

# وزارة التعليم العالي والبحث العلمي

UNIVERSITÉ de BADJI MOKHTAR –  
ANNABA



جامعة باجي مختار – عنابة

Faculté Sciences de l'Ingénierat  
Département d'Informatique

## Thèse

Présentée en vue de l'obtention du diplôme de  
Doctorat 3<sup>ème</sup> cycle

## Intitulée

# Approches Ensemblistes de Classification et Sélection des Prédicteurs pour la Reconnaissance de Formes

**Domaine :** Mathématiques et Informatique

**Filière :** Informatique

**Spécialité :** STIC (Sciences et Technologies de l'Information et de la  
Communication)

par

**M<sup>elle</sup> Soraya CHERIGUENE**

## Devant le jury

**Dr. Nabiha Azizi**

**Pr. Hasina Seridi**

**Pr. Hamid Seridi**

**Pr. Labiba Souici-Meslati**

**Pr. Yamina Mohammed Ben Ali**

Université Badji Mokhtar – Annaba

Université Badji Mokhtar – Annaba

Université 8 Mai 1945 – Guelma

Université Badji Mokhtar – Annaba

Université Badji Mokhtar – Annaba

**Rapporteur**

**Président**

**Examineur**

**Examineur**

**Examineur**

**Année 2018**

# Remerciements

*Je remercie tout d'abord Dieu tout puissant de m'avoir donné le courage, la force et la patience d'achever ce modeste travail*

*Je tiens à exprimer ma profonde gratitude à mon directeur de thèse Dr AZIZI Nabihha pour la confiance et l'intérêt qu'elle a porté à mes travaux depuis mes débuts dans le monde de la recherche ainsi que pour m'avoir fait profiter de sa sagesse et de son expérience durant ces quatre années de thèse.*

*Je tiens également à remercier Pr SERIDI Hasina pour avoir bien voulu accepter de présider le jury de cette thèse. Mes remerciements vont également à Pr MOHAMMED BEN ALI Yamina, Pr SOUICI-MESLATI Labiba et Pr SERIDI Hamid qui ont participé à cette soutenance en tant qu'examineurs.*

*Je voudrais également remercier tous les membres de laboratoire LABGED pour l'amitié et l'aide qu'ils m'ont témoignés durant cette thèse.*

*Les derniers mais non les moindres, j'adresse ici mes plus chaleureux remerciements à ma famille, mes proches et à mes amis qui m'ont toujours soutenu et encouragé pendant ce long travail avec une pensée particulière à Rana et Amel.*

*Merci papa pour être tellement fier de moi. Les mots me manquent pour exprimer à quel point je suis heureuse d'être à vos côtés. Merci Asma et Meryem, mes deux grandes sœurs adorées. Je peux dire sans hésitation que je n'aurais jamais réussie à en arriver là sans votre amour et votre soutien inconditionnels au cours des 27 dernières années. Merci à vous et à mes deux merveilles Frères, Zine Edinne et Amine pour m'avoir accompagnée dans cette aventure.*

*Enfin, bien que le mot merci ne soit pas suffisant ici pour celui qui est tous les jours à mes côtés depuis bientôt quatre ans, qui me soutient sans faiblir, qui a toujours su me reconforter dans les moments difficiles, par son écoute, son soutien et ses conseils : Merci Fouad.*

# المخلص

في السنوات الأخيرة، جذبت منهجيات مجموعة المصنفات أو النظام متعدد التصنيفات اهتمام مجتمع التعلم الآلي كأسلوب تصنيف مهم. و قد أظهرت الأبحاث أن الجمع بين العديد من المصنفات متوسطة الأداء هو أفضل وأسهل من البحث عن مؤشر واحد دقيق للغاية.

تتضمن عملية إنشاء نظام متعدد التصنيفات بشكل عام ثلاث مراحل رئيسية هي: الإنشاء، الاختيار و الدمج . اختيار المصنفات هو من أهم المراحل والذي يمكن تعريفه على أنه عملية انتقاء، وفقا لخوارزمية بحث ومعايير محددة، لمجموعة مثالية من المصنفات المؤهلة بدرجة كافية للتنبؤ بفئة العينات الغير مصنفة. و يتناول البحث الذي تم تقديمه في هذه الرسالة تطوير أساليب جمع و انتقاء المصنفات عن طريق تصميم تقنيات لاختيار المصنفات اعتمادا على بناء توافق بين "الدقة و التنوع"

تقترح المساهمة الأولى لهذه الأطروحة نظاما فعالا لتصنيف الصور الطبية، يجمع بين طريقة الفضاء العشوائي والاختيار الثابت للمصنفات باستخدام معايير تنوع مختلفة. أما مساهمتنا الثانية فتتعلق باقتراح منهج جديد للاختيار الثابت للمصنفات بالاعتماد على دمج مقياس الملاءمة و معايير التنوع باستخدام خوارزميات بحث مختلفة.

لقد لاحظنا أن نماذج التحديد المستعملة تختار مجموعة فرعية واحدة فقط لتصنيف كل عينات الاختبار في المستقبل. ومع ذلك، لا يعد اختيار مجموعة فرعية مثالية من المصنفات مناسبًا لبعض عينات الاختبار. لهذا السبب، نقترح في المساهمة النهائية منهج جديد للاختيار الديناميكي، يسمح باختيار المجموعة الفرعية الأكثر كفاءة لتصنيف كل عينة اختبار. يهدف هذا المنهج إلى تحسين أداء تصنيف قواعد البيانات الطبية خاصة غير المتوازنة من خلال الجمع بين طرق أخذ العينات والاستراتيجية الديناميكية لاختيار المصنفات.

تبين التجارب المختلفة التي أجريناها على عدة قواعد بيانات كفاءة المنهجيات المقترحة و أداءها الجيد، خاصةً فيما يتعلق بالاختيار الديناميكي للمصنفات.

**الكلمات المفتاحية:** التصنيف، التعرف على الأنماط، دمج المصنفات، التنوع، نظام متعدد التصنيفات، الدقة؛ الاختيار الثابت، الاختيار الديناميكي، توافق بين "الدقة و التنوع".

# Résumé

Ces dernières années, les approches d'Ensemble de Classifieurs (EdC) ou les systèmes multi-classifieurs ont attiré l'attention de la communauté de l'apprentissage automatique en tant que technique de classification significative. Trois mécanismes fondamentaux régissent la création d'un système multi-classifieurs : la génération d'un ensemble des classifieurs diversifiés, la sélection des classifieurs et finalement la combinaison des sorties des classifieurs sélectionnés. La sélection de classifieurs est une phase fondamentale dans le processus d'optimisation d'un système multi-classifieurs. Le travail de recherche présenté dans cette thèse concerne le développement d'approches de construction d'ensemble de classifieurs se basant sur le compromis « diversité-précision ».

Dans un premier temps, un système multi-classifieurs combinant la méthode Random Subspace et la sélection statique est proposé pour la classification des images médicales. L'étape de sélection de classifieurs de cette approche est basée sur l'utilisation de différentes mesures de diversité pour l'évaluation de la compétence des sous-ensembles candidats. Notre deuxième contribution concerne la proposition d'une nouvelle approche de sélection statique de classifieurs basée sur l'intégration de la méthode mRMR (minimum Redundancy Maximum Relevance) et les mesures de diversité en utilisant différents algorithmes de recherche.

Nous avons observé que cette approche de sélection a choisi un seul EdC pour classer tous les échantillons de test ou de généralisation. Cependant, il n'y a aucune garantie que l'EdC sélectionné est l'ensemble le plus susceptible pour classer un échantillon non-étiqueté. À ce titre, notre dernière contribution porte sur une nouvelle approche de sélection dynamique de prédicteurs permettant de choisir, pour chaque exemple à classer, l'EdC le plus confiant pour décider la classe d'appartenance. Cette approche vise à améliorer la performance de classification des bases de données médicales déséquilibrées en combinant les approches d'échantillonnage des données d'apprentissage et la sélection dynamique de prédicteurs.

Les différentes expérimentations que nous avons menées sur différents base de données montrent de très bonnes performances des approches proposées, plus particulièrement celle basée sur la sélection dynamique de classifieurs.

**Mots-Clés :** Classification ; Reconnaissance de Formes ; Combinaison de classifieurs ; Mesures de Diversité ; Système Multi-Classifieurs ; Précision ; Sélection de Classifieurs  
Sélection Statique ; Sélection Dynamique ; Relevance ; Compromis « Diversité-Performance».

# Abstract

In recent years, classifier ensembles approaches or Multiple Classifier Systems (MCSs) have attracted the attention of the machine learning community as a significant classification technique. In general, the MCS construction process consists of three main stages: generation, selection, and fusion. Classifier subset selection step plays a crucial role in the MCS construction process. The research work exposed in this thesis concerns the development of new schemes of classifier ensemble construction based on “diversity-accuracy” compromise.

First and foremost, a new multiple classifier system for medical images classification is proposed. This approach is based on combination of the Random Subspace method and the static classifier selection using different diversity measures to evaluate the competence of each candidate subset. Our second contribution concerns the proposition of a new static classifier selection approach based on the integration of the mRMR (minimum Redundancy Maximum Relevance) method and diversity measures using different search algorithms.

We have observed that these selection paradigms choose only one subset for all test or unseen instances. However, selecting a globally optimal subset of classifiers is not appropriate for some test samples. For this reason, the final contribution concerns a new dynamic classifier selection approach which allows choosing the most competent classifier subset for classifying each test sample. This approach aims to improve the classification performance of imbalanced medical databases by combining oversampling approaches and a dynamic classifier selection strategy.

Experimental results show that the proposed approaches have very good performances.

**Keywords:** Classification; Pattern Recognition; Classifier Combination; Diversity Measures; Multiple Classifier System; Accuracy; Classifier Selection; Static Selection; Dynamic Selection; “Diversity-Accuracy” Compromise.

# Table des Matières

<b>Remerciements</b> .....	I
<b>المخلص</b> .....	II
<b>Résumé</b> .....	III
<b>Abstract</b> .....	V
<b>Table des Matières</b> .....	VI
<b>Liste des Figures</b> .....	X
<b>Liste des Tableaux</b> .....	XII
<b>Liste des Abréviations</b> .....	XIV
<b>Introduction Générale</b> .....	1
<b>Chapitre 1. Concepts Généraux de Combinaison de Classifieurs</b> .....	8
1.1 Introduction .....	8
1.2 Formalisme de la classification supervisée .....	11
1.2.1 Définition d'un classifieur.....	11
1.2.2 Evaluation des classifieurs .....	13
1.3 Combinaison de classifieurs .....	15
1.4 Les architectures de combinaison.....	17
1.4.1 Combinaison séquentielle .....	17
1.4.2 Combinaison parallèle.....	18
1.4.3 Combinaison hybride .....	18
1.5 Combinaison parallèle de classifieurs .....	19

1.5.1	Problématique de la combinaison parallèle.....	20
1.5.2	Catégorisation des méthodes de combinaison.....	21
1.6	Méthodes de combinaison non paramétrique .....	23
1.6.1	Type classe .....	23
1.6.2	Type rang.....	24
1.6.3	Type mesure .....	25
1.7	Méthodes de combinaison paramétrique .....	28
1.7.1	Type classe .....	29
1.7.2	Type rang.....	32
1.7.3	Type mesure .....	34
1.8	Conclusion.....	36
<b>Chapitre 2. La Diversité pour la Construction et la Sélection d'Ensemble de Classifieurs.....</b>		<b>37</b>
2.1	Introduction .....	37
2.2	Diversité pour génération des classifieurs .....	38
2.2.1	Échantillonner les données d'apprentissage.....	39
2.2.2	Modifier l'espace de représentation .....	41
2.2.3	Manipuler les membres de l'ensemble .....	42
2.2.4	Manipuler les sorties .....	43
2.3	Quantification de la diversité.....	43
2.3.1	Mesures pairwise (par-paires) .....	43
2.3.2	Mesures globales (non- pairwise) .....	46
2.3.3	Limites de mesures de diversité .....	48
2.4	Sélection de classifieurs.....	49
2.5	Sélection statique de classifieurs .....	50
2.5.1	Regrouper et extraire.....	50
2.5.2	Surproduire et choisir .....	51

2.6	Sélection dynamique de classifieurs.....	59
2.6.1	Définition de la région de compétence.....	61
2.6.2	Critère de sélection.....	63
2.6.3	Schéma de sélection .....	64
2.6.4	Techniques de la sélection dynamique.....	65
2.7	Conclusion.....	70
<b>Chapitre 3. Approche Hybride basée sur Random Subspace et la Sélection Statique des Classifieurs .....</b>		<b>71</b>
3.1	Introduction .....	71
3.2	Système multi-classifieurs pour le diagnostic des images médicales.....	72
3.2.1	Les bases des images médicales.....	74
3.2.2	Prétraitement et extraction de caractéristiques.....	75
3.2.3	Approche basée sur les mesures de diversité pour la sélection d'ensemble de classifieurs.....	79
3.3	Etude expérimentale .....	83
3.3.1	Paramètres expérimentaux .....	83
3.3.2	Résultats expérimentaux .....	85
3.4	Conclusion.....	91
<b>Chapitre 4. Sélection Statique de Classifieurs Combinant la Diversité et la Méthode mRMR.....</b>		<b>93</b>
4.1	Introduction .....	93
4.2	La méthode mRMR.....	95
4.3	Sélection statique de classifieurs basé sur la méthode mRMR et les mesures de diversité.....	97
4.3.1	Génération des classifieurs.....	97
4.3.2	Sélection de classifieurs .....	97
4.3.3	Combinaison de classifieurs sélectionnés .....	102
4.4	Etude expérimental .....	103

4.4.1 Paramètres expérimentaux .....	103
4.4.2 Evaluation de la méthode MRMD .....	105
4.4.3 Evaluation de la méthode MRMD-II .....	109
4.5 Conclusion .....	120
<b>Chapitre 5. Sélection Dynamique de Classifieurs Combinant la Diversité et la Méthode mRMR .....</b>	<b>122</b>
5.1 Introduction .....	122
5.2 Apprentissage sensible au déséquilibre des classes.....	124
5.2.1 Stratégies algorithmiques .....	124
5.2.2 Stratégies d'échantillonnage .....	125
5.2.3 Stratégies ensemblistes.....	126
5.3 Sélection dynamique de classifieurs basée sur la méthode mRMR et les mesures de diversité.....	126
5.3.1 Génération d'ensemble de classifieurs.....	127
5.3.2 Sélection dynamique de classifieurs .....	128
5.4 Expérimentation .....	132
5.4.1 Les bases de données utilisées .....	132
5.4.2 Résultats expérimentaux .....	132
5.5 Conclusion.....	143
<b>Conclusion et Perspectives.....</b>	<b>145</b>
<b>Liste de Publications.....</b>	<b>149</b>
<b>Bibliographie.....</b>	<b>151</b>

# Liste des Figures

<b>Figure 1.1.</b>	Le schéma général d'un système de classification.....	10
<b>Figure 1.2.</b>	Représentations de Dietterich pour les EdCs .....	16
<b>Figure 1.3.</b>	Combinaison séquentielle de classifieurs .....	17
<b>Figure 1.4.</b>	Combinaison parallèle de classifieurs. ....	18
<b>Figure 1.5.</b>	Combinaison hybride de classifieurs.....	19
<b>Figure 1.6.</b>	Taxonomie des méthodes de combinaison parallèle de classifieurs.....	22
<b>Figure 2.1.</b>	Description schématique de l'apprentissage et la combinaison d'un ensemble de classifieurs par la méthode Bagging. ....	40
<b>Figure 2.2.</b>	Apprentissage et agrégation de classifieurs par la méthode Boosting.....	41
<b>Figure 2.3.</b>	Illustration du principe de la méthode Random Subspace pour un ensemble d'arbres de décision.....	42
<b>Figure 2.4.</b>	Schéma général du processus OCS. OCS est divisé en deux phases de surproduction et de sélection .....	52
<b>Figure 2.5.</b>	Processus de la sélection dynamique de classifieurs. ....	60
<b>Figure 3.1.</b>	Schéma général du système multi-classifieurs proposé pour le diagnostic des images médicales.....	73
<b>Figure 3.2.</b>	Un échantillon de la base DDSM : les images (a) et (b) représentent des images mammaires à masse maligne et les images (c) et (d) représentent des masses bénignes.....	74
<b>Figure 3.3.</b>	Images de fond d'œil typiques tirées de la base de données RIM-One: (a et b) Glaucome, (c et d) Normal.....	75

<b>Figure 3.4.</b> Les quatre directions d'adjacence pour calculer les caractéristiques de texture de Haralick.....	78
<b>Figure 3.5.</b> Précision moyenne du système proposé RSS-SCS en utilisant différentes mesures de diversité sur les bases de données DDSM et RIM-One.....	87
<b>Figure 3.6.</b> Précision moyenne du système proposé RSS-SCS sur les bases de données DDSM et RIM-One en utilisant différents types de classifieurs de base. ....	88
<b>Figure 3.7.</b> Précision moyenne du système proposé RSS-SCS sur les bases de données DDSM et RIM-One en utilisant différentes tailles de $C^*$ .....	89
<b>Figure 3.8.</b> Courbes ROC des bases de données DDSM et RIM-One.....	90
<b>Figure 4.1.</b> Schéma de génération d'ensemble préliminaire des classifieurs.....	97
<b>Figure 4.2.</b> Comparaison entre les résultats obtenus par MRMD utilisant différentes mesures de diversité, avec $L^* = 10$ . ....	106
<b>Figure 4.3.</b> Comparaison entre les résultats obtenus par MRMD utilisant différentes mesures de diversité, avec $L^* = 25$ . ....	107
<b>Figure 4.4.</b> Diagrammes Kappa-Erreur des bases : Breast-w, Credit-a, Credit-g, Colic, Contractions, Diabetes, Echocardiogram, Ecoli, Glass et Heart-statlog. ....	118
<b>Figure 4.5.</b> Diagrammes Kappa-Erreur des bases: Glass, Heart-statlog, Hepatitis, Ionosphere, Landsat, Parkinson's, Vote, Waveform, Yeast et Zoo.....	119
<b>Figure 5.1.</b> Un organigramme du système multi-classifieurs proposé.....	127
<b>Figure 5.2.</b> Les courbes ROC des bases Appendicitis, Diabetes, DDSM, Echocardiogram, Heart-disease et Hepatitis.....	137
<b>Figure 5.3.</b> Les courbes ROC des bases Laryngeal1, Parkinson's, Respiratory, RIM-One, Weaning et WBC.....	138
<b>Figure 5.4.</b> Les diagrammes kappa-erreur des méthodes : (a) Random Subspace, (b) DES-MRMD-II et (c) DES-MRMD-III sur les bases Appendicitis, Diabetes et Echocardiogram.....	142
<b>Figure 5.5.</b> Les diagrammes kappa-erreur des méthodes : (a) Random Subspace, (b) DES-MRMD-II et (c) DES-MRMD-III sur les bases Hepatitis et Parkinson's. ....	143

# Liste des Tableaux

<b>Tableau 1.1.</b>	Espace de connaissance du comportement BKS à 2 dimensions. ....	31
<b>Tableau 1.2.</b>	Exemple de recherche des tailles des listes des classifieurs dans l'intersection. ....	33
<b>Tableau 1.3.</b>	Procédure min-max pour déterminer les seuils dans l'union. ....	34
<b>Tableau 2.1.</b>	Relation entre les sorties de deux classifieurs $c_i$ et $c_j$ . ....	44
<b>Tableau 2.2.</b>	Résumé des mesures de diversité, où $\uparrow(\downarrow)$ indique que plus la mesure est grande (plus petite), plus la diversité est grande, "Connu" indique s'il nécessite de connaître la précision des classifieurs individuels. ....	49
<b>Tableau 2.3.</b>	Résumé de certains travaux de la sélection statique rapportés dans la littérature. ....	57
<b>Tableau 2.4.</b>	Catégorisation des méthodes de sélection dynamique. ....	65
<b>Tableau 3.1.</b>	Matrice de confusion. ....	84
<b>Tableau 3.2.</b>	Précision du système proposé RSS-SCS sur la base de données DDSM. .	86
<b>Tableau 3.3.</b>	Précision du système proposé RSS-SCS sur la base de données RIM-One. ....	87
<b>Tableau 3.4.</b>	Résultats obtenus sur les bases de données DDSM et RIM-One par différents EdC. ....	90
<b>Tableau 3.5.</b>	Différents systèmes de classification du cancer du sein. ....	91
<b>Tableau 3.6.</b>	Différents systèmes de classification pour le diagnostic du glaucome. ....	91
<b>Tableau 4.1.</b>	Caractéristiques des bases de données utilisées. ....	104

<b>Tableau 4.2.</b> Taux de classification des différentes méthodes de sélection, en utilisant différentes valeurs de $L^*$ .....	108
<b>Tableau 4.3.</b> Comparaison entre l'utilisation d'un classifieur unique et différents méthodes ensembliste utilisant le classifieur MLP.....	110
<b>Tableau 4.4.</b> Comparaison entre l'utilisation d'un classifieur unique et différents méthodes ensembliste utilisant le classifieur SVM. ....	111
<b>Tableau 4.5.</b> Comparaison entre l'utilisation d'un classifieur unique et différents méthodes ensembliste utilisant le classifieur C4.5. ....	112
<b>Tableau 4.6.</b> Comparaison entre la méthode RSS et différentes méthodes de sélection utilisant un ensemble de classifieurs hétérogènes. ....	114
<b>Tableau 4.7.</b> Comparaison des résultats de précision de différentes versions de schéma de sélection MRMD-II. ....	115
<b>Tableau 4.8.</b> Description des bases de données déséquilibrées.....	116
<b>Tableau 4.9.</b> Résultats de Sensibilité (SEN) et de Spécificité (SPE) obtenus sur des jeux de données déséquilibrées en utilisant différents ensembles de classifieurs. ....	116
<b>Tableau 5.1.</b> Caractéristiques des bases de données utilisées. ....	132
<b>Tableau 5.2.</b> Résultats de précision de l'approche DES-MRMD-II et différentes méthodes de sélection.....	134
<b>Tableau 5.3.</b> Résultats de sensibilité de différentes méthodes de sélection. ....	135
<b>Tableau 5.4.</b> Résultats de spécificité de différentes méthodes de sélection. ....	135
<b>Tableau 5.5.</b> Résultats d'AUC de différentes méthodes de sélection.....	136
<b>Tableau 5.6.</b> Description des bases de données déséquilibrées.....	139
<b>Tableau 5.7.</b> Taux de classification de différents ensembles de classifieurs sur des bases de données déséquilibrées. ....	140
<b>Tableau 5.8.</b> Résultats de sensibilité (SEN) et spécificité (SPE) de différents ensembles de classifieurs sur des bases de données déséquilibrées.....	140

## Liste des Abréviations

AG	Genetic Algorithm
AD	Arbre de Décision
AUC	L'air sous la courbe ROC (Area Under ROC Curve)
BK	Recherche en Arrière (Backward search)
BKS	Espace de Connaissance du Comportement (Behaviour Knowledge Space)
DCS	Sélection Dynamique de Classifieurs (Dynamic Classifier Selection)
DDSM	La base de données numérique pour la mammographie de dépistage (Digital Database for Screening Mammography)
DES	Sélection Dynamique d'Ensemble de Classifieurs (Dynamic Classifier Selection)
DSEL	base de données de la sélection dynamique (dynamic selection dataset)
EdC	Ensemble de Classifieurs
FD	Recherche en Avant (Forward search)
GANSEN	Genetic Algorithm based Selective Ensemble
GLCM	La matrice de co-occurrence des niveaux de gris (Gray Level Co occurrence Matrix)
HET	Hétérogènes
IM	Information Mutuelle
KNORA	les k plus proches oracles (k-Nearest Oracles)
KNOP	k plus proches profiles de sortie (k-Nearest Output Profiles)

## Liste des abréviations

---

kPPV	k- Plus Proche Voisins
KL	Kullback-Leibler
LCA	Local Classifier Accuracy
MCB	Comportement de Classifieur Multiple (Multiple Classifier Behavior)
MCS	Système Multi-Classifieurs (Multiple Classifier system)
MID	Mutual Information Difference
MIQ	Mutual Information Quotient
MLP	Réseau de Neurones Multi-Couches (Multi-Layer Perceptron)
mRMR	minimum Redundancy Maximum Relevance
MRMD	Maximum Relevance Maximum Diversity
OCS	Surproduction et choix (Overproduce and Select Strategy)
OLA	Overall Local Accuracy
PBIL	Population Based Incremental Learning
RBF	Réseaux à Fonction de Base Radiale (Radial Basis Function)
RIM-One	Une Base de données rétinienne pour l'évaluation du nerf optique (Retinal Image Database for Optic Nerve Evaluation)
ROC	Fonction d'efficacité du récepteur (Receiver Operating Characteristic)
RRC	Randomized Reference Classifier
RSS	Random Subspace
SCS	Sélection Statique de Classifieurs (Static Classifier Selection)
SVM	Machine à Vecteur de Supports (Support Vector Machine)
TS	Tabu Search

# Introduction Générale

La classification est l'une des techniques les plus anciennes d'analyse et de traitement de données qui compte parmi les plus grandes réussites de l'apprentissage automatique. La classification peut être définie comme le problème de la détermination de l'appartenance d'un ensemble d'observations à une nouvelle catégorie. Cette détermination est basée sur un ensemble de données d'apprentissage qui comprend des observations dont leurs appartenances à une catégorie sont connues. Plusieurs algorithmes de classification ont été proposés et utilisés dans différents domaines de la reconnaissance de formes mais les travaux récents de l'apprentissage automatique ont montré qu'aucune méthode individuelle ne peut être efficace pour traiter tous les problèmes de classification et que le choix d'une méthode de classification dépend du problème posé. Ceci a confirmé le théorème *No Free Lunch* de Wolpert stipulant que «*no algorithm is better than any other over all possible classes of problems*». Par conséquent, la question de savoir quelle méthode de classification pourrait convenir à une étude spécifique n'est pas facile à répondre. Pour résoudre ce problème complexe de classification, la combinaison de classifieurs a été proposée comme solution prometteuse.

