR : Utilisation de la méthode MRMR pour sélectionner un sous-ensemble de caractéristiques qui maximise l'information mutuelle tout en évitant la redondance.
8. Sélection de caractéristiques avec l'algorithme Relief : Utilisation de l'algorithme Relief, basé sur le poids des caractéristiques, pour estimer la qualité des caractéristiques en fonction de leur capacité à distinguer les exemples proches les uns des autres.

Résultats, comparaison et discussion de différentes approches proposées

5.1. Introduction

Dans ce chapitre, nous décrivons un nouveau cadre de classification des mammographies pour classer les tissus mammaires comme normaux, bénins ou malins. Dans un premier temps, nous établissons un ensemble de règles symboliques de pré-classification des mammographies à partir des particularités d'un tissu (le fond, la classe et la sévérité de l'anomalie) prenant en compte les différentes formes de sein selon leur anomalie. Un arbre de décision reflétant une attribution préliminaire à partir de la région d'intérêt (ROI) d'une mammographie représente cette pré-classification. Les images obtenues par cette technique sont traitées avec un filtre de diffusion anisotrope pour réduire le bruit et préserver les contours. Ensuite, une matrice de caractéristiques est générée en utilisant : d'une part, les caractéristiques texturales d'une matrice de cooccurrence de niveau de gris (GLCM) et d'une matrice de longueurs de parcours de niveau de gris (GLRLM) appliquées à toutes les régions d'intérêt filtrées (ROI) d'une mammographie ; et d'autre part, après la segmentation de la région de masse, les caractéristiques extraites de la limite de masse et la région de marge entre la masse et le fond pour la classification. Pour dériver les caractéristiques pertinentes de la matrice de caractéristiques, nous avons recours à la méthode de sélection des caractéristiques RELIEF et MRMR séparément.

La modélisation du BPNN (Back Propagation Neural Network) par un réseau de files d'attente est proposée pour représenter les propriétés texturales et morphologiques des masses. Les caractéristiques pertinentes et ordonnées sont injectées dans le classificateur BPNN à l'aide du réseau de file d'attente. La base de données standard MIAS est utilisée pour la validation du schéma proposé. Dans tous les cas, nous avons observé que l'ouverture du réseau de neurones fermé améliore la performance. Cependant, la méthode de sélection des caractéristiques MRMR surpasse RELIEF du point de vue de précision et aire sous la courbe (AUC) de la caractéristique de fonctionnement du récepteur (ROC). Ces mesures sont respectivement de 98,1 % et 0,9650 pour la classification normale vs anormale, alors qu'elles sont respectivement de 95,2 % et 0,9200 pour la classification bénigne vs maligne.

5.2. Modélisation, sélection et agrégation de caractéristiques avec un système de réseau de file d'attente

Cette section fournit une explication détaillée du processus de modélisation, de sélection et d'agrégation de caractéristiques en utilisant des réseaux de neurones artificiels et des réseaux de file d'attente. Il met en lumière l'importance des variables pertinentes et explique comment ces concepts sont appliqués dans le contexte de la reconnaissance de la masse mammaire.

Les réseaux neuronaux artificiels sont une tentative d'imiter la capacité de traitement des systèmes neuronaux biologiques. L'idée de base est de réaliser des systèmes capables d'effectuer des tâches de traitement complexes en interconnectant un grand nombre d'éléments de traitement agencés qui peuvent même fonctionner en parallèle [283]. La mise à jour des couches cachées des réseaux de neurones (couche par couche et motif par motif) a une relation étroite avec le pouvoir descriptif de l'entité dans la couche d'entrée. Une variable pertinente est une variable telle que son élimination entraîne une dégradation des performances du système d'apprentissage - le pouvoir discriminant en classification ou la qualité de prédiction en régression. Les algorithmes de sélection d'entités tentent de rechercher le sous-ensemble d'espace d'entités en insérant ou en supprimant progressivement les entités de leur ensemble actuel [284]. Cette notion de séquence permet, d'une part, d'organiser les traits selon leur poids du trait le plus descriptif au trait le plus faible ; et, d'autre part, d'ordonner l'état du réseau - vecteur de poids, terme d'erreur, temps de calcul, etc. - lors du processus d'apprentissage.

- Parmi les différents algorithmes d'apprentissage supervisé, un BPNN (Backward Propagation Neural Network) est un algorithme d'apprentissage supervisé qui utilise l'architecture MLP (Multiple Layer Perceptron) et un schéma de connexion feed-forward. Le BPNN est basé sur la règle d'apprentissage de correction d'erreurs et est basé sur la technique de descente de gradient pour résoudre un problème d'optimisation [283].

- À chaque nœud du réseau de neurones une file d'attente est associée pour stocker l'état actuel de ce réseau. Les réseaux de file d'attente constitués de plusieurs stations-service sont plus adaptés pour représenter la structure de nombreux systèmes avec un grand nombre de ressources que les modèles constitués d'une seule station-service. Dans un réseau de file d'attente, au moins deux stations-service sont connectées l'une à l'autre. Une station, c'est-à-dire un nœud, dans le réseau représente une ressource dans le système réel. Les travaux peuvent en principe être transférés entre deux nœuds quelconques du réseau ; en particulier, un travail peut être renvoyé directement au nœud qu'il vient de quitter [270].

- Deux types de réseaux ont été distingués, l'un est fermé, l'autre est ouvert. Un réseau de file d'attente est dit ouvert lorsque des travaux peuvent entrer dans le réseau depuis l'extérieur et que des travaux peuvent également quitter le réseau. Les travaux peuvent arriver de l'extérieur du réseau à chaque nœud et quitter le réseau à partir de n'importe quel nœud [270].

- Un réseau de file d'attente est dit fermé lorsque les travaux ne peuvent ni entrer ni sortir du réseau. Le nombre d'emplois dans un réseau fermé est constant. Un réseau dans lequel un nouvel emploi entre chaque fois qu'un emploi quitte le système peut être considéré comme un réseau fermé [270].

- Le réseau de neurones proposé dans ce travail est un réseau ouvert résultant d'une fusion (hybridation) de deux réseaux fermés ; le schéma de connexion suivant est nécessaire pour décrire les réseaux en file d'attente, à savoir :

• **BPNN-1 fermé et BPNN-2 fermé** : Ce schéma de connexion consiste à apprendre par l'algorithme feed-forward la rétro-propagation du gradient de deux réseaux de neurones (on les considère comme fermés) avec deux bases d'apprentissage différentes, en sauvegardant à chaque itération l'état du réseau en file d'attente : matrice de poids, l'erreur de sortie de chaque couche, etc. (Voir Algorithme 1 de la section 6 du quatrième chapitre)

• **Le résultat de la fusion est un BPNN ouvert** : Ce schéma de connexion consiste en l'apprentissage de l'un des deux réseaux précédents (est considéré comme ouvert) avec la

mise à jour de ses couches cachées avec une agrégation conjointe des états du deuxième réseau, qui sont disposés dans sa file d'attente. (Voir Algorithme 2 de la section 6 du quatrième chapitre)

L'objectif principal de ce travail est de proposer une approche de sélection de caractéristiques pour la reconnaissance de la masse mammaire. Ce travail présente les étapes de la démarche que nous avons suivie pour atteindre cet objectif (voir Figure 4.1 du quatrième chapitre), les expérimentations que nous avons menées et les résultats obtenus. Nous avons choisi un ensemble de caractéristiques de départ (voir les sections 5.2 et 3.4) pour représenter les classes de masse mammaire. Nous effectuons la reconnaissance de ces classes en utilisant des classificateurs de type QBPNN ayant en entrées toutes les caractéristiques sélectionnées. Nous avons opté pour la sélection des caractéristiques selon l'approche Symbiosis qui est basée sur les résultats obtenus par les classificateurs QBPNN utilisés. Pour chacune des catégories de ROI, les résultats de l'approche de sélection des caractéristiques fonctionnent comme suit :

1. La construction de la base de données ROI en utilisant une approche basée sur des règles et une représentation graphique d'un arbre de décision génère une base des ROI de 272 éléments distribués de (voir la section 4.6.1 du quatrième chapitre; et la Figure 4.7).
2. La filtration des ROI à l'aide d'une technique de diffusion anisotrope. Les paramètres de cette technique sont listés dans la section 5.3.2. Les résultats obtenus à ce stade sont présentés dans la Figure 5.7 (deuxième ligne).
3. Pour chaque catégorie filtrée des ROI, extrayez :
 - Caractéristiques texturales de premier ordre.
 - Caractéristiques texturales de second ordre à partir des matrices de cooccurrence de niveau de gris et de longueurs d'exécution en niveau de gris (GLRLM). (Voir la section 5.3.2 et les Tableaux 5.6 et 5.7)
 - Caractéristiques morphologiques issues du tracé des contours de masse, listés dans la section 5.3.2 et présentés dans la Figure 5.7.
4. Classez les caractéristiques en fonction de leur capacité de discrimination (de la plus discriminante à la moins discriminante) en utilisant l'un des algorithmes de sélection de caractéristiques individuelles MRMR ou RELIEF, en évaluant la corrélation entre les caractéristiques afin d'éliminer les redondances, en calculant :
 - Critère de dispersion.
 - Critère d'information mutuelle.
 - Corrélation linéaire absolue entre chaque caractéristique et toutes les autres.
 - Redondance des caractéristiques.

Les calculs et les résultats obtenus à ce stade sont présentés dans les Figures 5.1, 5.2.

5. Commencer avec un sous-ensemble vide et effectuer une agrégation progressive des caractéristiques (approche Forward, voir la section 5.3.3), en petits groupes (par exemple de 5) en commençant par les caractéristiques les plus discriminantes (ordre déterminé dans la première étape de sélection) et en éliminant les caractéristiques redondantes (les plus corrélées), c'est-à-dire conserver les caractéristiques les moins corrélées, donc les plus pertinentes.

A chaque agrégation de caractéristiques, l'apprentissage et le test de deux classificateurs fermés QBPNN correspondant à la catégorie de masse considérée sont effectués et leurs performances évaluées. Les détails de ce processus sont présentés dans la section 5.3.3. Les calculs et les résultats obtenus à ce stade sont présentés dans les Tableaux 5.1, 5.2, 5.3, 5.4 et 5.5.

6. Le processus d'apprentissage itératif consiste à :

• Fusionner chaque réseau QBPNN fermé en mettant en file d'attente (enfiler (pushback) dans la file d'attente respective) à la fois les poids de l'un et les erreurs de l'autre, (voir Algorithmes 1 et 2 de la section 4.7), et unifier les deux réseaux QBPNN en défilant (popFront de la file d'attente) en même temps les poids fusionnés de chacun (Voir Algorithme 3 de la section 4.7). Les calculs effectués et les résultats obtenus à ce stade sont présentés dans les Figures 5.7. Tout au long de ce travail, les résultats des classificateurs que nous avons proposés sont comparés avec SVM-RFE et CNN pour comparer les supériorités. Dans la section suivante, nous présentons sept sections, en commençant par avec la pré-classification basée sur des règles à partir d'un arbre de décision pour extraire les régions d'intérêt (ROI). Ensuite, la filtration de la diffusion anisotrope est présentée dans la section suivante. La troisième et la quatrième section décrivent l'étape d'extraction des caractéristiques texturales et morphologiques des mammographies qui sont déjà filtrés. La dernière section est la plus fondamentale de ce travail. Il décrit la modélisation, la sélection et l'agrégation de caractéristiques avec des systèmes de réseau de file d'attente. Enfin, nous présentons nos expériences ; discuter des résultats, suivis d'une conclusion et des perspectives d'extensions futures.

Tableau 5.1. Différentes valeurs des mesures de performance du classificateur à l'aide de deux méthodes de sélection de caractéristiques pour la classification du retour pour la classification normal-anormal avec n = 25.

Feature selection method	R best arranged features	Performances mesures (%)								
		First closed classifier			Second closed classifier			Unified open classifier		
		Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
	114	73.8	91.3	82.5	87.5	88.9	88.2	77.8	93.3	87.5
	92	88.2	93.3	90.7	89.2	94.0	91.2	100	91.2	95.0
	77	82.7	96.1	89.4	75.0	77.8	76.5	94.7	96.1	95.8
	47	94.0	93.0	93.5	77.8	87.5	96.4	98.0	97.0	98.1
	38	98.0	72.4	85.2	66.7	54.5	58.9	100	74.4	85.4
	104	71.0	90.8	87.5	100	66.7	82.4	75.0	93.8	87.5
	73	63.2	96	91.7	87.5	77.8	82.4	69.2	100	91.7
	56	83.8	97	95.8	88.9	87.5	88.2	89.2	100	96.8
	21	77.6	91.2	89.6	87.5	66.7	76.5	82.6	96.0	89.6
	20	63.5	92.5	83.3	62.5	77.8	70.6	64.7	93.5	83.3

Tableau 5.2. Différentes valeurs des mesures de performance du classificateur à l'aide de deux méthodes de sélection de caractéristiques pour la classification bénigne-maligne du retour sur investissement avec n = 25.

Feature selection method	R best arranged features	Performances mesures (%)								
		First closed classifier			Second closed classifier			Unified open classifier		
		Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
	114	77.8	93.3	87.5	87.5	88.9	88.2	82.3	9.01	85.6
	92	88.2	93.3	91.7	91.2	90.0	91.6	94.6	91.8	89.9
	77	92.0	95.0	93.5	89.8	90.6	90.4	100	90.2	95.2

57	85.7	97.1	93.8	75.0	77.8	76.5	80.3	87.9	86.8
47	100	77.4	85.4	66.7	54.5	58.9	84.2	66.3	73.9
101	69.2	100	91.7	87.5	77.8	82.4	78.6	89.2	87.0
73	88.2	100	95.8	88.9	87.5	88.2	88.2	94.0	93.1
42	75.0	93.8	87.5	100	66.7	82.4	87.5	79.9	85.1
21	82.6	96.0	89.6	87.5	66.7	76.5	85.1	82.2	84.2
20	64.7	93.5	83.3	62.5	77.8	70.6	64.3	85.9	77.1

5.3. Analyse des résultats expérimentaux

Afin d'évaluer l'efficacité du cadre de classification proposé pour l'amélioration de la prédiction des images mammographiques, plusieurs simulations ont été réalisées dans les environnements MATLAB et C++. Pour l'analyse de la méthode proposée, nous avons considéré la base de données MiniMIAS indiquée par[297], qui se compose d'images du sein gauche et droit de taille 1024×1024 pixels avec une résolution de tons de gris de 8 bits par pixel obtenu de 161 patients. Dans l'ensemble d'images qui contient des mammographies numérisées illustrant différents types d'anomalies, une lésion de masse fait référence à des masses (circonscrites ou spiculées), des distorsions architecturales, des asymétries et des densités en développement ; c'est-à-dire toute lésion non calcifiante, des exemples d'affections bénignes et malignes, ainsi qu'un ensemble complet de normales. Les images sont automatiquement catégorisées à l'aide d'un classement préliminaire par un arbre de décision (voir Figure 2 du quatrième chapitre). Les images mammographiques qui sont produites par l'arbre de décision sont soumises à une étape de recadrage, cette étape consiste à découper précisément l'image en limites de ROIs fournissant les coordonnées spatiales du centre de l'anomalie ainsi que le rayon du cercle qui l'entoure. Pour l'extraction du retour sur

investissement normal, la même procédure de recadrage est effectuée sur des images mammographiques normales avec une sélection aléatoire de l'emplacement (comme 60 x 60).

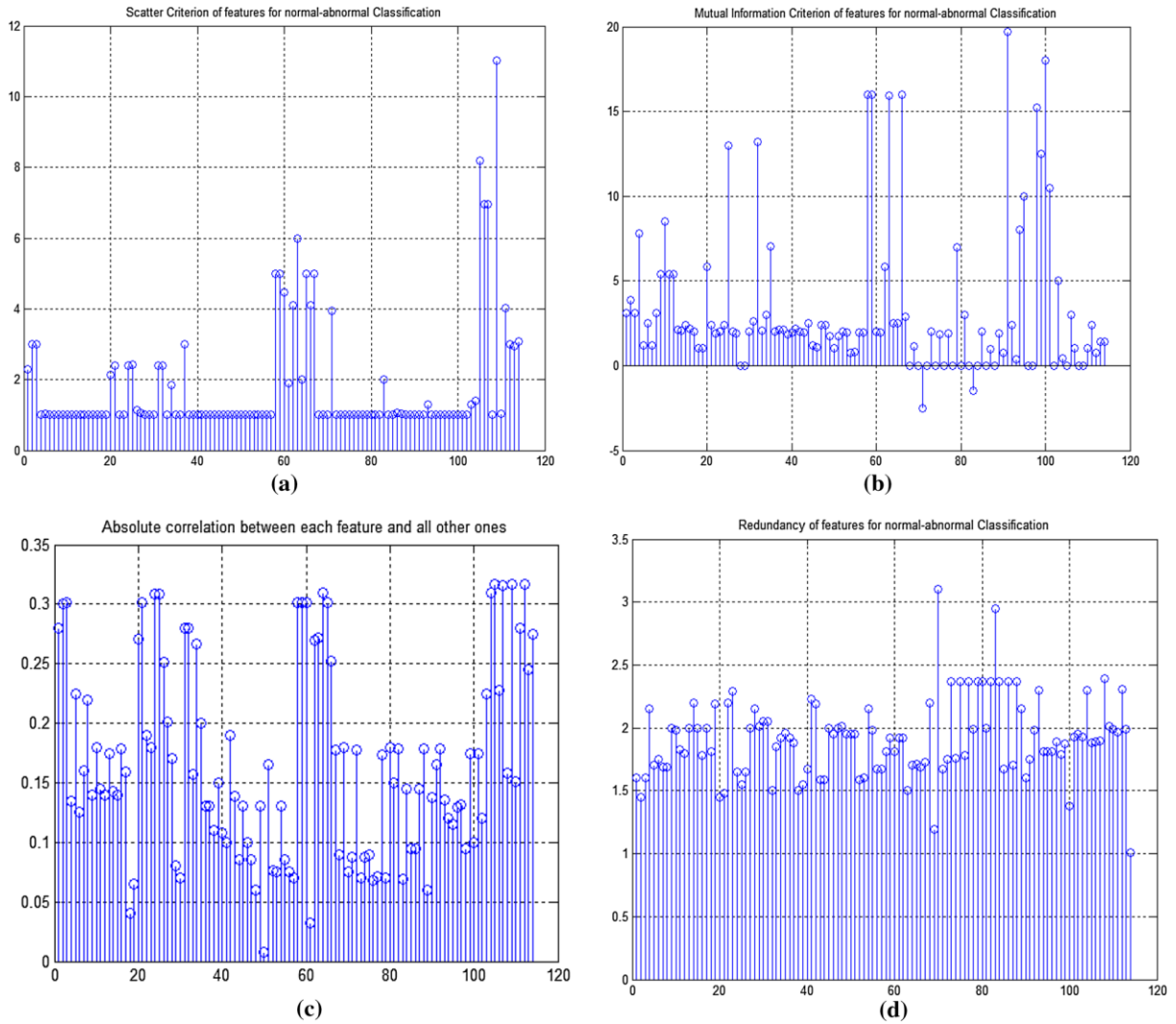


Figure 5.1. Évaluation des caractéristiques pour la classification des masses normales-anormales : (a) critère de dispersion, (b) critère d'information mutuelle, (c) corrélation absolue entre chaque caractéristique et toutes les autres, (d) redondance des caractéristiques

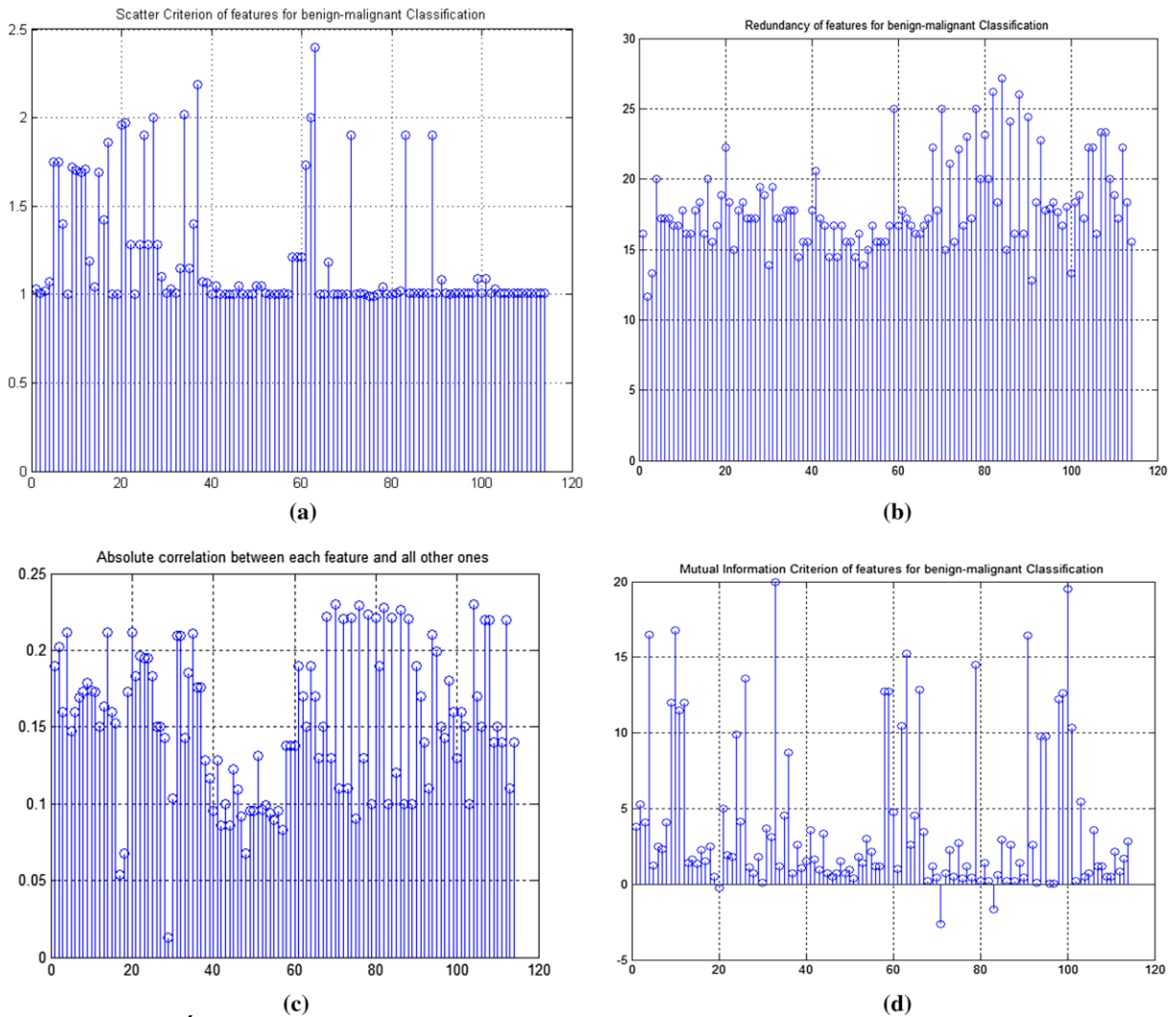


Figure 5.2. Évaluation des caractéristiques pour la classification des masses bénignes-malignes : (a) critère de dispersion, (b) critère d'information mutuelle, (c) corrélation absolue entre chaque caractéristique et toutes les autres, (d) redondance des caractéristiques

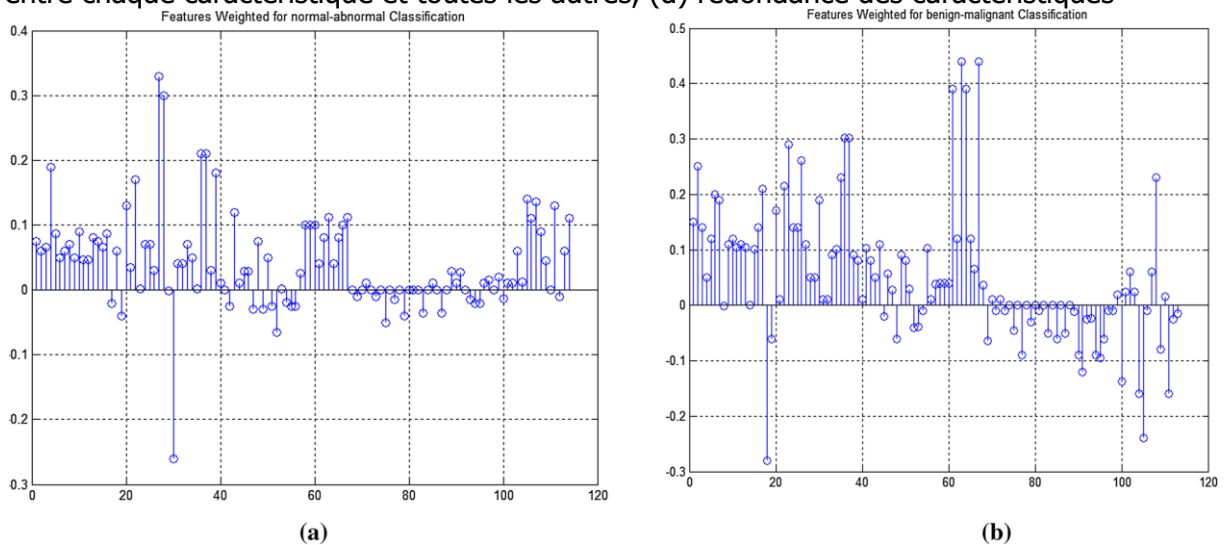


Figure 5.3. Caractéristiques pondérées pour une classification : (a) des masses normales-anormales et (b) des masses bénignes-malignes.