En général, les ensembles de classifieurs sont considérés comme un développement important dans le domaine de l'apprentissage automatique et de l'intelligence artificielle. Le concept intuitif de l'apprentissage d'ensemble est qu'aucune approche ou système unique ne peut prétendre être supérieur à un autre et que l'intégration de plusieurs approches améliorera la performance de classification. Différents scénarios et stratégies de combinaison ont été introduits dans la littérature, à savoir : séquentiel, parallèle et hybride. L'architecture parallèle est la stratégie la plus communément utilisée dans les travaux de la combinaison de classifieurs. Cette architecture est simple à mettre en œuvre et facilement

généralisable parce qu'elle ne nécessite pas de connaître le comportement de chacun des classifieurs. De nombreuses règles de combinaison ont été proposées pour combiner en parallèle les résultats d'un ensemble de classifieurs. Ces méthodes de fusion peuvent être catégorisée sur la base des sorties des classifieurs utilisés dans la combinaison qui sont généralement divisés en trois types, à savoir : classe, rang et mesure.

La plupart de ces méthodes de combinaison reposent sur l'hypothèse que tous les membres de l'ensemble font des erreurs indépendantes. L'efficacité des méthodes d'ensemble dépend généralement de l'indépendance des erreurs commises par les classifieurs de base pris individuellement. En outre, les experts du domaine ont prouvé que la performance de l'ensemble ne peut être meilleure que celle des classifieurs individuels si ses membres sont différents et pas nécessairement indépendants. Par conséquent, les membres de l'ensemble doivent reflètent un certain niveau de diversité. Ce concept de diversité est généralement considéré comme la capacité des classifieurs à faire des erreurs différentes sur les mêmes points de données. La diversité entre les classifieurs peut être générée de différentes manières, par exemple en utilisant des modèles de classification différents; manipulant les données d'apprentissage; manipulant les caractéristique d'entrée ou en utilisant différents types de sorties de classifieur.

Bien que le débat sur la nature de la diversité des classifieurs se poursuive, du point de vue technique, la notion de diversité ne peut être évidente que si elle est soutenue par des mesures appropriées. C'est là que les problèmes de diversité ont commencé à apparaître à grande échelle. Différentes mesures pour la quantification de la diversité ont été étudiées par Kuncheva et Whitaker [Kuncheva 2003a]. Ces mesures ont ouvert la voie à plusieurs travaux sur l'utilisation de la diversité de manière explicite durant le processus de construction d'un système multi-classifieurs tant qu'outil théorique, outil de visualisation ou comme un critère de sélection de classifieurs. Les études de Kuncheva et Whitaker [Kuncheva 2003a] ont montré également que ces mesures ne permettent pas de donner une idée claire sur la relation entre la performance de l'ensemble et la diversité de ces membres.

Généralement, la performance d'un ensemble des prédicteurs peut se varier selon le nombre et le type des classifieurs combinés. Lorsqu'on dispose d'un grand nombre de classifieurs, on se pose toujours la question de savoir combien et quels classifieurs choisir pour améliorer la performance de l'ensemble. Une phase intelligente de sélection de

classifieurs au cours de processus de construction d'un système d'ensemble est devenue cruciale pour aboutir au meilleur taux de classification avec un nombre minime de classifieurs. Deux catégories de techniques de sélection de classifieurs existent: statique et dynamique. Dans le premier cas, les classifieurs sont sélectionnés pendant la phase d'apprentissage et combinés pour classer tous les échantillons de test, tandis que dans le deuxième cas, elles sont définies pendant la phase de classification en tenant compte des caractéristiques de l'échantillon à classer. La sélection dynamique de classifieurs est traditionnellement définie comme une stratégie qui suppose que chaque membre de l'ensemble est un expert dans certaines régions locales de l'espace des caractéristiques. Le classifieur le plus localement compétant est sélectionné pour estimer la classe de chaque échantillon particulier de test.

### Problématique

En général, trois facteurs cruciaux interviennent dans l'optimisation d'un système ensembliste de classification : la construction d'un ensemble adéquat de classifieurs, la sélection des meilleurs membres et la conception d'une règle de fusion. La sélection de classifieurs est la phase fondamentale dans le processus d'optimisation d'un système multi-classifieurs, car elle se concentre sur la recherche du sous-ensemble des classifieurs, dont leurs combinaison produit une précision optimale. La performance d'une méthode de sélection dépend du choix de l'algorithme de recherche et du critère de sélection. Nous avons vu que, même si plusieurs contributions importantes ont été apportées à ce sujet, beaucoup de questions restent ouvertes :

- Comment définir un meilleur critère de recherche pour trouver le meilleur sous-ensemble de classifieurs ? En outre, la sélection des classifieurs ou prédicteurs les plus performants n'est pas forcément la bonne solution pour aboutir au meilleur résultat, est-ce que la diversité est un meilleur critère de sélection par rapport à la précision? et est-ce qu'on peut équilibrer ces deux facteurs en définissant un compromis entre eux ?
- Bien que la recherche du sous-ensemble optimal de classifieurs puisse être exhaustive, des algorithmes de recherche peuvent être utilisés lorsqu'un grand groupe initial de classifieurs est généré en raison de la complexité exponentielle d'une recherche exhaustive. La question qui se pose est est-ce que l'utilisation des

méthodes de recherche non-exhaustive peut trouver un sous-ensemble optimal de classifieurs ?

- La majeure limite de la sélection statique est qu'un sous-ensemble fixe de classifieurs défini à l'aide d'un ensemble de données de validation qui ne peut pas être bien adapté à l'ensemble de données de test. De plus, le concept de la sélection statique est assez similaire à l'utilisation d'un classifieur individuel pour classer tous les échantillons de test ; c'est-à-dire qu'en raison des différences entre les échantillons, il n'existe aucun classifieur individuel parfaitement adapté à tous les échantillons de test. Par conséquent, la question revenant est savoir est-ce que la sélection dynamique consistant à choisir un sous-ensemble de classifieurs pour chaque échantillon de test, est meilleure que la sélection statique ?

## Objectifs et Contributions

L'objectif principal de cette thèse est de développer une méthodologie pour améliorer la précision d'ensemble de classifieurs en concevant des techniques de sélection de classifieurs. Nos objectifs peuvent être cernés par les points suivants :

- Etudier le rôle des mesures diversité au cours de processus de sélection des classifieurs.
- Montrer expérimentalement le comportement de processus de choix de classifieurs en utilisant différents paramètres de sélection comme : le type des classifieurs, le nombre de classifieurs à sélectionner et le critère de recherche utilisé.
- Clarifier le lien existant entre la diversité et la performance des classifieurs en proposant un compromis entre eux et l'utiliser pour guider les différents algorithmes de recherche.
- Proposer une stratégie de sélection dynamique pour permettre le choix du sous-ensemble de classifieurs les plus compétents pour étiqueter chaque échantillon de test.

Afin de répondre aux besoins cités ci-dessus, trois contributions ont été proposées. La première contribution se fonde sur la notion de diversité et leur utilisation comme critère de construction et de sélection de classifieurs. Le système ensembliste proposé est basé sur

la combinaison de la méthode *Random Subspace* et la sélection statique de classifieurs. L'objectif de cette contribution est de montrer que la création d'un certain niveau de diversité pendant la phase de génération et de sélection de classifieurs permet de construire un système robuste de classification. La diversité a été renforcée entre les membres de l'ensemble initial de classifieurs en modifiant l'espace de caractéristiques avant l'apprentissage de chaque classifieur donné. Afin d'analyser l'influence de la diversité sur la performance globale du système, des mesures de diversité ont été introduites au cours de la phase de sélection. De plus, une recherche exhaustive est appliquée sur toutes les paires de mesures de diversité durant la phase de validation.

Afin de pallier des principaux inconvénients de l'approche précédente qui sont le coût de calcul, et plus particulièrement l'utilisation d'un seul critère de sélection (qui est la diversité), notre deuxième contribution suggère différents algorithmes de recherche en utilisant un nouveau compromis « diversité-précision » pour obtenir une génération optimale d'un système ensembliste. Le concept de la relevance entre les sorties des classifieurs et la classe cible ainsi qu'une combinaison des mesures de diversité sont utilisés en tant que critère de sélection. Ce critère a été utilisé comme un outil pour mesurer le lien entre la diversité d'opinions évaluées entre les membres de l'ensemble et la performance globale.

Nous avons intéressé plus particulièrement dans cette thèse des données médicales qui souffrent en majorité de déséquilibre entre classes. D'où notre dernière contribution qui consiste à proposer un système de sélection dynamique de classifieurs pour la classification des données déséquilibrées. La stratégie de sélection dynamique proposée dans cette thèse permet la sélection, pour chaque exemple à classer, d'un sous-ensemble de classifieurs le plus confiant pour décider la classe d'appartenance. Notre système vise à résoudre le problème de l'asymétrie des classes en deux niveaux. Premièrement, une approche d'échantillonnage est considérée pour rééquilibrer l'ensemble de données d'apprentissage, ensuite, une approche de sélection dynamique est appliquée sur l'ensemble de classifieurs générés sur les données équilibrées. Cette stratégie de sélection est basée sur un nouveau compromis diversité-performance proposé pour les données déséquilibrées. La méthode de sélection dynamique proposée domine les approches conventionnelles (approches statiques) sur les problèmes de classification des données médicales étudiés dans le cadre de cette thèse.

## Plan de la thèse

La présente thèse est organisée de la manière suivante :

Dans le premier chapitre, nous présentons un formalisme de la classification supervisée dans le cadre de la combinaison de classifieurs, et nous introduisons un ensemble de notations qui seront utilisées tout au long du présent manuscrit. En outre, Les principales architectures de combinaison (séquentielle, parallèle et hybride) sont présentées en revue. De plus, nous présentons un état de l'art varié et précis, bien que non-exhaustif, sur les méthodes classiques proposées pour la combinaison parallèle de classifieurs

Le chapitre 2 permet de répondre aux limitations des méthodes décrites dans le premier chapitre en introduisant la notion de la diversité. La première partie de ce chapitre décrit le problème lié à la diversité de classifieurs en illustrant les différentes stratégies de génération d'ensemble de classifieurs, ainsi que les mesures proposées pour estimer la diversité entre les membres de l'ensemble. Dans la deuxième partie, nous passons en revue les principales méthodes de sélection des classifieurs en présentant une taxonomie des systèmes de sélection (statique et dynamique), ainsi que le rôle que joue la diversité dans l'amélioration de performance durant le processus de sélection.

Le chapitre 3 présente notre première contribution qui est la proposition d'un système multi-classifieurs basé sur la méthode *Random Subspace* et la sélection statique de classifieurs pour la classification des images médicales. Dans ce chapitre, nous cherchons à clarifier le rôle de la diversité généralement dans le processus de construction d'ensemble de classifieurs et plus particulièrement dans la phase de sélection.

Le chapitre 4 aborde la seconde contribution basée sur la sélection statique de classifieurs en intégrant la méthode mRMR et les mesures de diversité. Dans cette contribution, nous mettons en œuvre deux approches originales de sélection basées sur un compromis diversité-précision pour tenir en compte simultanément de la diversité et de la performance des classifieurs en utilisant différents algorithmes de recherche de classifieurs. Les deux systèmes proposés seront détaillés successivement incluant les résultats expérimentaux et les bases de données intervenues dans la réalisation de ces systèmes.

Dans le chapitre 5, nous présentons notre dernière contribution qui est la proposition d'un système ensembliste basé sur la sélection dynamique de classifieurs. Ce système vise à traiter le problème des données déséquilibrées dans les cas médicales en combinant l'approche d'échantillonnage SMOTE et la sélection dynamique de classifieurs. Enfin, nous achevons cette thèse par une conclusion générale ainsi que les perspectives envisagées pour la poursuite de ce travail.

# Chapitre 1

## Concepts Généraux de Combinaison de Classifieurs

### 1.1 Introduction

Parmi les nombreuses disciplines représentées en informatique, l'Apprentissage Automatique, ou Machine Learning en anglais, domaine d'étude de l'intelligence artificiel, concerne la conception, la recherche, le développement et l'implémentation d'algorithmes permettant à une machine d'apprendre réaliser l'induction automatique de règles à partir d'un ensemble d'exemples [Gacquer 2008a]. De manière formelle, nous disons que la machine doit apprendre à produire la sortie désirée lorsqu'elle est face à une situation spécifique [Mitchell 1999]. En apprentissage automatique, on peut distinguer deux grandes familles dépendantes des informations disponibles sur les données à classer : l'apprentissage supervisé et l'apprentissage non supervisé.

L'apprentissage est dite supervisé [Vincent 2003 ; Oltean 2009] si les classes auxquelles appartiennent les exemples sont connues à l'avance et si la tâche d'apprentissage est guidée par la décision d'un superviseur ou d'un expert. Le but est de trouver un modèle qui décrit au mieux la relation entre les attributs et les classes afin de prédire la classe d'appartenance d'un nouvel exemple issu du système à surveiller.

L'apprentissage non-supervisé (ou clustering) [Jain 1999] se distingue de l'apprentissage supervisé par le fait que la nature et le nombre des classes d'appartenance sont déterminés sans aucune connaissance préalable [Hansen 1997]. L'objectif, dans ce cas, est de partitionner un ensemble de données hétérogènes en plusieurs sous-ensembles d'individus ayant des caractères similaires. Les objets sont attribués aux différentes classes

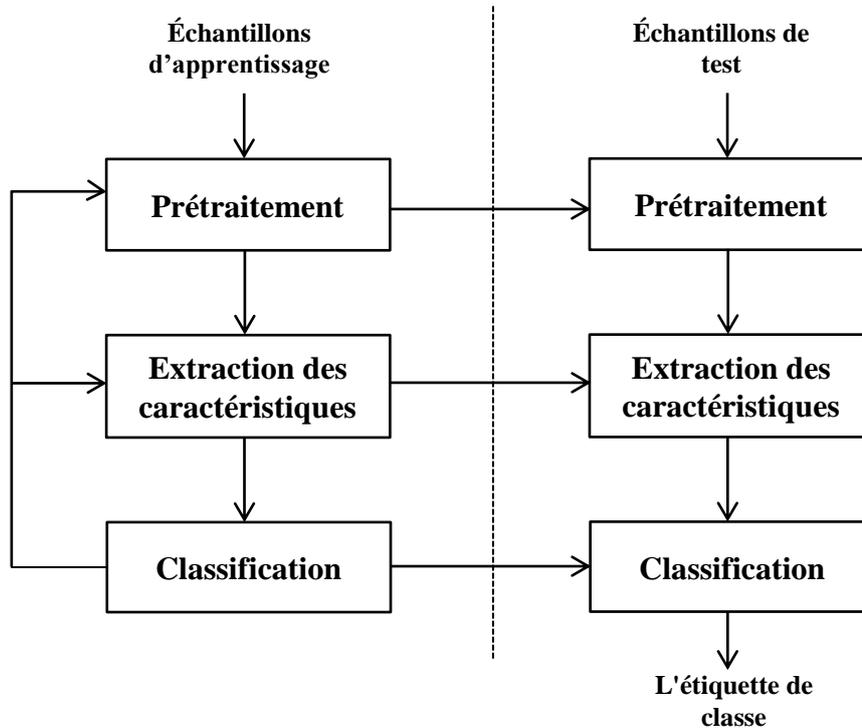
estimées selon deux critères essentiels qui sont la similitude entre les individus de chaque classe et la bonne séparation entre les classes.

Cette thèse concerne plus particulièrement le problème de la classification. La classification [Cormack 1971 ; Belacel 1999 ; Henriet 2000] est l'une des tâches d'apprentissage supervisé consistant à classer les objets dans un ensemble fini de classes. L'objectif de la classification est de construire un modèle qui sera ensuite utilisé pour la classification des nouveaux objets (ou exemples) dont la classe d'appartenance est inconnue. Ces objets représentent des unités de données compactes spécifiques à un problème particulier comme les images, les paroles, les caractères manuscrits et sont généralement appelés « échantillon » ou « instance ». Une instance est représentée par un ensemble de propriétés quantifiables appelées « caractéristiques » qui devraient contenir des informations pertinentes sur la structure de l'objet que nous souhaitons à classer [Zhang 2003]. Ces caractéristiques peuvent potentiellement être collectées à partir d'un grand nombre de sources de données, ce qui se traduirait par un vecteur de mesures de haute dimensionnalité.

Pour prendre une décision sur un échantillon non étiqueté, un système de classification doit se baser sur la connaissance a priori qui prend souvent la forme d'exemples connus [Vapnik 1998]. L'ensemble de ces exemples est appelé la base d'apprentissage ou la base d'entraînement. Cette base est fournie par l'expert humain qui structure les données et les étiquettes. Pour mesurer les performances d'un système de classification, une autre base de données étiquetée est utilisée. Cette base; appelée base de test ; ne doit pas avoir servi de connaissance préalable pour ne pas biaiser la mesure de performance. La classification est validée en évaluant sa pertinence à travers le taux des exemples de test bien classés.

Afin de garantir la performance globale d'un système de reconnaissance, le processus de classification doit être précédé par d'autres étapes préliminaires qui sont le prétraitement [Liu 2001 ; Zhang 2003] et l'extraction de caractéristiques [Liu 2013 ; El-Atlas 2014 ; Sundararaj 2014]. Un schéma général d'un système de classification est illustré à la figure 1.1. Lors de l'étape de prétraitement, les objets d'entrée sont mis en forme et traités avant toutes manipulations afin de faciliter la tâche d'extraction de caractéristiques. Par exemple, c'est à cette étape qu'un signal sera filtré, qu'une image sera binarisée, normalisée ou redimensionnée pour obtenir une image plus lisible.

Ensuite, ces objets seront représentés par des vecteurs de caractéristiques où chaque composante est un attribut qualitatif (vert, rond, chaud, etc.) ou une mesure faite sur les objets réels (hauteur d'un vérin dans un système mécanique, nombre de pixels noirs d'une image, amplitude, phase d'un signal électrique, etc.)



**Figure 1.1.** Le schéma général d'un système de classification.

L'un des défis de la reconnaissance de formes est de créer des systèmes de classification qui soient à la fois simples et performants. Différents algorithmes de classification (ou d'apprentissage) ont été proposés dans la littérature, tels que les arbres de décision [Quigley 1986], les réseaux de neurones [Tang 2016], les  $k$  plus proches voisins (kPPV) [Tremblay 2004], machines à vecteurs de support (*Support Vector Machine*, SVM) [Hearst 1998], etc. Étant donné l'échantillon  $x$  et sa classe  $y_k$  avec une fonction inconnue  $y_k = f(x)$ , tous ces algorithmes d'apprentissage cherchent à trouver dans l'espace d'hypothèses  $H$  la meilleure fonction d'approximation  $h$ , qui est un classifieur, à la fonction  $f(x)$ . Par conséquent, le but de ces algorithmes est la conception d'un classifieur unique robuste et bien adapté au problème concerné.

Il existe plusieurs façons pour améliorer la précision de ces algorithmes de classification. Les derniers travaux de l'apprentissage automatique [Woods 1997 ; Rahman

2003 ; Tulyakov 2008 ; Zhang 2015] ont proposé d'utiliser plusieurs classifieurs différents pour effectuer les tâches de reconnaissance de formes, ce qu'on appelle un système multi-classifieurs (*Multiple Classifier system, MCS*), système ensembliste ou un Ensemble de Classifieurs (EdC).

Dans cette thèse, les contributions apportées concernent les approches ensemblistes de classification. Cependant, nous présentons brièvement d'abord les différents aspects de la classification supervisée dans le cadre de la combinaison, pour éviter toute confusion éventuelle, et nous présentons ensuite les différentes architectures et stratégies de combinaison de classifieurs.

## 1.2 Formalisme de la classification supervisée

### 1.2.1 Définition d'un classifieur

Soit  $\psi = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_M)\}$  un ensemble de  $N$  objets.  $x_i$  est un exemple décrit par un vecteur d'attributs  $x_i \in X = [x_1^i, x_2^i, \dots, x_F^i]$  de  $F$  mesures,  $x_g^i \in \mathfrak{R}^n$ , ainsi que par un label  $y_k \in \Omega = \{y_1, y_2, \dots, y_M\}$  qui lui est associé, ou  $\Omega$  est l'ensemble des classes. Chaque exemple  $x_i$  est un vecteur de dimension  $F$  composé d'un ensemble des paramètres continus (numériques) et / ou discrets (nominaux). Le processus de classification a pour but d'établir une transformation  $h$  de l'espace d'attributs  $x_i$  vers l'espace de classes  $\Omega$  [Vapnik 1998]. C'est cette transformation  $h$  qui est souvent appelée classifieur.

$$x_i \xrightarrow{h} \Omega = \{y_1, \dots, y_M\} \quad (1.1)$$

Basé sur la définition de la classification, un classifieur (prédicteur ou expert) peut être vu comme une fonction  $c$  permettant d'assigner à chaque vecteur de caractéristiques d'un nouveau exemple  $x_i$  une des classes  $y_k$  pour lequel on ne dispose pas d'information a priori [Bishop 2006]:

$$c(x_i) \rightarrow y_k \quad (1.2)$$

La mise en œuvre d'un classifieur nécessite généralement de choisir une représentation pour décrire les données (caractéristiques) et une base d'apprentissage qui permet de fixer les paramètres de l'algorithme de décision. Introduire des modifications à un classifieur que ce soit au niveau des données qu'il traite ou au niveau de ses paramètres (type de

sorties, règles de décision, etc.) modifie ses performances. Une généralisation de la définition d'un classifieur est proposée dans [Duda 1973] : un classifieur  $c_j$  est un algorithme permettant d'associer à un exemple à reconnaître  $x_i$  un vecteur de niveau d'appartenance de cet exemple  $x_i$  à une classe  $y_k$  :

$$c_j(x_i) = [c_j^1(x_i), c_j^2(x_i), \dots, c_j^M(x_i)] \quad (1.3)$$

La fonction discriminantes  $c_j^k(x_i)$  avec  $y_k \in \Omega$  représente le degré d'appartenance d'un objet  $x_i$  à une classe  $y_k$ . Dans ce cas, l'exemple  $x_i$  peut appartenir à plusieurs classes si  $c_j^k(x_i) \neq 0$ . La règle d'agrégation maximale est utilisée pour définir la classification comme suit:

$$c_j(x_i) = \max_{j \in \Omega} c_j^k(x_i) \quad (1.4)$$

Selon Xu [1992], cette réponse  $c_j(x_i)$  peut être divisée en trois catégories suivant le type d'information fournis par le classifieur  $c_j$  :

- **Type classe:** c'est le type le plus général mais qui apporte le moins d'informations. Dans ce type, le classifieur ne fournis que l'étiquette de classe attribuée à l'exemple d'entrée. Il pointe vers une catégorie unique en tant que classe gagnante en ignorant les informations sur les chances de sélection des autres classes en tant que décision finale.

$$c_j(x_i) = y_k, k \in \{1, \dots, M\} \quad (1.5)$$

- **Type rang:** dans ce niveau, la sortie d'un classifieur  $c_j$  est un vecteur ordonné de classes candidates. La classe positionnée en premier est considérée comme le type de classe le plus probable, tandis que la classe placée à la fin du vecteur est la plus improbable. En outre, il n'y a aucune valeur de confiance incluse dans les étiquettes de classe dans ce type. Cependant, sa position dans la liste montre sa probabilité relative. Soit  $r_{k,j}$  le rang attribué à la classe  $y_k$  par le classifieur  $c_j$ , la sortie de  $c_j$  est définie par :

$$c_j(x_i) = (r_{1,j}, r_{2,j}, \dots, r_{M,j}) \quad (1.6)$$

- **Type mesure:** nombreuses méthodes de classification génèrent leurs décisions sur la base de certaines mesures obtenues par une fonction discriminante d'un classifieur ou de certaines mesures de distance parmi les données de l'espace d'entrée. Dans ce type, la sortie d'un classifieur est un vecteur des mesures de taille  $M$ , où chaque mesure  $m_{k,j}$  représente le niveau de confiance attribué à chaque classe  $y_k$ . Cette mesure de confiance, normalisée ou non, peut-être des nombres arbitraires selon la structure de classification utilisée (une probabilité a posteriori, un score, une distance, une possibilité, une crédibilité, etc.). Par conséquent, le type mesure contient plus d'informations par rapport aux autres types de sortie.

$$c_j(x_i) = (m_{1,j}, m_{2,j}, \dots, m_{M,j}) \quad (1.7)$$

### 1.2.2 Évaluation des classifieurs

L'évaluation des performances d'un classifieur est une phase importante dans le processus de conception d'un système de classification. Elle permet de savoir si ce système est suffisamment performant pour l'application visée et permet également de le comparer avec d'autres méthodes de classification. Une variété de mesures d'évaluation en fonction de différents critères a été proposée pour décrire la performance d'un classifieur, allant de la précision de classification et du taux d'erreur, en passant par la complexité du stockage et le temps de calcul jusqu'à la sensibilité et la robustesse [Vapnik 1998].

Pour évaluer le taux de classification ou la précision d'un classifieur, la méthode communément utilisée consiste à partitionner la base de données disponible en deux sous-ensembles  $BD_{App}$  et  $BD_{tes}$ . La base  $BD_{App}$  est dédiée à l'apprentissage du classifieur et la base  $BD_{tes}$  est utilisé comme un ensemble d'exemples de test inconnus pour l'évaluation de ses performances. La précision obtenue sur chacun des partitionnements est définie par :

$$Précision(C) = \frac{1}{|BD_{tes}|} \sum_{i=1}^{|BD_{tes}|} \{I(\hat{Y}_i - Y_i)\} \quad (1.8)$$

Où  $|BD_{tes}|$  représente la taille de l'ensemble de test  $BD_{tes}$ ,  $\hat{Y}_i = c_j(x_i)$  représente la classe prédit par le classifieur  $c_j$  pour l'échantillon  $x_i$  et  $Y_i$  est sa classe réelle. La fonction  $I$  est donnée par :

$$I(\hat{Y}_i - Y_i) = \begin{cases} 1 & \text{si } \hat{Y}_i = Y_i \\ 0 & \text{si } \hat{Y}_i \neq Y_i \end{cases} \quad (1.9)$$

Pour évaluer les qualités prédictives finales du modèle, on considère la précision moyenne des résultats obtenus sur les différents partitionnements réalisés. Cette méthode présente l'avantage de fournir une bonne estimation des facultés prédictives des modèles obtenus, en évitant une dépendance du résultat au choix particulier des échantillons d'apprentissage et de test. Le choix du partitionnement dépend principalement de la quantité de données disponibles. Une quantité très grande des exemples d'apprentissage peut induire un problème de sur-apprentissage du modèle, dont les facultés prédictives sur de nouveaux exemples sont diminuées. A l'inverse, un ensemble d'apprentissage de taille trop faible risque de provoquer une situation de sous-apprentissage, qui se traduit par un modèle incorrect et peu précis. Une étude des différentes approches de validation d'un modèle est présentée dans [Boser 1992] et [Gacquer 2008a]. Nous en présentons quelques-unes :

- La méthode **Hold-out** consiste à partitionner aléatoirement les données de départ en deux sous-ensembles, généralement de tailles égales. Un sous-ensemble est utilisé pour la phase d'apprentissage, et l'autre servant à tester ou évaluer la performance du modèle de classification. Un cas particulier de cette méthode, appelé **split-half** consiste à de permuter les deux parties et d'évaluer le classifieur par la moyenne des deux estimations ainsi obtenues.
- La **validation croisée** en  $k$  blocs (*k-fold cross validation*) partitionne l'ensemble de données aléatoirement en  $k$  sous-ensembles de taille approximativement égale. À chaque itération, un bloc est utilisé comme ensemble de test tandis que les autres blocs sont combinés pour former un ensemble d'apprentissage. L'algorithme d'apprentissage est exécuté  $k$  fois, en changeant de bloc pour tester le classifieur. La précision finale est estimée en moyennant les performances obtenues à chaque itération. Afin de garantir que chaque sous-ensemble est un bon représentatif de la base de départ, les données sont souvent stratifiées avant d'être divisées en  $k$  blocs de sorte que chaque bloc contient à peu près les mêmes proportions de classes que l'ensemble des données initial. Ceci est communément appelé validation croisée stratifiée en  $k$  blocs.

- La méthode *leave-one-out* est une version particulière de la validation croisée pour lequel  $k$  est égal au nombre des exemples disponibles  $N$ . L'ensemble test est alors constitué d'un seul exemple et la phase d'apprentissage est effectuée sur les  $(N - 1)$  exemples restants. Ce type de validation est plus long à mettre en œuvre du fait de la taille importante de l'ensemble d'apprentissage et du nombre d'itérations nécessaires mais elle s'avère particulièrement utile aux problèmes pour lequel la taille d'ensemble d'apprentissage est limitée. D'autre part, cette méthode donne une estimation précise du taux d'erreur du classifieur.
- Les méthodes de validation croisée décrites précédemment sont basées sur le partitionnement des données initiales en plusieurs sous-ensembles sans remplacement. Une approche alternative basée sur le *bootstrap* consiste à générer  $b$  ensembles d'apprentissage de taille  $N$ , appelés *bootstrap samples*, en effectuant des tirages aléatoires avec remise sur l'ensemble de données initiale. Un classifieur différent est alors entraîné sur chacun des  $b$  ensembles de données ainsi formés et les échantillons non inclus dans l'ensemble d'apprentissage sont utilisés pour tester le modèle. La moyenne des précisions de chaque classifieur constitue l'estimation par *bootstrap* ordinaire [Efron 1994].

### 1.3 Combinaison de classifieurs

L'idée de combinaison de classifieurs a été apparue dans les années 80 dans de nombreux travaux de la de reconnaissance de l'écriture manuscrite [Srihari 1982 ; Hull 1983 ; Hull, 1988 ; Mandler ; Lam 1994] qui ont montré l'importance de combinaison de différents prédicteurs pour l'amélioration de performances de ces problèmes de classification. Cependant, ce n'est qu'à la fin des années 90 et plus abondamment par la suite que les systèmes ensemblistes ont été concrètement mis en œuvre [Zouari 2004]. L'objectif de la combinaison de classifieurs est de produire un nouveau système de classification robuste, constitué d'un ensemble de prédicteurs de base afin de bénéficier de leur éventuelle complémentarité. Intuitivement, un modèle ensembliste permet de réduire le nombre d'exemples qui seraient mal classés par un modèle unique tout en conservant un temps de calcul raisonnable.

Ces dernières années, plusieurs travaux ont montré que l'utilisation d'un ensemble de classifieurs permet d'améliorer de manière significative les performances obtenues en

classification par rapport à l'utilisation d'un seul modèle. Les auteurs de [Fuchs 2015] et de [Verbiest 2016] ont prouvé que l'utilisation d'un ensemble de kPPVs donne de meilleur résultats que l'utilisation d'un kPPV unique ; les travaux de [Stork 2015 ; Zhou 2018] et ont prouvé la supériorité des systèmes utilisant des SVMs combinés par rapport à ceux basés sur un seul SVM et Pham et al. [Pham 2017] ont montré des améliorations de performance en combinant un ensemble de réseaux neurones, au lieu d'utiliser un seul réseau de neurones.

Selon Dietterich [2000a], il y a trois raisons principales pour lesquelles un ensemble de modèles peut donner souvent des résultats plus précis qu'un seul modèle, qui sont de nature statistique, computationnelle et représentationnelle. La figure 1.2 illustre ces raisons.

- **Aspect statique** : un algorithme d'apprentissage a une grande variance statique lorsqu'il trouve différentes hypothèses  $h_i$  qui semblent bien approximer la fonction recherchée  $f(x)$  durant la phase d'apprentissage mais qu'il choisit l'hypothèse la moins compétente lorsqu'il est testé sur des données inconnues. Cette variance peut être diminuée en combinant plusieurs hypothèses.
- **Aspect computationnelle** : les algorithmes d'apprentissage sont généralement basés sur des techniques exploitant les minimums locaux et, par conséquent, peuvent être bloqués dans de ces optimums. La combinaison de différents minimaux locaux peut fournir une meilleure approximation de  $f(x)$ .
- **Aspect représentationnel**: l'espace d'hypothèse  $H$  ne contient pas toujours de bonnes approximations de la fonction  $f(x)$ . Dans ce cas, les EdC permettent d'élargir l'espace des fonctions évaluées conduisant à une meilleure approximation de  $f(x)$ .

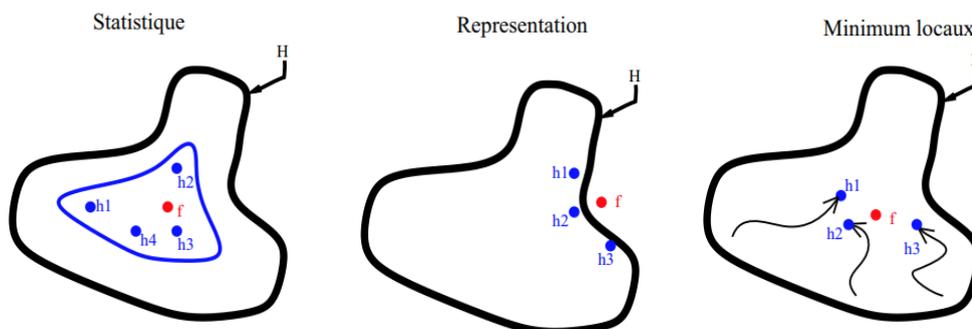


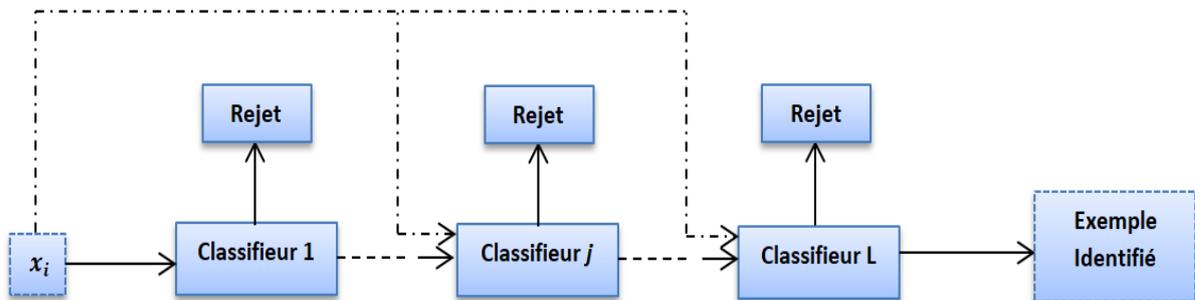
Figure 1.2. Représentations de Dietterich pour les EdCs [2000a].

## 1.4 Les architectures de combinaison

Pour intégrer les sorties des classifieurs dans une décision finale, trois architectures de combinaison peuvent être envisagées : séquentielle, parallèle, et hybride [Heutte 1994, Rahman 2003].

### 1.4.1 Combinaison séquentielle

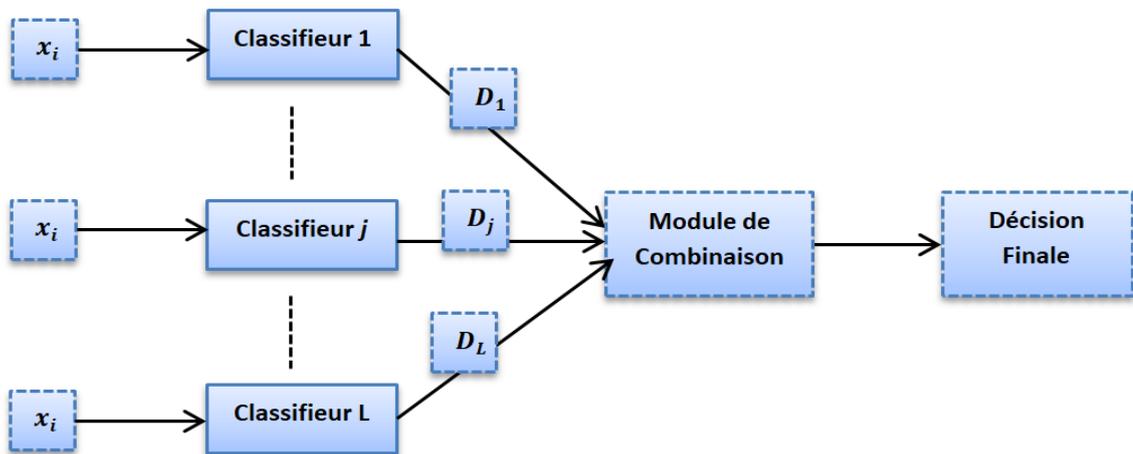
Les classifieurs de cette topologie sont organisés à des niveaux successifs de décision où chaque classifieur produit un ensemble réduit de classes possibles d'un objet à classer. Dans chaque niveau, il existe un seul classifieur qui prend en compte la décision prise auparavant afin de traiter les rejets ou confirmer la décision obtenue aux niveaux précédents (figure 1.3). L'efficacité de ce type de combinaison dépend généralement de l'ordre d'exécution des classifieurs. Les classifieurs doivent donc être ordonnés de telle sorte que les premiers classifieurs n'éliminent pas la bonne solution [Rahman 2003]. Intuitivement, les premiers classifieurs de la chaîne nécessitent d'être robustes plus que d'être performants car leur tâche est de sélectionner certaines classes candidates de sorte que la classe réelle de l'objet à classer doit apparaître dans les listes successives produites par ces classifieurs. En effet, si cette classe était éliminée dès les premiers filtrages, la décision finale serait erronée. La fin de la chaîne doit contenir des classifieurs plus performants qui ont la faculté de discriminer un nombre restreint de classes.



**Figure 1.3.** Combinaison séquentielle de classifieurs ( $x_i$  représente l'échantillon inconnue) [Zouari 2004].

### 1.4.2 Combinaison parallèle

Cette topologie est la plus communément utilisée dans les systèmes ensemblistes. Cette architecture est simple à mettre en œuvre et facilement généralisable parce qu'elle ne nécessite pas de connaître le comportement de chaque membre de l'ensemble [Zouari 2004]. Les classifieurs fonctionnent indépendamment les uns des autres puis leurs décisions sont combinées pour prédire l'étiquette d'un modèle inconnu (voir la figure 1.4). L'ordre des classifieurs n'est pas important dans ce type de configuration. La majorité des travaux les plus importants de la littérature appartiennent à cette catégorie comme : la reconnaissance de l'activité humaine [Nazabal 2016], la reconnaissance de caractères et de mots manuscrits [Ahmed 2017], l'identification de visages [Bashbaghi 2017] et la reconnaissance des images médicales [Choi 2016].

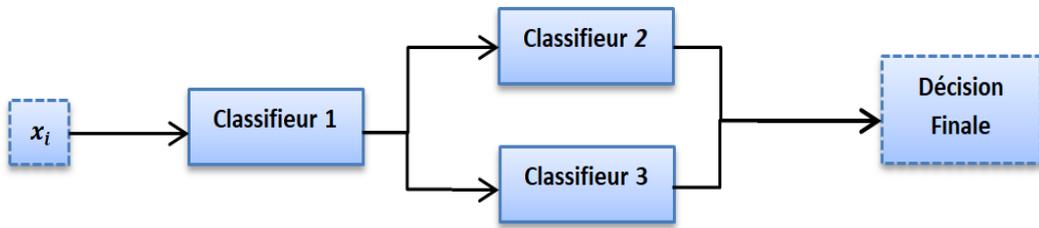


**Figure 1.4.** Combinaison parallèle de classifieurs ( $D_j$  représente la décision du classifieur  $c_j$  pour l'exemple  $x_i$ ) [Zouari 2004].

### 1.4.3 Combinaison hybride

Certains systèmes de combinaison effectuent à la fois des techniques séquentielles et parallèles afin de tirer parti des avantages de ces deux architectures simultanément. dans une architecture hybride de combinaison, les classifieurs individuels sont combinés en une structure similaire à celle d'un arbre de décision. Les nœuds d'arbre, cependant, peuvent maintenant être associés à des classifieurs complexes nécessitant un grand nombre de caractéristiques [Gosselin 1997]. Ce type de combinaison nécessite une bonne

connaissance de chacun des classifieurs utilisés de façon à choisir l'architecture la plus adaptée. La figure 1.5 fournit une représentation de la combinaison hybride de classifieurs.



**Figure 1.5.** Combinaison hybride de classifieurs [Zouari 2004].

### 1.5 Combinaison parallèle de classifieurs

Les approches séquentielles et hybrides de combinaison ont été largement ignorées par les chercheurs. La majorité des travaux dans le domaine d'ensemble de classifieurs se concentre principalement sur la combinaison parallèle de classifieurs. Plusieurs raisons peuvent expliquer l'intérêt porté par les chercheurs à la combinaison parallèle [Zouari 2004]:

- Deux membres différents du même ensemble peuvent présenter des performances globales équivalentes mais chacun d'eux est un expert dans certaines régions locales de l'espace de caractéristiques ou il est le plus performant.
- Dans un système ensembliste, il est possible d'utiliser un espace de caractéristiques de grande dimension mais en les repartant à des sous-ensembles différents ou chaque sous-espace est utilisé pour entraîner un classifieur individuel.
- La performance d'un classifieur peut être affectée par le choix de paramètres initiaux tels que le nombre de voisins ( $k$ ) pour un  $k$ PPV et les nombres de couches et de neurones par couche pour un réseau de neurones multicouches (*Multi Layer Perceptron*, MLP). Plutôt que de chercher la meilleure configuration des paramètres, la combinaison parallèle peut profiter des avantages de ces modèles appris différemment.
- Dans la combinaison, on peut avoir à notre disposition, différentes bases d'apprentissage où chacune a été capturée sur les mêmes données par différents

capteurs ayant des caractéristiques différentes. L'apprentissage d'un même modèle de classification sur ces bases peut produire des résultats différents.

Ce sont ces avantages qui nous ont conduits à focaliser notre travail sur la combinaison parallèle.

### 1.5.1 Problématique de la combinaison parallèle

Dans un système de combinaison parallèle, chaque membre participe de manière indépendante sur le même problème de classification et les sorties de tous les classifieurs sont combinée dans une dernière étape à une décision unique. Le problème principal de la combinaison est comment combiner les réponses de ces classifieurs sont combinées pour produire une décision finale ? Ce problème nécessite l'utilisation d'un opérateur de combinaison pour élaborer une réponse finale [Zouari 2004].

Considérons un système de classification composé d'un ensemble de  $L$  classifieurs  $C = [c_1, c_1, \dots, c_L]$ . Pour reconnaître la classe d'un exemple inconnu  $x_i$ , chaque classifieur  $c_j$  produit une sortie sous forme d'un vecteur des fonctions discriminantes  $c_j(x_i) = [c_{1,j}(x_i), c_{2,j}(x_i), \dots, c_{M,j}(x_i)]^T$ . Cette sortie peut être de type classe, rang ou mesure. La composante  $c_{k,j}(x_i)$  indique que le classifieur  $c_j$  a attribué la classe  $y_k$  à l'exemple  $x_i$  parmi l'ensemble de  $M$  classes possibles ( $y_k \in \Omega = \{y_1, y_2, \dots, y_M\}$ ,  $k = 1, \dots, M$ ). Toutes les réponses de l'ensemble des classifieurs pour un exemple  $x_i$  peuvent être présenter sous forme d'une matrice  $MD(x_i) = [c_1(x_i), c_2(x_i), \dots, c_L(x_i)]$ . Chaque ligne  $k$  de cette matrice représente les décisions de tous les classifieurs concernant la classe  $y_k$ . Chaque colonne  $j$  représente le vecteur de sorties d'un classifieur  $c_j$  pour l'échantillon  $x_i$ .

$$MD(x_i) = \begin{pmatrix} c_{1,1}(x_i) & \cdots & c_{1,j}(x_i) & \cdots & c_{1,L}(x_i) \\ \vdots & & \vdots & & \vdots \\ c_{k,1}(x_i) & \cdots & c_{k,j}(x_i) & \cdots & c_{k,L}(x_i) \\ \vdots & & \vdots & & \vdots \\ c_{M,1}(x_i) & \cdots & c_{M,j}(x_i) & \cdots & c_{M,L}(x_i) \end{pmatrix} \quad (1.10)$$

La décision de fusion est obtenue en combinant les vecteurs de  $s$   $c_j(x_i)$ , de chaque classifieur  $c_j$ :

$$E(x_i) = f(c_1(x_i), \dots, c_L(x_i)) \quad (1.11)$$

$f$  est appelé opérateur de fusion ou méthode de combinaison ou d'agrégation. De nombreux opérateurs de combinaison ont été proposés dans la littérature et ils ont été présentés par différentes taxonomies. Nous les passons maintenant en revue.

### 1.5.2 Catégorisation des méthodes de combinaison

Différentes taxonomies des méthodes de fusion des sorties des classifieurs ont été présentées dans la littérature [Xu 1992 ; Duin 2000 ; Jain 2000 ; Kuncheva 2000]. La première catégorisation a été proposée par Xu [1992] qui a construit une taxonomie simple basé sur le type de sorties des classifieurs (classe, rang, mesure). La faiblesse de cette catégorisation est qu'elle est trop générale et qu'au sein de chaque groupe, il existe encore une grande variété parmi les méthodes de combinaison de classifieurs.

Une autre taxonomie des méthodes de fusion basée sur la stratégie de combinaison est proposée par Kuncheva [2000] qui a fait la différence entre la fusion et la sélection de classifieurs. Rappelons que la fusion consiste à combiner toutes les réponses de classifieurs alors que la sélection consiste à choisir le meilleur sous-ensemble de classifieurs parmi l'ensemble initial pour identifier un échantillon inconnu.

Selon Duin [2000], la catégorisation des méthodes de combinaison parallèle est liée au type de classifieurs à combiner : classifieurs hétérogènes (différents) et classifieurs homogènes (faibles). En effet, même si ces dernières tiennent compte de toutes les informations produites par les classifieurs, l'intérêt des méthodes de combinaison de classifieurs faibles réside dans la combinaison de classifieurs ayant la même structure mais entraînés sur différents sous-espaces de caractéristiques ou leurs paramètres sont configurés de manière différente [Zouari 2004].

La taxonomie présenté par Jain [2000] a catégorisé les méthodes de fusion suivant trois critères d'égale importance : type de sorties des classifieurs, capacité d'exploitation des expertises des classifieurs et la capacité d'apprentissage de la combinaison. Ce travail a ouvert la voie à d'autres travaux sur l'utilisation de la capacité d'apprentissage pour séparer les méthodes de fusion comme le travail de Zouari [2004] qui a proposé une taxonomie des méthodes de fusion en fonction de deux critères: type de sorties des classifieurs à combiner et la capacité d'apprentissage.

Les méthodes avec apprentissage (ou paramétriques) permettent d'optimiser les paramètres à utiliser dans les fonctions d'agrégation sur la base d'apprentissage. Les méthodes de combinaison sans-apprentissage (ou non paramétriques) sont totalement indépendantes des données d'entraînements. Ces stratégies se contentent d'utiliser seulement les sorties des classifieurs sans intégrer d'autres informations a priori sur les performances de chacun des classifieurs, par conséquent, une fois que les classifieurs de base sont formés, l'ensemble est prêt à fonctionner. Dans les sections suivantes, nous présentons une brève description de méthodes de combinaison les plus utilisées en se basant sur la taxonomie de Zouari [2004].

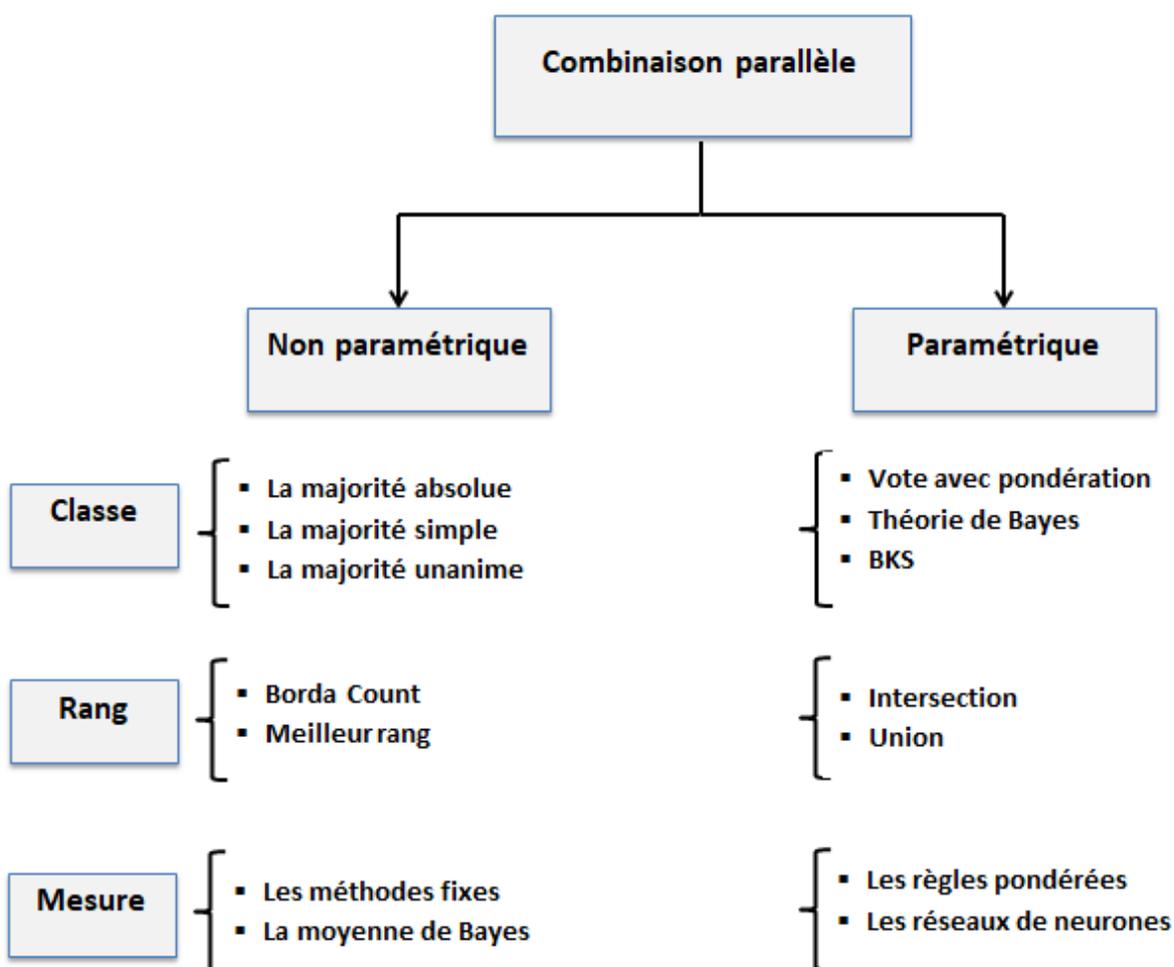


Figure 1.6. Taxonomie des méthodes de combinaison parallèle de classifieurs [Zouari 2004].