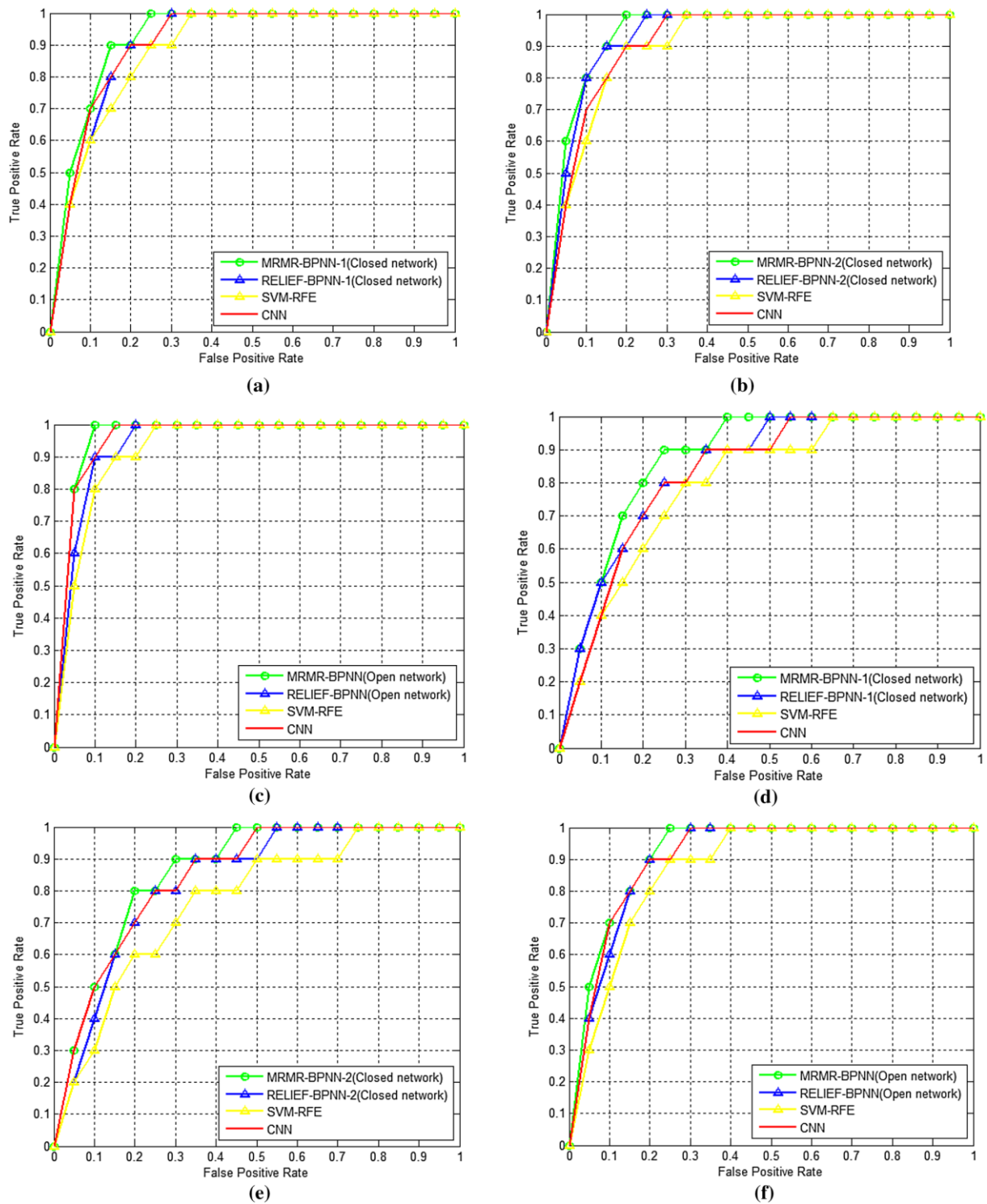


Figure 5.4. Comparaison des courbes ROC entre le schéma de classification des mammographies utilisant la caractéristique sélectionnée optimale et les méthodes de classification SVM-RFE et CNN. (a) Avec le premier réseau fermé pour la classification ROI normale-anormale. (b) Avec deuxième réseau fermé pour la classification ROI normal-anormal. (c) Avec réseau ouvert unifié pour la classification ROI normal-anormal. (d) Avec le premier réseau fermé pour la classification bénigne-maligne de la ROI. (e) Avec deuxième réseau fermé pour la classification bénigne-maligne de la ROI. (f) Avec réseau ouvert unifié pour la classification bénigne-maligne des ROI

5.3.1. La courbe ROC

La courbe ROC est un outil d'évaluation et de comparaison des modèles prédictifs. L'idée de la courbe ROC est de faire varier le "seuil" de 0 à 1 et, pour chaque seuil, la courbe ROC est construite en traçant la proportion de "vrais positifs" en termes de "faux positifs". Plus la courbe s'incurve vers le haut, plus le test est pertinent (le rapport "vrais positifs" sur "faux positifs" augmente). La pertinence est mesurée par l'aire sous la courbe (AUC) ; elle augmente avec sa courbure. La matrice de classification $M(i, j)$ pour une règle de classification à deux classes est une matrice 2×2 dont l'élément générique donne le nombre d'exemples de l'ensemble de test de la classe i qui ont été classés dans la classe j . La Figure 5.5 montre la matrice de classification pour cette situation, basée sur une règle de décision donnée, avec de vraies classes le long des lignes et des classifications décidées (prédites) le long des colonnes. Appelons la situation normale (N) et la situation anormale (A)

		Decision	
		A	N
A	a	b	
	N	c	d

Figure 5.5. La matrice de classification canonique pour la discrimination à deux classes de l'événement anormal (A) de l'événement normal (N).

A partir de cette matrice de classification, les paramètres suivants sont définis :

- Taux des vrais positif (également connu sous le nom de sensibilité) \equiv $TPR = a/(a+b)$. (5.1)

- Taux des vrais négatif (également connu sous le nom de spécificité) \equiv $TNR = d/(c+d)$. (5.2)

- Taux de faux positifs \equiv $FPR = c/(c+d) = 1 - \text{spécificité}$. (5.3)

- Taux de faux négatifs \equiv $FNR = a/(a+b) = 1 - \text{sensibilité}$. (5.4)

- La précision (ACC) pour estimer la performance du processus de classification est :

$$ACC = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% \quad (5.5)$$

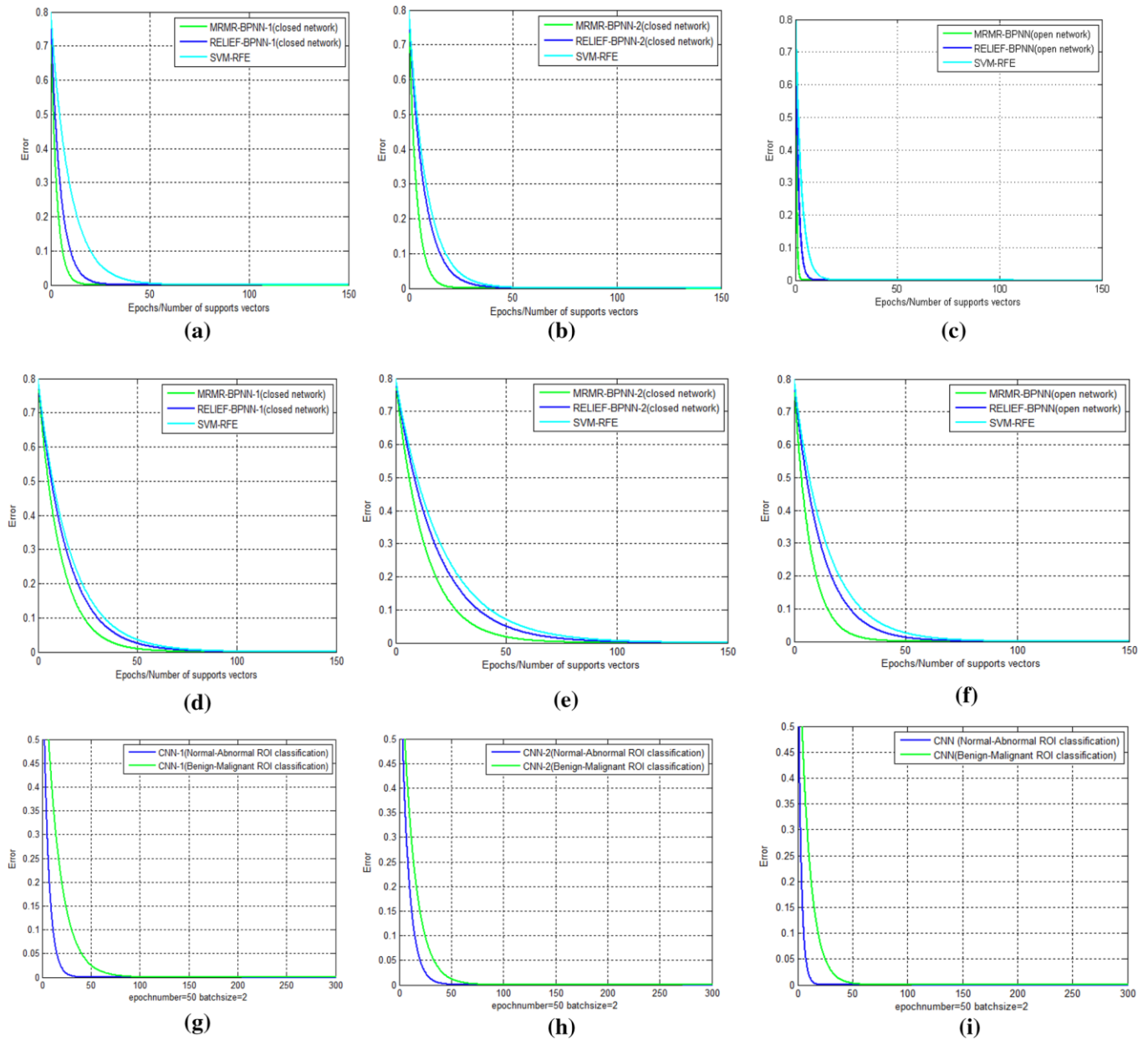


Figure 5.6. Comparaison de l'erreur d'apprentissage par le réseau neuro-files d'attente proposé en utilisant les algorithmes de sélection de caractéristiques MRMR ou RELIEF et les méthodes de classification SVM-RFE ou CNN. **(a)** Avec le premier réseau fermé pour la classification des ROI normal-anormale. **(b)** Avec le deuxième réseau fermé pour la classification des ROI normal-anormale. **(c)** Avec le réseau ouvert unifié pour la classification des ROI normal-anormale. **(d)** Avec le premier réseau fermé pour la classification des ROI bénin-malin. **(e)** Avec le deuxième réseau fermé pour la classification des ROI b bénin-malin. **(f)** avec le réseau ouvert unifié pour la classification des ROI bénin-malin. **(g)** CNN-1 pour la classification des ROI (en utilisant la demie base de données MIAS). **(h)** CNN-2 pour la classification des ROI (en utilisant l'autre demie base de données MIAS). **(i)** CNN pour la classification des ROI (en utilisant toute la base de données MIAS)

Tableau 5.3 Comparaison des taux de précision de test optimaux et des mesures AUC entre les méthodes proposées et SVM-RFE et CNN en utilisant le premier réseau fermé

Mammogram image class	Test accuracy rate (%)				AUC measurements			
	MRMR-BPNN-1 (closed network)	RELIEF-BPNN-1 (closed network)	SVM-RFE method	CNN method	MRMR-BPNN-1 (closed network)	RELIEF-BPNN-1 (closed network)	SVM-RFE method	CNN method
Normal–abnormal	95.3	94.2	92.8	94.9	0.9250	0.9050	0.9350	0.9075
Benign–malignant	92.1	88.7	81.6	80.9	0.8750	0.8250	0.8000	0.7575

Tableau 5.4 Comparaison des taux de précision de test optimaux et des mesures de l'AUC entre les méthodes proposées et SVM-RFE et CNN en utilisant un deuxième réseau fermé

Mammogram image class	Test accuracy rate (%)				AUC measurements			
	MRMR-BPNN-2 (closed network)	RELIEF-BPNN-2 (closed network)	SVM-RFE method	CNN method	MRMR-BPNN-2 (closed network)	RELIEF-BPNN-2 (closed network)	SVM-RFE method	CNN method
Normal–abnormal	94.1	93.8	92.0	91.2	0.8950	0.8850	0.8600	0.8375
Benign–malignant	93.5	88.1	80.1	83.7	0.8475	0.8150	0.7650	0.7975

Tableau 5.5 Comparaison des taux de précision de test optimaux et des mesures AUC entre les méthodes proposées et SVM-RFE et CNN en utilisant un réseau ouvert unifié

Mammogram image class	Test accuracy rate (%)				AUC measurements			
	MRMR-BPNN (opened network)	RELIEF-BPNN (opened network)	SVM-RFE method	CNN method	MRMR-BPNN (opened network)	RELIEF-BPNN (opened network)	SVM-RFE method	CNN method
Normal–abnormal	98.3	95.8	94.3	95.9	0.9650	0.9150	0.9070	0.9250
Benign–malignant	93.9	88.8	82.4	80.3	0.9200	0.8950	0.8725	0.8325

5.3.2. Résultats pour l'extraction de caractéristiques et le filtre de diffusion anisotrope

Dans ce travail, le filtre de diffusion anisotrope sur les images mammographiques a été utilisé pour réduire le bruit dans la zone et renforcer les bords d'une masse. L'application du filtre de diffusion anisotrope sur les images mammographiques est illustrée à la Figure 5.7 (la deuxième rangée). Au cours du processus de diffusion, afin de préserver la limite de masse, plusieurs itérations ont été effectuées pour définir la valeur de la constante de filtre K à 20. De plus, le GLCM et le NGLCM et le GLRLM correspondants et le NGLRLM correspondant sont calculés à partir de chaque mammographie filtrée. Pour améliorer les performances, plusieurs modifications de la matrice de cooccurrence et de la matrice de longueur de plage avec le choix du vecteur de déplacement (d) et des directions de pixel (θ) sont effectuées. Dans cette étude, les meilleurs paramètres du vecteur de déplacement et des directions de pixel sont définis avec $d = 1$ et $\theta = 90^\circ$ respectivement pour le calcul du GLCM, et $\theta = 0^\circ$ comme direction de balayage pour le calcul des matrices GLRLM.

- L'ensemble de $M=114$ caractéristiques texturales et morphologiques est extrait du nombre $N=272$ des régions d'intérêts(ROI). Le nombre M de caractéristiques est conservé dans des colonnes avec le nombre N correspondant de ROI dans des lignes pour générer une matrice de caractéristiques, qui doit être utilisée dans l'algorithme de sélection de caractéristiques. Les images obtenues sont classées en images normales/anormales et en images bénignes/malignes.
- Pour chaque frontière fermée (Figure 5.7, troisième ligne), les coefficients de Fourier du premier ordre (a_1, b_1, c_1, d_1) jusqu'à l'ordre 5 (a_5, b_5, c_5, d_5) sont extraits. Les valeurs des coefficients de Fourier qui correspondent à la plage (a_n, b_n, c_n, d_n) sont respectivement (0.26, 0.6, 0.695, 0.78), entre (-32535, -9.2072, -76.9844, -4.9934) et (9.5215, -0.9674, 11.6117, 62.1813), et entre (-23,8711, -15,0144, -66,4032, 0,0904) et (7,1004, 7,5912, 4,3160, 16,2538) pour les catégories de ROI normal, bénin et malin respectivement.
- Les moments de Zernike (ZM) ont également été calculés après le mappage des images de masse sur le disque unitaire (voir la Figure 5.7, dernière rangée). Les pixels tombant à l'extérieur du cercle unitaire ne sont pas pris en compte. Les ordres 4 à 14 sont extraits de chaque mammographie filtrée. Les magnitudes sont respectivement comprises entre (0,236 et 1,365), (1,742 et 1,965), (1,852 et 2,495) pour la catégorie ROI normal, bénin et malin, tandis que les angles sont respectivement (1,57), entre (0,236 et 1,365), et entre (0,893 et 2,536) pour la catégorie ROI normale, bénigne et maligne.

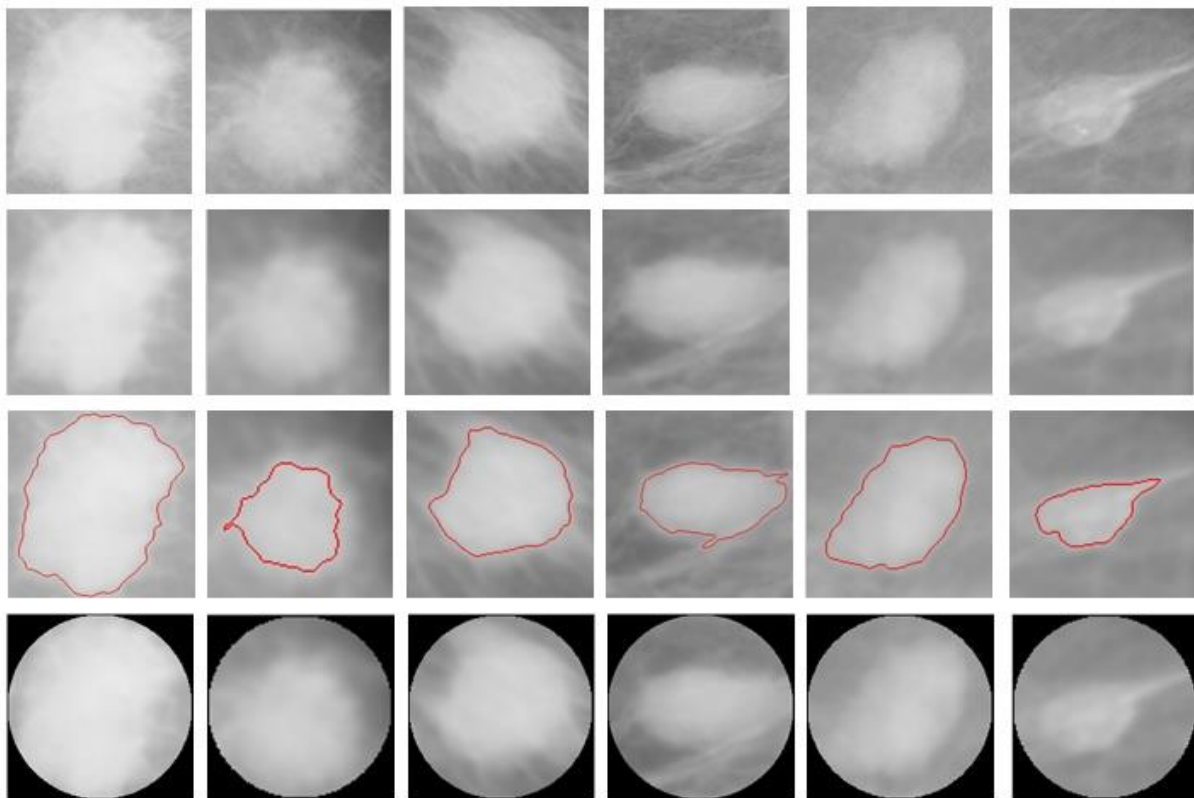


Figure 5.7. Exemples d'images de la base de données MiniMIAS utilisées pour l'évaluation expérimentale. La première rangée contient la masse originale initiale pour la mammographie (de gauche à droite, mdb184, mdb202, mdb028, mdb025, mdb015, mdb132), tandis que les rangées suivantes correspondent respectivement aux images traitées par le filtre de diffusion anisotrope, le traçage du contour de masse et les approches du cercle de masse.