## 1.6 Méthodes de combinaison non paramétrique

Les méthodes de fusion non paramétrique (ou sans-apprentissage) sont les plus utilisées car ils sont simples à mettre en œuvre et n'utilisent que les sorties de classifieurs. L'efficacité de ces règles d'agrégation a été analysée dans plusieurs travaux [Jain 2000 ; Ruta 2000 ; Shipp 2002 ; Rahman 2003 ; Tulyakov 2008]. Comme indiqué dans [Lam 1994], le problème avec ces règles de combinaison est qu'elles traitent les classifieurs de manière égale ce qui ne permet pas de tenir compte de sa capacité individuelle.

Cette catégorie peut être également divisée en trois groupes selon le type de sorties des classifieurs: classe, rang ou mesure. Un ensemble de classifieurs avec des membres hétérogènes peut comprend des classifieurs avec des sorties différentes, alors ces résultats devront être converti à l'un des trois niveaux de catégorisation mentionnés ci-dessus.

### 1.6.1 Type classe

Les méthodes de combinaison de type classe présentent certains avantages par rapport aux méthodes de combinaison conçues pour d'autres types de sorties des classifieurs [Xu 1992]. Les sorties de type classe peuvent être dérivées de n'importe quel type de classifieur quelle que soit sa structure. Les seules méthodes à appliquer pour combiner ces résultats non paramétriques sont basées sur le principe du vote.

Les méthodes de vote [Lam 1994] sont les opérateurs de fusion les mieux connus et les plus utilisés dans les systèmes ensemblistes de classification [Srihari 1982 ; Lam 1997 ; Ruta 2001a ; Lin 2003]. Ce sont les méthodes les plus simples à mettre en œuvre ou la sortie de chaque candidat de l'ensemble est interprétée comme un vote pour l'une des classes possibles. La classe ayant un nombre de votes supérieur à un seuil préfixé est retenue comme décision finale.

Malgré la simplicité de ce type, les méthodes de vote peuvent considérablement améliorer la précision des classifieurs individuels [Ruta 2001a ; Lin 2003]. Dans [Kuncheva 2002c], cet algorithme a été comparé à de nombreux algorithmes de type mesure. Les résultats de ces comparaisons ont montré que la performance de l'algorithme de vote est presque la même que celle du meilleur algorithme de type mesure. La plupart des méthodes de vote peuvent toutes être dérivées de la règle avec seuil définie ci-dessous :

$$E(x_i) = \begin{cases} y_k & \text{Si } \max_{t=1}^M \sum_{j=1}^L c_{t,j} \geq \alpha \cdot L \\ \text{rejet} & \text{Sinon} \end{cases} \quad (1.12)$$

$\alpha$  correspond à la proportion de candidats devant fournir la même classe pour que cette dernière soit retenue comme résultat final de combinaison. Selon la valeur de  $\alpha$ , on distingue généralement trois types de vote :

#### 1.6.1.1 La majorité absolue

Dans ce type de vote, la classe avec la majorité des votes est sélectionnée comme classe candidate [Lin 2003]. Si la classe candidate gagne plus de la moitié du total des votes ( $\alpha = 0.5$ ), elle sera sélectionnée comme résultat final. Sinon, le vecteur de caractéristiques de l'exemple à reconnaître est rejeté par le système. Le rejet se produit lorsque, pour chaque classe, le nombre de votes qui le soutiennent n'est pas dans la majorité absolue, ce qui reflète une grande confusion parmi les décisions des classifieurs.

#### 1.6.1.2 La majorité simple

Pour  $\alpha = 0$ , cette règle ne nécessite pas les sorties a posteriori pour chaque classe, et chaque classifieur ne donne qu'une seule sortie en tant que vote pour une classe. Ensuite, la sortie d'ensemble est assignée à la classe avec le nombre maximum de votes parmi toutes les classes. Néanmoins, il y a un rejet si toutes les classes ont le même nombre de votes, dans ce cas, les risques de conflits sont particulièrement importants.

#### 1.6.1.3 La majorité unanime

L'ensemble de classifieurs combinés décide qu'une instance  $x_i$  provient de la classe  $y_k$  si et seulement si tous les classifieurs décident que  $x_i$  appartient à la classe  $y_k$  ( $\alpha = 1$ ) sinon il rejette  $x_i$ .

### 1.6.2 Type rang

Les méthodes de type rang combinent des listes de propositions ou la sortie de chaque classifieur est présentée par une liste des étiquettes classée par ordre décroissant selon sa préférence [Xu 1992]. Les méthodes de combinaison sans-apprentissage les plus utilisées de ce type sont le borda count et le meilleur rang.

### 1.6.2.1 Borda Count

Cette méthode peut être considérée comme une généralisation de la règle du vote [Ho 1998] dans laquelle chaque classifieur arrange les  $M$  classes selon sa préférence en attribuant un nombre de points à chacune. La classe placée au début de la liste reçoit  $M$  points, la classe suivante reçoit  $M - 1$ , et ainsi de suite. La classe placée à la fin de la liste a un seul point. Le Borda Count  $BC(y_k)$  d'une classe  $y_k$  ( $k = 1, \dots, M$ ) est la somme de tous les rangs proposés par les classifieurs pour cette classe et il est défini comme suit :

$$BC(y_k) = \sum_{j=1}^L r_{k,j} \quad (1.13)$$

$r_{k,j}$  est le rang attribué par le classifieur  $c_j$  à la classe  $y_k$ ,  $r_{k,j} = M$  si la classe  $y_k$  est placée au début de la liste.  $r_{k,j} = 1$  si la classe  $y_k$  est placée à la fin de la liste.

### 1.6.2.2 Meilleur rang

Le principe de base derrière cette méthode est le suivant : le score attribué à chacune des classes est le meilleur rang  $MR(y_k)$  obtenu parmi les rangs proposés par tous les classifieurs et chaque classe  $y_k$  est positionnée dans la liste finale en fonction de leur  $MR(y_k)$ . La classe positionnée en haut de la liste finale est choisie comme étant classe recherchée [Ho 1998].

$$MR(y_k) = \max_{j=1}^L r_{k,j} \quad (1.14)$$

Cette méthode est adaptée pour des problèmes à plusieurs classes et peu de classifieurs non corrélés. Son avantage est qu'elle utilise la meilleure réponse de chaque classifieur. Cependant, il apparaît que cette méthode peut facilement provoquer des conflits lorsque plusieurs classes obtiennent le même meilleur rang. Ce problème peut être résolu en utilisant des critères supplémentaires [Zouari 2004].

### 1.6.3 Type mesure

Le degré de confiance d'un classifieur sur l'appartenance de l'exemple à reconnaître à certaine classe fournit souvent des informations bénéfiques. Cependant, il n'a aucun impact sur le résultat final des méthodes de combinaison de types classe et rang parce que la confiance ne joue aucun rôle dans le processus de décision. Ceci conduit à utiliser des

méthodes de combinaison de type mesure qui ont capables de bénéficier de la confiance apportée à chaque classe candidate. En outre, il convient de noter que chaque classifieur peut donner des différentes valeurs de confiance avec différents paramètres tels que la gamme, l'échelle, la moyenne, etc. Toutefois, une normalisation des sorties des classifieurs est souvent nécessaire pour les rendre comparables [Huang 1995]. Les méthodes non paramétriques de type mesure les plus reconnues sont les règles " fixes " et la moyenne de Bayes.

### 1.6.3.1 Méthodes fixes

Pour reconnaître une échantillon  $x_i$ , on utilise une règle de décision  $E(x_i)$  qui revient à choisir la classe  $y_k$  qui maximise la probabilité a posteriori  $P_k$  :

$$E(x_i) = \begin{cases} y_k & \text{Si } \max_{k=1}^M P_k = \max_{m=1}^M P_m \\ \text{rejet} & \text{Sinon} \end{cases} \quad (1.15)$$

La probabilité a posteriori  $P_m$  peut être calculée par l'une des règles suivantes [Duin 2000]:

#### a) La règle Somme (Moyenne):

L'idée de base de cette règle consiste simplement à calculer la somme, ou la moyenne, des scores de confiance d'une classe particulière à travers les classifieurs de base pour obtenir le score final de cette classe.

$$P_m = \frac{1}{L} \sum_{j=1}^L m_{k,j} \quad (1.16)$$

#### b) La règle Produit :

Les mesures issues de chaque classifieur pour une classe  $y_k$  sont multipliées pour calculer le score final. Cette méthode est très sensible aux valeurs aberrantes présentées dans les sorties de classifieurs, car un score très faible ou très élevé a un effet énorme sur le résultat obtenu.

$$P_m = \prod_{j=1}^L m_{k,j} \quad (1.17)$$

c) **La règle maximum:**

Le score final est juste le maximum des scores fournis par les classifieurs. Avec cette combinaison, si un classifieur insiste sur une classe particulière pour un exemple donnée, la décision finale affecte l'exemple à cette classe; même si tous les autres classifieurs sont en désaccord.

$$P_m = \max_{j=1}^L m_{k,j} \quad (1.18)$$

d) **La règle minimum:**

Le concept de cette règle est exactement l'inverse de celle du *maximum*. On désigne donc comme score final d'une classe  $y_k$  le score le plus faible obtenu parmi l'ensemble de classifieurs.

$$P_m = \min_{j=1}^L m_{k,j} \quad (1.19)$$

e) **La règle médiane :**

Cette règle génère la sortie combinée en calculant la médiane des scores des classifieurs. la combinaison médiane est utile pour réduire l'effet des valeurs aberrantes.

$$P_m = \begin{cases} \frac{m_{k,\frac{L}{2}} + m_{k,\frac{L+1}{2}}}{2} & \text{si } L \text{ est pair} \\ m_{k,\frac{L+1}{2}} & \text{si } L \text{ est impair} \end{cases} \quad (1.20)$$

Plusieurs comparaisons ont été effectuées entre ces règles de combinaison alternatives, telles que les règles de somme et de produit [Kittler 1997a ; Bilmes 2000 ; Tax 2000]. Les travaux de [Bilmes 2000] et [Tax 2000] soulignent que la règle de produit fonctionne bien si les classifieurs sont indépendants, tandis que la règle de somme (ou moyenne) est préférée si les classifieurs combinés sont entraîné sur différentes sous-espaces de caractéristiques. Les auteurs de [Kittler 1997a] ont prouvé que les règles d'opérateurs d'ordre statique (minimum et maximum) sont plus robustes aux valeurs aberrantes que la règle somme, mais n'offrent pas généralement autant d'amélioration par rapport à la variance d'erreur.

### 1.6.3.2 Moyenne de bayes

La règle moyenne de Bayes est utilisée pour combiner les sorties des classifieurs exprimées en probabilités a posteriori [Ruta 2000]. Si chaque classifieur  $c_j$  propose une probabilité a posteriori  $P_j(x_i \in y_k / x_i)$  attribuée à la classe  $y_k$  ( $k = 1, \dots, M$ ) pour un échantillon à reconnaître  $x_i$ , il est alors possible de calculer la probabilité a posteriori moyenne de cette classe à partir de tous les classifieurs :

$$P_{moy}(x_i \in y_k / x_i) = \frac{1}{L} \sum_{j=1}^L P_j(x_i \in y_k / x_i) \quad (1.21)$$

La théorie de bayes consiste à déterminer la classe  $y_k$  pour laquelle la probabilité a posteriori moyenne  $P_{moy}(x_i \in y_k / x_i)$  est maximale. La probabilité a posteriori  $P_j(x_i \in y_k / x_i)$  peut être directement fournie par les classifieurs de type bayésien. Pour d'autres types de classifieurs, les probabilités  $P_j(x_i \in y_k / x_i)$  peuvent être déterminées comme suit :

$$P_j(x_i \in y_k / x_i) = \frac{n_{y_k, y_l}^j}{n_{., y_l}^j} \quad (1.22)$$

Où  $n_{y_k, y_l}^j$  est le nombre des exemples pour lesquels le classifieur  $c_j$  attribue les exemples de la classe  $y_k$  à la classe  $y_l$  et  $n_{., y_l}^j$  le nombre total des exemples assignés par  $c_j$  à la classe  $y_l$ .

La qualité de cette méthode dépend de l'estimation des probabilités a posteriori et de la corrélation des classifieurs utilisés [Ruta 2000]

## 1.7 Méthodes de combinaison paramétrique

Contrairement aux stratégies précédentes, les méthodes paramétriques (avec apprentissage) sont plus complexes à mettre en œuvre. Elles utilisent des paramètres supplémentaires calculés sur la base d'apprentissage. Le rationnel pour ces méthodes de fusion est qu'au lieu d'utiliser seulement une règle de combinaison fixe, la combinaison peut être adaptée à la sensibilité du problème de classification. Pour éviter d'induire un sur-apprentissage, cette catégorie nécessite d'utiliser une deuxième base de données au niveau de la combinaison et d'éviter de réutiliser la même base d'entraînement.

### 1.7.1 Type classe

Les algorithmes de vote sont peu coûteux en temps de calcul, simples à mettre en œuvre et applicables dans la plupart des problèmes de classification [Duin 2000]. Cependant, ces méthodes souffrent d'un inconvénient majeur : l'heuristique de décision est strictement démocratique, ce qui signifie que les votes des classifieurs sont toujours pondérés de façon égale et indépendamment des performances des classifieurs individuels. Par conséquent, les votes des classifieurs faibles peuvent diminuer la performance globale du système quand ils ont la majorité des votes. Pour surmonter cet inconvénient de l'algorithme de vote majoritaire, nombreuses méthodes de combinaison paramétriques de type classe ont été proposées. Les méthodes les plus utilisées sont principalement le vote pondéré, la théorie de Bayes, la méthode de Dempster-Shafer et la méthode d'espace de connaissance du comportement.

#### 1.7.1.1 Vote avec pondération

Si les classifieurs combinés ont des performances inégales, intuitivement, il est raisonnable de donner plus de pouvoir aux classifieurs les plus forts dans le vote; ceci est réalisé par un vote pondéré [Stefano 2002]. Dans cette méthode, la réponse  $c_{k,j}$  de chaque classifieur  $C_j$  est pondérée par un coefficient  $w_j$  indiquant son importance dans la combinaison.

$$E(x_i) = \begin{cases} y_k & \text{Si } \sum_{j=1}^L w_j c_{k,j} = \max_{t=1}^M \sum_{j=1}^L w_j c_{t,j} \\ \text{rejet} & \text{Sinon} \end{cases} \quad (1.23)$$

où  $w_j$  est le coefficient qui détermine l'importance attribuée au classifieur  $c_j$  dans la combinaison et  $c_{t,j}$  est la réponse du classifieur  $c_j$  pour la classe  $y_t$ . Il existe deux façons pour déterminer les poids des classifieurs :

- **Poids de performance** : Le poids  $w_j$  est égal au taux de classification du classifieur  $c_j$  sur la base d'apprentissage. Dans cette combinaison, le classifieur le plus performant obtient le poids le plus élevé. Si les classifieurs ont des performances similaires, il s'agit probablement que la règle du vote pondérée produit des résultats meilleurs que ceux de vote simple. Dans ce cas, ce système peut être considéré comme une légère modification des méthodes de vote.

- **Poids optimisés** : Les poids  $w_1, \dots, w_L$  des classifieurs  $c_1, \dots, c_L$  sont définis sur des valeurs pour lesquelles une performance optimale e l'ensemble est atteinte sur la base d'apprentissage. Pour optimiser ces poids, un algorithme génétique [Baluja 1994 ; Lam 1997 ; Günter 2004] peut être utilisé. Dans l'algorithme génétique, les génomes sont des tableaux de nombres réels compris entre 0 et 1, où la valeur en position  $j$  du tableau correspond au poids du  $j^{ieme}$  classifieur.

### 1.7.1.2 Théorie de Bayes

Cette méthode, également appelée règle de combinaison bayésienne [Suen 2000], suppose que les membres de l'ensemble sont mutuellement indépendants et estiment des probabilités a posteriori des classes. Selon le critère de Bayes, l'exemple  $x_i$  doit être assigné à la classe qui maximise la probabilité a posteriori  $P_j(x_i \in y_k / x_i)$ , c'est-à-dire:

$$E(x_i) = y_k \text{ Si } P(y_k / c_1 = y_1, \dots, c_L = y_L) = \max_{\sum_{m=1}^M} P(y_m / c_1 = y_1, \dots, c_L = y_L) \quad (1.24)$$

Pour estimer la probabilité a posteriori, cette règle supposent que les classifieurs sont indépendants [Xu 1992]. Sous cette hypothèse, la probabilité a posteriori est déterminée de la manière suivante:

$$P(y_k / c_1 = y_1, \dots, c_L = y_L) = P(y_k) \prod_{j=1}^L \frac{P(y_k / c_j = y_j)}{P(y_k)} \quad (1.25)$$

### 1.7.1.3 Méthode d'espace de connaissance du comportement

La plupart des méthodes de fusion supposent l'indépendance des décisions prises par les classifieurs individuels et ceci n'est pas toujours vrai. Pour cette raison, les auteurs [Huang 1995] ont proposé une autre méthode de type classe qui ne requise pas cette condition, appelée méthode d'espace de connaissance du comportement (*Behaviour-Knowledge Space Method, BKS*). la méthode *BKS* fournit un espace de connaissances contenant les décisions de tous les classifieurs pour chaque exemple de la base d'apprentissage.

Si le problème de fusion de décision est défini comme un mappage de  $L$  classifieurs:  $c_1, \dots, c_L$  en  $M$  classes :  $y_1, \dots, y_M$ , la méthode fonctionne sur un espace à  $L$  dimensions.

Chaque dimension correspond à la réponse d'un classifieur qui peut produire  $M + 1$  décisions ( $M$  classes et une décision de rejet). Une unité de de l'espace  $BKS$  ( $c_1, \dots, c_L$ ) est une intersection des décisions des classifieurs individuels. Elle accumule le nombre d'échantillons d'apprentissage reconnus par les  $L$  classifieurs pour chaque classe. Un exemple d'espace de  $BKS$  à 2 dimensions est présenté dans le tableau 1.1.  $(p, m)$  représente une unité focale avec  $c_1 = y_p$  et  $c_2 = y_m$ .

**Tableau 1.1.** Espace de connaissance du comportement  $BKS$  à 2 dimensions.

$C(1)/C(2)$	1	...	$m$	...	$M$
1	(1,1)	...	(1, $m$ )	...	(1, $M$ )
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$p$	( $p$ , 1)	...	( $p$ , $m$ )	...	( $p$ , $M$ )
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$M$	( $M$ , 1)	...	( $M$ , $m$ )	...	( $M$ , $M$ )

Chaque unité focale comporte trois types de d'informations :

- $n_{c_1, \dots, c_L}(y_k)$  : le nombre total d'échantillon appartenant à la classe  $y_k$ .
- $T_{c_1, \dots, c_L}$  : le nombre total d'échantillons présents dans l'unité.

$$T_{c_1, \dots, c_L} = \sum_{k=1}^M n_{c_1, \dots, c_L}(y_k) \quad (1.26)$$

- $R_{c_1, \dots, c_L}$  : La classe la plus représenté dans l'unité

$$R_{c_1, \dots, c_L} = \{y_k / n_{c_1, \dots, c_L}(y_k) = \max_{1 \leq y_m \leq M} n_{c_1, \dots, c_L}(y_k)\} \quad (1.27)$$

La méthode  $BKS$  est constituée de deux étapes : une étape d'apprentissage et une étape de décision. Dans la première étape, les données d'apprentissage sont largement exploitées pour construire le  $BKS$ . Les valeurs  $n_{c_1, \dots, c_L}$  sont déterminées après un passage complet de la base d'apprentissage. Les valeurs de  $T_{c_1, \dots, c_L}$  et  $R_{c_1, \dots, c_L}$  de chaque unité de  $BKS(c_1, \dots, c_L)$  sont ensuite calculées en utilisant les équations (1.26) et (1.27) ,

respectivement. Ensuite, la décision de classification finale pour un échantillon d'entrée repose sur l'utilisation des réponses des  $L$  classifieurs et les informations de comportement enregistrées pour choisir l'unité focale correspondante. La règle suivante est ensuite utilisée pour donner le résultat final à propos de la classe  $y_k$  :

$$E(x_i) = \begin{cases} R_{c_1, \dots, c_L} & \text{Si } T_{c_1, \dots, c_L} \geq 0 \text{ et } \frac{n_{c_1, \dots, c_L}(R_{c_1, \dots, c_L})}{T_{c_1, \dots, c_L}} \geq \lambda \\ \text{rejet} & \text{Sinon} \end{cases} \quad (1.28)$$

Où  $\lambda$  est un seuil qui contrôle la fiabilité de la décision finale. Le processus de réglage du modèle devrait inclure la recherche automatique du seuil.

L'efficacité de la méthode *BKS* reste dépendante de la taille de la base d'apprentissage. Une base de données riche et représentative est nécessaire pour estimer la distribution d'ordre élevé des sorties de classifieurs. Ainsi, cette règle de fusion souffre d'un problème de taille des échantillons. Si l'ensemble d'apprentissage est trop petit ou si la répartition des classes n'est pas uniforme, un sur-apprentissage est susceptible de se produire et la capacité de généralisation devient très faible.

### 1.7.2 Type rang

Bien que les méthodes non paramétriques de ce type sont simples à mettre en œuvre, elles ne tiennent pas compte de la différence dans la performance des classifieurs. Tous les classifieurs de cette méthode participent de manière égale dans la prise de décision même s'il existe des informations préalables sur l'efficacité de certains classifieurs. Donc, il est important d'intégrer le degré de crédibilité des classifieurs dans la combinaison. Les méthodes de type rang les plus citées dans ce cas sont l'intersection et l'union [Zouari 2004].

#### 1.7.2.1 Intersection

Cette méthode calcule l'intersection de listes des classes ordonnées par chaque classifieur. Une classe n'appartient à la liste finale que si elle existe dans toutes les listes des classifieurs. Les tailles de ces listes peuvent être différentes et elles sont déterminées par les rangs des classes réelles en apprentissage. Le rang le plus élevé est considéré comme le seuil à utiliser dans la phase de test ou de généralisation.

Un exemple de calcul [Ho 1992] de seuil est illustré dans le tableau 1.2 présentant les rangs attribués par 4 classifieurs à 5 données de la base d'apprentissage  $B^a$ . Pour un échantillon de test, les classes avant les seuils sont prises en compte dans l'intersection. Dans cette méthode, un classifieur est redondant si son seuil est égal à la taille de l'ensemble de classes.

**Tableau 1.2.** Exemple de recherche des tailles des listes des classifieurs dans l'intersection [Ho 1992 ; Zouari 2004].

$B^a / C_j$	$c_1$	$c_2$	$c_3$	$c_4$
$x_1$	13	23	12	5
$x_2$	7	9	24	17
$x_3$	4	15	6	2
$x_4$	32	1	3	16
$x_5$	4	11	5	8
Seuil	32	23	24	17

### 1.7.2.2 Union

Cette règle calcule l'union des listes proposées des rangs par les classifieurs. Une classe qui existe au moins une fois dans l'une des listes des classifieurs, appartient également à la liste finale. Les seuils de ces listes sont sélectionnés par une procédure min-max. Cette procédure est illustrée dans le tableau 1.3. La moitié gauche du tableau 1.3 présente les rangs de la vraie classe de chaque échantillon d'apprentissage. Le meilleur (minimum) rang de chaque ligne est déterminé et marqué à la case de classifieur qui le produit dans la moitié droite du tableau. Le maximum de tous ces minima est calculé pour chaque colonne. Contrairement à l'intersection, ici le nombre de classes de chaque classifieur doit être suffisamment petit pour ne pas avoir un ensemble trop important dans la liste finale [Ho 1992].