Tableau 5.6. Valeurs moyennes de certaines caractéristiques statistiques pour les formulaires des trois mammographies

No	Feature	Normal	Benign	Malignant
1	Area A (in pixels)	10362	802	19598
2	Width W (in pixels)	120	36	157
3	Height H (in pixels)	120	32	177
4	Ratio W/H	1	1.125	0.887
5	UR/A	0.25	0.2095	0.2791
6	LR/A	0.25	0.2668	0.1876
7	LL/A	0.25	0.318	0.2922
8	UL/A	0.25	0.2843	0.2546
9	U/A	0.5	0.4738	0.531
10	R/A	0.5	0.4589	0.4638
11	Lo/A	0.5	0.5636	0.4755
12	Lt/A	0.5	0.4738	0.531
13	\bar{x}_N	-0.975	-0.0064	0.1046
14	\bar{y}_N	-0.975	-0.0842	-0.1519
15	Elongation E	1.7326	1.2149	2.068
16	Orientation θ (in degrees)	90	34.3436	23.516
17	Roundness	0.3331	0.6775	0.2338

Tableau 5.7. Valeurs moyennes des caractéristiques limites pour les quatre formes de mammographie

No	Feature	Normal	Benign	Malignant
1	Perimeter Length	188.5	382.81	572.74
2	Perimeter to Diagonal Ratio	1.15	0.9077	1.2036
3	Compactness Ratio	1.0	1.0968	2.0292
4	Bending Energy	0.41	0.49	0.42

5.3.3. Résultats de la modélisation, de la sélection et de l'agrégation de caractéristiques à l'aide de la mise en file d'attente et du réseau de neurones

Nous avons choisi d'agréger successivement les caractéristiques dans une file d'attente et de contrôler l'effet de leur ajout par le taux d'erreur obtenu par le QBPN (Queuing Back Propagation Neural Network). Le principe de construction du vecteur caractéristique est le suivant :

Dans un premier temps, nous ajoutons au vecteur de caractéristiques une seule caractéristique, qui est testée dans l'apprentissage/test par le NN. On ajoute alors, à la meilleure combinaison de l'étape précédente, une des caractéristiques restantes (celle qui a le meilleur pouvoir discriminant). La meilleure combinaison est choisie pour l'étape suivante. Par la suite, pour chaque groupe de cinq meilleures caractéristiques agrégées de la base d'apprentissage (Voir Figure 5.1a), les redondantes sont éliminées afin de garantir l'unicité d'une caractéristique (Voir Figure 5.1d) en utilisant le critère d'information mutuelle (Voir Figure 5.1b) dans l'algorithme mRMR et le critère de corrélation linéaire (voir Figure 8c) dans l'algorithme RELIEF. Les caractéristiques sont sélectionnées après que leurs poids sont mis à jour après plusieurs itérations pour l'algorithme RELIEF (voir Figure 5.3). On calcule les performances du système sur la base de test et on relance l'apprentissage/test pour la meilleure combinaison obtenue précédemment.

Nous avons relancé le processus d'apprentissage/test plusieurs fois, et nous nous sommes limités à atteindre une solution sous-optimale dès que le taux de classification se détériore. Cependant, l'étape de construction du vecteur caractéristique est réalisée après application d'un filtre de diffusion anisotrope et réglage des paramètres du réseau de neurones. Les valeurs des paramètres de ce classificateur (fonction d'activation, 114 neurones dans la couche d'entrée, nombre de neurones de la couche cachée et un seul neurone en sortie) doivent être optimisées. Afin de déterminer le bon paramètre, le nombre de neurones de la couche cachée (n) des réseaux de neurones, nous avons réalisé plusieurs expériences avec plusieurs choix manuels des valeurs des neurones de la couche cachée en contrôlant leurs effets par le taux de classification retenu par le NN. Cette partie du choix des paramètres nous a permis de fixer certaines valeurs du paramètre ($n=25$) tout en préservant la distribution originale des données d'apprentissage. Ces valeurs seront utilisées pour sélectionner les caractéristiques en se rapprochant de la distribution d'origine de la classe.

- Les performances de notre système de détection assistée par ordinateur sont évaluées à l'aide de plusieurs critères, notamment l'analyse des caractéristiques de l'opérateur du récepteur (ROC : A Receiver Operator Characteristic), les valeurs d'aire sous la courbe (AUC : Area Under Curve), la sensibilité, la spécificité et la précision de la classification des tests à l'aide de deux méthodes de sélection des caractéristiques. Les trois derniers critères sont présentés dans les tableaux 5.1 et 5.2. Les résultats obtenus en termes de taux de précision illustrent que la méthode de sélection des caractéristiques MRMR surpasse RELIEF. Ces valeurs sont de 98,1 % (normal-anormal) et 95,2 % (bénigne-maligne).

- De plus, R représente la cardinalité du sous-ensemble de caractéristiques pertinentes sélectionnées soit par l'algorithme MRMR, soit par RELIEF. Les Tableaux 5.1 et 5.2 illustrent l'étape de réglage du paramètre R . À chaque fois que nous choisissons le nombre (R), les réseaux fermés et ouverts renvoient des mesures accrues de sensibilité et de spécificité. Cependant, ces mesures s'améliorent progressivement lorsqu'on ouvre l'un des réseaux fermés par l'effet des poids mis à jours de l'autre. Le meilleur sous-ensemble discriminant qui reflète les meilleures performances, est sélectionné par l'algorithme MRMR. La sensibilité et la

spécificité sont calculées par rapport à la normale vs. anormal et bénin vs. malin. Dans le premier cas, $R = 47$ caractéristiques pertinentes décrivent mieux les données d'entraînement et le taux de sensibilité de l'ouverture qui atteint respectivement 98,0 % et 97 %. Alors que, pour $R = 77$ caractéristiques, ces mesures atteintes sont à 100 % et 90,2 % pour le deuxième cas.

- Nous avons également comparé les performances obtenues par le classificateur QBPNN du point de vue de la fermeture ou de l'ouverture du réseau de file d'attente en utilisant soit l'algorithme de sélection de caractéristiques MRMR ou RELIEF et la méthode SVM-RFE [211] et le paradigme d'apprentissage profond CNN[298]. Les courbes ROC obtenues à l'aide du schéma proposé et des méthodes SVM-RFE et CNN sont présentées à la Figure 5.4. On en déduit que le modèle unifié ouvert QBPNN surpasse les méthodes SVM-RFE et CNN. Les Tableaux 5.3, 5.4 et 5.5 présentent la comparaison des taux de précision des tests et des mesures AUC pour le premier classificateur QBPNN fermé, le deuxième classificateur QBPNN fermé, le classificateur QBPNN ouvert unifié et les méthodes SVM-RFE et CNN. Les valeurs maximales de l'AUC obtenues par le QBPNN avec la méthode de sélection des caractéristiques MRMR sont respectivement de 0,9650 et 0,9200 pour les classifications de mammographie normale-anormale et bénigne-maligne. Dans tous les cas, il est clairement observé que le modèle ouvert unifié de QBPNN utilisant la méthode de sélection des caractéristiques MRMR a des valeurs de performance plus élevées que les autres schémas mentionnés. Généralement, on observe également que les performances s'améliorent progressivement lorsque l'on ouvre le réseau de neurones fermé en sélectionnant et agrégeant les états (les matrices de poids) d'un autre réseau de neurones fermé.

- De plus, nous avons comparé les performances d'apprentissage d'un point de vue de l'erreur de généralisation retenue par le classificateur QBPNN. La fonction de coût utilisée est l'erreur quadratique moyenne (MSE). Le problème est d'optimiser les huit en minimisant cette fonction de coût (Eq. 4.18-4.21 du quatrième chapitre). Généralement, les courbes d'erreur d'apprentissage convergent plus rapidement lorsque nous ouvrons le réseau neuronal fermé. Cependant, la méthode de sélection des caractéristiques MRMR atteint des performances supérieures par rapport aux autres méthodes. La figure 5.6 montre la convergence de l'erreur d'apprentissage en comparant le schéma proposé et les méthodes SVM-RFE et CNN.

- Si des caractéristiques très pertinentes (sélectionnées par les méthodes de sélection des caractéristiques MRMR ou RELIEF) sont supprimées après le processus de sélection, les performances du système de classification diminuent complètement pour la catégorie normale-anormale en raison de l'éloignement des classes de données d'entraînement, ce sont des caractéristiques qui récupèrent une grande partie des performances initiales en conservant la distribution originale des données d'entraînement dans leurs classes, contrairement aux performances bénignes-malignes à chaque élimination de ces caractéristiques. Enfin, le Tableau 5.8 présente une évaluation comparative du schéma proposé par rapport aux autres schémas existants.

Tableau 5.8. Comparaison des performances de différentes méthodes avec le schéma proposé

Méthode référence	Database	Mesure de performance declassification			
		TPR	TNR	AUC	ACC
Liu et al. (2001) [223]	–	–	–	–	84.2%
Rashed et al. (2007)[224]	MIAS	B/M	Les résultats obtenus ont montré que la classification en utilisant l' ondelettes Daubechies-8 est plus performante que les autres		
Yoon and Kim (2009)[225]	DDSM	–	La précision Az obtenue de mRMR avec la corrélation absolue de Pearson est relativement plus performante que les autres schémas par l'utilisation des SVM avec RBF		
Lladó et al. (2009) [226]	DDSM	–	–	–	–
Braz et al. (2009) [227]	DDSM	N/A	–	–	0.946 (Geary'scoefficient) 96.04% (Geary's coefficient)
Tahmasbi et al. (2011) [228]	MIAS	–	–	–	0.976 (MLP – BP) 0.975 (MLP – OBL)
Ramirez-Villegas and Ramirez-Moreno (2012)[229]	MIAS	–	–	–	93.75%
Wen-Jie et al. (2012) [230]	–	–	–	–	95.24% (with SVM-RBF)
Kanwal and Hussain (2015) [231]	MIAS IN breast	–	–	–	1.0000 100% (SVM + linear kernel)
Vaidehi and Subashini (2015) [232]	MIAS	–	90.62%	99.99%	93.75%
Zyout et al. (2015) [233]	MIAS DDSM	–	–	–	98.88%
de Oliveira et al. (2015) [234]	–	–	–	–	98.88%
Subashini et al. (2009) [235]	Wisconsin DataBase	–	–	–	92.1316% (SVM)
Beura et al. (2015) [236]	MIAS	N/A	100%	97.0%	0.9899 98.0%
Shi et al. (2016) [237]	ultrasound dataset	B/M	100%	90.0%	0.9504 94.2%
Kong et al. (2016) [238]	BCW WPBC BC	–	–	–	92.40 ± 1.1%
Xie et al. (2016) [239]	DDSM MIAS	–	–	–	JSDA est plus performant que les autres méthodes d'apprentissage sur des données de cancer 96.02%
Khan et al. (2017) [240]	DDSM	–	–	–	98.48%
Jadoon et al. (2017) [242]	IRMA	–	–	–	81.83% (CNN-DW) 83.74% (CNN-CT)
Soriano et al. (2017) [241]	DDSM	–	–	–	85%
Proposed scheme [41]	MIAS	N/A	98%	97.0%	0.9650 98.1%
		B/M	100%	92.2	% 0.9200 95.2%

N/A Normal/abnormal, **B/M** benign/malignant

5.4. Conclusion et recherches futures

Dans ce travail, nous proposons un cadre de classification efficace des mammographies pour appuyer la décision des radiologues. Ce cadre construit un arbre de décision pour une classification préliminaire du ROI (région d'intérêt) à partir des particularités des tissus (fond, classe et anomalie). Les différentes images générées par cet arbre de décision sont ensuite filtrées à l'aide d'une technique de diffusion anisotrope. La matrice de caractéristiques à partir de telles images est générée à l'aide de caractéristiques de texture provenant de la matrice de cooccurrence de niveaux de gris (GLCM) et de la matrice de longueur de plage de niveaux de gris (GLRLM) ; et caractéristiques morphologiques à partir du tracé du contour des différentes masses.

Afin de sélectionner les caractéristiques pertinentes à partir de la matrice de caractéristiques, les algorithmes MRMR et RELIEF sont utilisés pour la sélection des caractéristiques et sont comparés à la méthode SVM-RFE. L'approche proposée (modélisation du réseau de mise en file d'attente) fait partie du cadre d'évaluation des performances basé sur un modèle, c'est-à-dire que toute méthode d'apprentissage automatique en général peut être entrée dans cette proposition. Nous avons utilisé comme exemple les réseaux de neurones. On peut superposer et unifier soit par plusieurs arbres de décision C4.5, soit par plusieurs modèles SVM, soit par l'unification de modèles générés par KNN, etc.

C'est vrai que les méthodes d'apprentissage automatique sont déjà modélisées et s'inspirent soit de la biologie comme les réseaux de neurones, les algorithmes génétiques, soit du système visuel (CNN) soit de la logique et des mathématiques comme les SVM,... etc. fusionner et unifier ces modèles de même type, taille et architecture. Pour atteindre cet objectif, le système doit manipuler ces différents modèles pour qu'il mémorise (conserve), qu'il supprime, qu'il récupère ; et il change leurs états de leurs variations internes au cours de l'étape d'apprentissage. Les réseaux de file d'attente considèrent tout modèle d'apprentissage comme un réseau et peuvent y entrer ou en sortir par leurs nœuds (réseau ouvert ou réseau fermé). Cet aspect est exprimé en terme d'absorption : C'est l'ouverture de l'un des réseaux fermés pour absorber les états de tous les autres réseaux fermés, c'est-à-dire que le réseau ouvert met à jour ses matrices de poids intermédiaires par l'effet des autres réseaux de neurones fermés. La fonction de coût du réseau de neurones QBPNN unifié ouvert est représentée par le produit des deux fonctions de coût des réseaux QBPNN fermés. Le produit de deux fonctions décroissantes est une fonction plus décroissante dans l'intervalle]0, 1[. Lorsque la matrice de poids du réseau ouvert unifié se déplace avec une estimation de ΔW , l'erreur totale E est minimisée par l'effet de déplacement ΔW_1 du premier réseau fermé fusionné et le déplacement ΔW_2 du second.

Pour valider l'efficacité du schéma proposé, la simulation a été réalisée à l'aide de la base de données MIAS. Une analyse comparative en ce qui concerne la précision et l'AUC du ROC révèle que le schéma suggéré surpasse relativement d'autres schémas. Une précision de 98,1 % et 95,2 % a été obtenue respectivement pour le normal-anormal et le bénin-malin.

De plus, la connaissance a priori du radiologue est représentée par un arbre de décision. Cette dernière nous permet une analyse préliminaire du problème et conduit au choix de la mise en œuvre d'une méthode de classification statistique. Dans le cadre de la classification statistique, les formules de Bayes, la fonction de densité de probabilité (pdf), l'estimation de densité par noyau (KDE) et l'estimation de densité récursive (RDE)[299,300] permettent le calcul de la probabilité que le nouvel objet décrit par le vecteur descripteur appartienne à la classe correspondante. Nous pouvons utiliser ces classificateurs pour estimer la probabilité a posteriori d'une nouvelle masse de ROI décrite par x à l'aide de la formule de Bayes, suivie

d'une prise de décision selon la règle de décision de Bayes : affecter la masse de ROI à une catégorie dont la probabilité a posteriori est la plus grande.

Une extension immédiate de ce travail, que nous sommes en train de finaliser, consiste à appliquer la méthodologie proposée à d'autres schémas d'apprentissage automatique/sélection de caractéristiques et à utiliser d'autres opérateurs de fusion de données plus précis [301]. Les performances peuvent être comparées grâce à ce processus de modélisation. Il est également intéressant d'injecter d'autres mammographies ayant les mêmes caractéristiques de base de données grâce à des algorithmes d'apprentissage incrémental pour atteindre des performances plus élevées dans le domaine de la classification des masses mammaires. De plus, nous prévoyons d'implanter les techniques de sélection de caractéristiques dans les réseaux de neurones convolutifs (CNN) pour connaître le chemin des caractéristiques qui attribuent les exemples d'apprentissage à leurs classes correspondantes ; et modéliser ce nouveau paradigme par notre système de réseau de file d'attente, car nous pensons que le réseau de neurones profond pourrait nous fournir de meilleurs résultats dans ce domaine que les méthodes classiques d'apprentissage automatique[298].

Conclusion générale et perspectives

L'étude menée dans le cadre de cette thèse a exploré divers aspects de l'application de l'apprentissage automatique et de l'analyse d'images médicales, en mettant particulièrement l'accent sur l'amélioration des diagnostics et des traitements dans le domaine de la santé, notamment dans le contexte du dépistage et de la classification du cancer du sein. Cette thèse a présenté un cadre complet pour la classification des mammographies, combinant des techniques de pré-classification, de filtrage, d'extraction de caractéristiques, de sélection de caractéristiques et de modélisation de réseau de neurones par les systèmes de réseaux de files d'attente. Un réseau de neurones BPNN est modélisé à l'aide d'un réseau de files d'attente pour représenter les propriétés texturales et morphologiques des masses mammaires. Les caractéristiques pertinentes sont injectées dans le classificateur BPNN via le réseau de files d'attente. Les résultats de validation sur la base de données MIAS montrent que l'approche proposée améliore significativement la précision de la classification des mammographies, en particulier pour la distinction entre les tissus normaux et anormaux ainsi que entre les tissus bénins et malins. L'approche a été validée sur une base de données standard et a montré des performances prometteuses en termes de précision et d'aire sous la courbe ROC. Ces résultats suggèrent que cette méthode pourrait être un outil précieux pour aider les radiologues dans leur processus de diagnostic du cancer du sein.

À travers cinq chapitres distincts, nous avons examiné des techniques, des méthodes et des approches visant à optimiser la qualité des images médicales, à extraire des caractéristiques pertinentes et à développer des modèles de classification précis.

Dans le premier chapitre, nous avons introduit les concepts fondamentaux de l'apprentissage automatique et de l'analyse d'images médicales, en mettant en évidence l'importance de la sélection de caractéristiques pour améliorer la précision des modèles de classification. Nous avons souligné les défis spécifiques rencontrés dans le domaine de l'imagerie médicale, tels que la complexité des données et la nécessité d'identifier des biomarqueurs pertinents pour la détection des maladies.

Le deuxième chapitre s'est concentré sur les techniques de prétraitement et d'amélioration d'images, en mettant en évidence l'importance de ces étapes pour obtenir des images de haute qualité, facilitant ainsi l'interprétation clinique et améliorant la précision des diagnostics. Nous avons examiné différentes méthodes, telles que le filtrage anisotrope et l'extraction de caractéristiques texturales et morphologiques, pour améliorer la qualité des images médicales.

Dans le troisième chapitre, nous avons exploré les techniques de modélisation probabiliste pour évaluer la performance des systèmes de traitement de l'information, en mettant en évidence l'importance des réseaux de files d'attente pour comprendre et optimiser les performances des systèmes informatiques. Nous avons examiné les principaux indicateurs de performance utilisés pour évaluer les systèmes, tels que le débit, l'utilisation des ressources et la probabilité de perte.

Le quatrième chapitre a présenté une méthode de classification des images de mammographies en utilisant différentes techniques, telles que la construction d'arbres de

décision, l'extraction de caractéristiques texturales et morphologiques, et l'utilisation d'algorithmes de sélection de caractéristiques pour identifier les caractéristiques les plus pertinentes. Nous avons souligné l'importance de ces techniques pour améliorer la précision des diagnostics de cancer du sein.

Enfin, le cinquième chapitre a décrit un nouveau cadre de classification des mammographies, en mettant en évidence l'efficacité de l'utilisation de réseaux de files d'attente pour représenter les propriétés texturales et morphologiques des masses mammaires. Nous avons comparé l'efficacité des algorithmes de sélection de caractéristiques pour améliorer la précision de la classification des tissus mammaires.

Sur le plan de la recherche, cette thèse apporte une contribution significative à la littérature en proposant de nouvelles approches pour l'analyse d'images médicales et la classification des tissus mammaires. En explorant différentes techniques et en comparant leur efficacité, nous avons pu identifier les meilleures pratiques pour optimiser la qualité des images et améliorer la précision des diagnostics. Ces résultats peuvent servir de base pour de futures études visant à développer des méthodes encore plus avancées pour l'analyse d'images médicales. Malgré les avancées significatives réalisées dans cette thèse, il reste encore beaucoup à faire pour améliorer les diagnostics et les traitements des maladies.

- 1. Intégration de l'apprentissage profond (Deep Learning) :** Bien que nous ayons utilisé des techniques d'apprentissage automatique classiques dans cette thèse, l'intégration de méthodes d'apprentissage profond pourrait offrir des avantages supplémentaires en permettant une représentation hiérarchique des caractéristiques et une meilleure adaptation aux données complexes de l'imagerie médicale.
- 2. Utilisation de données multi-modales :** Au lieu de se concentrer uniquement sur les images de mammographie, il serait intéressant d'explorer l'intégration de données multi-modales, telles que l'imagerie par résonance magnétique (IRM) et l'échographie, pour améliorer la précision des diagnostics en fournissant des informations complémentaires sur la structure et la composition des tissus.
- 3. Développement de modèles interprétables :** Alors que les modèles de classification basés sur l'apprentissage automatique peuvent être très performants, leur opacité peut être un obstacle à leur adoption clinique. Le développement de modèles interprétables, capables d'expliquer leurs décisions de manière compréhensible pour les cliniciens, pourrait favoriser leur acceptation et leur utilisation dans la pratique médicale.
- 4. Exploration de nouvelles applications :** En dehors du domaine spécifique du cancer du sein, les techniques et les méthodes développées dans cette thèse pourraient être étendues à d'autres domaines de la santé, tels que la détection précoce d'autres types de cancer, la surveillance des maladies chroniques et le suivi de la progression des maladies.
- 5. Intégration de données longitudinales :** Plutôt que de se limiter à des données statiques à un moment donné, l'intégration de données longitudinales sur la progression des maladies pourrait fournir des informations précieuses pour la prédiction du risque, la détection précoce et la personnalisation des traitements.
- 6. Synergie entre la Sélection Complètement/partiellement linéaire de Caractéristiques et l'Équilibre Local : Vers une Modélisation Avancée des Systèmes de Files d'Attente**

L'intégration de la sélection complètement/partiellement linéaire des caractéristiques et de l'équilibre local dans les systèmes de réseaux de files d'attente pourrait ouvrir de nouvelles perspectives pour l'optimisation et la gestion efficace des systèmes complexes. En combinant ces deux approches, nous pourrions développer des méthodes plus robustes et plus efficaces pour la modélisation et l'analyse des systèmes dynamiques, en particulier dans des domaines tels que la gestion des réseaux de télécommunications, des systèmes de production et des services informatiques.

L'idée de fusionner ces deux approches réside dans leur capacité à simplifier la modélisation et l'analyse des systèmes complexes tout en préservant leur précision et leur pertinence. La sélection complètement/partiellement linéaire des caractéristiques permettrait de réduire la dimension de l'espace des caractéristiques en identifiant les variables les plus pertinentes pour la modélisation du système de files d'attente. En sélectionnant les caractéristiques les plus importantes de manière linéaire, cette approche pourrait permettre une réduction significative de la complexité du modèle, tout en préservant les informations essentielles pour la prise de décision.