**Tableau 1.3.** Procédure min-max pour déterminer les seuils dans l'union  
[Ho 1992 ; Zouari 2004].

$B^a / C_j$	$c_1$	$c_2$	$c_3$	$c_4$		$c_1$	$c_2$	$c_3$	$c_4$
$x_1$	13	23	12	5		0	0	0	5
$x_2$	7	9	24	17		7	0	0	0
$x_3$	4	15	6	2		0	0	0	2
$x_4$	32	1	3	16		0	1	0	0
$x_5$	4	11	5	8		4	0	0	0
Colmax	/	/	/	/		7	1	0	5

### 1.7.3 Type mesure

D'autres méthodes paramétriques qui ont bénéficié de l'essor des travaux sur la classification dans plusieurs domaines sont utilisées pour combiner les classifieurs de type mesure tels que les règles pondérées et les réseaux de neurones [Zouari 2004].

#### 1.7.3.1 Les règles pondérées

##### a) La moyenne pondérée :

L'idée de base de cette règle repose sur l'application des pondérations aux sorties des classifieurs. Nous attribuons à chaque classifieur un poids en fonction de ses performances. La moyenne pondérée  $P_k$  d'une classe  $y_k$  peut être obtenue par la règle suivante:

$$P_k = \lambda \sum_{j=1}^L \omega_j m_{k,j} \quad (1.29)$$

$\omega_j$  est le coefficient qui détermine l'importance attribuée au  $j^{\text{ème}}$  classifieur dans la combinaison ( $j = 1, \dots, L$ ). Avec  $\lambda = 1/L$  . Pour  $\lambda = 1$ , il s'agit de la somme pondérée.

**b) *Le produit pondéré :***

Le concept de cette règle est exactement le même du produit non paramétrique discuté dans les règles fixes de combinaison. Les sorties des classifieurs pour chaque classe sont multipliées pour calculer le score final, mais en prenant en compte les poids comme performance du classifieur. le score final de la classe  $y_k$  est estimé comme suit:

$$P_k = \prod_{j=1}^L \omega_j m_{k,j} \quad (1.30)$$

Dans [Kittler 1997b ; Kittler 1998], l'efficacité des règles pondérées en termes de taux de classification a été montrée par rapport à d'autres méthodes comme les règles de moyenne, minimum et maximum. Lorsque les classifieurs sont utilisés sur des représentations de données identiques, les classifieurs estiment la même probabilité postérieure. Pour supprimer les erreurs dans les estimations, les sorties des classifieurs doivent être moyennées. D'un autre côté, lorsque des représentations de données indépendantes sont disponibles, les résultats des classifieurs doivent être multipliés pour tirer le maximum des représentations non-corrélés.

**1.7.3.2 *Les réseaux de neurones***

L'idée d'un opérateur d'agrégation basé sur un réseau de neurones est assez intuitive [Wolpert 1992 ; Guo 2004]. Cette méthode utilise les scores obtenus par les classifieurs comme entrées pour opérer lui-même à une classification. Un réseau de neurones conçu pour la fusion de classifieurs doit avoir une sortie nette ou alternativement un nombre de sorties égal au nombre de classes s'il est nécessaire de produire des valeurs d'assignation qualitatives pour chaque classe. Les scores des classifieurs sont calculés sur une base de validation diffère de la base d'apprentissage. La sortie de classifieurs pour chaque échantillon est considérée comme une nouvelle donnée à apprendre par le réseau de neurones ce qui lui permet d'apprendre le comportement des classifieurs [Zouari 2004].

Il est assez commun qu'un ensemble des réseaux de neurones soit combiné en utilisant un autre réseau de neurones. Dans [Lee 1995], trois réseaux de type MLP ayant des architectures différentes ont été combinés en utilisant un autre réseau de neurones. Les auteurs ont montré que le réseau de neurones donne de meilleures performances, que celles

obtenues avec la moyenne de bayes et le vote à la majorité, si les classifieurs sont dépendants.

### 1.8 Conclusion

Dans ce chapitre, nous avons présenté un état de l'art sur la combinaison de classifieurs. Nous avons rappelé ce qu'est un classifieur dans le cadre de la combinaison. Nous avons ensuite passé en revue les différentes architectures de combinaison (séquentielle, parallèle et hybride) qui se différencient essentiellement par l'agencement des classifieurs. Parmi les différentes stratégies de combinaison présentées dans ce chapitre, nous avons focalisé en particulier sur la combinaison parallèle de classifieurs qui a donné lieu aux travaux les plus importants dans le domaine de la reconnaissance de formes.

Dans le contexte de combinaison de classifieurs, les opérateurs de fusion parallèle les plus utilisées dans la littérature, qui diffèrent essentiellement par leur capacité d'apprentissage et le type de sortie des classifieurs, ont été brièvement décrites suivant la présentation de Zouari [2004].

Généralement, les méthodes de combinaison présentées dans la littérature exigent l'indépendance des décisions prises par les classifieurs combinés. L'indépendance des classifieurs est cruciale car la plupart des règles de combinaison connues ne peuvent améliorer les performances des classifieurs individuels que si elles sont utilisées avec un ensemble de classifieurs indépendants. Par conséquent, pour assurer l'efficacité d'un système ensembliste, les membres de l'ensemble doivent présenter des différences, que nous désignerons par le terme de *diversité*. Le chapitre suivant s'appuie sur le concept de la diversité et leur rôle dans les processus de génération et de sélection des classifieurs.

# Chapitre 2

## La Diversité pour la Construction et la Sélection d'Ensemble de Classifieurs

### 2.1 Introduction

Dans la classification, il a été démontré que la performance de la combinaison de classifieurs a surpassé le meilleur classifieur individuel dans un certain nombre d'applications [Radová 1997 ; Rahman 2003 ; Roller 2013 ; Sun 2017]. Cependant, la littérature a montré aussi que la diversité entre les classifieurs est un facteur important pour la réduction de l'erreur de l'ensemble globale [Dietterich 2000b ; Kuncheva 2007 ; Brown 2009 ; Brown 2010]. Intuitivement, les membres de l'ensemble doivent être différents les uns des autres, présentant des erreurs particulièrement diverses. De plus, si tous les classifieurs fournissent le même résultat, la correction d'une certaine erreur sera impossible. Par conséquent, les classifieurs individuels doivent faire des erreurs différentes sur les mêmes exemples, puis une combinaison appropriée de ces classifieurs peut alors réduire l'erreur totale.

Néanmoins, une classification très précise et fiable est nécessaire dans les applications pratiques d'apprentissage automatique et de reconnaissance de formes [Roller 2013 ; Sun 2017]. Ainsi, les membres de l'ensemble doivent être diversifiés et maintenir une certaine performance individuelle pour assurer une amélioration des performances. Par conséquent, les questions revenant les plus souvent sont savoir comment générer cette diversité lorsque tous les classifieurs sont basés sur le même algorithme et entraînés sur les mêmes données d'apprentissage ? et comment la mesurer afin d'établir le compromis parfait entre la corrélation et la précision ? Plusieurs travaux [Kittler 1998 ; Golestani 2007 ;

Woloszynski 2009] ont montré que la diversité peut être créée entre des modèles de classification distincts, essayant de résoudre le même problème en utilisant différents principes, concepts, représentations ou même méthodologies.

Lorsqu'un certain nombre de classifieurs a été généré, l'ensemble final peut contenir des membres identiques et imprécis, ce qui dégrade la performance de classification. L'étape de sélection d'un sous-ensemble de classifieurs joue un rôle important durant le processus de construction des systèmes ensemblistes [Ruta 2005]. La sélection de classifieurs peut être définie comme le processus de choix d'un sous-ensemble optimal de classifieurs, suffisamment qualifié pour prédire la classe d'un échantillon de test. Dans la littérature, les travaux qui traitent le problème de sélection de classifieurs peuvent être répartis en deux catégories: statique et dynamique. Dans les approches statiques [Giacinto 2001a; Santos 2006; Cruz 2013], la sélection est effectuée pendant la phase d'apprentissage du système. Ensuite, le classifieur ou l'EdC sélectionné est utilisé pour classer tous les échantillons de test. A l'opposé, les approches dynamiques [Ko 2008; Dos Santos 2008; Xiao 2012; Cavalin 2013; Bashbaghi 2017] sélectionnent un classifieur ou un EdC différent pour chaque exemple de test.

Au sein de ce chapitre, nous décrivons le problème lié à la diversité de classifieurs en abordant l'axe de construction d'ensemble de classifieurs et illustrant les différentes stratégies existantes. Nous présentons ensuite les différents moyens issus de la littérature pour quantifier cette diversité. Ce chapitre permettra également de réaliser un état de l'art des différentes stratégies de sélection de classifieurs.

### **2.2 Diversité pour génération des classifieurs**

La construction d'un ensemble de classifieurs peut être réalisée en adoptant différentes stratégies [Wang 2006]. La diversité entre les classifieurs peut être obtenue en utilisant (1) différents modèles de classification et la même base d'apprentissage; (2) le même modèle de classification et différentes données d'apprentissage; (3) différents modèles utilisant différents types de sorties de classifieur.

Certains auteurs [Golestani 2007] ont proposé de diviser ces méthodes de création en différentes catégories. Sharkey [1997] a montré que les quatre aspects suivants peuvent être manipulés pour créer des ensembles de réseaux de neurones: les paramètres initiaux, les données d'apprentissage, la topologie des réseaux et l'algorithme d'apprentissage. Plus

récemment, Gacquer [2008a] a proposé que les méthodes de création peuvent être catégorisées en trois groupes selon les différents niveaux de la construction :

- Le niveau *Données* : les différences entre les classifieurs à ce niveau s'effectuent sur les exemples d'apprentissage eux-mêmes. Des traitements peuvent être appliqués de différentes façons à la base de données d'apprentissage en utilisant différentes sources de données, différentes méthodes de prétraitement ou différentes méthodes d'échantillonnage.
- Le niveau *Classifieur* : il désigne les modifications apportées à l'algorithme d'apprentissage utilisé en lui-même. Ces modifications peuvent être appliquées en utilisant différents types de classifieurs [Gacquer 2008b], différentes architectures des classifieurs [Galán 2002] et différentes initialisations de paramètres d'apprentissage.
- Le niveau *Superviseur* : ce niveau concerne la méthode d'agrégation utilisée pour fusionner les réponses des membres de l'ensemble afin d'obtenir une décision final.

Dans les sous sections suivantes, nous présentons les principales méthodes de création d'un ensemble de classifieurs qui sont divisées en quatre groupes suivant les catégorisations de Gacquer [2008a] et de Sharkey [1997].

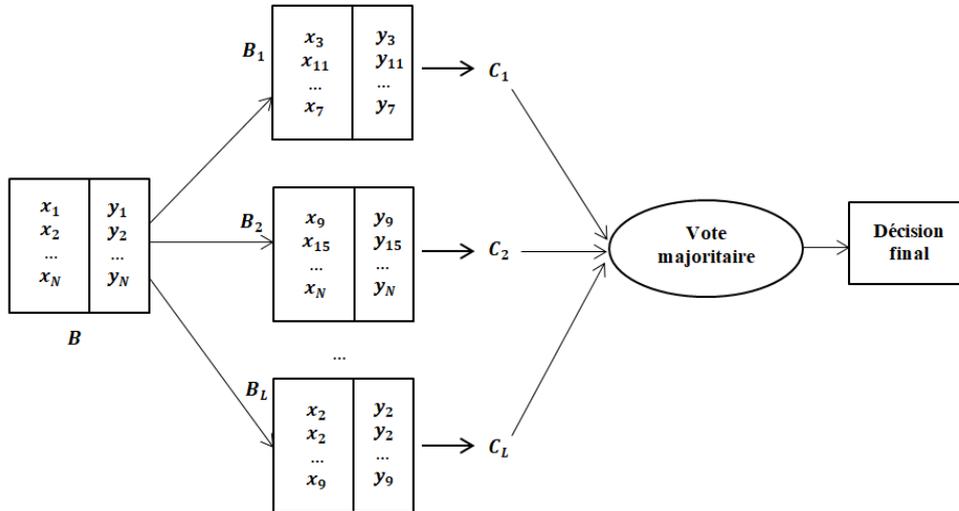
### 2.2.1 Échantillonner les données d'apprentissage

Ce groupe contient des méthodes qui construisent des ensembles de classifieurs en faisant varier les données d'apprentissage afin de générer différentes sous bases d'entraînement des membres de l'ensemble. Les méthodes de construction ensembliste suivantes sont des exemples de ce type d'approche.

#### 2.2.1.1 Bagging

Le Bagging [Breiman 1996a], signifiant « Bootstrap Aggregating », a été l'une des premières méthodes proposées pour la création d'un ensemble de classifieurs. La technique Bagging redéfinit l'ensemble d'apprentissage pour chaque classifieur. Elle consiste à générer plusieurs versions de la même base de données d'apprentissage (appelés bootstraps) et utiliser chacun d'eux pour entraîner un membre de l'ensemble en utilisant le même algorithme d'apprentissage. Les différents bootstraps sont construits par tirages aléatoires avec remise des exemples de l'ensemble de données d'apprentissage d'origine.

Ainsi, les sorties des classifieurs sont ensuite combinées par un vote majoritaire ou tout autre règle de fusion. . La figure 2.1 illustre le principe de cet algorithme. Il est prévu que 63,2% des données originales d'apprentissage seront inclus dans chaque bootstrap [Breiman 1996a].



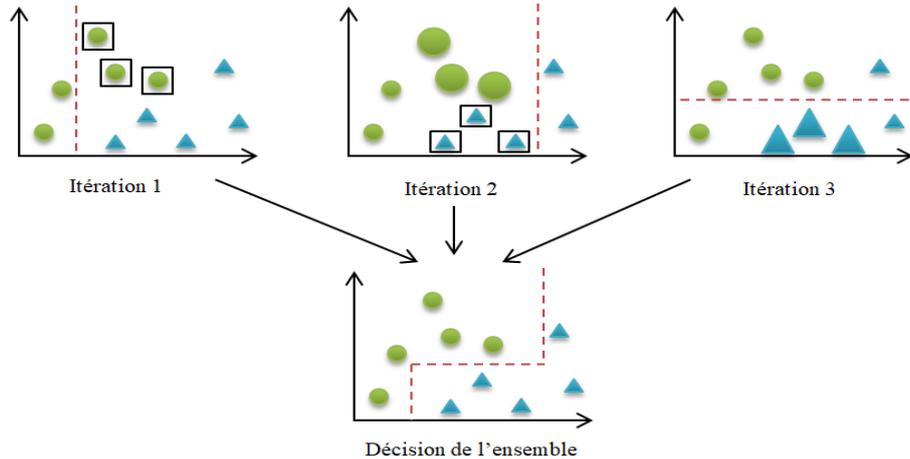
**Figure 2.1.** Description schématique de l'apprentissage et la combinaison d'un ensemble de classifieurs par la méthode Bagging [Gacquer 2008a].

Breiman [1996a ; 1996b ; 1997] signale qu'il est important que le type de classifieur de base doit être instable pour que le Bagging fonctionne efficacement. Un algorithme de classification instable est sensible aux changements des données d'apprentissage. Une légère modification sur la base d'apprentissage peut avoir un impact important sur le comportement d'un classifieur instable.

### 2.2.1.2 Boosting

Boosting est un algorithme itératif attribuant des poids à chaque exemple de la base d'apprentissage afin de se focaliser sur les exemples « difficiles » (i.e. les exemples mal classés par les classifieurs précédents). À chaque itération, l'algorithme de Boosting augmente les poids des échantillons mal classés de sorte que le classifieur suivant se concentre en priorité sur ces exemples difficiles. La décision finale est donc obtenue par la combinaison des sorties de tous les classifieurs générés à chaque itération. Ces classifieurs sont généralement combinées par un vote majoritaire ou par un vote pondéré. La figure 2.2 décrit le fonctionnement de cet algorithme.

Plusieurs études comparatives [Breiman 1996a ; Schapire 1998 ; Bauer 1999] ont montré que le Boosting exige moins d'instabilité que le Bagging parce que cette dernière permet d'engendrer de beaucoup plus grands changements dans l'ensemble d'apprentissage.



**Figure 2.2.** Apprentissage et agrégation de classifieurs par la méthode Boosting [Gacquer 2008a].

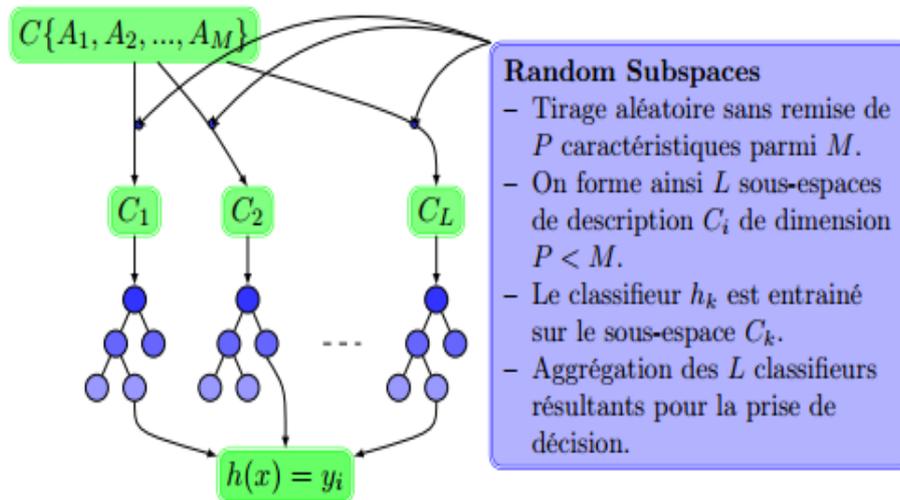
## 2.2.2 Modifier l'espace de représentation

Ces méthodes construisent des ensembles de classifieurs en modifiant l'ensemble original de caractéristiques. L'objectif est de fournir une vue partielle et différente des données d'apprentissage à chaque membre de l'ensemble pour les amener à être différents les uns des autres [Tsymbal 2005]. L'une des méthodes ensemblistes les plus réussies dans cette catégorie est la méthode proposée par Ho [1998], nommée *Random Subspace*. Cette approche consiste à modifier les données d'apprentissage comme le Bagging et le Boosting. Cependant, ces modifications sont réalisées sur l'espace des caractéristiques.

### 2.2.2.1 *Random Subspace*

Random Subspace consiste à construire un ensemble de classifieurs sur la base d'une méthode de classification et d'une sélection aléatoire des sous-espaces de représentation. Supposons que  $F$  est la dimension de l'espace de représentation initial, la méthode RSS consiste à tirer aléatoirement  $D$  membres parmi  $F$  afin d'obtenir un sous-espace de  $D$  caractéristiques ( $D < F$ ). La dimension  $D$  d'un sous-espace de représentation est fixe et chaque caractéristique ne peut être sélectionnée qu'une seule fois. Chaque sous-espace de

représentation est utilisé pour entraîner un classifieur individuel en utilisant le même algorithme d'apprentissage. Les sorties des classifieurs sont généralement combinés par la règle de vote majoritaire. Ho [1998] a montré que l'utilisation d'un grand nombre de classifieurs reposant sur des espaces de petite dimension peut être préférable à l'apprentissage d'un classifieur unique dans l'espace initial. La figure 2.3 illustre l'application de cette procédure à l'induction de forêts de décision [Bernard 2009].



**Figure 2.3.** Illustration du principe de la méthode Random Subspace pour un ensemble d'arbres de décision [Bernard 2009].

### 2.2.3 Manipuler les membres de l'ensemble

Dans les méthodes de ce niveau, la structure ou les paramètres de l'architecture du classifieur sont variés. Ces méthodes tentent de générer des classifieurs diversifiés en utilisant différents modèles de classification [Ruta 2005], différentes architectures de classifieurs [Roli 2001] ou différentes configurations des paramètres de l'algorithme d'apprentissage [Zhu 2004], tout en conservant les mêmes données d'apprentissage. Par exemple, différents poids initiaux peuvent être attribués à des réseaux de neurones individuels [Kolen 1991] ou différentes sélections fractionnées peuvent être appliquées aux arbres de décision [Kwok 1990 ; Liu 2008], etc. afin de composer un ensemble de 15 classifieurs hétérogènes, Ruta et Gabrys [Ruta 2005] ont utilisé 15 algorithmes d'apprentissage, à savoir : le discriminant quadratique, les réseaux de neurones, le kPPV, les arbres de décision et d'autres.

## 2.2.4 Manipuler les sorties

Ce groupe contient des méthodes basées sur la manipulation des sorties des exemples d'apprentissage. Ces méthodes tentent de construire un ensemble de divers classifieurs en utilisant différentes représentations des sorties. Dans la technique, de codage des sorties, utilisée par Dietterich et Bakiri [Dietterich 1994], un problème multi-classes est transformé en un ensemble de problèmes binaires. À chaque itération, une nouvelle division binaire de la base de données d'apprentissage est utilisée pour former un nouveau classifieur. Un autre exemple est la méthode proposée par Breiman [2000], qui introduit le bruit pour modifier certaines classes des exemples d'apprentissage.

## 2.3 Quantification de la diversité

D'un point de vue technique, la notion de diversité ne peut être évidente que s'elle est soutenue par des mesures appropriées permettant de quantifier la dépendance entre les classifieurs. Différentes mesures statistiques pour la quantification de la diversité ont été étudiées par Kuncheva et Whitaker [Kuncheva 2003a]. Ces mesures ont ouvert la voie à plusieurs travaux sur l'utilisation de la diversité de manière explicite durant le processus de construction d'ensembles de classifieurs. Les mesures de diversité peuvent être utilisées en tant qu'outil théorique, outil de visualisation ou comme un critère de sélection de classifieurs [Gacquer 2008a].

Il existe de nombreuses mesures de diversité dans la littérature qui peuvent être classées en deux catégories : les mesures pairwise ou par-paires, qui se calculent pour chaque paire de classifieurs et les mesures globales ou non-pairwise, qui quantifient la diversité de tout l'ensemble de classifieurs.

### 2.3.1 Mesures pairwise (par-paires)

Soit  $A = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_M)\}$  un ensemble d'apprentissage de  $N$  objets et  $C = [c_1, c_1, \dots, c_L]$ , un ensemble de  $L$  classifieurs. Dans le cas des mesures par-paires, les sorties de chaque classifieur  $c_j$  sont représentées sous forme d'un vecteur binaire de taille  $N$  ( $N$  est le nombre de données de validation)  $S_j = [S_1^j, \dots, S_N^j]$ , sachant que  $S_n^j = 1$  si le classifieur  $c_j$  propose la bonne solution pour l'exemple  $x_n$ , et 0 sinon ( $n = 1$  à  $N, j = 1, \dots, L$ ). La diversité globale d'un ensemble de  $L$  classifieurs est obtenue en faisant la moyenne des mesures de diversité calculées entre chaque couple de classifieurs [Kuncheva

2003a]. La moyenne de diversité  $M$  des mesures "pairwise"  $M_{i,j}$  pour toutes les paires  $(c_i, c_j)$  est calculée par :

$$M = \frac{2}{N * L(L - 1)} \sum_{i=1}^{L-1} \sum_{j=i+1}^L M_{i,j} \quad (2.1)$$

La relation entre chaque paire de classifieurs  $(c_i, c_j)$  peut alors être décrite par un tableau de contingence (voir tableau 2.1). Les termes de ce tableau correspondent aux nombres des exemples pour lesquelles les classifieurs  $c_i$  et  $c_j$  sont en accord ou en désaccord concernant la classe attribuée à les mêmes exemples, avec  $N^{00} + N^{01} + N^{10} + N^{11} = N$ . Par exemple,  $N^{01}$  est le nombre de sorties pour lesquelles le classifieur  $c_i$  est correcte ( $S_n^i = 1$ ) et le classifieur  $c_j$  est incorrect ( $S_n^j = 0$ ). Sur la base de ce tableau, plusieurs mesures peuvent être calculées pour deux classifieurs donnés.

**Tableau 2.1.** Relation entre les sorties de deux classifieurs  $c_i$  et  $c_j$ .

	$c_j$ correct (1)	$c_j$ incorrect (0)
$c_i$ correct (1)	$N^{11}$	$N^{01}$
$c_i$ incorrect (0)	$N^{10}$	$N^{00}$

### 2.3.1.1 $Q$ -statistique ( $Q$ )

La mesure  $Q$ -statistique proposée par Yule [1900] pour deux classifieurs  $c_i$  et  $c_j$  est définie par :

$$Q_{i,j} = \frac{N^{11}N^{00} - N^{01}N^{10}}{N^{11}N^{00} + N^{01}N^{10}} \quad (2.2)$$

On peut voir que  $Q_{i,j}$  varie entre  $-1$  et  $1$ .  $Q_{i,j}=0$  si  $c_i$  et  $c_j$  sont statistiquement indépendants;  $Q_{i,j} = 1$  indique que  $c_i$  et  $c_j$  proposent des prédictions similaires;  $Q_{i,j} = -1$  indique que  $c_i$  et  $c_j$  proposent des prédictions différentes.

### 2.3.1.2 Coefficient de corrélation ( $\rho$ )

La corrélation  $\rho$  entre deux classifieurs  $c_i$  et  $c_j$  est définie par [Kuncheva 2003a]:

$$\rho_{i,j} = \frac{N^{11}N^{00} - N^{01}N^{10}}{\sqrt{(N^{11} + N^{10})(N^{01} + N^{00})(N^{11} + N^{01})(N^{10} + N^{00})}} \quad (2.3)$$

Pour deux classifieurs donnés,  $Q$  et  $\rho$  ont le même signe et  $|\rho| \leq |Q|$ .

### 2.3.1.3 Mesure de désaccord ( $D$ )

La mesure de désaccord, notée  $D_{i,j}$ , a été utilisée par Skalak [1996] pour déterminer la diversité entre deux classifieurs. Elle correspond au rapport entre le nombre des exemples pour lesquels les classifieurs  $c_i$  et  $c_j$  n'étaient pas d'accord, et le nombre total des exemples  $N$ . Cela peut être écrit comme suit:

$$D_{i,j} = \frac{N^{01} + N^{10}}{N^{00} + N^{01} + N^{10} + N^{11}} \quad (2.4)$$

$D \in [0, 1]$  ; Une valeur élevée de  $D$  indique une diversité accrue.

### 2.3.1.4 Double-faute ( $DF$ )

Cette mesure représente le pourcentage d'échantillons qui sont mal classés par les deux classifieurs  $c_i$  et  $c_j$ . La diversité augmente avec la diminution de la valeur de  $DF$  [Giacinto 2001a].

$$DF_{i,j} = \frac{N^{00}}{N^{00} + N^{01} + N^{10} + N^{11}} \quad (2.5)$$

### 2.3.1.5 Mesure de kappa ( $k$ )

La mesure de kappa  $k$  introduite par Cohen [1960] a été utilisée par [Margineantu 1997 ; Dietterich 2000b] en tant que mesure de diversité pour évaluer le désaccord respectif entre deux prédicteurs. La mesure kappa prend en compte des informations supplémentaires. Soit  $N^{p,*}$ , le nombre d'exemples de la classe  $y_p$  reconnus par le premier classifieur,  $N^{*,p}$ , le nombre d'exemples reconnus de la classe  $y_p$  par le deuxième classifieur et  $N^{p,p}$  le nombre d'exemples de la classe  $y_p$  reconnus par les deux classifieurs en même temps. On peut alors définir  $k$  comme suit:

$$k_{i,j} = \frac{\theta_1 - \theta_2}{1 - \theta_2} \quad (2.6)$$

$\theta_1$  estime l'accord observé des deux classifieurs et  $\theta_2$  estime l'accord par chance (la probabilité que les deux classifieurs soient d'accord simplement par chance).

$$\theta_1 = \frac{\sum_{p=1}^M N^{p,p}}{N} \quad (2.7)$$

$$\theta_2 = \sum_{p=1}^M \left( \frac{N^{p,*}}{N} \cdot \frac{N^{*,p}}{N} \right) \quad (2.8)$$

La valeur de kappa est comprise entre -1 et 1.  $k = -1$  indique un grand désaccord entre  $c_i$  et  $c_j$  et par conséquent une diversité élevée et  $k = 1$  indique le contraire.

### 2.3.1.6 Weighted Count of Errors and Correct results (WCEC)

Cette mesure tient en compte en même temps les résultats corrects et incorrects des classifieurs. Une valeur élevée de *WCEC* signifie que les classifieurs sont différents.

$$WCEC_{i,j} = N^{11} + \frac{1}{2}(N^{01} + N^{10}) - N_{different}^{00} - 5N_{meme}^{00} \quad (2.9)$$

Où,  $N_{different}^{00}$  représente le nombre d'échantillons incorrectement classés par les deux classifieurs avec des erreurs différentes, et  $N_{meme}^{00}$  est le nombre d'échantillons incorrectement classés par les deux classifieurs, mais avec les mêmes résultats de classification.

### 2.3.2 Mesures globales (non- pairwise)

Les mesures globales tentent d'évaluer la diversité de l'ensemble directement plutôt que de faire la moyenne des mesures par-paires.

#### 2.3.2.1 L'entropie (E)

La mesure d'entropie [Cunningham 2000] repose sur le constat, que pour un exemple donné  $x_i$ , le désaccord sera maximisé s'il y a une égalité dans les votes des  $L$  classifieurs individuels. C'est-à-dire  $\left\lfloor \frac{L}{2} \right\rfloor$  classifieurs attribuent la bonne réponse à  $x_i$  et les  $L - \left\lfloor \frac{L}{2} \right\rfloor$  restants sont incorrects. L'entropie peut être calculée comme suit:

$$E = \frac{1}{N} \frac{2}{L-1} \sum_{i=1}^N \min\{\bar{L}(x_i), L - \bar{L}(x_i)\} \quad (2.10)$$

Avec,  $\bar{L}(x_i)$  est le nombre de classifieurs proposant, pour un échantillon à reconnaître  $x_i$  ( $i$  de 1 à  $N$ ), une réponse correcte.  $E \in [0, 1]$ ,  $E = 1$  indique une diversité maximale de l'ensemble et  $E = 0$  indique les classifieurs sont identiques.

### 2.3.2.2 Difficulté ( $\theta$ )

Le principe est de définir une variable aléatoire discrète  $X$  ayant une valeur dans  $\{\frac{0}{L}, \frac{1}{L}, \dots, 1\}$  indiquant la proportion de classifieurs de  $C$  qui classifient correctement un exemple  $x_i$  tiré aléatoirement de la distribution du problème [Hansen 1990].

$$\theta(C) = var(X) = \sum_{j=1}^L \left(\frac{L^*}{L} - \bar{p}\right)^2 p\left(\frac{L^*}{L}\right) \quad (2.11)$$

$\bar{p}$  est la moyenne des taux de classification des classifieurs.  $p\left(\frac{L^*}{L}\right)$  est la probabilité que exactement  $L^*$  classifieurs parmi les  $L$  proposent la bonne réponse pour un échantillon de test sélectionné aléatoirement.  $\theta = 1$  indique que les classifieurs sont identiques.

### 2.3.2.3 Diversité généralisé (DG)

Soient  $p(1)$  la probabilité qu'un classifieur tiré au hasard échoue sur un échantillon choisi aléatoirement et  $p(2)$  la probabilité qu'au moins deux classifieurs tirés aléatoirement dans l'ensemble, échouent sur un échantillon tiré au hasard [Partridge 1996].

$$DG(C) = 1 - \frac{p(2)}{p(1)} \quad (2.12)$$

$$p(1) = \sum_{k=1}^L \frac{k}{L} p_k, \quad (2.13)$$

$$p(2) = \sum_{k=1}^L \frac{k}{L} \frac{k-1}{L-1} p_k \quad (2.14)$$

$p_k$  indique la probabilité que exactement  $k$  classifieurs parmi  $L$  choisis aléatoirement proposent des réponses incorrectes pour un exemple de test sélectionnée aléatoirement.

$DG = 1$  signifie que la diversité est maximale. Dans ce cas, la probabilité que les deux classifieurs proposent une réponse incorrecte pour le même exemple à reconnaître est  $p(2) = 0$ .  $DG = 0$  signifie que la diversité est minimale c'est à dire que la probabilité qu'un classifieur tiré aléatoirement échoue  $p(1)$  est la même que la probabilité que deux classifieurs se trompent  $p(2)$  [Zouari 2004].

#### 2.3.2.4 Coincident Failure (CF)

Cette mesure [Partridge 1997] est simplement une modification de la diversité généralisée  $DG$ , afin d'obtenir une valeur minimale de 0 lorsque tous les classifieurs ont toujours raison, ou lorsqu'ils donnent tous simultanément soit une réponse juste, soit une réponse fausse [Zouari 2004]:

$$CF = \begin{cases} \frac{1}{1-p_0} \sum_{j=1}^L \frac{L-j}{L-1} p_j & \text{si } p_0 < 1 \\ 0 & \text{si } p_0 = 1 \end{cases} \quad (2.15)$$

Le tableau 2.2 résume les 10 mesures de diversité présentées ci-dessus. Le tableau montre si une mesure est par-paire ou non et si elle nécessite de connaître la précision des classifieurs.

### 2.3.3 Limites de mesures de diversité

Tang et al. [Tang 2006] ont analysé théoriquement six mesures de diversité. Ils ont montré que si la précision moyenne des apprenants individuels est fixe et que la diversité maximale est réalisable, la maximisation de la diversité parmi les apprenants individuels est équivalente à la maximisation de la marge minimale de l'ensemble sur les exemples d'apprentissage. Cependant, ils ont montré empiriquement que la diversité maximale n'est pas réalisable généralement, et que la marge minimale d'un ensemble n'augmente pas de façon monotone par rapport aux mesures de diversité existantes. En particulier, Tang et al. [Tang 2006] ont prouvé que la modification des mesures de diversité ne permet pas de savoir si un ensemble atteint une bonne performance de généralisation.

Les études exhaustives de Kuncheva et Whitaker [Kuncheva 2003a] sur la diversité ont montré que l'efficacité des mesures de diversité est décourageante car il est constaté qu'il n'existe aucune relation claire entre ces mesures et la performance globale de l'ensemble.

Ruta [2001a] a montré également que les mesures de diversité n'aident pas directement à créer des ensembles performants.

**Tableau 2.2.** Résumé des mesures de diversité, où  $\uparrow$  ( $\downarrow$ ) indique que plus la mesure est grande (plus petite), plus la diversité est grande, "Connu" indique s'il nécessite de connaître la précision des classifieurs individuels.

Mesure de diversité	Symbole	$\uparrow/\downarrow$	Par-paire	Connu
Q-statistique	$Q$	$\downarrow$	Oui	Non
Corrélation	$\rho$	$\downarrow$	Oui	Non
Désaccord	$D$	$\uparrow$	Oui	Non
Double-faute	$DF$	$\downarrow$	Oui	Oui
Kappa	$K$	$\downarrow$	Oui	Non
Weighted Count of Errors and Correct results	$WCEC$	$\uparrow$	Oui	Non
Entropie	$E$	$\uparrow$	Non	Non
Difficulté	$\theta$	$\downarrow$	Non	Oui
Diversité généralisé	$DG$	$\uparrow$	Non	Oui
Coincident Failure	$CF$	$\uparrow$	Non	Oui

## 2.4 Sélection de classifieurs

Comme nous le constatons, plusieurs mesures de diversité ont été proposées dans la littérature mais ces mesures ne permettent pas de donner une idée claire sur la relation entre la performance de l'ensemble et la diversité de ces membres. Pour cette raison, d'autres travaux ont cherché de clarifier le lien existant entre la précision et la diversité, en abordant la construction d'un ensemble de classifieur par le choix d'un sous-ensemble optimal de ces derniers [Aksela 2006 ; Azizi 2010 ; Yang 2011 ; Cheriguene 2016d ;

Visentini 2016]. Le processus de recherche d'un tel sous-ensemble s'appelle *sélection de classifieurs*.

Le paradigme de sélection de classifieurs consiste à choisir, parmi un ensemble de grande taille appelé *pool de classifieurs*, le sous-ensemble de classifieurs qui favorisent la précision, la diversité, ou toute combinaison de ces paramètres. Ainsi, la sélection de classifieurs vise à rechercher le meilleur sous-ensemble de classifieurs, dont leurs combinaison produit le meilleur résultat. La sélection de classifieurs est une stratégie utile pour réduire la taille d'un EdC, en choisissant seulement une fraction de classifieurs de l'ensemble original. Cette stratégie a pour but de pouvoir améliorer la performance et l'efficacité d'un système ensembliste en réduisant l'espace de stockage et le coût de calcul [Yang 2011 ; Cheriguene 2016e].

Généralement, les techniques de sélection de classifieurs sont réparties en deux catégories : sélection statique et sélection dynamique. Selon le premier type [Dos Santos 2006 ; Partalas 2008], le meilleur classifieur ou le meilleur sous-ensemble de classifieurs est choisi pendant la phase d'apprentissage et sera utilisé lors de la combinaison. Dans le deuxième type, le choix est fait dynamiquement durant la phase de généralisation, en sélectionnant un EdC différent pour chaque échantillon de test [Cruz 2011]. Les différentes stratégies de sélection de classifieurs sont décrites dans les sections suivantes.

### 2.5 Sélection statique de classifieurs

Les techniques de sélection statique de classifieurs (Static Classifier Selection, SCS) proposées dans la littérature peuvent être divisées en deux groupes généraux selon la stratégie de conception:

#### 2.5.1 Regrouper et extraire

Cette approche se base sur le même principe de clustering et fonctionne en deux étapes. La première étape consiste à partitionner l'ensemble initial de classifieurs à des sous-ensembles qui diffèrent selon le niveau de diversité. L'objectif de cette phase est de regrouper les classifieurs de telle façon que la combinaison des membres du même groupe ou cluster permet d'avoir un pourcentage d'erreur élevé. Ceci permet de dire que les classifieurs de chaque sous-ensemble sont très corrélés. Dans la deuxième étape, un classifieur représentatif de chaque sous-ensemble ou groupe est sélectionné de façon que la combinaison de classifieurs représentatifs permette de minimiser le taux d'erreur globale.

Différentes techniques ont été exploitées pour cet objectif. Par exemple, les auteurs de [Lazarevic 2001] ont utilisé le clustering k-moyenne basé sur la distance euclidienne et ont retiré itérativement les apprenants individuels du moins au plus précis de chaque groupe jusqu'à ce que la précision de l'ensemble commence à diminuer. La méthode proposée par Bakker [2003] consiste à utiliser un recuit déterministe pour le clustering et sélectionner le centre de chaque groupe.

D'autres travaux ont utilisé l'algorithme de clustering hiérarchique [Partridge 1996 ; Giacinto 2000 ; Ripley 2007]. Cet algorithme commence par placer chaque classifieur  $c_j$  dans un groupe individuel  $G_j$  et une matrice de dissimilarité  $D$  de dimension  $L \times L$  entre les groupes est définie avec les éléments  $D(i, j) = dis(G_i, G_j)$ .

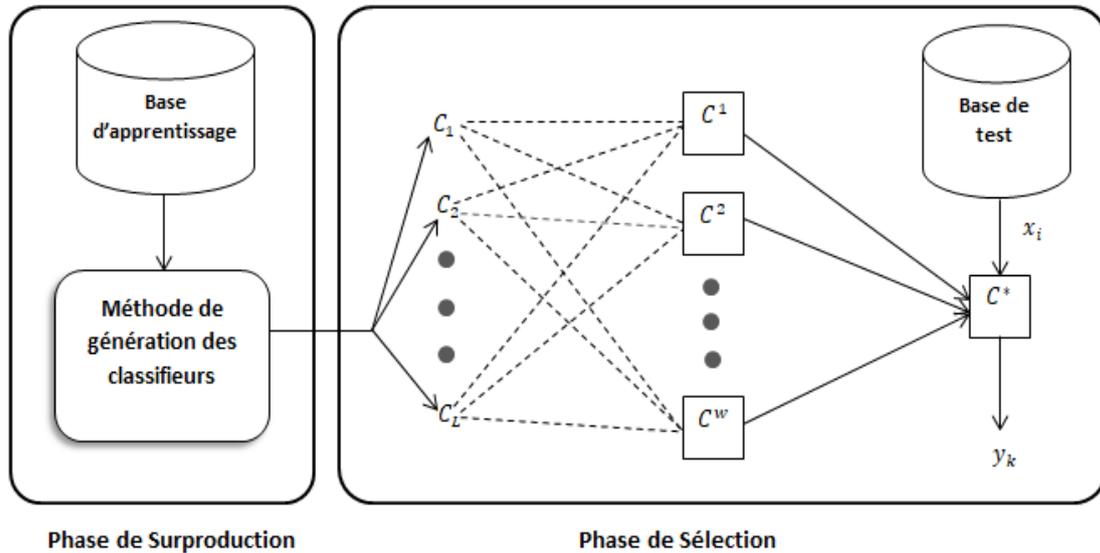
A chaque étape, les deux groupes les plus proches sont rassemblés dans un même groupe  $G_h$ . La matrice de dissimilarité  $D$  est mise à jour pour refléter la suppression de  $G_i$  and  $G_j$ , ainsi que les nouvelles dissimilarités entre  $G_h$  et tous les clusters restants  $G_k$  ( $k \neq i, j$ ). Ici, le sens du plus proche change d'une étape à une autre puisque les distances entre les groupes deviennent plus grandes. Ce processus est répété jusqu'à ce qu'on obtienne un seul groupe. Il est possible ensuite de choisir le sous-ensemble optimal en sélectionnant un classifieur de chaque groupe (celui dont la distance moyenne avec les autres groupes est maximale) [Zouari 2004].

Dans [Partridge 1996], deux approches basées sur principe ont été proposées. La première méthode consiste à sélectionner le classifieur qui a la performance la plus élevée dans une certaine région de l'espace. L'autre méthode consiste à sélectionner le sous-ensemble de classifieurs qui ont les performances les plus élevées parmi l'ensemble de tous les classifieurs disponibles. Les résultats de comparaison de ces méthodes avec l'approche de Giacinto [2000] ont montré que cette dernière permet de sélectionner plus de classifieurs indépendants que les deux autres approches.

### 2.5.2 Surproduire et choisir

La plupart des méthodes proposées de la sélection statique dans la littérature sont basés sur le paradigme de *surproduction et choix* (*Overproduce and Select Strategy*, OCS). L'objectif de ce paradigme est de trouver le sous-ensemble de prédicteurs les plus pertinents, en supposant que les classifieurs de l'ensemble original sont redondants [Zhou 2002]. Une fois que le meilleur sous-ensemble de classifieurs a été sélectionné, les sorties

de ses membres doivent être combinées. Les méthodes basées sur OCS sont divisées en deux phases [Sharkey 2000]: (1) surproduction; et (2) sélection. La figure 2.4 illustre les phases d'OCS.



**Figure 2.4.** Schéma général du processus OCS. OCS est divisé en deux phases de surproduction et de sélection [Dos Santos 2011].

La phase de surproduction consiste à générer un grand nombre de classifieurs pouvant être obtenus par les différentes stratégies énoncées à la section 2.2, en utilisant la base de données d'apprentissage. Tandis que la phase de sélection vise à tester différents sous-ensembles de classifieurs afin de sélectionner l'EdC le plus performant, qui est ensuite utilisé pour classer tous les échantillons de test.

La sélection du sous-ensemble optimal de classifieurs consiste à rechercher dans toutes les possibilités de combinaison des classifieurs disponibles. Le problème majeur de cette approche est que le nombre de combinaisons croît exponentiellement en fonction du nombre de classifieurs [Álvarez 2015]. Cependant, pour un ensemble de  $L$  classifieurs obtenus à l'étape de surproduction, le nombre de sous-ensembles possibles est égal à  $2^L - 1$ , ce qui donc n'est pas applicable en pratique du fait du grand nombre de solutions à traiter (un problème NP-complet). La performance d'une méthode de sélection est généralement dépend du choix de l'algorithme de recherche et du critère de sélection. Différents algorithmes de recherche ont été proposés dans la littérature pour sélectionner le meilleur sous-ensemble de classifieurs [Partridge 1996 ; Ruta 2001 ; Mao 2011]. En

général, les stratégies de recherche peuvent être classées en trois catégories : recherche par ordonnancement, algorithmes gloutons et les algorithmes évolutionnistes.

### 2.5.2.1 Les algorithmes de recherche

#### a) Recherche par ordonnancement

La recherche par ordonnancement, aussi appelée méthode des  $L^*$ -meilleurs classifieurs ou méthode de recherche heuristique, est une stratégie très peu coûteuse en calcul, donnant des résultats rapidement, mais ne garantit pas nécessairement l'obtention de la solution optimale. Cette méthode se base sur une heuristique simple et déterministe pour former un ensemble de classifieurs [Ruta 2005]. Il s'agit, en premier lieu, de trier les classifieurs par ordre décroissant en fonction de leurs performances sur la base de validation. En second lieu, les  $L^*$  meilleurs classifieurs sont sélectionnés pour former l'ensemble final. Le nombre de classifieurs à sélectionner  $L^*$  peut être déterminé a priori selon la complexité désirée du système ou en choisissant la valeur de  $L^*$  pour laquelle la performance de l'EdC sélectionné sur la base de validation est maximale. En effet, cela nécessite d'examiner le meilleur classifieur, une paire de meilleurs classifieurs, les 3 meilleurs et ainsi de suite jusqu'à compléter l'ensemble de  $L$  classifieurs. Ce processus étant entièrement déterministe, il n'est pas nécessaire de refaire l'expérience plusieurs fois.

Plusieurs méthodes basées sur la recherche par ordonnancement ont été proposées. Partridge et Yates [Partridge 1996] ont généré un grand nombre de réseaux de neurones en utilisant deux types de réseaux : Les réseaux multicouches (MLP) et les réseaux à fonction de base radiale (*Radial Basis Function*, RBF). Ces réseaux de neurones ont été entraînés en les initialisant différemment (avec différent nombre de couches cachées et différentes bases d'apprentissage). Les auteurs ont proposé l'utilisation de deux heuristiques lors de la phase de sélection. La première heuristique repose sur la sélection de  $L^*$  meilleurs classifieurs candidats en fonction de leurs performances. La deuxième heuristique choisit le classifieur ayant la plus grande précision, à partir de chaque type de classifieur pour composer l'ensemble final. Visentini et al. [Visentini 2016] ont suggéré la réduction de d'un ensemble de 50 réseaux de neurones entraînés avec la méthode Bagging, en utilisant le F-score comme critère de sélection. Chaque classifieur a été classé en mesurant son F-score et les  $L^*$  meilleurs membres ont été sélectionnés. Ces méthodes ne prennent pas en compte la diversité entre les membres sélectionnés.

D'autres chercheurs ont considéré la diversité lors de la sélection de classifieurs. Canuto et al. [Canuto 2007] ont généré un ensemble de prédicteurs en utilisant des classifieurs hétérogènes (HET) tels que MLP, RBF, SVM, kPPV et autres, et ont testé différentes tailles fixes d'ensembles candidats. La précision individuelle des classifieurs et la diversité de chaque ensemble généré sont pris en compte durant la phase de sélection. Dans la même optique, les auteurs de [Aksela 2006] ont également généré un grand ensemble de classifieurs hétérogènes pendant la phase de surproduction. Les réseaux de neurones et les SVM sont des exemples de classifieurs utilisés. Les auteurs ont appliqué avec succès un OCS basé sur la diversité des erreurs en utilisant des tailles fixes des ensembles candidats.

Bien que ces approches heuristiques conduisent à réduire la complexité de la phase de sélection, il n'y a aucune garantie que la solution optimale sera trouvée. Une alternative de ces méthodes est d'utiliser des algorithmes de recherche séquentielle.

### *b) Les algorithmes gloutons*

Les algorithmes gloutons (*Greedy algorithms*) ou les méthodes de recherche séquentielle sont des algorithmes pour lesquels, à chaque itération, un classifieur ou un petit nombre de classifieurs sont ajoutés ou retirés itérativement du sous-ensemble choisi dans l'objectif d'améliorer le critère de sélection [Banfield 2003]. Parmi les principaux algorithmes de recherche gloutons, on trouve:

- **Recherche en avant (ou forward en anglais, FD)** : Cette méthode de sélection part d'un sous-ensemble vide et ajoute les classifieurs un par un. A chaque itération, on ajoute à notre nouvel ensemble de prédicteurs, le classifieur nous permettant d'obtenir les meilleures performances lors de la combinaison de celle-ci avec notre sous-ensemble final selon la mesure considérée. Lorsque l'ajout de nouvelles classifieurs ne permet plus d'améliorer la performance du meilleur sous-ensemble de l'itération précédente, l'algorithme s'arrête.
- **Recherche en arrière (ou backward en anglais, BK)** : Cette méthode fonctionne de façon inversée. Elle considère un ensemble initial contenant tous les classifieurs et enlève tour à tour le classifieur le moins utile à l'ensemble. Les classifieurs sont retirés jusqu'à ce que les performances cessent de converger.

Plusieurs approches basées sur les algorithmes gloutons ont été proposées. Partalas et al. [Partalas 2008] ont proposé un algorithme de recherche glouton pour trouver le sous-ensemble optimal à partir d'un ensemble initial composé de classifieurs de types MLP et SVM. Les classifieurs ont été formés en utilisant différentes configurations des paramètres initiaux. Ainsi, les performances des ensembles candidats ont été utilisées pour guider le processus d'optimisation de la phase de sélection. Cette méthode ne tient pas en compte le critère de diversité.

A l'opposé, les auteurs de [Margeintu 1997] ont proposé un OCS basé sur les mesures de diversité pour réduire les coûts de calcul du Boosting. L'ensemble initial de classifieurs a été généré en utilisant la méthode Boosting et les arbres de décision (AD). Ensuite, des algorithmes gloutons basés sur la mesure de diversité  $kappa$  ont été appliqués pour sélectionner le sous-ensemble optimal de classifieurs. Cette méthode ne tient pas en compte du lien existant entre la précision individuelle des classifieurs et la diversité entre eux. Dans [Mao 2011], les auteurs ont utilisé également un algorithme de recherche séquentielle durant la phase de sélection. Le critère de sélection est basé sur la performance de la combinaison ainsi que la diversité entre les sorties des classifieurs. Les résultats expérimentaux ont indiqué que cette méthode augmente la diversité entre les membres et améliore également la précision par rapport aux méthodes Bagging et Random Subspace.