D'autre part, l'intégration de l'équilibre local dans la modélisation des réseaux de files d'attente offrirait une approche alternative pour la résolution des équations d'équilibre globales. En décomposant les équations d'équilibre globales en équations d'équilibre local, cette approche permettrait de simplifier considérablement l'analyse des systèmes complexes en évitant la génération et la résolution des équations d'équilibre globales coûteuses en calcul. De plus, en utilisant des algorithmes de résolution efficaces basés sur l'équilibre local, nous pourrions obtenir des solutions précises et rapides pour une large classe de réseaux de files d'attente.

En combinant la sélection complètement/partiellement linéaire des caractéristiques avec l'équilibre local, nous pourrions bénéficier d'une approche holistique pour la modélisation et l'analyse des systèmes de réseaux de files d'attente, en tirant parti des avantages de chaque méthode tout en atténuant leurs limitations respectives. Cette approche intégrée pourrait offrir une meilleure compréhension des comportements dynamiques des systèmes, une optimisation plus efficace des performances et une prise de décision plus éclairée dans des environnements complexes et en évolution constante. En fin de compte, cela pourrait conduire à des systèmes plus robustes, plus fiables et plus adaptatifs, capables de répondre aux exigences croissantes de diverses applications industrielles et commerciales.

ANNEXE 1

A.1. Approches d'apprentissage automatique pour la détection et la localisation des tissus/maladies anatomiques

La segmentation multi-organes dans les scanners abdominaux et corps entier est difficile car il existe différents organes et structures qui doivent être capturés simultanément.

La taille, la forme et l'apparence des organes abdominaux varient considérablement d'un patient à l'autre, mais les positions relatives changent également dans une certaine mesure. Les petits organes sont moins souvent étudiés que les grands organes, bien que les petits organes présentent un intérêt pour le diagnostic et les applications cliniques telles que le dépistage du cancer.

La localisation automatique de l'anatomie est utile pour une navigation visuelle efficace, pour initialiser un traitement spécifique à un organe (par exemple, la détection de tumeurs du foie) et un marquage sémantique des balayages de patients pour faciliter leur tri et leur récupération.

- Dans [215], les auteurs utilisent des forêts de régression dans le domaine de l'imagerie médicale et propose un nouvel algorithme parallèle pour la détection et la localisation efficaces des structures anatomiques («organes») dans les études de tomographie assistée par ordinateur (CT).

La contribution principale est une nouvelle paramétrisation de la tâche de localisation de l'anatomie en tant que problème d'estimation de paramètres multiples et continus. Ceci est traité efficacement via une régression non linéaire basée sur une arborescence.

La base de données étiquetée. Les structures anatomiques que nous souhaitons reconnaître sont $C = \{\text{cœur, foie, rate, poumon gauche, poumon droit, l. rein, r. rein, vésicule biliaire, l. bassin, r. pelvis}\}$. La base de données contient 100 scans qui ont été annotés manuellement avec des cadres de délimitation 3D étroitement dessinés autour des structures d'intérêt. Le cadre de sélection de l'organe $c \in C$ est paramétré comme un vecteur $6 \text{ bc} = (bLc, bRc, bAc, bPc, bHc, bFc)$ où chaque composant représente la position (en mm) de la paroi correspondante alignée sur l'axe.

La clé de l'algorithme réside dans le fait que tous les voxels d'un volume de test CT contribuent avec une confiance variable à l'estimation de la position des six parois des boîtes de sélection de toutes les structures. Intuitivement, certains groupes de voxels distincts (par exemple, des côtes ou des vertèbres) peuvent prédire la position d'un organe (par exemple le cœur).

Apprentissage de forêt. Le processus d'apprentissage construit chaque arbre de régression et décide au niveau de chaque nœud de la meilleure manière de fractionner les voxels entrants. Le sous-ensemble de tous les volumes de CT étiquetés (le jeu d'entraînement) est donné, ainsi que les boîtes de liaison d'organes de vérité au sol associées. La taille de la forêt T est fixe et tous les arbres sont dressés en parallèle. Chaque voxel est poussé à travers chacun des arbres en partant de la racine.

l'optimisation des nœuds consiste à maximiser une mesure du gain d'informations, définie comme suit:

$IG = H(S) - \sum_{i \in \{L,R\}} w_i H(S_i)$ où H désigne une entropie, S est l'ensemble des points d'apprentissage atteignant le nœud et L, R représente les enfants gauche et droit. Dans la classification, l'entropie est définie sur les distributions d'étiquettes de classe discrètes. Dans

la régression, nous mesurons plutôt la pureté de la densité de probabilité des prédictions à valeur réelle.

Les erreurs résultantes (calculées sur le même ensemble de test) indiquent qu'une moyenne d'erreur et un écart beaucoup plus grands (environ le double) que l'approche proposée. La performance en fonction de la moyenne et l'écart type sont respectivement (mean= 16.79 mm, std= 17.3 mm) et (mean= 31.30 mm, std= 50.01 mm) de l'approche proposée et celle basée sur le recalage. (Voir Tableau A.1.1 et A.1.2)

Table A.1.1. Erreurs de localisation de la forêt de régression(en mm).

<i>organ</i>	heart	liver	spleen	left lung	right lung	left kidney	right kidney	gall bladder	left pelvis	right pelvis	<i>across all organs</i>
mean	15.4	17.1	20.7	17.0	15.6	17.3	18.5	18.5	13.2	12.8	16.7
std	15.5	16.5	22.8	17.2	16.3	16.5	18.0	13.2	14.0	13.9	17.0
median	9.3	13.2	12.9	11.3	10.6	12.8	12.3	14.8	8.8	8.4	11.5

Table A.1.2. Erreurs de localisation basée sur l'Atlas(en mm).

<i>organ</i>	heart	liver	spleen	left lung	right lung	left kidney	right kidney	gall bladder	left pelvis	right pelvis	<i>across all organs</i>
mean	24.4	34.2	36.0	27.8	27.0	39.1	28.3	27.6	23.4	22.4	31.3
std	27.0	59.3	57.2	29.9	27.6	55.6	53.3	26.7	43.3	43.5	50.0
median	15.5	16.4	20.1	15.7	18.0	25.7	15.4	19.8	10.9	11.8	17.2

La précision en fonction des paramètres de la forêt montre l'effet de la profondeur des arbres et de la taille de la forêt sur la précision. Les arbres plus profonds que 7 niveaux mènent au sur-ajustement. Cela n'est pas surprenant puisque le surentraînement avec de grands arbres a été rapporté dans la littérature. En outre, comme prévu, l'augmentation de la taille de la forêt T produit une amélioration monotone sans surajustement.

Par rapport à la méthode de classification décrite dans [214], les auteurs ont constaté que les techniques de régression généraient des erreurs inférieures à la moitié de celles rapportées dans [214] (sur des ensembles d'apprentissage et du test identiques), ce qui a démontré une meilleure précision que GMM et les approches basées sur les modèles (template-based approaches).

- Dans [216], les auteurs se concentrent sur le problème de la segmentation 3D dans l'imagerie de tomodesitométrie volumétrique pour identifier les organes de l'abdomen. Ils ont évalué également l'utilité des caractéristiques basées sur des histogrammes de gradients orientés ou des caractéristiques HOG, sur un ensemble plus large de structures anatomiques trouvées dans une base de données sur le foie

Ils ont utilisé 22 volumes 3D de patients pour réaliser les expériences. 19 de ces volumes sont tirés de l'ensemble de données sur le foie pour lesquels des résultats hépatiques segmentés ont été fournis. L'objectif est de segmenter cinq organes du foie : le foie, deux reins, la rate et la vésicule biliaire.

Une plage de valeurs d'intensité plus élevée (-180 à 1200HU) a été utilisée afin de pouvoir s'adapter à toute amélioration du contraste ou à la vésicule biliaire ou aux calculs rénaux. Les valeurs d'intensité ont été quantifiées dans des bacs et modélisées sous forme d'histogramme ou de distribution discrète. Le gradient entre deux pixels voisins est également modélisé sous la forme d'un histogramme de valeurs de gradient. Le gradient est la différence des intensités de pixel des pixels voisins. La localisation spatiale des organes a été modélisée à l'aide d'un mélange de gaussiennes et le choix du nombre de gaussiennes dans le mélange est déterminé par la vraisemblance de l'ensemble de validation croisée, les résultats obtenus ont tendance à dépendre de la taille et de la compacité de l'organe. Les caractéristiques d'apparence sont modélisées sous la forme d'un patch 5x5 autour de chaque pixel et le regroupement est utilisé. Les caractéristiques HOG sont également regroupées pour créer des livres de codes (code

books). Pour calculer les caractéristiques HOG, ils ont utilisé des blocs non superposés où chaque bloc est composé de 3x3 cellules et chaque cellule de 8x8 pixels a 10 cases d'orientation.

Deux modèles probabilistes différents ont été examinés pour la segmentation d'images. Les modèles sont soit des CRF (Conditional Random Fields) entièrement discriminatifs, soit contiennent à la fois des composants génératifs et discriminatifs (CRF-MRF).

Nous examinons deux modèles probabilistes différents pour la segmentation d'images. Les modèles sont soit des CRF entièrement discriminatifs, soit contiennent à la fois des composants génératifs et discriminatifs (CRF-MRF).

L'apprentissage par maximum de vraisemblance (ML) ou ML est effectué en utilisant une descente de gradient standard.

Les expériences montrent une augmentation de la précision des pixels et des classes lorsque HOG est utilisé dans la plupart des cas. L'expérience de configuration interactive montre l'utilisation d'un modèle structuré discriminant comme le CRF faisant mieux que la régression logistique plus simple. Les caractéristiques d'apparence n'ont pas tendance à bien fonctionner et il est possible que davantage de groupes d'apparences ou de mots de code soient nécessaires.

L'utilisation de la régression logistique modifiée a amélioré la précision moyenne des classes sans les classes de la vésicule biliaire et les reins qui ont des taux d'erreur très importants. Il est noté également que les CRF améliorent la précision des pixels dans une certaine mesure tandis que la précision de la classe reste presque la même.

La précision de la classe passe de 30 % (uniquement l'intensité) à 60 % (IH : intensité et HOG). La fonction de localisation dépend cependant davantage de l'étape de recalage manuel. Les performances améliorées lors de l'utilisation du modèle CRF discriminant par rapport au MRF-CRF, même avec des caractéristiques plus simples.

A.2. Approches d'apprentissage automatique pour la classification et prédiction images du poumon

Dans la pratique clinique, les radiologues doivent généralement prendre en compte les ganglions lymphatiques, en particulier dans la région du médiastin (la zone située entre les poumons). Ils sont très pertinents en cas de cancer. Les ganglions lymphatiques affectés sont généralement hypertrophiés.

Les ganglions lymphatiques doivent systématiquement être pris en compte dans la pratique clinique dans tous les types d'examens oncologiques. La détection automatique des ganglions lymphatiques à partir des données du scanner thoracique est cependant difficile en raison du faible contraste et de l'encombrement. Les détecteurs de fenêtres coulissantes utilisant des caractéristiques traditionnelles sont facilement confondus avec des structures similaires comme les muscles et les vaisseaux.

La classification des pixels basée sur la texture dans les images de tomodensitométrie (CT) du poumon est un outil émergent et prometteur pour l'analyse quantitative des maladies pulmonaires telles que l'emphysème, l'une des principales composantes de la maladie pulmonaire obstructive chronique (MPOC).

- L'emphysème, qui se caractérise par la perte de tissu pulmonaire, est visible sur les images CT sous forme de motifs texturaux comprenant des taches de faible intensité ou de faibles zones d'atténuation de tailles variables. Dans [217], les auteurs proposent d'utiliser des signatures du texton basées sur une représentation brute des pixels ainsi qu'un système de classificateurs multiples parallèles pour la classification de l'emphysème dans les images tomodensitométriques du poumon.

Le système de classificateurs multiples est composé de machines à vecteurs de support sur les signatures texton en tant que classificateurs de base et combine leurs décisions à l'aide d'une règle de produit.

Le calcul des caractéristiques est basé sur les Texton. Le codebook des textons est construit en extrayant 500 patches aléatoires de chaque ROI dans l'ensemble d'apprentissage. Des tailles de patch de 3×3 jusqu'à 9×9 pixels sont utilisées dans les expériences. La représentation brute des pixels est utilisée. Les patches extraits des différentes ROI de chaque classe sont donnés à l'algorithme de clustering k-means pour trouver le livre de codes.

L'emphysème est souvent classé en différents sous-types en fonction de la morphologie. Ce travail se concentre sur les deux sous-types liés au tabagisme, à savoir l'emphysème centrolobulaire (CLE), défini comme de multiples petites zones de faible atténuation et l'emphysème paraseptal (PSE), défini comme de multiples zones de faible atténuation dans une seule couche le long de la plèvre souvent entourée de cloisons interlobulaires visibles sous forme de fines parois blanches. Les données utilisées pour les expérimentations comprennent 168 ROI, de 50×50 pixels, représentant les trois classes suivantes :

Tissu normal (NT) (59 ROI), CLE (50 ROI) et PSE (59 ROI). Les ROIs sont extraits de 75 images CT pulmonaires en tranches minces de 25 sujets différents, où le modèle principal a été obtenu en tant qu'évaluation visuelle consensuelle par deux lecteurs expérimentés.

Les meilleurs résultats du système de classification basé sur le texton sur ces images CT du poumon sont trouvés par une agrégation des tailles de texton (TS) de 3×3 jusqu'à 9×9 , et pour des valeurs de k de 10 jusqu'à 50 de l'algorithme K-Moyenne en utilisant un seul SVM. La performance du système proposé, avec une précision de 95%, est similaire à une approche récemment proposée basée sur des modèles binaires locaux, qui fonctionne presque comme la meilleure parmi d'autres approches dans la littérature.

- Dans [219], les auteurs ont proposé de combiner la segmentation et la détection pour améliorer les performances de détection des ganglions lymphatiques (Lymph nodes). Les caractéristiques extraites d'une segmentation qui est initialisée avec un candidat de détection peuvent être utilisées pour l'apprentissage d'un classifieur qui décide si la détection est un vrai ou un faux positif. Dans cet article, la méthode des coupes de graphes (graph cut) est adaptée au problème de segmentation des ganglions lymphatiques. Il est proposé un cadre qui ne nécessite qu'une seule graine positive et résout en même temps le petit problème de coupure des coupes de graphe.

De plus, ils proposent un ensemble de caractéristiques extrait pour la segmentation candidate. Un classificateur est formé sur cet ensemble de caractéristiques et utilisé pour rejeter les fausses alarmes.

Contrairement aux travaux précédents, qui n'adaptent pas une sphère à un ganglion lymphatique. Les auteurs adaptent les coupes graphiques au problème de la segmentation des ganglions lymphatiques. La connaissance préalable de l'apparence des ganglions lymphatiques est incluse en sélectionnant les poids du graphe en fonction des annotations manuelles des données.

La segmentation est initialisée avec un point unique à partir du résultat de la détection. Pour surmonter un problème majeur de la segmentation des coupes de graphe, le problème des petites coupes, nous introduisons une pondération radiale supplémentaire du graphe qui est bien adaptée à la segmentation des structures de type blob. La méthode proposée pour la détection et la segmentation Ganglions lymphatiques est passée par deux étapes :

- 1. Génération de candidats :** en cours de cette étape, le classificateur d'arbre de renforcement probabiliste (PBT : probabilistic boosting tree) est formé avec des caractéristiques de type Haar 3-D afin de générer les modèles de détection. Ce processus est complètement séquentiel, dans la phase de l'apprentissage, les exemples négatifs de toutes les étapes sauf la première sont générés à partir de fausses alarmes de la précédente. Ainsi, les classificateurs se spécialisent sur les exemples concrets.

Un PBT est un arbre de décision binaire avec un puissant classificateur Ada-Boost à chaque nœud. Les caractéristiques de type Haar sont des caractéristiques de boîte simples mais

puissantes car elles peuvent être calculées très efficacement. A partir de la deuxième étape, seules les détections de l'étape précédente sont prises en compte lors du test.

Le troisième classificateur utilise les caractéristiques alignées sur le gradient (GAF). L'idée est d'extraire les caractéristiques aux emplacements des sauts de gradient, car ce sont des indices pour la limite de l'objet. À partir du candidat t qui correspond au centre ganglionnaire, les gradients sont calculés le long des rayons dans la direction radiale. Aux maximums locaux de l'amplitude du gradient, les caractéristiques ponctuelles simples sont calculées.

2- détection conjointe et segmentation des structures de type nœud à l'aide de coupes de graphe : après la génération du centre du candidat de ganglion lymphatique détecté, à partir de l'étape précédente, une sous-image recadrée d'une taille de sous-image relativement grand et garantit que presque tous les ganglions lymphatiques (par exemple $4 \times 4 \times 4$ cm). Dans la sous-image, le ganglion lymphatique est segmenté à l'aide de la méthode des coupes de graphe pour la segmentation d'image ensemencée. Les voxels forment les sommets du graphe et les voxels voisins sont connectés.

2-Segmentation based features. Comme dernière étape de la cascade de détection, un classificateur Ada-Boost est formé avec des caractéristiques extraites de la segmentation qui a été initialisée avec le centre du ganglion lymphatique détecté pour savoir s'il s'agit d'un vrai ganglion lymphatique ou d'une fausse détection.

La méthode proposée a été évaluée sur 54 ensembles de données CT montrant la région thoracique de patients atteints de lymphome. L'espacement des voxels était généralement de $0,8 \times 0,8 \times 1$ mm³.

La validation croisée sur 54 ensembles de données CT a montré que le système proposé atteint un taux de détection de 60,9 % avec seulement 6,1 fausses alarmes par image de volume, ce qui est meilleur que l'état actuel de l'art de la détection des ganglions lymphatiques médiastinaux.

- Un nouveau cadre implémenté sur les unités de traitement graphique NVIDIA (GPU) utilisant CUDA pour la segmentation précise des tissus pulmonaires à partir de la tomographie par ordinateur (CT) est proposé []. Les images CT, la génération de données de l'espace à l'échelle gaussienne (GSS) à l'aide de noyaux gaussiens (GK) et les cartes souhaitées des régions (poumon et autres tissus thoraciques) sont décrites par un modèle conjoint de champ aléatoire de Markov-Gibbs (MGRF) de signaux d'image indépendants et des étiquettes de régions interdépendantes implémentées sur GPU. La segmentation initiale à partir des images GSS CT originales et générées est basée sur les modèles de combinaison linéaire de gaussiens discrets (LCDG) ; La segmentation initiale est obtenue à partir des images GSS CT originales et générées ; puis ils sont affinés de manière itérative à l'aide d'un modèle MGRF parallèle implémenté sur GPU avec des potentiels estimés analytiquement. Enfin, ces segmentations initiales sont fusionnées à l'aide d'une approche de fusion bayésienne pour obtenir la segmentation finale de la région pulmonaire. Des expériences sur onze ensembles de données réels basés sur la métrique Dice Similarity Coefficient (DSC) confirment la grande précision de l'approche proposée. Les résultats du temps d'exécution montrent que cet algorithme prend environ trois secondes, ce qui est environ 103 fois plus rapide par rapport à une implémentation naïve à un seul thread sur CPU.

La segmentation des tissus pulmonaires à partir d'images CT est une étape essentielle dans le développement de tout système de diagnostic assisté par ordinateur (CAD) non invasif pour le diagnostic précoce du cancer du poumon. Le tissu pulmonaire, dans les images CT, apparaît sous forme de régions sombres. Ainsi, ce contraste entre le poumon et le thorax est à la base de la plupart des schémas de segmentation ; cependant, certains tissus pulmonaires tels que les artères, les veines, les bronches, les bronchioles et les nodules pulmonaires (s'ils existent) ont des niveaux de gris très proches des tissus thoraciques.

Ce chevauchement de la poitrine et certains des signaux pulmonaires doivent être différenciés afin d'obtenir une segmentation précise du tissu pulmonaire.

Cet article détaille les images CT, les données GSS générées et les cartes de régions souhaitées par un modèle MGRF conjoint parallèle de signaux d'image indépendants, mis en œuvre à l'aide de CUDA, qui tient compte des interactions spatiales entre les étiquettes de région. Nous suivons une approche de programmation parallèle dans la mise en œuvre du modèle MGRF conjoint. Récemment, la théorie GSS a reçu beaucoup d'attention en raison de son cadre mathématique solide d'analyse des données 2D/3D dans le domaine spatial. Ils ont utilisé trois échelles différentes de GK 3D pour réduire l'inhomogénéité qui existe dans les images du scanner thoracique. Des images CT originales et lissées à l'aide des GK 3D sont montrés. Les données d'image d'origine et les données GSS générées sont segmentées par les modèles LCDG, et les segmentations sont affinées de manière itérative par un modèle MGRF parallèle implémenté sur GPU avec des potentiels estimés analytiquement. Enfin, ces segmentations initiales sont fusionnées par une approche de fusion bayésienne pour obtenir la segmentation finale des régions pulmonaires.

Quatre modèles probabilistes ont été proposés :

- **Modèle conjoint de Markov-Gibbs des images pulmonaires CT**

Le modèle d'images MGRF est donné par une distribution de probabilité conjointe des images CT et des cartes de régions souhaitées

- **Modèle d'interaction spatiale des images CT**

Conformément à un MGRF générique avec des interactions de paires de voxels [9], nous restreignons le voisinage des interactions aux 26 voxels les plus proches ; par symétrie, ils sont indépendants des orientations des régions.

- **Modèle d'intensité des images pulmonaires CT**

La distribution empirique des niveaux de gris pour les images CT est étroitement approchée à l'aide de la combinaison linéaire de gaussiennes discrètes (LCDG) avec des composants C_p positifs et C_n négatifs. Un algorithme séquentiel basé sur EM est adapté pour résoudre les paramètres du modèle d'intensité LCDG **Fusion bayésienne des résultats de la segmentation initiale**

La méthode de fusion bayésienne peut être appliquée pour fusionner les résultats de segmentation de différents classificateurs sous une condition ; que la sortie de chaque classifieur est exprimée par une probabilité a posteriori. Les sorties de N classificateurs sont fusionnées en calculant la moyenne de leur probabilité a posteriori.

La décision finale pour chaque pixel à classer comme poumon ou autres tissus sera basée sur la nouvelle probabilité a posteriori estimée PF

Des expériences ont été menées sur onze ensembles de données avec les images CT.