### *c) Les algorithmes évolutionnistes*

Les algorithmes évolutionnistes sont des méthodes heuristiques qui s'inspirent de la théorie de l'évolution pour résoudre des problèmes d'optimisation. Ces algorithmes sont utilisés pour obtenir une solution, non nécessairement optimale, à un problème donné lorsque l'espace des solutions possibles est trop important pour être parcouru de manière exhaustive [Gacquer 2008a].

Plusieurs algorithmes de recherche évolutionnistes ont été appliqués dans la littérature pour la sélection des ensembles de classifieurs. Gacquer et al. [Gacquer 2009] ont proposé un algorithme génétique (AG) pour la sélection d'ensemble de classifieurs afin de contrôler le compromis entre la diversité et la précision. À l'aide d'un processus itératif, chaque solution a été évaluée avec la fonction de fitness qui définit le compromis entre la précision et la diversité. Le sous-ensemble qui maximise cette fonction de fitness a été sélectionné comme solution optimale. Les auteurs de [Zhou 2002] ont proposé GANSEN (Genetic Algorithm based Selective Ensemble), qui utilise les algorithmes génétiques pour attribuer

un poids aléatoire à chaque réseau de neurones et évolue ces poids en évaluant la performance de la combinaison obtenue lorsque chaque réseau de neurones est inclus dans l'ensemble. Ensuite, chaque réseau de neurones dont le poids est supérieur à un seuil fixe est sélectionné et ajouté à l'ensemble final.

Dans [Ruta 2001b], une comparaison de trois algorithmes évolutionnistes a été réalisée, à savoir : l'algorithme génétique (AG) [Davis 1991], la méthode PBIL (*Population Based Incremental Learning*) [Baluja 1994] et la méthode TS (*Tabu Search*) [Glover 1998]. Les auteurs ont utilisé ces algorithmes pour rechercher efficacement une population unique de meilleures de classifieurs basé sur leurs taux de classification et les combiner utilisant le vote majoritaire pour améliorer la fiabilité des solutions obtenues.

Dans [Roli 2001], les auteurs ont réalisé une étude comparative des algorithmes de recherche séquentielle FD et BD et d'algorithme évolutionniste TS. Pour sélectionner le meilleur ensemble de classifieurs, deux critères ont été utilisés: la précision de l'ensemble et le désaccord entre les classifieurs en utilisant les trois mesures de diversité suivantes: diversité généralisée (*DG*), *Q*-statistique (*Q*) et double-faute (*DF*). L'ensemble initial est composé de classifieurs hétérogènes tels que MLP, kPPV et RBF. Les auteurs ont conclu que les critères de recherche et les algorithmes de recherche étudiés présentaient des résultats équivalents.

Le tableau 2.3 présente certaines approches proposées dans la littérature pour la sélection statique de classifieurs. Nous constatons de ce tableau que les classifieurs hétérogènes en phase de surproduction sont le plus souvent appliqués. De plus, la performance des classifieurs est le critère de recherche le plus souvent cité comme la meilleure fonction d'évaluation pour la sélection de l'ensemble optimal. D'autre part, peu de travaux ont été consacrés la combinaison de différents critères de recherche dans une phase de sélection multi-objective. Ces résultats montrent que la définition du meilleur algorithme de recherche et du meilleur critère de recherche pour la phase de sélection d'OCS reste un problème ouvert.

**Tableau 2.3.** Résumé de certains travaux de la sélection statique rapportés dans la littérature.

Reference de l'approche	Classifieur individuelle	Méthode ensembliste	Critère de Sélection	Algorithme de recherche
[Partridge 1996]	MLP, RBF	HET	Précision	heuristique
[Visentini 2016]	RN	BG	F-score	heuristique
[Canuto 2007]	MLP, RBF, SVM, kPPV	HET	Précision, Diversité	heuristique
[Aksela 2006]	RN, SVM	HET	Diversité	heuristique
[Partalas 2008]	MLP, SVM	HET	Précision	Séquentielle
[Margineantu 1997]	AD	BST	Diversité	Séquentielle
[Mao 2011]	AD, SVM	BG, RSS	Précision, Diversité	Séquentielle
[Gacquer 2009]	/	/	Précision, Diversité	AG
[Zhou 2004]	RN	HET	Précision globale	AG
[Ruta 2001]	/	/	Précision	AG, TS, PBIL
[Roli 2001]	MLP, kPPV, RBF	HET	Précision, Diversité	FD, BD, TS

### 2.5.2.2 Relation entre la diversité et la performance

La qualité d'un système de combinaison de classifieurs OCS dépend généralement de la qualité du critère de sélection [Aksela 2006]. Dans la plupart des cas, le critère de sélection est utilisé directement dans le processus de sélection pour guider la recherche du sous-ensemble optimal. L'interprétation la plus naturelle de l'optimalité d'un sous-ensemble de classifieurs est donc sa meilleure performance possible. L'utilisation directe de la performance globale de l'ensemble en tant que critère de sélection est précise, significative et permet des comparaisons cohérentes des différents sous-ensembles de classifieurs, quel que soit le nombre de classifieurs et leurs performances individuelles [Aksela 2003].

Cependant, il a été démontré que la recherche d'un sous-ensemble optimal de classifieurs, basée sur leurs précisions en utilisant les mêmes données d'entraînement, peut produire certain sur-apprentissage [Reunanen 2003 ; Loughrey 2004]. Notons tout d'abord

que même si l'algorithme de sélection est capable de choisir une combinaison optimale, leur optimalité ne peut être évaluée que pour l'ensemble d'apprentissage pour lequel les étiquettes sont connues et donc la performance peut être obtenue. Cela signifie qu'une fois la combinaison optimale sélectionnée, elle ne reste pas nécessairement optimale pour les données non étiquetées.

Pour limiter son effet, la performance réelle du sélecteur doit être estimée sur une base de validation (autre que la base de test et d'apprentissage) qui n'a pas été utilisée pour entraîner les classifieurs individuels. Le maintien d'une fiabilité raisonnable lors de la sélection statique de classifieurs nécessite donc des ensembles d'apprentissage plus grands que d'habitude. Sharkey et al. [Sharkey 2000] ont proposé une méthode OCS qui se base sur ce principe et ont déduit que si le sous-ensemble sélectionné donne une meilleure performance sur une base de validation, alors cet ensemble se comporte aussi de la même façon sur la base de test.

Au lieu de sélectionner les classifieurs en fonction de leur performance, d'autres chercheurs ont étudié la sélection de prédicteurs en se basant sur les mesures de diversité [Aksela 2006; Azizi 2010; Li 2012b; Visentini 2016]. Aksela [2006], par exemple, compare six mesures de diversité pour la sélection de 4 membres différents parmi un ensemble de 8 classifieurs. Cependant, des preuves expérimentales détaillées présentées dans [Kuncheva 2001; Ruta 2001b; Shipp 2002; Kuncheva 2003a] ont montré une très faible corrélation entre les mesures de diversité et la performance de l'ensemble combiné. En outre, l'utilisation d'un seul critère de sélection peut conduire à une mauvaise estimation des compétences des classifieurs candidats et par conséquent, à une sélection des membres incompetents. Les recherches décrites dans [Rodriguez 2006] ont prouvé que la diversité n'est pas le seul élément à prendre en compte pour assurer l'efficacité d'un ensemble de classifieurs. Les ensembles de classifieurs qui ont une grande valeur de diversité ne sont pas nécessairement ceux qui produisent le taux de classification le plus élevé.

La combinaison de la performance de l'ensemble et de la diversité entre ces membres en tant que critères de recherche a prouvé son efficacité au processus de sélection [Opitz 1996; Zenobi 2001; Oliveira 2005]. Opitz et Shavlik [Opitz 1996] ont appliqué un OCS combinant à la fois la mesure de diversité *ambiguïté* et le taux d'erreur pour la recherche d'un sous-ensemble optimal dans une population des réseaux de neurones. Ils ont montré

que cet OCS permet d'améliorer la performance de la combinaison de tous les classifieurs de l'ensemble original.

Dans [Zenobi 2001], les auteurs ont créé durant la phase de surproduction un ensemble de kPPV en appliquant une approche de sélection de caractéristiques. Lors de la phase de sélection, la diversité et le taux d'erreur ont été utilisés pour guider la recherche. Ils ont montré que leur approche a amélioré la performance des ensembles sélectionnés en utilisant le taux d'erreur comme critère de sélection. Enfin, dans l'OCS proposé par Oliveira et al. [Oliveira 2005], un processus de sélection de caractéristiques a été appliqué pour générer une population de réseaux de neurones lors de la phase de surproduction. La phase de sélection est également guidée par la diversité et le taux d'erreur en tant que fonctions objectives.

Il est important de mentionner que les travaux précédents ont une caractéristique commune : la solution optimale, trouvée et analysée lors de la phase de sélection, permet de classer tous les échantillons de la base de test. Cependant, il n'y a aucune garantie que cet ensemble optimal est la bonne solution pour classer chaque échantillon de test. Pour cette raison, d'autres chercheurs ont proposé de sélectionner les classifieurs dynamiquement pour chaque exemple inconnu. Les classifieurs sont choisis en fonction de l'exemple à reconnaître.

### 2.6 Sélection dynamique de classifieurs

Les techniques de sélection dynamique (*Dynamic Selection*) consistent, sur la base d'un grand ensemble de classifieurs  $\mathcal{C}$ , à trouver un seul classifieur  $c_j$ , ou un sous-ensemble de classifieurs  $\mathcal{C}_i^*$ , ayant les membres les plus compétents de  $\mathcal{C}$  pour prédire la classe d'appartenance d'un exemple spécifique  $x_i$  [Dos Santos 2008 ; Cruz 2011 ; Cavalin 2013 ; Philipose 2018].

La sélection dynamique de classifieurs est traditionnellement définie comme une stratégie qui suppose que chaque membre de l'ensemble est un expert dans certaines régions locales de l'espace de caractéristiques [Cruz 2017]. Ainsi, étant donné un nouvel échantillon  $x_i$ , les techniques dynamiques visent à sélectionner les classifieurs les plus compétents dans la région locale de l'espace de caractéristiques où cet échantillon de test est situé. Seuls les classifieurs qui atteignent un certain niveau de compétence, selon un critère de sélection, sont choisis pour classer  $x_i$ .

L'efficacité des approches de sélection dynamiques prouvée dans de nombreux problèmes de classification expliquent ses succès notamment sur les approches statiques [Singh 2005 ; Cavalin 2013 ; Souza 2018], en particulier dans le cas de problèmes mal définis, c'est-à-dire lorsque la taille de la base de données est petite et qu'il n'y a pas suffisamment de données pour former un classifieur fort ayant beaucoup de paramètres à apprendre. En outre, la distribution réelle du problème ne peut être représentée adéquatement en raison de l'insuffisance des données d'apprentissage. Par conséquent, les classifieurs ne peuvent pas apprendre la séparation entre les classes dans ces cas.

Dans la sélection dynamique de classifieurs, la classification d'un nouvel exemple  $x_i$  implique généralement trois étapes [Cruz 2014], comme illustré à la figure 2.5:

- **Définition de la région de compétence** : Cette étape consiste à définir la région locale entourant l'échantillon  $x_i$ , noté  $\theta_i = \{x_1, \dots, x_k\}$ , dans laquelle le niveau de compétence des classifieurs de base est estimé.
- **Détermination des critères de sélection** : Ces critères sont utilisés pour estimer le niveau de compétence des classifieurs, par exemple, précision, probabiliste, diversité, etc.
- **Détermination du schéma de sélection** : cette étape consiste à choisir un classifieur unique (*Dynamic Classifier Selection*, DCS) ou un ensemble de classifieurs (*Dynamic Ensemble Selection*, DES) en fonction de son niveau de compétence estimé.

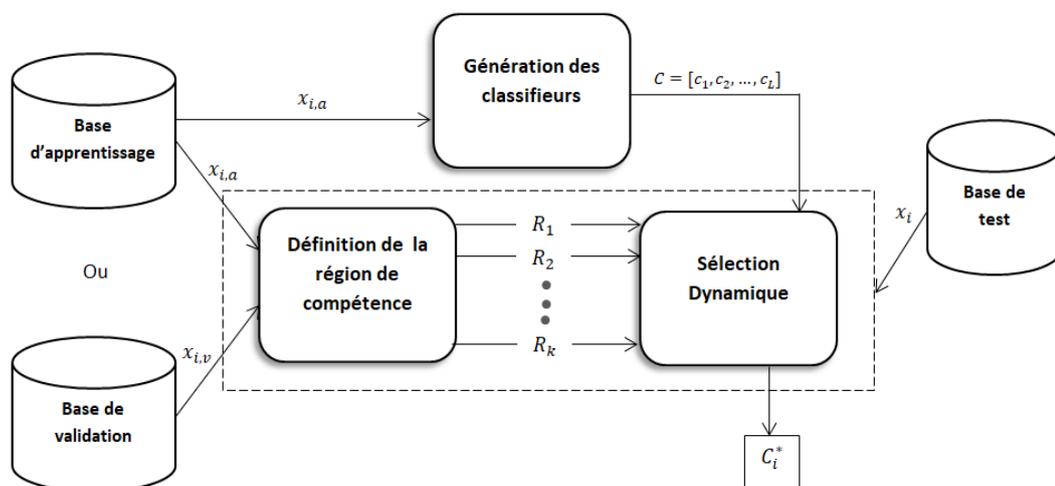


Figure 2.5. Processus de la sélection dynamique de classifieurs [Dos Santos 2011].

### 2.6.1 Définition de la région de compétence

Généralement, les régions locales de compétence sont définies à l'aide de la technique kPPV [Woods 1997, Ko 2008], via des méthodes de clustering (par exemple K-Means) [Kuncheva 2000 ; Soares 2006], en utilisant les décisions des classifieurs de base [Giacinto 2001b ; Cavalin 2012 ; Cavalin 2013] ou par l'utilisation d'une fonction potentielle [Woloszynski 2011]. Dans tous les cas, un ensemble d'exemples étiquetés, qui peut être la base d'apprentissage ou de validation, est requis. Cet ensemble est appelé la base de données de la sélection dynamique (*dynamic selection dataset* , *DSEL*) [Cruz 2017 ; Oliveira 2017].

#### 2.6.1.1 Clustering

Dans les techniques qui utilisent le clustering pour définir la région de compétence [Kuncheva 2000 ; Soares 2006 ; de Souto 2008 ; Lin 2014], la première étape consiste à partitionner la base *DSEL* en groupes ou en clusters. Ensuite, la compétence de chaque classifieur est estimée pour tous les clusters. Au cours de la phase de généralisation, pour chaque exemple de test  $x_i$ , la distance entre  $x_i$  et le centre de chaque cluster est calculée et le cluster le plus proche à  $x_i$  est choisis comme sa région de compétence . La compétence des classifieurs est ensuite mesurée sur cette région de compétence.

L'avantage de l'utilisation de la technique de clustering est que tous les classements et sélections des classifieurs sont estimés pendant la phase d'apprentissage. Pour chaque cluster, l'EdC est défini a priori. Ainsi, les techniques de sélection dynamique basées sur le clustering sont beaucoup plus rapides pendant de la phase de généralisation. En outre, seulement la distance entre l'échantillon de test et le centre de chaque cluster doit être estimée, plutôt que toutes les instances de *DSEL*.

#### 2.6.1.2 *K-Plus proches voisins (kPPV)*

Dans le cas de la technique kPPV [Woods 1997 ; Ko 2008], les  $k$  plus proches voisins de l'exemple de test  $x_i$  sont estimés à l'aide de la base *DSEL*. Ensuite, les compétences des classifieurs de base sont évaluées en ne tenant compte que les instances appartenant à cette région de compétence.

L'avantage d'utilisation de kPPV par rapport au clustering est que la technique kPPV permet une estimation plus précise de la région locale, ce qui conduit à des configurations

différentes d'EdC selon la classification des nouvelles instances [Soares 2006 ; de Souto 2008]. Cependant, cette approche contrairement au clustering est plus couteuse en temps de calcul, puisque la distance entre l'exemple  $x_i$  et le  $DSEL$  doit être calculée avant d'estimer les compétences des classifieurs. Ce problème apparait en particulier lorsqu'il s'agit d'une base de données de grande taille [Cruz 2015].

### 2.6.1.3 *Modèle de fonction potentielle*

Ces méthodes sont inspirées du travail de Rastrigin et Erenstein [Rastrigin 1981], qui est l'un des premiers travaux concernant la sélection dynamique de classifieurs. Ils diffèrent de la majorité des autres techniques de la sélection dynamique en ce qui concerne le calcul de la compétence des classifieurs. Plutôt que d'utilisation de voisinage de l'échantillon de test pour l'estimation de la compétence, les techniques basées sur un modèle de fonction potentielle utilisent tous l'ensemble de données de sélection dynamique ( $DSEL$ ).

Cependant, l'influence de chaque point de données  $x_k \in DSEL$  est pondérée par sa distance euclidienne à l'échantillon  $x_i$  en utilisant un modèle de fonction potentielle. Généralement, une fonction potentielle gaussienne est considérée (équation 2.16). Par conséquent, les points qui sont plus proches de l'échantillon  $x_i$  ont une grande influence sur l'estimation de la compétence des classifieurs.

$$K(x_k, x_i) = \exp(-d(x_k, x_i)^2) \quad (2.16)$$

Plusieurs techniques de sélection ont été proposées en utilisant le modèle de fonction potentiel: sélection dynamique d'ensemble de classifieurs basé sur la divergence de Kullback-Leibler (DES-KL) [Woloszynski 2012], la technique basée sur le classifieur de référence randomisé (*Randomized Reference Classifier*, RRC) [Woloszynski 2012] et les méthodes DCS basées sur les fonctions logarithmiques et exponentielles [Woloszynski 2009].

L'utilisation de ces méthodes pour définir les régions de compétence présente l'avantage de ne pas nécessiter de définir de la taille du voisinage a priori car la fonction potentielle  $K(x_k, x_i)$  est utilisée pour réduire l'influence de chaque point de donnée en se basant sur sa distance euclidienne à l'échantillon à classer. Cependant, son inconvénient est

le coût de calcul augmenté puisque toutes les données de la base *DSEL*, et pas seulement le voisinage de  $x_i$ , sont utilisées pour estimer la compétence des classifieurs candidats.

### 2.6.1.4 Espace de décisions

Les techniques de sélection de cette catégorie sont basées sur le comportement de l'ensemble de classifieurs en utilisant les prédictions des classifieurs comme informations. Ils sont inspirés de l'espace de connaissance du comportement (BKS) [Huang 1995], souvent appelé «espace de décisions», car il est basé sur les décisions prises par les classifieurs de base [Giacinto 2001b ; Cavalin 2013].

Un aspect important de cette classe de techniques est la transformation des exemples de test et d'apprentissage en profils de sortie. Cette transformation peut être effectuée en utilisant les décisions difficiles des classifieurs de base (par exemple, les étiquettes de classe prédites), comme dans la méthode BKS, ou en utilisant les probabilités postérieures estimées des classifieurs de base, comme suggéré dans [Giacinto 2001b ; Cavalin 2013]. Le profil de sortie d'un échantillon  $x_j$  est noté  $\bar{x}_i = \{\bar{x}_{i,1}, \bar{x}_{i,2}, \dots, \bar{x}_{i,L}\}$  où chaque  $\bar{x}_{i,j}$  est la décision obtenue par le classifieur  $c_j$  ( $j = 1, \dots, L$ ) pour l'exemple  $x_i$ .

Ensuite, la région de compétence est calculée par la similarité entre le profil de sortie  $\bar{x}_i$ , et les profils de sortie des échantillons de la base *DSEL*. L'ensemble avec les profils de sortie les plus similaires, noté  $\theta_i$ , est utilisé pour estimer le niveau de compétence des classifieurs de base. Des exemples de techniques qui utilisent une région de compétence définie dans l'espace de décision sont : le comportement de classifieur multiple (*Multiple Classifier Behavior*, MCB) [Giacinto 2001b], les k-plus proches profils de sorties (*K-Nearest Output Profiles*, KNOP) [Batista 2011 ; Cavalin 2013] et META-DES [Cruz 2015].

### 2.6.2 Critère de sélection

Le critère utilisé pour mesurer le niveau de compétence des classifieurs candidats est un élément principal de toute technique de sélection dynamique. Selon [Britto 2014], ces critères peuvent être organisés en deux catégories: les mesures individuelles et les mesures de l'ensemble. La première catégorie présente les mesures se basant sur la performance individuelle d'un classifieur pour estimer son niveau de compétence. La compétence de chaque classifieur  $c_j$  est mesurée indépendamment des performances des autres membres

du même ensemble. Cette catégorie peut être aussi divisée en plusieurs autres sous-groupes selon le type d'information utilisé pour mesurer la compétence des classifieurs, à savoir : le classement [Sabourin 1993 ; Smits 2002], la précision [Woods 1997 ; Cruz 2011], le probabiliste [Giacinto 2001a ; Kurzynski 2010 ; Woloszynski 2011], le comportement [Giacinto 2001b ; Cavalin 2013], Oracle [Ko 2008] et le méta-apprentissage [Cruz 2014 ; Cruz 2015].

Les mesures de l'ensemble sont composées de métriques qui prennent en compte l'interaction entre les classifieurs de l'ensemble. Cette catégorie peut être divisée en trois sous-groupes [Britto 2014]: la diversité [Soares 2006 ; Dos Santos 2009], la manipulation des données [Xiao 2012] et l'ambiguïté [Dos Santos 2008]. Ces mesures ne sont pas directement liées à la notion de compétence d'un classifieur de base, mais plutôt à la notion de pertinence, c'est-à-dire si le classifieur de base fonctionne bien en conjonction avec les autres classifieurs de l'ensemble.

### 2.6.3 Schéma de sélection

Concernant le schéma de sélection, les techniques de la sélection dynamique peuvent choisir soit un classifieur unique (sélection dynamique de classifieur, DCS) ou un ensemble de classifieurs (DES) [Roy 2018]. Les premiers travaux de la sélection dynamique ont commencé avec la sélection d'un seul classifieur plutôt que la sélection d'un EdC. Dans ces techniques, le seul classifieur qui atteint le niveau de compétence le plus élevé est utilisé pour classifier l'exemple de test donné. Des exemples des méthodes DCS sont les technique A Priori et A Posteriori [Giacinto 1999], ainsi que le comportement de classifieur multiple (MCB) [Giacinto 2001b].

Cependant, étant donné que la sélection d'un seul classifieur peut être très sujette aux erreurs, certains chercheurs ont décidé de sélectionner un sous-ensemble de classifieurs plutôt qu'un seul classifieur de base. Tous les classifieurs de base qui ont obtenu un certain niveau de compétence sont utilisés pour composer l'EdC optimal, et leurs sorties sont combinées pour prédire la classe de  $x_i$ . Des exemples de techniques DES sont : le DES-kPPV [Soares 2006 ; de Souto 2008], les  $k$  plus proches oracles (*k-Nearest Oracles*, KNORA) [Ko 2008] et les  $k$  plus proches profils de sortie (*k-Nearest Output Profiles*, KNOP) [Cavalin 2013].

## 2.6.4 Techniques de la sélection dynamique

Dans cette section, nous présentons une revue des algorithmes de sélection dynamique les plus pertinents. Les techniques de la sélection dynamique ont été choisies en tenant compte leur importance dans la littérature par l'introduction de nouveaux concepts dans le domaine (c'est-à-dire des méthodes qui ont introduit différentes façons pour définir la région de compétence ou les critères de sélection) et leur nombre de citations. Le tableau 2.4 catégorise les principales techniques de sélection dynamique décrites dans cette revue.

**Tableau 2.4.** Catégorisation des méthodes de sélection dynamique.

Référence	Technique	Définition de $\theta_i$	Critère de Sélection	Approche de sélection
[Sabourin 1993]	Classement des classifieurs	kPPV	Classement	DCS
[Woods 1997]	OLA	kPPV	Précision	DCS
[Woods 1997]	LCA	kPPV	Précision	DCS
[Giacinto 1999]	A Priori	kPPV	Probabiliste	DCS
[Giacinto 1999]	A Posteriori	kPPV	Probabiliste	DCS
[Giacinto 2001b]	MCB	Espace de décisions	Comportement	DCS
[Soares 2006 ; de Souto 2008]	DES-Clustering	Clustering	Précision, Diversité	DES
[Soares 2006 ; de Souto 2008]	DES-kPPV	kPPV	Précision, Diversité	DES
[Ko 2008]	KNORA-E	kPPV	Oracle	DES
[Ko 2008]	KNORA-U	kPPV	Oracle	DES
[Cavalin 2013]	KNOP	Espace de décisions	Comportement	DES

### 2.6.4.1 Classement des classifieurs modifiés

Dans la méthode de classement des classifieurs modifiés [Sabourin 1993], le classement d'un seul classifieur  $c_j$  est simplement estimé par le nombre d'échantillons correctement classés consécutivement dans la région de compétence  $\theta_i$ . Le classifieur qui

classe correctement le plus grand nombre d'exemples consécutifs est sélectionné comme le classifieur le plus compétent pour la classification de  $x_i$ .