La taille de chaque ensemble de données 3D est de $512 \times 512 \times 390$. Les images CT contiennent deux classes ($K = 2$) : le tissu pulmonaire plus foncé et la région thoracique plus claire. Un CT 3D typique et ses images 3D lissées utilisant $33 \times 33 \times 5$ GK (Gaussian Kernels). Pour mettre en évidence l'avantage d'intégrer le GSS avec l'intensité et les informations spatiales dans l'approche de segmentation, les données de test CT ont été segmentées en utilisant les informations d'intensité uniquement (Ionly) et les interactions combinées d'intensité et spatiales (I&S), l'intégration de l'ESG dans l'approche proposée a amélioré les résultats de la segmentation, comme en témoigne le plus grand DSC. De plus, les tests t statistiques appariés entre les valeurs moyennes des DSC de l'approche proposée et celles des approches (I) et (I&S) montrent une différence significative entre cette approche et les autres approches (valeurs de p inférieures à 0,05). Ces résultats mettent en évidence les avantages d'utiliser l'ESG en plus des caractéristiques d'intensité et d'interaction spatiale.

Pour vérifier la force de l'approche proposée par rapport aux algorithmes actuels, la méthode proposée est comparée à trois algorithmes de pointe : IT suivi d'une séquence d'opérations morphologiques (IT), segmentation à résolution multiple (MRS), un modèle déformable basé sur Gradient Vector Flow (GVF).

L'approche proposée montre la meilleure précision de segmentation (le plus grand DSC). De plus, cet incrément est statistiquement significatif selon le test t apparié statistique sur les DSC

des approches comparables. Ces résultats mettent en évidence les avantages et les forces de l'approche proposée.

A.3. Classification et prédication des images du foie

• Dans [220], les auteurs proposent une méthode semi-automatique de la réponse aux traitements locorégionaux préopératoires (PLT) pour le carcinome hépato-cellulaire (CHC), qui nécessite de faibles efforts d'interaction pour segmenter le parenchyme, les tissus tumoraux actifs et nécrotiques. La méthode proposée estime le taux de nécrose tumorale (TN : tumor necrosis) à partir de tomodensitogrammes dynamiques à contraste amélioré. En combinant les supervoxels SLIC (SLIC : simple linear iterative clustering) et la forêt de décision aléatoire, il implique des caractéristiques multiphases discriminatives par cluster extraites de tomodensitogrammes dynamiques enregistrés à contraste amélioré. L'évaluation quantitative sur les annotations expertes de la vérité au sol confirme les avantages de l'exploitation des informations multiphases des régions sémantiques pour segmenter avec précision les tumeurs hépatiques HCC. La méthode interactive de segmentation tumorale se traduit par une classification du forêt aléatoire multiphase par grappes pour l'estimation du taux de TN. Trois classes sont considérées : parenchyme, zones actives et nécrotiques. Ils ont également proposé de tirer pleinement parti des scanners multiphases avec injection de contraste. Pour un examen donné, un produit de contraste est injecté au patient et des tomodensitogrammes sont acquis à différentes phases : avant injection (WI) mais aussi après au niveau artériel (AR), veineux précoce (EV) et veineux tardif (LV) phases. La précision de la méthode est liée à la capacité des caractéristiques à discriminer les différents tissus. Nous attribuons à chaque région sémantique, r , 20 caractéristiques visuelles en grappes multiphases divisées en 3 groupes. Les deux premiers groupes introduisent des caractéristiques spatiales en termes d'intensité et d'amplitude du gradient à l'étendue spatiale du cluster. Dans les deux cas et pour chaque phase, les valeurs moyennes et d'écart type sont calculées parmi tous les voxels de r . Ainsi, nous quantifions l'intensité intrinsèque, l'homogénéité visuelle, l'information texturale et la répartition des échelles de texture. Le troisième groupe exploite pleinement les données dynamiques à contraste amélioré en combinant plusieurs informations d'intensité multiphase moyennées parmi les voxels de r : caractéristiques dynamiques utilisées dans, y compris l'amélioration des pics (PE) et l'aire sous la courbe (AUC). Le protocole a été testé sur des données collectées à partir de 7 examens réalisés sur des patients atteints de CHC. La comparaison entre les approches monophasées et multiphases confirme que l'exploitation des images CT multiphases au lieu d'un seul scanner améliore significativement les résultats avec des gains de DICE pour les stratégies basées sur les clusters et les voxels.

A.4. Classification et prédication des images du cœur

• La détection précise et cohérente des frontières endocardiques en échocardiographie 3D est une tâche difficile. Cela s'explique en partie par le fait que la structure trabéculée de la limite endocardique conduit à des caractéristiques de bord alternées qui varient au cours d'un cycle cardiaque. Les détecteurs de bords à gradient maximum (MG), critère de pas (STEP) et flux max/coupe min (MFMC) ont été précédemment appliqués pour la détection des bords endocardiques. Dans [221], les auteurs ont introduit les méthodes MLE et JS qui utilisent l'inférence statistique pour la détection des bords endocardiques. Les estimateurs du maximum de vraisemblance (MLE) et de James-Stein (JS) sont utilisés pour combiner plusieurs détecteurs de bord dans le contexte de l'imagerie médicale. Les réponses de ces détecteurs de bord en fonction de leurs confiances à l'aide des MLE et JS. Ils ont présenté également une méthode pour utiliser les estimations basées sur la confiance en tant que mesures dans un cadre de suivi du ventricule gauche (VG) basé sur un filtre de Kalman. L'efficacité des méthodes introduites est validée par des analyses comparatives entre le MG, le STEP, le MFMC, le MLE et le JS. Un ensemble de 18 enregistrements d'échocardiographie 3D (3D

echocardiography recordings) apicale a été acquis à l'aide d'un échographe Vivid 7 (GE Vingmed Ultrasound, Norvège) et d'un transducteur matriciel.

- La délimitation des frontières du myocarde à partir de l'échocardiographie 3D est une étape critique pour le diagnostic des maladies cardiaques. Dans [222], Une nouvelle approche d'apprentissage structuré empruntée à la vision par ordinateur est montrée pour effectuer une détection rapide et robuste des bords du myocarde. Les auteurs adaptent les forêts aléatoires structurées, empruntées à la vision par ordinateur, pour une détection rapide et robuste des contours du myocarde.

Les résultats qualitatifs et quantitatifs démontrent que la méthode proposée surpasse les détecteurs de bord standard, et suppriment efficacement la prédiction des réponses de bord endocardique non pertinentes et permettant aux modèles déformables et aux approches basées sur les contours de converger de manière plus stable vers les limites myocardiques détectées, permettant le calcul d'indices cliniques LV plus précis.

Cette méthode est évaluée sur un ensemble de données composé de tranches à petit axe de 25 volumes d'échocardiographie en fin de diastolique. Les résultats montrent que le modèle d'ensemble proposé surpasse les détecteurs de bord standard basés sur l'intensité et sur la phase locale, tout en supprimant ou en supprimant de manière significative les bords non pertinents déclenchés par les artefacts d'image ultrasonore et les structures anatomiques du pool sanguin.

A.5. Classification et prédication des images des artères coronaires

- Le travail exploré par [223] intègre la sélection des caractéristiques pour améliorer les performances dans le domaine de la détection des anomalies vasculaires. En partant d'un cadre de classification défini précédemment et basé sur des machines à vecteurs de support (SVM), ils ont essayé de déterminer les caractéristiques qui améliorent les performances de classification et de définir des directives pour la sélection des caractéristiques. Trois stratégies différentes ont été utilisées à l'étape de sélection des caractéristiques, tandis qu'un SVM de détection de la densité (DLD-SVM) a été utilisé pour valider les performances des caractéristiques sélectionnées par rapport aux données de test. Les résultats montrent qu'une sélection minutieuse des caractéristiques permet une bonne performance de classification. DLD-SVM affiche des performances médiocres lorsque toutes les caractéristiques sont utilisées ensemble, en raison de la malédiction de la dimensionnalité.

Quatre Strategies de Feature Selection ont été implémentés : 1) Aucune sélection. La première stratégie consiste à utiliser toutes les caractéristiques définies pour la classification à l'aide des SVM définis en 2.4. 2) F-score. Le score F [15] est une technique qui mesure la discrimination de deux ensembles de nombres réels. Forêt aléatoire (RF) et SVM - Élimination de caractéristiques récursives (SVM-RFE). Au total, 14 ensembles de données expérimentales MDCT cardiaques ont été utilisés. Les performances des données de test, en utilisant le meilleur ensemble de caractéristiques sélectionnées par chaque stratégie, sont résumées dans le tableau 3. Les résultats obtenus avec F-score et RF sont assez proches, alors que ceux obtenus avec SVM-RFE présentent un BER (Balanced Error Rate) supérieur.

ANNEXE2

La fonction d'erreurs en termes de la matrice de poids $E(W)$ est une fonction cout. La descente de gradient est une technique qui permet le poids se déplace progressivement avec une estimée ΔW . Les poids optimums peuvent être trouvés itérativement dans l'espace des poids par la formule :

$$W_{new} = W_{old} + \Delta W$$

Supposons deux réseaux de neurones représentés par les fonctions d'erreurs $E_1(W)$ et $E_2(W)$, la fusion (the merge) de deux réseaux de neurones est représentés par $E(W) = E_1(W) \times E_2(W)$. Le produit de deux fonctions décroissantes est une fonction plus décroissante dans l'intervalle]0,1[. Lorsque la matrice de poids se déplace avec une estimée ΔW , l'erreur total E est minimisée par l'effet de déplacement ΔW_1 du premier réseau et le déplacement ΔW_2 du deuxième réseau. La façon qui détermine la mise à jours est de déplacer les poids dans la direction opposée du gradient de la fonction d'erreur:

$$\Delta W = -\eta \frac{\partial E(W)}{\partial W}$$

En plus, la fonction d'erreur à chaque nœud dans le réseau de neurones, avec la somme nette ζ et les entrées u_0, u_1, \dots, u_q avec les poids w_0, w_1, \dots, w_q , est exprimés par dérivation partielle en term des poids suivants :

$$\begin{aligned} \frac{\partial E}{\partial w_j} &= \frac{\partial E}{\partial \zeta} \frac{\partial \zeta}{\partial w_j} = \frac{\partial E}{\partial \zeta} u_j \\ \frac{\partial E}{\partial w_j} &= \delta \frac{\partial \zeta}{\partial w_j} = \delta u_j \end{aligned}$$

$\delta = \frac{\partial E}{\partial \zeta}$ est le terme d'erreur pour le nœud et est utilisé pour la mise à jours de tous les poids d'entrées entrants à partir de couche précédente.

Plusieurs modèles des architectures identiques des réseaux nde neurones (Nombre des couches, Nombre des nouds dans chaque couche, fonction d'activation) peuvent être superposés et unifiés par la somme de leurs erreurs partielles en termes des poids.

- Réseau de neurones fermé :

Dans la suite, W represente un ensemble de poids interconnectés d'un réseau de neurones et $W^{(l)}(t)$ la valeur au moment t des poids reliant les neurones des $(l-1)$ couches avec la $(l)^{\text{ème}}$ couche. L'application de la méthode de gradient pour les poids de la couche (l) conduit à un algorithme:

$$W^{(l)}(t+1) = W^{(l)}(t) - \eta(t) \frac{\partial E(t)}{\partial W^{(l)}}$$

Considérons le réseau de neurones de la figure 3.9 dans lequel $\text{sum}_i^l(t)$ est le total de tous signaux atteignant le $i^{\text{ème}}$ neurone de la couche l . l'optimisation global de la method du gradient par retro-propagation conduit à un algorithme de la forme (en utilisant la règle de chain):

$$w_{ij}^{(l)}(t+1) = w_{ij}^{(l)}(t) - \eta \left(\frac{\partial E(t)}{\partial z_i^{(l+1)}(t)} \right) \left(\frac{\partial z_i^{(l+1)}(t)}{\partial \text{sum}_i^{(l+1)}(t)} \right) \left(\frac{\partial \text{sum}_i^{(l+1)}(t)}{\partial w_{ij}^{(l)}} \right)$$

Soit pour la couche de sortie $(l = q)$:

$$w_{ij}^{(q-1)}(t+1) = w_{ij}^{(q-1)}(t) - \eta(t) \delta_i^{(q)}(t) z_j^{(q)}(t)$$

Voici la démonstration de cette formule : (Pour la couche de sortie) ($l = q$) :

$$\left\{ \begin{array}{l} \frac{\partial E}{\partial w_{1j}^{(q-1)}} = \frac{\partial E}{\partial z_1^{(q)}} \frac{\partial z_1^{(q)}}{\partial y_1^{(q)}} \frac{\partial y_1^{(q)}}{\partial w_{1j}^{(q-1)}} \\ \frac{\partial E}{\partial w_{2j}^{(q-1)}} = \frac{\partial E}{\partial z_2^{(q)}} \frac{\partial z_2^{(q)}}{\partial y_2^{(q)}} \frac{\partial y_2^{(q)}}{\partial w_{1j}^{(q-1)}} \\ \dots \\ \dots \\ \dots \\ \frac{\partial E}{\partial w_{nqj}^{(q-1)}} = \frac{\partial E}{\partial z_{nq}^{(q)}} \frac{\partial z_{nq}^{(q)}}{\partial y_{nq}^{(q)}} \frac{\partial y_{nq}^{(q)}}{\partial w_{nqj}^{(q-1)}} \end{array} \right. \text{ on a:}$$

$$\left\{ \begin{array}{l} \delta_i^{(q)} = \frac{\partial E}{\partial z_i^{(q)}} \frac{\partial z_i^{(q)}}{\partial y_i^{(q)}} \\ \text{and} \\ \frac{\partial y_i^{(q)}}{\partial w_{ij}^{(q-1)}} = z_j^{(q-1)} \end{array} \right. \text{ donc :} \left\{ \begin{array}{l} w_{1j}^{(q-1)}(t+1) = w_{1j}^{(q-1)}(t) - \eta(t) \delta_1^{(q)}(t) z_j^{(q-1)}(t) \\ w_{2j}^{(q-1)}(t+1) = w_{2j}^{(q-1)}(t) - \eta(t) \delta_2^{(q)}(t) z_j^{(q-1)}(t) \\ \dots \\ \dots \\ \dots \\ w_{nqj}^{(q-1)}(t+1) = w_{nqj}^{(q-1)}(t) - \eta(t) \delta_{nq}^{(q)}(t) z_j^{(q-1)}(t) \end{array} \right.$$

D'où:

$$w_{ij}^{(q)}(t+1) = w_{ij}^{(q)}(t) - \eta(t) \delta_i^{(q)}(t) z_j^{(q-1)}(t)$$

Avec : $\delta_i^{(q)}(t) = \frac{\partial E}{\partial z_i^{(q)}} \frac{\partial z_i^{(q)}}{\partial y_i^{(q)}} = (z_i^{(p)} - s_i(t)) \frac{\partial f(y_1)}{\partial y_1}$

Pour la couche cachée qui précède immédiatement la couche de sortie: ($l = q - 1$)

$$\left\{ \begin{array}{l} \frac{\partial E}{\partial w_{ij}^{(q-2)}} = \frac{\partial E}{\partial z_1^{(q)}} \frac{\partial z_1^{(q)}}{\partial y_1^{(q)}} \frac{\partial y_1^{(q)}}{\partial z_i^{(q-1)}} \frac{\partial z_i^{(q-1)}}{\partial y_i^{(q-1)}} \frac{\partial y_i^{(q-1)}}{\partial w_{ij}^{(q-2)}} = \delta_1^{(q)} \frac{\partial z_1^{(q)}}{\partial y_1^{(q)}} w_{1i}^{(q-1)} \frac{\partial z_i^{(q-1)}}{\partial y_i^{(q-1)}} z_j^{(q-2)} = \delta_i^{(q-2)} z_j^{(q-2)} \\ \text{or} \\ \frac{\partial E}{\partial w_{ij}^{(q-2)}} = \frac{\partial E}{\partial z_2^{(q)}} \frac{\partial z_2^{(q)}}{\partial y_2^{(q)}} \frac{\partial y_2^{(q)}}{\partial z_i^{(q-1)}} \frac{\partial z_i^{(q-1)}}{\partial y_i^{(q-1)}} \frac{\partial y_i^{(q-1)}}{\partial w_{ij}^{(q-2)}} = \delta_2^{(q)} \frac{\partial z_2^{(q)}}{\partial y_2^{(q)}} w_{2i}^{(q-1)} \frac{\partial z_i^{(q-1)}}{\partial y_i^{(q-1)}} z_j^{(q-2)} = \delta_i^{(q-2)} z_j^{(q-2)} \\ \text{or} \\ \dots \\ \text{or} \\ \frac{\partial E}{\partial w_{ij}^{(q-2)}} = \frac{\partial E}{\partial z_{nq}^{(q)}} \frac{\partial z_{nq}^{(q)}}{\partial y_{nq}^{(q)}} \frac{\partial y_{nq}^{(q)}}{\partial z_i^{(q-1)}} \frac{\partial z_i^{(q-1)}}{\partial y_i^{(q-1)}} \frac{\partial y_i^{(q-1)}}{\partial w_{ij}^{(q-2)}} = \delta_{nq}^{(q)} \frac{\partial z_{nq}^{(q)}}{\partial y_{nq}^{(q)}} w_{nqi}^{(q-1)} \frac{\partial z_i^{(q-1)}}{\partial y_i^{(q-1)}} z_j^{(q-2)} = \delta_i^{(q-2)} z_j^{(q-2)} \end{array} \right.$$

Tel que : $\left\{ \begin{array}{l} y_1^{(q)} = y_1(w_{ij}^{(q-1)}) \\ y_2^{(q)} = y_2(w_{ij}^{(q-1)}) \\ \dots \\ y_{nq}^{(q)} = y_{nq}(w_{ij}^{(q-1)}) \end{array} \right.$

En utilisant la règle de chain, puisque $y_1^{(q)}, y_2^{(q)}, \dots, y_{nq}^{(q)}$ sont des variables indépendantes, alors on peut écrire :

$$\frac{\partial E}{\partial w_{ij}^{(q-2)}} = \delta_1^{(q)} \frac{\partial z_1^{(q)}}{\partial y_1^{(q)}} w_{1i}^{(q-1)} \frac{\partial z_i^{(q-1)}}{\partial y_i^{(q-1)}} z_j^{(q-2)} + \delta_2^{(q)} \frac{\partial z_2^{(q)}}{\partial y_2^{(q)}} w_{2i}^{(q-1)} \frac{\partial z_i^{(q-1)}}{\partial y_i^{(q-1)}} z_j^{(q-2)} + \dots$$

$$+ \delta_{n_q}^{(q)} \frac{\partial z_{n_q}^{(q)}}{\partial y_{n_q}^{(q)}} w_{n_q i}^{(q-1)} \frac{\partial z_i^{(q-1)}}{\partial y_i^{(q-1)}} z_j^{(q-2)}$$

$$\frac{\partial E}{\partial w_{ij}^{(q-2)}} = \left[\delta_1^{(q)} \frac{\partial z_1^{(q)}}{\partial y_1^{(q)}} w_{1i}^{(q-1)} \frac{\partial z_i^{(q-1)}}{\partial y_i^{(q-1)}} + \delta_2^{(q)} \frac{\partial z_2^{(q)}}{\partial y_2^{(q)}} w_{2i}^{(q-1)} \frac{\partial z_i^{(q-1)}}{\partial y_i^{(q-1)}} + \dots \right. \\ \left. + \delta_{n_q}^{(q)} \frac{\partial z_{n_q}^{(q)}}{\partial y_{n_q}^{(q)}} w_{n_q i}^{(q-1)} \frac{\partial z_i^{(q-1)}}{\partial y_i^{(q-1)}} \right] z_j^{(q-2)}$$

$$\frac{\partial E}{\partial w_{ij}^{(q-2)}} = \delta_i^{(q-2)} z_j^{(q-2)}$$

Tel que: $\delta_i^{(q-2)} = \delta_1^{(q)} \frac{\partial z_1^{(q)}}{\partial y_1^{(q)}} w_{1i}^{(q-1)} \frac{\partial z_i^{(q-1)}}{\partial y_i^{(q-1)}} + \delta_2^{(q)} \frac{\partial z_2^{(q)}}{\partial y_2^{(q)}} w_{2i}^{(q-1)} \frac{\partial z_i^{(q-1)}}{\partial y_i^{(q-1)}} + \dots + \delta_{n_q}^{(q)} \frac{\partial z_{n_q}^{(q)}}{\partial y_{n_q}^{(q)}} w_{n_q i}^{(q-1)} \frac{\partial z_i^{(q-1)}}{\partial y_i^{(q-1)}}$

$$\delta_i^{(q-2)} = \frac{\partial z_i^{(q-1)}}{\partial y_i^{(q-1)}} \sum_{m=1}^{n_q} \delta_i^{(q-1)} w_{mi}^{(q-1)}$$

Donc:

$$w_{ij}^{(q-2)}(t+1) = w_{ij}^{(q-2)}(t) - \eta(t) \delta_i^{(q-2)}(t) z_j^{(q-2)}(t)$$

Avec:

$$\delta_i^{(q-2)} = \frac{\partial z_i^{(q-1)}}{\partial y_i^{(q-1)}} \sum_{m=1}^{n_{q-1}} \delta_i^{(q-1)} w_{mi}^{(q-1)}$$

Récursivement, avec la même façon on peut démontrer pour les couches cachées ($2 \leq l \leq q-1$) que:

$$w_{ij}^{(l)}(t+1) = w_{ij}^{(l)}(t) - \eta(t) \delta_i^{(l)}(t) z_j^{(l-1)}(t)$$

Avec : $\delta_i^{(l)}(t) = \frac{\partial f(\text{sum}_i^{(l)}(t))}{\partial \text{sum}_i^{(l)}(t)} \sum_{k=1}^{n_{l+1}} \delta_k^{(l+1)} w_{ki}^{(l+1)}$

Nous avons associé deux files d'attentes afin d'enfiler (ajouter au début de la file) et de défiler (supprimer le début à partir de la file) les termes d'erreurs $\delta_i^{(l)}$ et les poids $w_{ij}^{(l)}$.

BIBLIOGRAPHIE

- [1] Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical pattern recognition: A review. *IEEE Transactions on pattern analysis and machine intelligence*, 22(1), 4-37.
- [2] Walck, C. (2007). Handbook on statistical distributions for experimentalists. University of Stockholm, 10, 96-01.
- [3] Duda RO, Hart PE, Stork DG (1973) Pattern classification, vol 2. Wiley, New York.
- [4] Bishop. (2006) "Pattern Recognition and Machine Learning". Information Science and Statistics, Springer, Heidelberg.
- [4a] Kaster, F. O., Menze, B. H., Weber, M. A., & Hamprecht, F. A. (2011). Comparative validation of graphical models for learning tumor segmentations from noisy manual annotations. In *Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging: International MICCAI Workshop, MCV 2010, Beijing, China, September 20, 2010, Revised Selected Papers 1* (pp. 74-85). Springer Berlin Heidelberg.
- [5] Sokolovska, N. (2010). Contributions to the estimation of probabilistic discriminative models: semi-supervised learning and feature selection (Doctoral dissertation, TélécomParisTech).
- [5a] Ramirez, O. D. Z. (2018). *Bayesian statistics and modeling for the prediction of radiotherapy outcomes: an application to glioblastoma treatment* (Doctoral dissertation, Normandie Université).
- [5b] Xia, T. (2022). Deep generative models for medical image synthesis and strategies to utilise them.
- [5c] Celard, P., Iglesias, E. L., Sorribes-Fdez, J. M., Romero, R., Vieira, A. S., & Borrajo, L. (2023). A survey on deep learning applied to medical images: from simple artificial neural networks to generative models. *Neural Computing and Applications*, 35(3), 2291-2323.
- [6] Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. MIT press.
- [6a] Zhang, B., Qi, S., Monkam, P., Li, C., Yang, F., Yao, Y. D., & Qian, W. (2019). Ensemble learners of multiple deep CNNs for pulmonary nodules classification using CT images. *IEEE Access*, 7, 110358-110371.
- [7] Geman, Bienenstock et Doursat (1992) Geman, S., E. Bienenstock, and R. Doursat. 1992. "Neural Networks and the Bias/Variance Dilemma." *Neural Computation* 4: 1-58.
- [7a] Desir, C. (2013). *Classification automatique d'images, application à l'imagerie du poumon profond* (Doctoral dissertation, Université de Rouen).
- [8] Moon, T. K. (1996). The expectation-maximization algorithm. *IEEE Signal processing magazine*, 13(6), 47-60.
- [8a] Chung, A. C., Noble, J. A., & Summers, P. (2002). Fusing speed and phase information for vascular segmentation of phase contrast MR angiograms. *Medical Image Analysis*, 6(2), 109-128.
- [10] Härdle, W. (1990). *Applied nonparametric regression* (No. 19). Cambridge university press.
- [11] Wettschereck, D., Aha, D. W., & Mohri, T. (1997). A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. *Artificial Intelligence Review*, 11, 273-314.
- [12] Webb, A. (1999). *Statistical Pattern Recognition* (Arnold, London). *Search in*.
- [13] Fan, J., & Fan, Y. (2008). High dimensional classification using features annealed independence rules. *Annals of statistics*, 36(6), 2605.
- [14] Braun, A. C., Weidner, U., & Hinz, S. (2012). Classification in high-dimensional feature spaces—Assessment using SVM, IVM and RVM with focus on simulated EnMAP data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2), 436-443.
- [15] Jensen, R. (2005). *Combining rough and fuzzy sets for feature selection* (Doctoral dissertation, University of Edinburgh).

- [16] Jain, A., & Zongker, D. (1997). Feature selection: Evaluation, application, and small sample performance. *IEEE transactions on pattern analysis and machine intelligence*, 19(2), 153-158.
- [17] Pudil, P., Novovičová, J., & Kittler, J. (1994). Floating search methods in feature selection. *Pattern recognition letters*, 15(11), 1119-1125.
- [18] Kudo, M., & Sklansky, J. (2000). Comparison of algorithms that select features for pattern classifiers. *Pattern recognition*, 33(1), 25-41.
- [19] Dong, G., & Li, J. (1997). Feature selection methods for classification. *Intelligent Data Analysis: An International Journal*, 1.
- [20] Hall, M. A. (1999). *Correlation-based feature selection for machine learning* (Doctoral dissertation, The University of Waikato).
- [21] Guyon, I., & Elisseeff, A. (2003). An introduction to feature and variable selection. *Journal of Machine Learning Research*, 3(1157-1182), 100.
- [22] Grandidier, F. (2003). *Un nouvel algorithme de sélection de caractéristiques: application à la lecture automatique de l'écriture manuscrite* (Doctoral dissertation, École de technologie supérieure).
- [23] Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial intelligence*, 97(1-2), 273-324.
- [24] Guyon, I., & Elisseeff, A. (2003). An introduction to feature and variable selection. *Journal of Machine Learning Research*, 3(1157-1182), 100.
- [25] Cakmakov D, Bennani Y, (2002). Feature selection for pattern recognition. Skopje, Informa.
- [26] Jourdan, L. (2003). *Métaheuristiques pour l'extraction de connaissances: application à la génomique* (Doctoral dissertation, Université des Sciences et Technologie de Lille-Lille I).
- [27] Martínez-Trinidad, J. F., & Guzmán-Arenas, A. (2001). The logical combinatorial approach to pattern recognition, an overview through selected works. *Pattern Recognition*, 34(4), 741-751.
- [28] Skellam, J. G. (1952). Studies in statistical ecology: I. Spatial pattern. *Biometrika*, 39(3/4), 346-362.
- [29] Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1, 81-106.
- [30] Breiman L., Friedman J., Olshen R., and Stone C. (1984). *Classification and Regression Trees*, and Hall/CRC, pp. 81-106.
- [31] Gherghout Y, Tlili Y (2014) Toward a generic model of machine learning algorithms. In: Laouar MR, Eom SB (eds) *Business intelligence and mobile technology: an information systems engineering perspective*. Cambridge Scholars Publishing, Cambridge.
- [32] Härdle, W. (1990). *Applied nonparametric regression* (No. 19). Cambridge university press.
- [33] Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.
- [34] Tu, Z. (2005, October). Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1* (Vol. 2, pp. 1589-1596). IEEE.
- [35] van den Bosch, A. (2000). Using induced rules as complex features in memory-based language learning. In *Fourth Conference on Computational Natural Language Learning and the Second Learning Language in Logic Workshop*.
- [36] Witten, I. H., & Frank, E. (2002). Data mining: practical machine learning tools and techniques with Java implementations. *Acm Sigmod Record*, 31(1), 76-77.
- [37] Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160(1), 3-24.
- [38] Angelov P, Gu X (2017) MICE: multi-layer multi-model images classifier ensemble. In: Cybernetics (CYBCONF), 2017 3rd IEEE international conference on. IEEE, pp 1-8
- [39] Angelov PP, Gu X (2018) Deep rule-based classifier with human-level performance and characteristics. *InfSci* 463-464:196-213

- [40] Gu X, Angelov PP, Zhang C, Atkinson PM (2018) A massively parallel deep rule-based ensemble classifier for remote sensing scenes. *IEEE Geosci Remote Sens Lett* 15(3):345–349
- [41] Gu X, Angelov PP (2018) Semi-supervised deep rule-based approach for image classification. *Appl Soft Comput* 68:53–68
- [42] Gherghout, Y., Tlili, Y., &Souici, L. (2021). Classification of breast mass in mammography using anisotropic diffusion filter by selecting and aggregating morphological and textural features. *Evolving Systems*, 12(2), 273-302.
- [43] Patgiri, C., &Ganguly, A. (2021). Adaptive thresholding technique based classification of red blood cell and sickle cell using Naïve Bayes Classifier and K-nearest neighbor classifier. *Biomedical Signal Processing and Control*, 68, 102745.
- [44] Cortes, C., &Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20, 273-297.
- [45] Hric, M., Chmulík, M., &Jarina, R. (2011, April). Model parameters selection for SVM classification using Particle Swarm Optimization. In *Proceedings of 21st International Conference Radioelektronika 2011* (pp. 1-4). IEEE.
- [46] Lee, K. C., & Cho, H. (2010). Performance of ensemble classifier for location prediction task: emphasis on Markov Blanket perspective. *International Journal of u-and e-Service, Science and Technology*, 3(3), 2010.
- [47] Bartlett, P., Freund, Y., Lee, W. S., &Schapire, R. E. (1998). Boosting the margin: A new explanation for the effectiveness of voting methods. *The annals of statistics*, 26(5), 1651-1686.
- [48] Barros, R. C., Basgalupp, M. P., De Carvalho, A. C., &Freitas, A. A. (2011). A survey of evolutionary algorithms for decision-tree induction. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(3), 291-312.
- [49] Eiben, A. E., & Smith, J. E. (2003). Introduction to evolutionary computing (SpringerVerlag).
- [50] Krishna, K., &Murty, M. N. (1999). Genetic K-means algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 29(3), 433-439.
- [51] Malhotra, R., Singh, N., & Singh, Y. (2011). Genetic algorithms: Concepts, design for optimization of process controllers. *Computer and information science*, 4(2), 39.
- [52] Tjong, A. S., &Monteiro, S. T. (2011, June). Feature selection with PSO and kernel methods for hyperspectral classification. In *2011 IEEE Congress of Evolutionary Computation (CEC)* (pp. 1762-1769). IEEE.
- [53] Schuh, M. A., Angryk, R., & Sheppard, J. (2012, May). Evolving kernel functions with particle swarms and genetic programming. In *Twenty-Fifth International FLAIRS Conference*.
- [54] Korürek, M., &Doğan, B. (2010). ECG beat classification using particle swarm optimization and radial basis function neural network. *Expert systems with Applications*, 37(12), 7563-7569.
- [55] Mitchell, M. (1998). *An introduction to genetic algorithms*. MIT press.
- [56] Moraglio, A., Di Chio, C., Togelius, J., &Poli, R. (2008). Geometric particle swarm optimization. *Journal of Artificial Evolution and Applications*, 2008.
- [57] Russ, J. C. (2006). *The image processing handbook*. CRC press.
- [58] Lanca, L., & Silva, A. (2012). *Digital imaging systems for plain radiography*. Springer Science & Business Media.
- [59] Frei, W. (1977). Image enhancement by histogram hyperbolization. *Computer Graphics and Image Processing*, 6(3), 286-294.
- [60] Ketcham, D. J. (1976, July). Real-time image enhancement techniques. In *Image processing* (Vol. 74, pp. 120-125). SPIE.
- [61] Gonzalez, R. C., & Woods, R. E. (2002). Digital image processing second edition. *Beijing: Publishing House of Electronics Industry*, 455, 6.
- [62] Rosenfeld, A. (1976). Digital picture processing. Academic press.
- [63] Pizer, S. M., Amburn, E. P., Austin, J. D., Cromartie, R., Geselowitz, A., Greer, T., ...&Zuiderveld, K.

- (1987). Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3), 355-368.
- [64] Restrepo, A., and Bovik, A., An adaptive trim mcd mean filter for image restoration. *IEEE Trans. Acouics, Speech, Signal Process. ASSP-36(8)*, 8813—8818 (1988).
- [65] Low, A., Introduction to Computer Vision and Image Processing. McGraw-Hill, U.K., 1991.
- [66] Jain, A. K. (1989). *Fundamentals of digital image processing*. Prentice-Hall, Inc..
- [67] Perona, P., & Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on pattern analysis and machine intelligence*, 12(7), 629-639.
- [68] Hau CC (2015) Handbook of pattern recognition and computer vision. World Scientific, Singapore
- [69] Shen, L., Rangayyan, R. M., &Desautels, J. L. (1994). Application of shape analysis to mammographic calcifications. *IEEE transactions on medical imaging*, 13(2), 263-274.
- [70] Pohlman, S., Powell, K. A., Obuchowski, N. A., Chilcote, W. A., &Grundfest-Broniatowski, S. (1996). Quantitative classification of breast tumors in digitized mammograms. *Medical Physics*, 23(8), 1337-1345.
- [71] Rangayyan, R. M., El-Faramawy, N. M., Desautels, J. L., &Alim, O. A. (1997). Measures of acutance and shape for classification of breast tumors. *IEEE Transactions on medical imaging*, 16(6), 799-810.
- [72] Bruce, L. M., &Kallergi, M. (1999, May). Effects of image resolution and segmentation method on automated mammographic mass shape classification. In *Medical Imaging 1999: Image Processing* (Vol. 3661, pp. 940-947). SPIE.
- [73] Hu, M. K. (1962). Visual pattern recognition by moment invariants. *IRE transactions on information theory*, 8(2), 179-187.
- [74] Lin, W., Xiao, J., &Micheli-Tzanakou, E. (1998, May). A computational intelligence system for cell classification. In *Proceedings. 1998 IEEE International Conference on Information Technology Applications in Biomedicine, ITAB'98 (Cat. No. 98EX188)* (pp. 105-109). IEEE.
- [75] Coakley, K., & Van Doorn, T. (1995). Invariant moment shape description of microcalcifications in digital mammograms. *Australasian physical & engineering sciences in medicine*, 18(2), 114-118.
- [76] Mertzios, B. G., &Tsirikolias, K. (1993). Statistical shape discrimination and clustering using an efficient set of moments. *Pattern recognition letters*, 14(6), 517-522.
- [77] Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7), 674-693.
- [78] Gopinath, R. A., &Burrus, C. S. (1993). Wavelet transforms and filter banks. In *Wavelets: A Tutorial in Theory and Applications* (pp. 603-654).
- [79] Bruce, L. M., &Adhami, R. R. (1999). Classifying mammographic mass shapes using the wavelet transform modulus-maxima method. *IEEE transactions on medical imaging*, 18(12), 1170-1177.
- [80] Shen, L., Rangayyan, R. M., &Desautels, J. L. (1994). Application of shape analysis to mammographic calcifications. *IEEE transactions on medical imaging*, 13(2), 263-274.
- [81] Ichimura, N. (1996, September). Volume data coding based on region segmentation using finite mixture model. In *Proceedings of 3rd IEEE International Conference on Image Processing* (Vol. 3, pp. 363-366). IEEE.
- [82] Vaidehi, K., &Subashini, T. S. (2015). Automatic characterization of benign and malignant masses in mammography. *Procedia Computer Science*, 46, 1762-1769.
- [83] Sun, B., &Xu, J. (1991, November). Image processing system for brain and neural tissue. In *Visual Communications and Image Processing'91: Image Processing* (Vol. 1606, pp. 1022-1026). SPIE.
- [84] Bengtsson, E., Eriksson, O., Holmquist, J., Jarkrans, T., Nordin, B., &Stenkvis, B. (1981). Segmentation of cervical cells: Detection of overlapping cell nuclei. *Computer Graphics and Image Processing*, 16(4), 382-394.
- [85] Song, L., & Donovan, R. M. (1988, November). Segmentation of cell images using an expert

- system. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 1383-1384). IEEE.
- [86] Bombardier, V., Levrat, E., & Bremont, J. (1990). A powerful algorithm for contour tracing. In *Application to the counting of nerve fibers, "Proceedings of the ISSM International Symposium. Mini and Microcomputers and their Applications* (pp. 25-28).
- [87] Goggin, S. D., & Janson, S. D. (1996, October). Shape features for recognition of Pap smear cells. In *Statistical and Stochastic Methods for Image Processing* (Vol. 2823, pp. 224-235). SPIE.
- [88] Ichimura, N. (1996, September). Volume data coding based on region segmentation using finite mixture model. In *Proceedings of 3rd IEEE International Conference on Image Processing* (Vol. 3, pp. 363-366). IEEE.
- [89] Zheng, Y., Du, J., Qi, L., & Yu, D. (1996, December). Echocardiographic visual processing and quantitative analysis. In *International Conference on Holography and Optical Information Processing (IHOIP'96)* (Vol. 2866, pp. 46-49). SPIE.
- [90] Granlund, G. H. (1972). Fourier preprocessing for hand print character recognition. *IEEE transactions on computers*, 10(2), 195-201.
- [91] Sharafeddin, H., Parnianpour, M., Hemami, H., Hanson, T., Goldman, S., & Madson, T. (1996, March). Computer aided diagnosis of low back disorders using the motion profile. In *Proceedings of the 1996 Fifteenth Southern Biomedical Engineering Conference* (pp. 431-432). IEEE.
- [92] Sanchez-Marin, F. J. (2000). Automatic recognition of biological shapes with and without representations of shape. *Artificial Intelligence in Medicine*, 18(2), 173-186.
- [93] Wang, P. C., Lin, K. P., Lou, S. A., Lin, H. D., & Chen, T. S. (1999, May). 3D MR image segmentation of prostate gland using two scan planes and Fourier descriptor technique. In *Medical Imaging 1999: Image Processing* (Vol. 3661, pp. 1047-1054). SPIE.
- [94] Granlund, G. H. (1976). Identification of human chromosomes by using integrated density profiles. *IEEE Transactions on Biomedical Engineering*, (3), 182-192.
- [95] Delibasis, K., Undrill, P. E., & Cameron, G. G. (1997). Designing Fourier descriptor-based geometric models for object interpretation in medical images using genetic algorithms. *Computer Vision and Image Understanding*, 66(3), 286-300.
- [96] Blum, H. (1967). A transformation for extracting new descriptions of shape. *Models for the perception of speech and visual form*, 362-380.
- [97] Davidson, J. (1991). Thinning and skeletonizing: A tutorial and overview. *Digital Image Processing: Fundamental and Applications*, E. Dougherty ed., Marcel Dekker, New York.
- [98] Lam, L., Lee, S. W., & Suen, C. Y. (1992). Thinning methodologies-a comprehensive survey. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 14(9), 869-885.
- [99] Zhang, T. Y., & Suen, C. Y. (1984). A fast parallel algorithm for thinning digital patterns. *Communications of the ACM*, 27(3), 236-239.
- [100] Sharafeddin, H., Parnianpour, M., Hemami, H., Hanson, T., Goldman, S., & Madson, T. (1996, March). Computer aided diagnosis of low back disorders using the motion profile. In *Proceedings of the 1996 Fifteenth Southern Biomedical Engineering Conference* (pp. 431-432). IEEE.
- [101] Sanchez-Marin, F. J. (2000). Automatic recognition of biological shapes with and without representations of shape. *Artificial Intelligence in Medicine*, 18(2), 173-186.
- [102] Rangayyan, R. M., El-Faramawy, N. M., Desautels, J. L., & Alim, O. A. (1997). Measures of acutance and shape for classification of breast tumors. *IEEE Transactions on medical imaging*, 16(6), 799-810.
- [103] Granlund, G. H. (1976). Identification of human chromosomes by using integrated density profiles. *IEEE Transactions on Biomedical Engineering*, (3), 182-192.
- [104] Delibasis, K., Undrill, P. E., & Cameron, G. G. (1997). Designing Fourier descriptor-based geometric models for object interpretation in medical images using genetic algorithms. *Computer Vision and*

Image Understanding, 66(3), 286-300.

- [105] Haris, K., Efstratiadis, S. N., Maglaveras, N., Pappas, C., Gourassas, J., & Louridas, G. (1999). Model-based morphological segmentation and labeling of coronary angiograms. *IEEE Transactions on Medical Imaging*, 18(10), 1003-1015.
- [106] Kumar, S., Asari, K. V., & Radhakrishnan, D. (1999). Real-time automatic extraction of lumen region and boundary from endoscopic images. *Medical & biological engineering & computing*, 37, 600-604.
- [107] Gregson, P. H., Shen, Z., Scott, R. C., & Kozousek, V. (1995). Automated grading of venous beading. *Computers and Biomedical Research*, 28(4), 291-304.
- [108] Wiffen, L. N., & Kent, J. T. (1993, September). The automatic classification of normal and abnormal chromosomes using image analysis. In *International Conference on Computer Analysis of Images and Patterns* (pp. 693-697). Berlin, Heidelberg: Springer Berlin Heidelberg.
- [109] Deng, W., Iyengar, S. S., & Brener, N. E. (2000). A fast parallel thinning algorithm for the binary image skeletonization. *The International Journal of High Performance Computing Applications*, 14(1), 65-81.
- [110] Xu, M., & Pycock, D. (1999). A scale-space medialness transform based on boundary concordance voting. *Journal of Mathematical Imaging and Vision*, 11, 277-299.
- [111] Morse, B. S., Pizer, S. M., & Liu, A. (1994). Multiscale medial analysis of medical images. *Image and Vision Computing*, 12(6), 327-338.
- [112] Dufresne, T. E., Sarwal, A., & Dhawan, A. P. (1994). A gray-level thinning method for delineation and representation of arteries. *Computerized medical imaging and graphics*, 18(5), 343-355.
- [113] Näf, M., Székely, G., Kikinis, R., Shenton, M. E., & Kübler, O. (1997). 3D Voronoi skeletons and their usage for the characterization and recognition of 3D organ shape. *Computer Vision and Image Understanding*, 66(2), 147-161.
- [114] Palagyi, K., & Kuba, A. (1998). A hybrid thinning algorithm for 3D medical images. *Journal of Computing and Information technology*, 6(2), 149-164.
- [115] Zhou, Y., & Toga, A. W. (1999). Efficient skeletonization of volumetric objects. *IEEE Transactions on visualization and computer graphics*, 5(3), 196-209.
- [116] Pavlopoulos, S., Kyriacou, E., Koutsouris, D., Blekas, K., Stafylopatis, A., & Zoumpoulis, P. (2000). Fuzzy neural network-based texture analysis of ultrasonic images. *IEEE Engineering in Medicine and Biology Magazine*, 19(1), 39-47.
- [117] Haralick, R. M. (1979). Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5), 786-804.
- [118] Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6), 610-621.
- [119] Handels, H., Ross, T., Kreusch, J., Wolff, H. H., & Pöpl, S. J. (1999). Computer-supported diagnosis of melanoma in profilometry. *Methods of information in medicine*, 38(01), 43-49.
- [120] Valckx, F. M., & Thijssen, J. M. (1997). Characterization of echographic image texture by cooccurrence matrix parameters. *Ultrasound in medicine & biology*, 23(4), 559-571.
- [121] Tsai, D. Y., Watanabe, S., & Tomita, M. (1996, September). Computerized analysis for classification of heart diseases in echocardiographic images. In *Proceedings of 3rd IEEE International Conference on Image Processing* (Vol. 3, pp. 283-286). IEEE.
- [122] Basset, O., Sun, Z., Mestas, J. L., & Gimenez, G. (1993). Texture analysis of ultrasonic images of the prostate by means of co-occurrence matrices. *Ultrasonic Imaging*, 15(3), 218-237.
- [123] Wu, C. M., Chen, Y. C., & Hsieh, K. S. (1992). Texture features for classification of ultrasonic liver images. *IEEE Transactions on medical imaging*, 11(2), 141-152.
- [124] Thiele, D. L., Kimme-Smith, C., Johnson, T. D., McCombs, M., & Bassett, L. W. (1996). Using tissue texture surrounding calcification clusters to predict benign vs malignant outcomes. *Medical Physics*, 23(4), 549-555.

- [125] Garra, B. S., Krasner, B. H., Horii, S. C., Ascher, S., Mun, S. K., & Zeman, R. K. (1993). Improving the distinction between benign and malignant breast lesions: the value of sonographic texture analysis. *Ultrasonic imaging*, 15(4), 267-285.
- [126] Parkkinen, J., Selkäinaho, K., & Oja, E. (1990). Detecting texture periodicity from the cooccurrence matrix. *Pattern Recognition Letters*, 11(1), 43-50.
- [127] Galloway, M. M. (1975). Texture analysis using gray level run lengths. *Computer graphics and image processing*, 4(2), 172-179.
- [128] Tang, X. (1998). Texture information in run-length matrices. *IEEE transactions on image processing*, 7(11), 1602-1609.
- [129] Gangeh, M. J., Sørensen, L., Shaker, S. B., Kamel, M. S., & De Bruijne, M. (2011). Multiple classifier systems in texton-based approach for the classification of CT images of lung. In *Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging: International MICCAI Workshop, MCV 2010, Beijing, China, September 20, 2010, Revised Selected Papers 1* (pp. 153-163). Springer Berlin Heidelberg.
- [130] Uppaluri, R., Mitsa, T., Sonka, M., Hoffman, E. A., & McLennan, G. (1997). Quantification of pulmonary emphysema from lung computed tomography images. *American journal of respiratory and critical care medicine*, 156(1), 248-254.
- [131] Sluimer, I. C., Prokop, M., Hartmann, I., & van Ginneken, B. (2006). Automated classification of hyperlucency, fibrosis, ground glass, solid, and focal lesions in high-resolution CT of the lung. *Medical Physics*, 33(7Part1), 2610-2620.
- [132] Chabat, F., Yang, G. Z., & Hansell, D. M. (2003). Obstructive lung diseases: texture classification for differentiation at CT. *Radiology*, 228(3), 871-877.
- [133] Xu, Y., Sonka, M., McLennan, G., Guo, J., & Hoffman, E. A. (2006). MDCT-based 3-D texture classification of emphysema and early smoking related lung pathologies. *IEEE transactions on medical imaging*, 25(4), 464-475.
- [134] Ojala, T., Pietikainen, M., & Maenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7), 971-987.
- [135] Varma, M., & Zisserman, A. (2008). A statistical approach to material classification using image patch exemplars. *IEEE transactions on pattern analysis and machine intelligence*, 31(11), 2032-2047.
- [136] D. G. Ballard, C. M. Brown, *Computer Vision*, Englewood Cliffs: Prentice Hall, 1982.
- [137] Bankman, I. N., Nizialek, T., Simon, I., Gatewood, O. B., Weinberg, I. N., & Brody, W. R. (1997). Segmentation algorithms for detecting microcalcifications in mammograms. *IEEE Transactions on Information Technology in Biomedicine*, 1(2), 141-149.
- [138] Bezdek, J. C., Hall, L. O., & Clarke, L. (1993). Review of MR image segmentation techniques using pattern recognition. *Medical physics*, 20(4), 1033-1048.
- [139] Bomans, M., Hohne, K. H., Tiede, U., & Riemer, M. (1990). 3-D segmentation of MR images of the head for 3-D display. *IEEE transactions on medical imaging*, 9(2), 177-183.
- [140] Brandt, M. E., Bohant, T. P., Kramer, L. A., & Fletcher, J. M. (1994). Estimation of CSF, white and gray matter volumes in hydrocephalic children using fuzzy clustering of MR images. *Computerized medical imaging and graphics*, 18(1), 25-34.
- [141] Brenner, J. F., Lester, J. M., & Selles, W. D. (1981). Scene segmentation in automated histopathology: techniques evolved from cytology automation. *Pattern Recognition*, 13(1), 65-77.
- [142] Brzakovic, D., Luo, X. M., & Brzakovic, P. (1990). An approach to automated detection of tumors in mammograms. *IEEE Transactions on Medical Imaging*, 9(3), 233-241.
- [143] K. R. Castleman, *Digital Image Processing*. Upper Saddle River: Prentice Hall, 1996.
- [144] C. K. Chow, T. Kaneko, "Automatic boundary detection of the left ventricle from cineangiograms,"

- Comput. Biomed. Res., vol. 5, pp. 388–410, 1972.
- [146] Clarke, L. P., Velthuizen, R. P., Camacho, M. A., Heine, J. J., Vaidyanathan, M., Hall, L. O., ...&Silbiger, M. L. (1995). MRI segmentation: methods and applications. *Magnetic resonance imaging*, 13(3), 343-368.
- [147] Cline, H. E., Dumoulin, C. L., Hart Jr, H. R., Lorensen, W. E., &Ludke, S. (1987). 3D reconstruction of the brain from magnetic resonance images using a connectivity algorithm. *Magnetic Resonance Imaging*, 5(5), 345-352.
- [148] E. R. Davies, *Machine Vision*, San Diego: Academic Press, 1997.
- [149] Fletcher, L. M., Barsotti, J. B., &Hornak, J. P. (1993). A multispectral analysis of brain tissues. *Magnetic Resonance in Medicine*, 29(5), 623-630.
- [150] Fu, K. S., &Mui, J. K. (1981). A survey on image segmentation. *Pattern recognition*, 13(1), 3-16.
- [151] Golston, J. E., Moss, R. H., &Stoecker, W. V. (1990). Boundary detection in skin tumor images: an overall approach and a radial search algorithm. *Pattern Recognition*, 23(11), 1235-1247.
- [152] Goshtasby, A., &Turner, D. A. (1995). Segmentation of cardiac cine MR images for extraction of right and left ventricular chambers. *IEEE transactions on medical imaging*, 14(1), 56-64.
- [153] Haralick, R. M., &Shapiro, L. G. (1985). Image segmentation techniques. *Computer vision, graphics, and image processing*, 29(1), 100-132.
- [154] Heine, J. J., Deans, S. R., Cullers, D. K., Stauduhar, R., & Clarke, L. P. (1997). Multiresolution statistical analysis of high-resolution digital mammograms. *IEEE Transactions on medical imaging*, 16(5), 503-515.
- [155] B. Jahne, *Practical Handbook on Image Processing for Scientific Applications*, Boca Raton, FL: CRC Press, 1997. Jahne, B., &Jaehne, B. (1995). *Practical handbook on image processing for scientific applications*. CRC Press, Inc..
- [156] Jain, A. K. (1989). *Fundamentals of digital image processing*. Prentice-Hall, Inc..
- [157] Johnson, L. A., Pearlman, J. D., Miller, C. A., Young, T. I., &Thulborn, K. R. (1993). MR quantification of cerebral ventricular volume using a semiautomated algorithm. *American journal of neuroradiology*, 14(6), 1373-1378.
- [158] Kikinis, R., Shenton, M. E., Iosifescu, D. V., McCarley, R. W., Saiviroonporn, P., Hokama, H. H., ... &Jolesz, F. A. (1996). A digital brain atlas for surgical planning, model-driven segmentation, and teaching. *IEEE Transactions on visualization and computer graphics*, 2(3), 232-241.
- [159] Kittler, J., Illingworth, J., &Föglein, J. (1985). Threshold selection based on a simple image statistic. *Computer vision, graphics, and image processing*, 30(2), 125-147.
- [160] Lester, J. M., Brenner, J. F., & Selles, W. D. (1980). Local transforms for biomedical image analysis. *Computer Graphics and Image Processing*, 13(1), 17-30.
- [161] Liang, Z. (1993). Tissue classification and segmentation of MR images. *IEEE Engineering in Medicine and Biology Magazine*, 12(1), 81-85.
- [162] F. A. Lucas-Quesada, U. Sinha, S. Sinha, "Segmentation strategies for breast tumors from dynamic MR images," *JMRI*, vol. 6, pp. 753–763, 1996.
- [163] Mitiche, A., &Aggarwal, J. K. (1985). Image segmentation by conventional and information-integrating techniques: a synopsis. *Image and Vision Computing*, 3(2), 50-62.
- [164] Pal, N. R., & Pal, S. K. (1993). A review on image segmentation techniques. *Pattern recognition*, 26(9), 1277-1294.
- [165] Rajapakse, J. C., Giedd, J. N., &Rapoport, J. L. (1997). Statistical approach to segmentation of single-channel cerebral MR images. *IEEE transactions on medical imaging*, 16(2), 176-186.
- [166] Rajapakse, J. C., Giedd, J. N., DeCarli, C., Snell, J. W., McLaughlin, A., Vauss, Y. C., ... &Rapoport, J. L. (1996). A technique for single-channel MR brain tissue segmentation: application to a pediatric sample. *Magnetic resonance imaging*, 14(9), 1053-1065.
- [167] Reddick, W. E., Glass, J. O., Cook, E. N., Elkin, T. D., & Deaton, R. J. (1997). Automated

- segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks. *IEEE Transactions on medical imaging*, 16(6), 911-918.
- [168] Rogowska, J., Preston, K., Hunter, G. J., Hamberg, L. M., Kwong, K. K., Salonen, O., & Wolf, G. L. (1995). Applications of similarity mapping in dynamic MRI. *IEEE Transactions on medical imaging*, 14(3), 480-486.
- [169] A. Rosenfeld, A. C. Kak, Digital Image Processing. New York: Academic Press, 1982.
- [170] Sahoo, P. K., Soltani, S. A. K. C., & Wong, A. K. (1988). A survey of thresholding techniques. *Computer vision, graphics, and image processing*, 41(2), 233-260.
- [171] Schmid, P. (1999). Segmentation of digitized dermatoscopic images by two-dimensional color clustering. *IEEE Transactions on Medical Imaging*, 18(2), 164-171.
- [172] Shareef, N., Wang, D. L., & Yagel, R. (1999). Segmentation of medical images using LEGION. *IEEE transactions on medical imaging*, 18(1), 74-91.
- [173] Sonka, M., Reddy, G. K., Winniford, M. D., & Collins, S. M. (1997). Adaptive approach to accurate analysis of small-diameter vessels in cineangiograms. *IEEE transactions on medical imaging*, 16(1), 87-95.
- [174] M. Sonka, V. Hlavac, R. Boyle, Image Processing, Analysis, and Machine Vision. CA: PWS Publishing, Pacific Grove, 1999.
- [175] D. Spielman, M. Sidhu, R. Herfkens, L. Shortlife, "Correlation imaging of the kidney," Proceedings of the International SMRM Conference, Nice, France, p. 373, 1995.
- [176] Suetens, P., Bellon, E., Vandermeulen, D., Smet, M., Marchal, G., Nuyts, J., & Mortelmans, L. (1993). Image segmentation: methods and applications in diagnostic radiology and nuclear medicine. *European journal of radiology*, 17(1), 14-21.
- [177] Udupa, J. K., Wei, L., Samarasekera, S., Miki, Y., van Buchem, M. A., & Grossman, R. I. (1997). Multiple sclerosis lesion quantification using fuzzy-connectedness principles. *IEEE Transactions on medical imaging*, 16(5), 598-609.
- [178] Vannier, M. W., Butterfield, R. L., Jordan, D., Murphy, W. A., Levitt, R. G., & Gado, M. (1985). Multispectral analysis of magnetic resonance images. *Radiology*, 154(1), 221-224.
- [179] Weszka, J. S. (1978). A survey of threshold selection techniques. *Computer graphics and image processing*, 7(2), 259-265.
- [180] Zhu, Y., & Yan, Z. (1997). Computerized tumor boundary detection using a Hopfield neural network. *IEEE transactions on medical imaging*, 16(1), 55-67.
- [181] Brenner, J. F., Lester, J. M., & Selles, W. D. (1981). Scene segmentation in automated histopathology: techniques evolved from cytology automation. *Pattern Recognition*, 13(1), 65-77.
- [182] Lester, J. M., Brenner, J. F., & Selles, W. D. (1980). Local transforms for biomedical image analysis. *Computer Graphics and Image Processing*, 13(1), 17-30.
- [183] Jahne, B., & Jaehne, B. (1995). *Practical handbook on image processing for scientific applications*. CRC Press, Inc..
- [184] Lim, J. S. (1990). Two-dimensional signal and image processing. *Englewood Cliffs*.
- [185] Westin, C. F., Warfield, S., Bhalerao, A., Mui, L., Richolt, J., & Kikinis, R. (1998). Tensor controlled local structure enhancement of CT images for bone segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI'98: First International Conference Cambridge, MA, USA, October 11–13, 1998 Proceedings 1* (pp. 1205-1212). Springer Berlin Heidelberg.
- [186] Adams, R., & Bischof, L. (1994). Seeded region growing. *IEEE Transactions on pattern analysis and machine intelligence*, 16(6), 641-647.
- [187] Sowmya, A., & Trinder, J. (2000). Modelling and representation issues in automated feature extraction from aerial and satellite images. *ISPRS journal of photogrammetry and remote sensing*, 55(1), 34-47.
- [188] W. K. Pratt, Digital Image Processing. New York: John Wiley & Sons, 1991.

- [189] Shen, L., Rangayyan, R. M., & Desautels, J. L. (1993). Detection and classification of mammographic calcifications. *International Journal of Pattern Recognition and Artificial Intelligence*, 7(06), 1403-1416.
- [190] Singleton, H. R., & Pohost, G. M. (1997). Automatic cardiac MR image segmentation using edge detection by tissue classification in pixel neighborhoods. *Magnetic resonance in medicine*, 37(3), 418-424.
- [191] Torre, V., & Poggio, T. A. (1986). On edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (2), 147-163.
- [192] Umbaugh, S. E. (1997). *Computer vision and image processing: a practical approach using cviptools with cdrom*. Prentice Hall PTR.
- [193] T. Pavlidis, *Structural Pattern Recognition*, Berlin: Springer-Verlag, 1977.
- [194] Zucker, S. W. (1976). Region growing: Childhood and adolescence. *Computer graphics and image processing*, 5(3), 382-399.
- [196] Cline, H. E., Dumoulin, C. L., Hart Jr, H. R., Lorensen, W. E., & Ludke, S. (1987). 3D reconstruction of the brain from magnetic resonance images using a connectivity algorithm. *Magnetic Resonance Imaging*, 5(5), 345-352.
- [197] Meyer, F., & Beucher, S. (1990). Morphological segmentation. *Journal of visual communication and image representation*, 1(1), 21-46.
- [198] Vincent, L., & Soille, P. (1991). Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 13(06), 583-598.
- [199] Vincent, L. (1993). Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. *IEEE transactions on image processing*, 2(2), 176-201.
- [200] Higgins, W. E., & Ojard, E. J. (1993). Interactive morphological watershed analysis for 3D medical images. *Computerized medical imaging and graphics*, 17(4-5), 387-395.
- [201] Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 207(1167), 187-217.
- [202] Wang, J. Z., Turner, D. A., & Chutuape, M. D. (1992). Fast, interactive algorithm for segmentation of a series of related images: Application to volumetric analysis of MR images of the heart. *Journal of Magnetic Resonance Imaging*, 2(5), 575-582.
- [203] Lester, J. M., Williams, H. A., Weintraub, B. A., & Brenner, J. F. (1978). Two graph searching techniques for boundary finding in white blood cell images. *Computers in Biology and Medicine*, 8(4), 293-308.
- [204] Nilsson, N. J. (1971). Problem-solving methods in. *Artificial Intelligence*, 5.
- [205] Jahne, B., & Jaehne, B. (1995). *Practical handbook on image processing for scientific applications*. CRC Press, Inc..
- [206] Jain, A. K. (1989). *Fundamentals of digital image processing*. Prentice-Hall, Inc..
- [207] Johnson, L. A., Pearlman, J. D., Miller, C. A., Young, T. I., & Thulborn, K. R. (1993). MR quantification of cerebral ventricular volume using a semiautomated algorithm. *American journal of neuroradiology*, 14(6), 1373-1378.
- [208] Kikinis, R., Shenton, M. E., Iosifescu, D. V., McCarley, R. W., Saiviroonporn, P., Hokama, H. H., ... & Jolesz, F. A. (1996). A digital brain atlas for surgical planning, model-driven segmentation, and teaching. *IEEE Transactions on visualization and computer graphics*, 2(3), 232-241.
- [155] Kittler, J., Illingworth, J., & Föglein, J. (1985). Threshold selection based on a simple image statistic. *Computer vision, graphics, and image processing*, 30(2), 125-147.
- [156] Lester, J. M., Brenner, J. F., & Selles, W. D. (1980). Local transforms for biomedical image analysis. *Computer Graphics and Image Processing*, 13(1), 17-30.

- [157] Liang, Z. (1993). Tissue classification and segmentation of MR images. *IEEE Engineering in Medicine and Biology Magazine*, 12(1), 81-85.
- [158] Lucas-Quesada, F. A., Sinha, U., & Sinha, S. (1996). Segmentation strategies for breast tumors from dynamic MR images. *Journal of Magnetic Resonance Imaging*, 6(5), 753-763.
- [159] Mitiche, A., & Aggarwal, J. K. (1985). Image segmentation by conventional and information-integrating techniques: a synopsis. *Image and Vision Computing*, 3(2), 50-62.
- [160] Pal, N. R., & Pal, S. K. (1993). A review on image segmentation techniques. *Pattern recognition*, 26(9), 1277-1294.
- [161] Rajapakse, J. C., Giedd, J. N., & Rapoport, J. L. (1997). Statistical approach to segmentation of single-channel cerebral MR images. *IEEE transactions on medical imaging*, 16(2), 176-186.
- [162] Rajapakse, J. C., Giedd, J. N., DeCarli, C., Snell, J. W., McLaughlin, A., Vauss, Y. C., ... & Rapoport, J. L. (1996). A technique for single-channel MR brain tissue segmentation: application to a pediatric sample. *Magnetic resonance imaging*, 14(9), 1053-1065.
- [163] Reddick, W. E., Glass, J. O., Cook, E. N., Elkin, T. D., & Deaton, R. J. (1997). Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks. *IEEE Transactions on medical imaging*, 16(6), 911-918.
- [164] Rogowska, J., Preston, K., Hunter, G. J., Hamberg, L. M., Kwong, K. K., Salonen, O., & Wolf, G. L. (1995). Applications of similarity mapping in dynamic MRI. *IEEE Transactions on medical imaging*, 14(3), 480-486.
- [165] Roerdink, J. B. T. M. (1998). *An introduction to digital image processing* (pp. 175-196). John Wiley & Sons, Ltd., England.
- [166] Sahoo, P. K., Soltani, S., Wong, A. K. C., & Chen, Y. C. (1988). Comp. Vision Graph. In *Image Proc* (Vol. 41, p. 233).
- [167] Schmid, P. (1999). Segmentation of digitized dermatoscopic images by two-dimensional color clustering. *IEEE Transactions on Medical Imaging*, 18(2), 164-171.
- [168] Shareef, N., Wang, D. L., & Yagel, R. (1999). Segmentation of medical images using LEGION. *IEEE transactions on medical imaging*, 18(1), 74-91.
- [169] Sonka, M., Reddy, G. K., Winniford, M. D., & Collins, S. M. (1997). Adaptive approach to accurate analysis of small-diameter vessels in cineangiograms. *IEEE transactions on medical imaging*, 16(1), 87-95.
- [170] Mendoza, F., & Lu, R. (2015). Basics of image analysis. *Hyperspectral imaging technology in food and agriculture*, 9-56.
- [171] D. Spielman, M. Sidhu, R. Herfkens, L. Shortlife, "Correlation imaging of the kidney," Proceedings of the International SMRM Conference, Nice, France, p. 373, 1995.
- [172] Suetens, P., Bellon, E., Vandermeulen, D., Smet, M., Marchal, G., Nuyts, J., & Mortelmans, L. (1993). Image segmentation: methods and applications in diagnostic radiology and nuclear medicine. *European journal of radiology*, 17(1), 14-21.
- [173] Udupa, J. K., Wei, L., Samarasekera, S., Miki, Y., van Buchem, M. A., & Grossman, R. I. (1997). Multiple sclerosis lesion quantification using fuzzy-connectedness principles. *IEEE Transactions on medical imaging*, 16(5), 598-609.
- [174] Vannier, M. W., Butterfield, R. L., Jordan, D., Murphy, W. A., Levitt, R. G., & Gado, M. (1985). Multispectral analysis of magnetic resonance images. *Radiology*, 154(1), 221-224.
- [175] Weszka, J. S. (1978). A survey of threshold selection techniques. *Computer graphics and image processing*, 7(2), 259-265.
- [176] Zhu, Y., & Yan, Z. (1997). Computerized tumor boundary detection using a Hopfield neural network. *IEEE transactions on medical imaging*, 16(1), 55-67.
- [177] Brenner, J. F., Lester, J. M., & Selles, W. D. (1981). Scene segmentation in automated histopathology: techniques evolved from cytology automation. *Pattern Recognition*, 13(1), 65-77.

- [178] Lester, J. M., Brenner, J. F., & Selles, W. D. (1980). Local transforms for biomedical image analysis. *Computer Graphics and Image Processing*, 13(1), 17-30.
- [179] Jahne, B., & Jaehne, B. (1995). *Practical handbook on image processing for scientific applications*. CRC Press, Inc..
- [180] Lim, J. S. (1990). Two-dimensional signal and image processing. *Englewood Cliffs*.
- [181] Westin, C. F., Warfield, S., Bhalerao, A., Mui, L., Richolt, J., & Kikinis, R. (1998). Tensor controlled local structure enhancement of CT images for bone segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI'98: First International Conference Cambridge, MA, USA, October 11–13, 1998 Proceedings 1* (pp. 1205-1212). Springer Berlin Heidelberg.
- [182] Adams, R., & Bischof, L. (1994). Seeded region growing. *IEEE Transactions on pattern analysis and machine intelligence*, 16(6), 641-647.
- [183] D. G. Ballard, C. M. Brown, *Computer Vision*, Englewood Cliffs: Prentice Hall, 1982.
- [184] W. K. Pratt, *Digital Image Processing*. New York: John Wiley & Sons, 1991.
- [185] Shen, L., Rangayyan, R. M., & Desautels, J. L. (1993). Detection and classification of mammographic calcifications. *International Journal of Pattern Recognition and Artificial Intelligence*, 7(06), 1403-1416.
- [186] Singleton, H. R., & Pohost, G. M. (1997). Automatic cardiac MR image segmentation using edge detection by tissue classification in pixel neighborhoods. *Magnetic resonance in medicine*, 37(3), 418-424.
- [187] Torre, V., & Poggio, T. A. (1986). On edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (2), 147-163.
- [188] Umbaugh, S. E. (1997). *Computer vision and image processing: a practical approach using cviptools with cdrom*. Prentice Hall PTR. Umbaugh, S. E. (1997). *Computer vision and image processing: a practical approach using cviptools with cdrom*. Prentice Hall PTR.
- [189] Pavlidis, T. (2013). *Structural pattern recognition* (Vol. 1). Springer.
- [190] Zucker, S. W. (1976). Region growing: Childhood and adolescence. *Computer graphics and image processing*, 5(3), 382-399.
- [191] Cline, H. E., Dumoulin, C. L., Hart Jr, H. R., Lorensen, W. E., & Ludke, S. (1987). 3D reconstruction of the brain from magnetic resonance images using a connectivity algorithm. *Magnetic Resonance Imaging*, 5(5), 345-352.
- [192] Meyer, F., & Beucher, S. (1990). Morphological segmentation. *Journal of visual communication and image representation*, 1(1), 21-46.
- [193] Vincent, L., & Soille, P. (1991). Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 13(06), 583-598.
- [194] Vincent, L. (1993). Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. *IEEE transactions on image processing*, 2(2), 176-201.
- [196] Higgins, W. E., & Ojard, E. J. (1993). Interactive morphological watershed analysis for 3D medical images. *Computerized medical imaging and graphics*, 17(4-5), 387-395.
- [197] Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 207(1167), 187-217.
- [198] Wang, J. Z., Turner, D. A., & Chutuape, M. D. (1992). Fast, interactive algorithm for segmentation of a series of related images: Application to volumetric analysis of MR images of the heart. *Journal of Magnetic Resonance Imaging*, 2(5), 575-582.
- [199] Lester, J. M., Williams, H. A., Weintraub, B. A., & Brenner, J. F. (1978). Two graph searching techniques for boundary finding in white blood cell images. *Computers in Biology and Medicine*, 8(4), 293-308.
- [200] N. J. Nilsson, *Problem Solving Methods in Artificial Intelligence*. New York: McGraw-Hill, 1971.

- [201] Jahne, B., & Jaehne, B. (1995). *Practical handbook on image processing for scientific applications*. CRC Press, Inc..
- [202] Jain, A. K. (1989). *Fundamentals of digital image processing*. Prentice-Hall, Inc..
- [203] Johnson, L. A., Pearlman, J. D., Miller, C. A., Young, T. I., & Thulborn, K. R. (1993). MR quantification of cerebral ventricular volume using a semiautomated algorithm. *American journal of neuroradiology*, *14*(6), 1373-1378.
- [204] Kikinis, R., Shenton, M. E., Iosifescu, D. V., McCarley, R. W., Saiviroonporn, P., Hokama, H. H., ... & Jolesz, F. A. (1996). A digital brain atlas for surgical planning, model-driven segmentation, and teaching. *IEEE Transactions on visualization and computer graphics*, *2*(3), 232-241.
- [205] Sonka, M., Zhang, X., Siebes, M., Bissing, M. S., DeJong, S. C., Collins, S. M., & McKay, C. R. (1995). Segmentation of intravascular ultrasound images: A knowledge-based approach. *IEEE Transactions on Medical Imaging*, *14*(4), 719-732.
- [206] Thedens, D. R., Skorton, D. J., & Fleagle, S. R. (1995). Methods of graph searching for border detection in image sequences with applications to cardiac magnetic resonance imaging. *IEEE transactions on medical imaging*, *14*(1), 42-55.
- [207] Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *207*(1167), 187-217.
- [208] Sun, Y., Lucariello, R. J., & Chiaramida, S. A. (1995). Directional low-pass filtering for improved accuracy and reproducibility of stenosis quantification in coronary arteriograms. *IEEE Transactions on Medical Imaging*, *14*(2), 242-248.
- [209] Davis, L. S. (1975). A survey of edge detection techniques. *Computer graphics and image processing*, *4*(3), 248-270.
- [210] Askari, S. (2021). Fuzzy C-Means clustering algorithm for data with unequal cluster sizes and contaminated with noise and outliers: Review and development. *Expert Systems with Applications*, *165*, 113856.
- [211] Guyon, I., & Elisseeff, A. (2003). An introduction to feature and variable selection. *Journal of Machine Learning Research*, *3*(1157-1182), 100.
- [212] Zhou, S. K., Zhou, J., & Comaniciu, D. (2007, June). A boosting regression approach to medical anatomy detection. In *2007 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1-8). IEEE.
- [213] Fenchel, M., Thesen, S., & Schilling, A. (2008, September). Automatic labeling of anatomical structures in MR FastView images using a statistical atlas. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 576-584). Berlin, Heidelberg: Springer Berlin Heidelberg.
- [214] Criminisi, A., Shotton, J., & Bucciarelli, S. (2009, September). Decision forests with long-range spatial context for organ localization in CT volumes. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (pp. 69-80).
- [215] Criminisi, A., Robertson, D., Konukoglu, E., Shotton, J., Pathak, S., White, S., & Siddiqui, K. (2013). Regression forests for efficient anatomy detection and localization in computed tomography scans. *Medical image analysis*, *17*(8), 1293-1303.
- [216] Bhole, C., Morsillo, N., & Pal, C. (2011). 3D segmentation in CT imagery with conditional random fields and histograms of oriented gradients. In *Machine Learning in Medical Imaging: Second International Workshop, MLMI 2011, Held in Conjunction with MICCAI 2011, Toronto, Canada, September 18, 2011. Proceedings 2* (pp. 326-334). Springer Berlin Heidelberg.
- [217] Gangeh, M. J., Sørensen, L., Shaker, S. B., Kamel, M. S., & De Bruijne, M. (2011). Multiple classifier systems in texton-based approach for the classification of CT images of lung. In *Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging: International MICCAI Workshop, MCV 2010, Beijing, China, September 20, 2010, Revised Selected Papers 1* (pp. 153-

- 163). Springer Berlin Heidelberg.
- [218] Feulner, J., Zhou, S. K., Hammon, M., Hornegger, J., &Comaniciu, D. (2011). Segmentation based features for lymph node detection from 3-d chest ct. In *Machine Learning in Medical Imaging: Second International Workshop, MLMI 2011, Held in Conjunction with MICCAI 2011, Toronto, Canada, September 18, 2011. Proceedings 2* (pp. 91-99). Springer Berlin Heidelberg.
- [219] Conze, P. H., Rousseau, F., Noblet, V., Heitz, F., Memeo, R., &Pessaux, P. (2015). Semi-automatic liver tumor segmentation in dynamic contrast-enhanced CT scans using random forests and supervoxels. In *Machine Learning in Medical Imaging: 6th International Workshop, MLMI 2015, Held in Conjunction with MICCAI 2015, Munich, Germany, October 5, 2015, Proceedings 6* (pp. 212-219). Springer International Publishing.
- [220] Dikici, E., &Orderud, F. (2011). Maximum likelihood and James-Stein edge estimators for left ventricle tracking in 3d echocardiography. In *Machine Learning in Medical Imaging: Second International Workshop, MLMI 2011, Held in Conjunction with MICCAI 2011, Toronto, Canada, September 18, 2011. Proceedings 2* (pp. 43-50). Springer Berlin Heidelberg.
- [221] Domingos, J. S., Stebbing, R. V., Leeson, P., & Noble, J. A. (2014). Structured random forests for myocardium delineation in 3D echocardiography. In *Machine Learning in Medical Imaging: 5th International Workshop, MLMI 2014, Held in Conjunction with MICCAI 2014, Boston, MA, USA, September 14, 2014. Proceedings 5* (pp. 215-222). Springer International Publishing.
- [222] Zuluaga, M. A., Delgado Leyton, E. J., Hernández Hoyos, M., &Orkisz, M. (2011). Feature selection for SVM-based vascular anomaly detection. In *Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging: International MICCAI Workshop, MCV 2010, Beijing, China, September 20, 2010, Revised Selected Papers 1* (pp. 141-152). Springer Berlin Heidelberg
- [223] Liu S, Babbs CF, Delp EJ (2001) Multiresolution detection of speculated lesions in digital mammograms. *IEEE Trans Image Process* 10(6):874–884
- [224] Rashed EA, Ismail IA, Zaki SI (2007) Multiresolution mammogram analysis in multilevel decomposition. *Pattern RecognLett* 28(2):286–292
- [225] Yoon S, Kim S (2009) Mutual information-based SVM-RFE for diagnostic classification of digitized mammograms. *Pattern RecognLett* 30(16):1489–1495
- [226] Lladó X, Oliver A, Freixenet J, Martí R, Martí J (2009) A textural approach for mass false positive reduction in mammography. *Comput Med Imaging Graph* 33(6):415–422
- [227] BrazJr G, Cardoso de Paiva A, Corrêa Silva A, Cesar Muniz de Oliveira A (2009) Classification of breast tissues using Moran’s index and Geary’s coefficient as texture signatures and SVM. *ComputBiol Med* 39(12):1063–1072
- [228] Tahmasbi A, Saki F, Shokouhi SB (2011) Classification of benign and malignant masses based on Zernike moments. *ComputBiol Med* 41(8):726–735
- [229] Ramirez-Villegas JF, Ramirez-Moreno DF (2012) Wavelet packet energy, Tsallis entropy and statistical parameterization for support vector-based and neural-based classification of mammographic regions. *Neurocomputing* 77(1):82–100
- [230] Wu W-J, Lin S-W, Moon WK (2012) Combining support vector machine with genetic algorithm to classify ultrasound breast tumor images. *Comput Med Imaging Graph* 36(8):627–633
- [231] Wajid SK, Hussain A (2015) Local energy-based shape histogram feature extraction technique for breast cancer diagnosis. *Expert SystAppl* 42(20):6990–6999
- [232] Vaidehi K, Subashini TS (2015) Automatic characterization of benign and malignant masses in mammography. *ProcComputSci* 46:1762–1769
- [233] Zyout I, Czajkowska J, Grzegorzec M (2015) Multi-scale textural feature extraction and particle swarm optimization based model selection for false positive reduction in mammography. *Comput Med Imaging Graph* 46:95–107
- [234] de Oliveira FSS, de CarvalhoFilho AO, Silva AC, de Paiva AC, Gattass M (2015) Classification of

- breast regions as mass and non-mass based on digital mammograms using taxonomic indexes and SVM. *Comput Biol Med* 57:42–53
- [235] Subashini TS, Ramalingam V, Palanivel S (2009) Breast mass classification based on cytological patterns using RBFNN and SVM. *Expert SystAppl* 36(3):5284–5290
- [236] Beura S, Majhi B, Dash R (2015) Mammogram classification using two dimensional discrete wavelet transform and gray-level cooccurrence matrix for detection of breast cancer. *Neurocomputing* 154:1–14
- [237] Shi J, Zhou S, Liu X, Zhang Q, Lu M, Wang T (2016) Stacked deep polynomial network based representation learning for tumor classification with small ultrasound image dataset. *Neurocomputing* 194:87–94
- [238] Kong H, Lai Z, Wang X, Liu F (2016) Breast cancer discriminant feature analysis for diagnosis via jointly sparse learning. *Neurocomputing* 177:198–205
- [239] Xie W, Li Y, Ma Y (2016) Breast mass classification in digital mammography based on extreme learning machine. *Neurocomputing* 173:930–941
- [240] Khan MM, Mendes A, Zhang P, Chalup SK (2017) Evolving multidimensional wavelet neural networks for classification using Cartesian Genetic Programming. *Neurocomputing* 247:39–58
- [241] Soriano D, Aguilar C, Ramirez-Morales I, Tusa E, Rivas W, Pinta M (2017) Mammogram classification schemes by using convolutional neural networks. In: *International conference on technology trends*. Springer, pp 71–85
- [242] Jadoon MM, Zhang Q, Haq IU, Butt S, Jadoon A (2017) Three-class mammogram classification based on descriptive CNN features. *BioMed Res Int* 2017:3640901
- [243] Jain, R. (1991). *The art of computer systems performance analysis: techniques for experimental design, measurement, simulation, and modeling* (Vol. 1). New York: Wiley.
- [244] Trivedi, K. S. (2008). *Probability & statistics with reliability, queuing and computer science applications*. John Wiley & Sons.
- [245] ERBAŞ, M. K., AKTAŞ, G. R., ŞANVER, S., DEMİR, E., TOKTAŞ, İ., ALGIN, O., & OFLAZ, H. (2017). MR GÖRÜNTÜLERİNDEN BEYİN DAMARININ MODELLENMESİ VE 3 BOYUTLU BASKISI. *Karabük Üniversitesi Teknoloji Fakültesi Endüstriyel Tasarım Mühendisliği Bölümü www.3dexpoturkey.com*, 187(188), 627-630.
- [246] Chandy, K. M., & Martin, A. J. (1983). A characterization of product-form queuing networks. *Journal of the ACM (JACM)*, 30(2), 286-299.
- [247] Mat05. Mathematica: <http://www.wolfram.com/>. Last verified on E'ebruary 08, 2005, 2005.
- [248] C. Hirel, R. A. Sahner, X. Zarig, ant1 K. S. Trivedi Hirel, C., Sahner, R., Zang, X., & Trivedi, K. (2000). Reliability and performability modeling using SHARPE 2000. In *Computer Performance Evaluation. Modelling Techniques and Tools: 11th International Conference, TOOLS 2000 Schaumburg, IL, USA, March 27–31, 2000 Proceedings 11* (pp. 345-349). Springer Berlin Heidelberg. 2000. In *Proc. 11th International Conference on Computer Performance Evaluation: Modelling Techniques and Tools*, pages 345-349. Springer-Verlag, London, UK, 2000.
- [249] Hirel, C., Tuffin, B., & Trivedi, K. S. (2000). Spnp: Stochastic petri nets. version 6.0. In *Computer Performance Evaluation. Modelling Techniques and Tools: 11th International Conference, TOOLS 2000 Schaumburg, IL, USA, March 27–31, 2000 Proceedings 11* (pp. 354-357). Springer Berlin Heidelberg.
- [250] Zimmermann, A., Freiheit, J., German, R., & Hommel, G. (2000). Petri net modelling and performability evaluation with TimeNET 3.0. In *Computer Performance Evaluation. Modelling Techniques and Tools: 11th International Conference, TOOLS 2000 Schaumburg, IL, USA, March 27–31, 2000 Proceedings 11* (pp. 188-202). Springer Berlin Heidelberg.
- [251] Sahner, R. A., Trivedi, K., & Puliafito, A. (2012). *Performance and reliability analysis of computer systems: an example-based approach using the SHARPE software package*. Springer Science &

Business Media.

- [252] Sargent, R. G. (1994, December). A historical view of hybrid simulation/analytic models. In *Proceedings of Winter Simulation Conference* (pp. 383-386). IEEE.
- [253] Shanthikumar, J. G., & Sargent, R. G. (1983). A unifying view of hybrid simulation/analytic models and modeling. *Operations research*, *31*(6), 1030-1052.
- [254] Haverkort, B. R., & Trivedi, K. S. (1993). Specification techniques for Markov reward models. *Discrete Event Dynamic Systems*, *3*(2-3), 219-247.
- [255] Ciardo, G., & Trivedi, K. S. (1993). A decomposition approach for stochastic reward net models. *Performance Evaluation*, *18*(1), 37-59.
- [256] BOYD, M., Veeraraghavan, M., DUGAN, J., & TRIVEDI, K. (1988, October). An approach to solving large reliability models. In *Digital Avionics Systems Conference* (p. 3905).
- [257] A. Goyal, W. Carter, E. de Souza e Silva, S. Lavenberg, and K. S. Trivedi. The System Availability Estimator (SAVE). In *Proc. 16th Int. Symp. on Fault-Tolerant Computing, FTCS*, pages 84-89, Vienna, Austria, Los Alamitos, CA, July 1986. IEEE Computer Society Press.
- [258] e Silva, E. D. S., & Gail, H. R. (1992). Performability analysis of computer systems: from model specification to solution. *Performance Evaluation*, *14*(3-4), 157-196.
- [259] Wirth, N. (1971). Program development by stepwise refinement. *Communications of the ACM*, *14*(4), 221-227.
- [260] Morris, J. M. (1987). A theoretical basis for stepwise refinement and the programming calculus. *Science of Computer programming*, *9*(3), 287-306.
- [261] T. Naylor and J. Finger. Verification of Computer Simulation Models. *Management Science*, 14:92-101, October 1967.
- [262] D. Heimann, N. Mittal, and K. S. Trivedi. Dependability Modeling for Computer Systems. In *Proc. Annual Reliability and Maintainability Symposium*, pages 120-128, Orlando, FL, January 1991.
- [263] Baynat, B. (2000). *La théorie des files d'attente: des chaînes de Markov aux réseaux à forme produit* (p. 330). Hermès.
- [264] P. Brémaud. *Markov chains : Gibbs fields, Monte Carlo simulation and queues*. Texts in applied mathematics. Springer, New York, Berlin, Heidelberg, 1999.
- [265] R. B. Cooper. *Introduction to Queueing Theory*. North-Holland, New York, NY, second edition, 1981.
- [265a] Gherghout, Y., Tlili. (2013) Modeling of the machine learning algorithms by features/instances aggregation and selection. (N2C'13). Jaune 14-15, 2013. Hammamet, Tunisia
- [266] Bruell, S. C., & Balbo, G. (1980). *Computational algorithms for closed queueing networks* (pp. 1-206). Elsevier North-Holland.
- [267] Chandy, K. M. (1972, March). The analysis and solutions for general queueing networks. In *Proceedings of the Sixth Annual Princeton Conference on Information Sciences and Systems* (pp. 224-228).
- [268] Jackson, J. R. (1963). Jobshop-like queueing systems. *Management science*, *10*(1), 131-142.
- [269] Gordon, W. J., & Newell, G. F. (1967). Closed queueing systems with exponential servers. *Operations research*, *15*(2), 254-265.
- [270] Bolch, G., Greiner, S., De Meer, H., & Trivedi, K. S. (2006). *Queueing networks and Markov chains: modeling and performance evaluation with computer science applications*. John Wiley & Sons.
- [271] Baskett, F., Chandy, K. M., Muntz, R. R., & Palacios, F. G. (1975). Open, closed, and mixed networks of queues with different classes of customers. *Journal of the ACM (JACM)*, *22*(2), 248-260.
- [272] Chandy, K. M., Howard Jr, J. H., & Towsley, D. F. (1977). Product form and local balance in queueing networks. *Journal of the ACM (JACM)*, *24*(2), 250-263.
- [273] Dash, M., & Liu, H. (2003). Consistency-based search in feature selection. *Artificial intelligence*, *151*(1-2), 155-176.
- [274]

Liu, H., & Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on knowledge and data engineering*, 17(4), 491-502.

- [275] Gherghout Y, Souici-Meslati L, "Vers une sélection automatique de caractéristiques pour la reconnaissance de l'écriture arabe", Journée des jeunes chercheurs en informatique, 2008.
- [276] Chen, K., & Liu, H. (1999, July). Towards an evolutionary algorithm: a comparison of two feature selection algorithms. In *Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406)* (Vol. 2, pp. 1309-1313). IEEE.
- [277] Pudil, P., Novovičová, J., & Kittler, J. (1994). Floating search methods in feature selection. *Pattern recognition letters*, 15(11), 1119-1125.
- [278] Pudil, P., Ferri, F. J., Novovicova, J., & Kittler, J. (1994, October). Floating search methods for feature selection with nonmonotonic criterion functions. In *Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol. 3-Conference C: Signal Processing (Cat. No. 94CH3440-5)* (Vol. 2, pp. 279-283). IEEE.
- [279] Ferri, F. J., Pudil, P., Hatef, M., & Kittler, J. (1994). Comparative study of techniques for large-scale feature selection. In *Machine intelligence and pattern recognition* (Vol. 16, pp. 403-413). North-Holland.
- [280] Jain, A., & Zongker, D. (1997). Feature selection: Evaluation, application, and small sample performance. *IEEE transactions on pattern analysis and machine intelligence*, 19(2), 153-158.
- [281] Kudo, M., & Sklansky, J. (2000). Comparison of algorithms that select features for pattern classifiers. *Pattern recognition*, 33(1), 25-41.
- [282] Hall, M. A. (1999). *Correlation-based feature selection for machine learning* (Doctoral dissertation, The University of Waikato).
- [283] Pal, D. S. K., Gu, B. H. K. J., & Kim, H. K. T. H. (1998). Image Processing, and Pattern Recognition.
- [284] Bennani, Y. (2001). Systèmes d'apprentissage connexionnistes: Sélection de variables. *Revue d'intelligence artificielle*, 15(3-4), 303-316.
- [285] Cancer IA for R. on (2012) GLOBOCAN 2012: estimated cancer incidence, mortality and prevalence worldwide in 2012
- [286] Tao Z, Shi A, Lu C, Song T, Zhang Z, Zhao J (2015) Breast cancer: epidemiology and etiology. *Cell BiochemBiophys* 72(2):333–338
- [287] Report (2017) American Cancer Society, 2017. Breast cancer facts and figures. 2017–2018, Atlanta
- [288] Bick U, Diekmann F (2010) Digital mammography. Springer Science & Business Media, New York
- [289] Teertstra HJ, Loo CE, van den Bosch MAAJ, van Tinteren H, Rutgers EJT, Muller SH, Gilhuijs KGA (2010) Breast tomosynthesis in clinical practice: initial results. *EurRadiol* 20(1):16–24
- [290] Freer TW, Ulissey MJ (2001) Screening mammography with computeraided detection: prospective study of 12,860 patients in a community breast center. *Radiology* 220(3):781–786
- [291] Angelov P, Gu X (2017) MICE: multi-layer multi-model images classifier ensemble. In: Cybernetics (CYBCONF), 2017 3rd IEEE international conference on. IEEE, pp 1–8
- [292] Angelov PP, Gu X (2018) Deep rule-based classifier with human-level performance and characteristics. *InfSci* 463–464:196–213
- [293] Liu X, Liu J, Zhou D, Tang J (2010) A benign and malignant mass classification algorithm based on an improved level set segmentation and texture feature analysis. In: Science and technology, IEEE, pp 0–3
- [294] Guliato D, Rangayyan RM, Carvalho JD, Santiago SA (2008) Polygonal modeling of contours of breast tumors with the preservation of spicules. *IEEE Trans Biomed Eng* 55(1):14–20
- [295] Freeman H (1961) On the encoding of arbitrary geometric configurations. *IRE Trans Electron*

Comput 2:260–268

- [296] Peng H, Long F, Ding C (2005) Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell* 27(8):1226–1238
- [297] Suckling J, Parker J, Dance D, Astley S, Hutt I, Boggis C, Ricketts I, Stamatakis E, Cerneaz N, Kok S, Taylor P, Betal D, Savage J (1994) The mammographic image analysis society digital mammogram database. In: *Excerptamedica. International congress series, vol 1069*. Elsevier, Amsterdam, pp 375–378
- [298] Palm RB (2012) Prediction as a candidate for learning deep hierarchical models of data, vol 5. Technical University of Denmark
- [299] Angelov P, Ramezani R, Zhou X (2008) Autonomous novelty detection and object tracking in video streams using evolving clustering and Takagi-Sugeno type neuro-fuzzy system. In: *Neural networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence)*. IEEE international joint conference on. IEEE, pp 1456–1463
- [300] Angelov P, Sadeghi-Tehran P, Ramezani R (2011) An approach to automatic real-time novelty detection, object identification, and tracking in video streams based on recursive density estimation and evolving Takagi–Sugeno fuzzy systems. *Int J IntellSyst* 26(3):189–205
- [301] Angelov P, Yager R (2013) Density-based averaging—a new operator for data fusion. *InfSci* 222:163–174