#### 2.6.4.2 Overall Local Accuracy (OLA)

Dans cette méthode [Woods 1997], le niveau de compétence,  $\delta_{i,j}$ , d'un classifieur  $c_j$  est simplement calculé par l'estimation de sa précision de classification dans la région de compétence  $\theta_i$  (équation 2.17). Le classifieur présentant le niveau de compétence le plus élevé est sélectionné pour prédire la classe de  $x_i$ .

$$\delta_{i,j} = \frac{1}{|\theta_i|} \sum_{q=1}^{|\theta_i|} P(y_k | x_q \in y_k, c_j) \quad (2.17)$$

#### 2.6.4.3 Local Classifier Accuracy (LCA)

La technique LCA [Woods 1997] est similaire à OLA, la seule différence étant que la précision locale de la méthode LCA est estimée par rapport à les exemples de la classe  $y_k$  ( $y_k$  est la classe assignée pour  $x_i$  par  $c_j$ ) de la région de compétence  $\theta_i$  (Équation 2.18). Le classifieur présentant le niveau de compétence le plus élevé,  $\delta_{i,j}$ , est sélectionné pour prédire l'étiquette de  $x_i$ .

$$\delta_{i,j} = \frac{\sum_{x_q \in y_k} (y_k | x_q \in y_k, c_j)}{\sum_{q=1}^{|\theta_i|} (y_k | x_q \in y_k, c_j)} \quad (2.18)$$

#### 2.6.4.4 Les méthodes A priori et A Posteriori

La méthode A Priori [Giacinto 1999] considère la probabilité d'une classification correcte du classifieur  $c_j$ , en  $\theta_i$  prenant en compte les supports obtenus par le classifieur  $c_j$ . Par conséquent, le vecteur contenant les probabilités a posteriori pour chaque classe est considéré à la place de l'étiquette assignée à chaque  $x_q \in \theta_i$ . De plus, cette méthode pondère également l'influence de chaque exemple  $x_q$  dans la région de compétence selon la distance euclidienne  $d(x_q, x_i)$  entre  $x_q$  et l'échantillon  $x_i$ . Les exemples les plus proches ont une influence plus importante sur le calcul du niveau de compétence  $\delta_{i,j}$ . L'équation (2.19) démontre le calcul du niveau de compétence  $\delta_{i,j}$  en utilisant la méthode A Priori:

$$\delta_{i,j} = \frac{\sum_{q=1}^{|\theta_i|} P(y_k | x_q \in y_k, C_j) W_q}{\sum_{q=1}^{|\theta_i|} W_q} \quad (2.19)$$

Avec  $W_q = 1/d(x_q, x_i)$ . Le classifieur avec la valeur la plus élevée de  $\delta_{i,j}$  est sélectionné. Cependant, ce classifieur sélectionné est utilisé uniquement pour prédire l'étiquette de  $x_i$  si son niveau de compétence est significativement meilleur que celui des autres classifieurs de l'ensemble initial (c'est-à-dire lorsque la différence de niveau de compétence est supérieure à un seuil prédéfini). Sinon, tous les classifieurs de l'ensemble initial sont combinés en utilisant la règle de vote majoritaire.

La méthode A Posteriori [Giacinto 1999] fonctionne de la même manière que la méthode A Priori. La seule différence est qu'il prend en compte la classe proposée par le classifieur  $c_j$ , pour l'échantillon de test  $x_i$  pendant l'estimation de sa compétence (équation 2.20).

$$\delta_{i,j} = \frac{\sum_{x_p \in y_k} P(y_k | x_q \in y_k, c_j) W_q}{\sum_{q=1}^{|\theta_i|} P(y_k | x_q \in y_k, c_j) W_q} \quad (2.20)$$

#### 2.6.4.5 Comportement de classifieur multiple

La technique de comportement de classifieur multiple (MCB) est basée sur l'espace de connaissance du comportement (BKS) [Huang 1995] et la précision locale du classifieur. Étant donné un exemple de test  $x_i$  sa région de compétence  $\theta_i$  est estimée. Ensuite, le profil de sortie  $\tilde{x}_i$  de l'exemple  $x_i$ , ainsi que ceux de la région de compétence, sont calculés en utilisant l'algorithme BKS.

La similarité entre le profil de sortie de l'exemple de test  $x_i$ , et ceux de sa région de compétence,  $x_k \in \theta_i$ , est calculée (équation 2.21). Les exemples ayant des similarités inférieures à un seuil prédéfini sont retirés de la région de compétence  $\theta_i$ . Par conséquent, la taille de la région de compétence est variable, car elle dépend également du degré de similarité entre  $x_i$  et les échantillons de sa région de compétence. après la sélection exemples similaires, la compétence du classifieur de base,  $\delta_{i,j}$ , est estimée par sa précision de classification dans la région de compétence résultante.

$$S(\tilde{x}_i, \tilde{x}_q) = \frac{1}{L} \sum_{j=1}^L T(x_i, x_q) \quad (2.21)$$

$$T(x_i, x_q) = \begin{cases} 1 & \text{if } c_j(x_i) = c_j(x_q), \\ 0 & \text{if } c_j(x_i) \neq c_j(x_q) \end{cases} \quad (2.22)$$

La fonction  $S(x_i, x_q)$  prend des valeurs dans  $[0,1]$ ,  $S(x_i, x_q) = 1$  si les  $L$  décisions de classification sont les mêmes pour  $x_i$  et  $x_q$ , et  $S(x_i, x_q) = 0$  si toutes les décisions sont totalement différentes pour les deux individus. La décision final du classifieur sélectionné est prise de la même façon des méthodes A Priori et A Posteriori.

#### 2.6.4.6 *DES-Clustering (DES-K-Moyenne)*

Dans cette méthode [Soares 2006 ; de Souto 2008], l'algorithme K-Moyenne est appliqué à DSEL afin de subdiviser cet ensemble en plusieurs clusters. Pour chaque cluster, les classifieurs sont classés par ordre de précision décroissante et par ordre de diversité croissante. Ensuite, les  $J$  classifieurs les plus précis et les  $K$  classifieurs les plus divers sont associés à chaque cluster. Dans [Giacinto 2001a], la mesure Double-Faute a été utilisée pour mesurer la diversité des classifieurs de base, puisque elle a présenté la corrélation la plus élevée avec la précision d'ensemble dans l'étude présenté par Shipp et Kuncheva [Shipp 2002].

La distance euclidienne entre l'échantillon de test  $x_i$  et le centre de chaque cluster est d'abord calculée. Ensuite, l'ensemble des  $J$  classifieurs les plus précis et  $K$  les plus divers associés au cluster le plus proche est utilisé pour composer l'ensemble des classifieurs  $C_i^*$ . Les valeurs de  $J$  et  $K$ , ( $K \leq J$ ) doivent être définies avant d'appliquer cette méthode.

#### 2.6.4.7 *DES-kPPV*

La première étape [Soares 2006 ; de Souto 2008] de cette technique consiste à calculer la région de compétence  $\theta_i$  de l'exemple de test  $x_i$  en utilisant l'algorithme kPPV. Ensuite, les classifieurs de l'ensemble initial  $C$  sont classés par ordre de précision décroissant et par ordre de diversité croissante basé sur les exemples appartenant à  $\theta_i$ . Ensuite, Les  $J$  classifieurs les plus précis et les  $K$  classifieurs les plus divers sont sélectionnés pour composer le sous-ensemble  $C_i^*$  pour classer  $x_i$ .

#### 2.6.4.8 *KNORA-Eliminate*

La technique KNORA-Eliminate proposée par Ko et al. [Ko 2008] explore les propriétés du concept d'oracle pour sélectionner le sous-ensemble de classifieurs le plus approprié pour chaque échantillon de test. Pour chaque exemple  $x_i$ , seuls les classifieurs qui reconnaissent correctement tous les exemples appartenant à sa région de compétence  $\theta_i$  sont sélectionnés. En d'autres termes, tous les classifieurs qui ont atteint une précision de 100% dans cette région (c'est-à-dire, qui sont des oracles locaux) sont choisis pour composer l'ensemble  $C_i^*$ . Ensuite, les décisions des classifieurs de  $C_i^*$  sont combinées en utilisant la règle de vote majoritaire. Si aucun classifieur de base n'est sélectionné, la taille de la région de compétence sera réduite et la recherche des classifieurs compétents est redémarrée.

#### 2.6.4.9 *KNORA-Union*

La technique KNORA-Union [Ko 2008] sélectionne tous les classifieurs qui sont capables de reconnaître correctement au moins un échantillon de la région de compétence. Cette méthode considère également qu'un classifieur peut participer plus d'une fois au système de vote lorsqu'il classe correctement plus d'une instance de la région de compétence. Le nombre de votes d'un classifieur  $c_j$  est égal au nombre d'échantillons de  $\theta_i$  correctement classés par ce classifieur. Les votes collectés par tous les classifieurs sont combinés pour obtenir la décision final de l'ensemble  $C_i^*$ .

#### 2.6.4.10 *K-Nearest Output Profiles (KNOP)*

La technique KNOP [Cavalin 2013] fonctionne de manière similaire à celle de la technique KNORA-U, la seule différence étant que KNORA-U fonctionne dans l'espace de caractéristiques, alors que KNOP fonctionne dans l'espace de décision. Tout d'abord, la transformation des profils de sortie est appliquée sur l'entrée  $x_i$ , donnant son profil de sortie  $\tilde{x}_i$ . Ensuite, la similarité entre  $\tilde{x}_i$  et les profils de sortie de la base de données de sélection dynamique (*DSEL*) est calculée et stockée dans l'ensemble  $\theta_i$ . De manière similaire à la règle KNORA-U, chaque fois qu'un classifieur effectue une prédiction correcte, pour un échantillon appartenant à  $\theta_i$ , il gagne un vote. Les votes obtenus par tous les classifieurs de base sont combinés pour obtenir la décision d'ensemble  $C_i^*$ .

## 2.7 Conclusion

Le concept de diversité a été utilisé dans plusieurs problèmes différents comme l'analyse de la relation entre l'erreur de classification et la performance des classifieurs, l'identification du sous-ensemble de classifieurs donnant la performance prédictive la plus élevée, ou la construction des ensembles de classifieurs corrélés.

Dans ce chapitre, nous avons présenté une taxonomie des mesures de diversité et des méthodes de construction utilisées, à savoir : les méthodes Bagging, Boosting et Random Subspace qui ont pour principe de générer et renforcer la diversité entre les sorties des classifieurs afin d'améliorer les performances de l'ensemble.

Ce chapitre a présenté également un état de l'art des différentes stratégies de sélection de classifieurs. Nous avons dressé tout au long de ce chapitre un panorama des deux principales catégories de sélection : statique et dynamique, ainsi que le rôle qui joue la diversité dans l'amélioration des performances des méthodes de sélection de classifieurs. Nous avons vu que, même si plusieurs contributions importantes ont été faites dans le domaine de la sélection de classifieur, la définition du meilleur critère de recherche pour trouver le meilleur sous-ensemble de classifieurs reste une question ouverte.

Le chapitre suivant portera sur notre première contribution qui vise à combiner la méthode Random Subspace et la sélection statique de classifieurs basant sur les mesures de diversité; afin d'analyser le rôle de la notion de diversité dans le processus de construction et de sélection des ensembles de classifieurs.

# Chapitre 3

## Approche Hybride basée sur Random Subspace et la Sélection Statique de Classifieurs

### 3.1 Introduction

Avec le développement des traitements numériques de l'image, de la reconnaissance des formes et de l'intelligence artificielle, les médecins ont la possibilité d'améliorer leur diagnostic à l'aide des systèmes informatiques [Chabat 2000 ; Rao 2012]. Les avantages de l'utilisation de ces systèmes intelligents comprennent l'amélioration de la précision du diagnostic tout en réduisant le temps et les coûts associés au traitement du patient [Çomak 2007 ; Prabusankarlal 2015].

Différents modèles d'apprentissage automatique ont été développés pour soutenir les tâches de prise de décision médicale [Markowska-Kaczmar 2006 ; Kahramanli 2009 ; Luukka 2011 ; Sekar 2012]. Les dernières recherches ont proposé la combinaison de plusieurs modèles de classifieurs afin d'améliorer la précision et l'efficacité d'un système de classification [Stoean 2013 ; Huang 2017 ; Zilly 2017]. Pendant l'apprentissage, les classifieurs de base sont créés séparément sur l'ensemble de données d'apprentissage, ensuite, une règle de combinaison efficace est utilisée pour combiner les sorties et produire la décision finale de l'ensemble sur un échantillon de test. La diversité entre les membres de l'ensemble est reconnue comme l'une des caractéristiques requises pour atteindre un taux de précision élevé [Kuncheva 2003b ; Tsybal 2005]. En d'autres termes, il n'y a aucun gain à combiner des composants identiques.

Dans ce chapitre, nous présentons notre première contribution : la proposition d'un système de classification des images médicales, basé sur la combinaison de la méthode Random Subspace et la sélection statique de classifieurs. Ce système passe par une phase d'extraction des caractéristiques en employant différentes familles de caractéristiques. Notre objectif est de représenter les images par un ensemble riche de primitives afin de bénéficier de leurs avantages et la possibilité d'avoir une complémentarité entre eux. Ces familles sont basées sur des caractéristiques de texture et de forme qui sont: la matrice de cooccurrence qui vise à extraire les caractéristiques de texture, les moments de Hu et les moments centraux qui visent à décrire la forme de l'image.

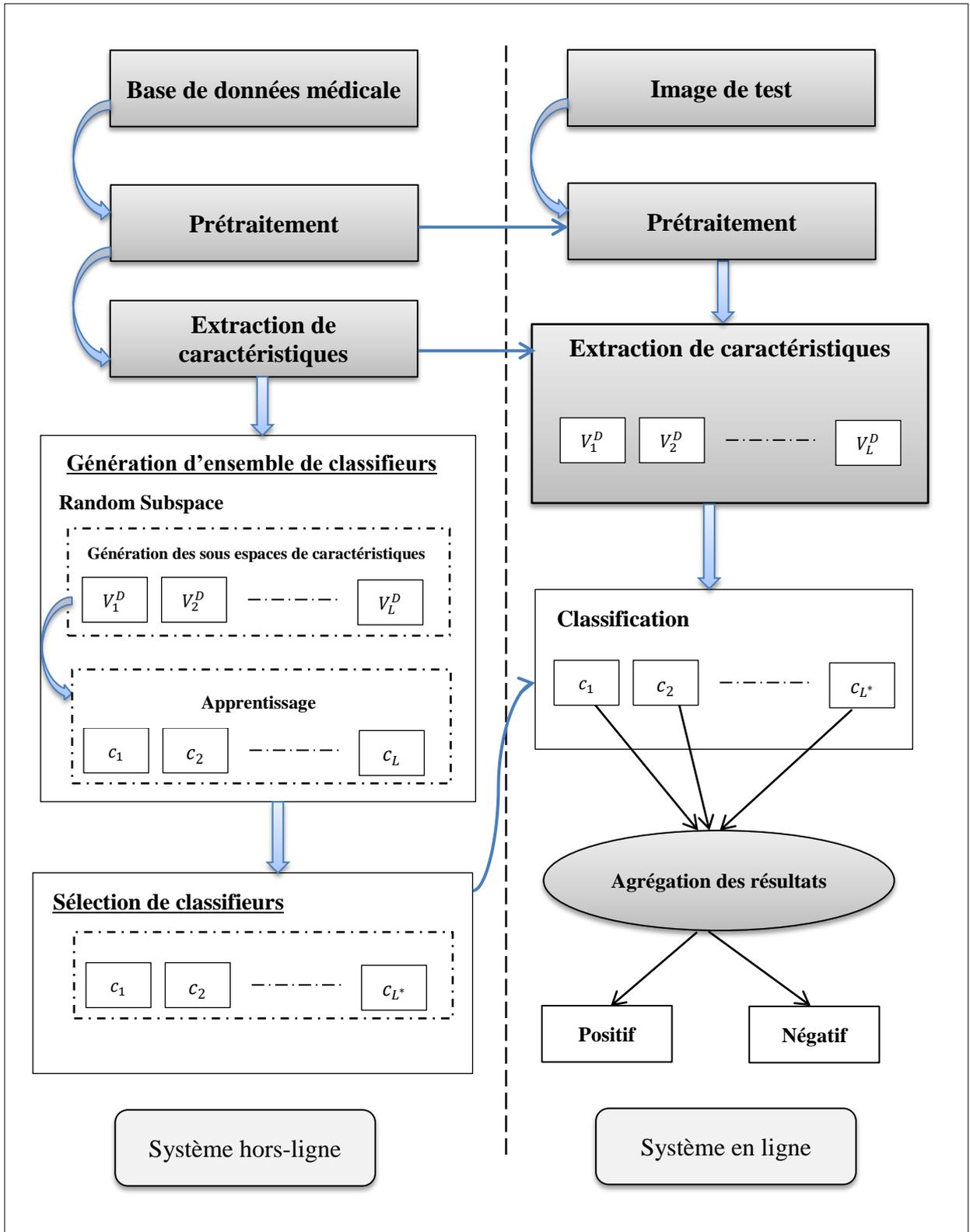
Un nouveau système de classification multiple est ensuite exploité pour classifier et analyser automatiquement ces images médicales en utilisant une combinaison de la méthode Random Subspace et la sélection statique de classifieurs se basant sur l'utilisation des mesures de diversité. Le sous-ensemble de classifieurs sélectionnés dans la procédure de sélection statique est combiné en utilisant le vote majoritaire.

L'objectif de cette proposition est de clarifier le rôle de la diversité durant le processus de construction des systèmes multi-classifieurs. Pour cette raison, ce chapitre se concentre sur le côté d'application de la diversité en tant que critère de sélection, mené par des résultats expérimentaux extensifs sur la sélection statiques de classifieurs guidés par différentes mesures de diversité.

Nous présentons dans un premier temps les différentes bases de données médicales et les méthodes d'extraction de caractéristique utilisées dans notre proposition ainsi que le modèle de classification proposé se basant sur le paradigme de surproduction et choix. Dans un deuxième temps, nous évaluons l'approche proposée sur des bases de données de référence. Enfin, dans la dernière section, nous concluons sur quelques extensions pouvant améliorer l'approche proposée.

## **3.2 Système multi-classifieurs pour le diagnostic des images médicales**

Notre objectif est de concevoir un système d'aide au diagnostic des images médicales bénéficiant des avantages des approches ensemblistes. Notre processus de diagnostic des images médicales comprend trois étapes principales [Cheriguene 2016b ; Cheriguene 2016c]: le prétraitement, l'extraction des caractéristiques et la prise de décision ou de classification. La figure 3.1 illustre le schéma général du système proposé.



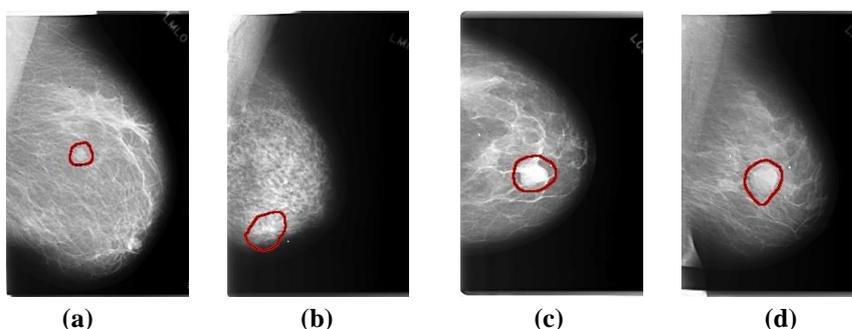
**Figure 3.1.** Schéma général du système multi-classifieurs proposé pour le diagnostic des images médicales.

### 3.2.1 Les bases des images médicales

#### 3.2.1.1 La base DDSM d'images mammographiques

La première base de données établie pour la validation du système proposé est la base d'images mammographiques DDSM (*Digital Database for Screening Mammography*) pour le diagnostic du cancer du sein [Heath 1998 ; Heath 2000]. Le cancer du sein est l'un des cancers les plus communs chez les femmes [Lanyi 1985]. Cette pathologie représente le cancer le plus répandu en Algérie. Son incidence ne cesse d'augmenter d'année en année de 7%, ce qui fait de lui un véritable problème de santé publique.

La base DDSM a été rassemblée par un groupe de chercheurs de l'université de Floride du sud et a été largement utilisée par la communauté scientifique dans le domaine du cancer du sein [Cheng 2006 ; Dominguez 2009]. La base de données DDSM contient 2620 dossiers de patients répartis en trois catégories: normal, bénin et malin. Ces cas sont recueillis auprès de l'hôpital "Massachusetts General Hospital" (MGH), l'Université "Wake Forest University" (WFU) et l'hôpital "Washington University of St. Louis School of Medicine"(WUSTL). La figure 3.2 présente quelques échantillons de la base DDSM contenant des masses bénignes ou malignes.



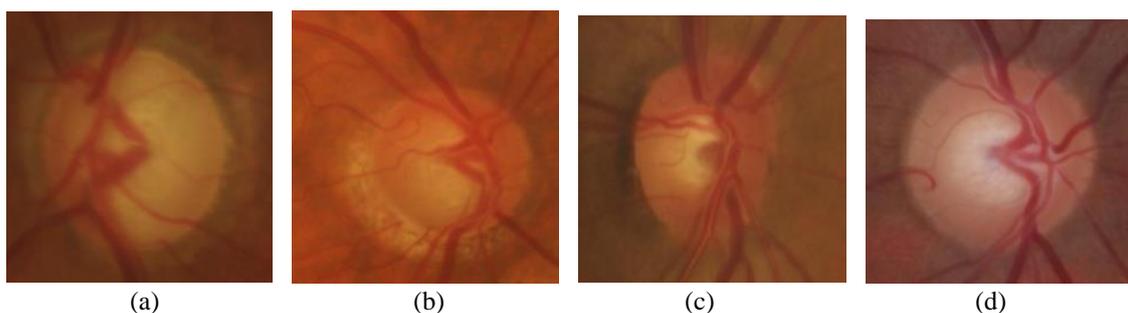
**Figure 3.2.** Un échantillon de la base DDSM : les images (a) et (b) représentent des images mammaires à masse maligne et les images (c) et (d) représentent des masses bénignes [Zemmal 2018].

#### 3.2.1.2 La base RIM-One du glaucome

Le glaucome est une maladie oculaire causée par une pression élevée à l'intérieur de l'œil. Cette pression accrue dans l'œil est généralement causée par des lésions nerveuses entraînant une perte de vision progressive et permanente, commençant par des points

aveugles imperceptibles sur les bords du champ de vision, atteignant la vision en tunnel, jusqu'à la cécité [Mookiah 2012]. Il a été rapporté que près de 30 millions de personnes dans le monde souffrent de glaucome; et ce nombre devrait augmenter au cours de la présente décennie. D'ici 2020, on estime qu'environ 80 millions de personnes dans le monde recevront un diagnostic de glaucome [Quigley 2006].

Cette étude utilise des images rétiniennes à partir de la base de données RIM-ONE (*An Open Retinal Image Database for Optic Nerve Evaluation*). La base RIM-ONE [Fumero 2011] est collecté par la collaboration de trois organisations espagnoles, à savoir l'Hôpital Miguel Servet, l'Université National de Education a Distancia et l'Université Complutense. Cette base se compose de 169 images ONH obtenues à partir de 169 images de fond d'œil classées en deux catégories : glaucome (74 images) et normal (95 images) [Zemmal 2018]. La figure 3.3 présente quelques échantillons de la base RIM-One.



**Figure 3.3.** Images de fond d'œil typiques tirées de la base de données RIM-One: (a et b) Glaucome, (c et d) Normal [Zemmal 2018].

#### 3.2.2 Prétraitement et extraction de caractéristiques

Comme nous l'avons déjà mentionné, le prétraitement des données constitue une étape préliminaire de la phase de classification ou de prise de décision. Dans notre système, nous avons utilisé les mêmes types de prétraitement appliqués à [Azizi 2013]. Dans la proposition de [Azizi 2013], les auteurs ont adopté la conversion en niveau de gris en utilisant l'outil ImageJ parce que la matrice de co-occurrence est déterminée d'une image à niveaux de gris. D'autre part, la méthode Otsu a été utilisée pour effectuer un seuillage d'une image à niveaux de gris en une image binaire.

L'extraction de caractéristiques est une tâche très importante dans tous systèmes de reconnaissance de formes [Liu 2013 ; Sundararaj 2014]. Elle vise à extraire les informations pertinentes d'une image et les transformer à des mesures distinctes pour les utiliser à l'étape de classification. Plusieurs techniques d'extraction des caractéristiques ont été proposées pour l'analyse des images [El-Atlas 2014]. Les méthodes d'extraction des caractéristiques sont variantes et dépendantes du choix des descripteurs employés (descripteur bas niveau, descripteur haut niveau) ainsi que du type d'image à analyser (image binaire, image en niveaux de gris et image couleur). Dans notre proposition, nous avons choisi de décrire nos images par une combinaison des informations de bas niveau comme : la forme et la texture. Les moments de Hu et les moments centraux ont été utilisés pour représenter la forme, tandis que la matrice de co-occurrence a été utilisée pour assurer la texture. Nos images ont été éventuellement représentées par un grand vecteur de caractéristiques incluant les valeurs de ces trois familles de caractéristiques (ce vecteur contient plus de 24 caractéristiques).

### 3.2.2.1 Les moments de Hu

Hu [1962] a introduit un ensemble des moments invariants en utilisant les moments géométriques. En traitement d'image, un moment est la moyenne pondérée des intensités des pixels en fonction de leurs positions dans l'image. Comme son nom l'indique, les moments centraux sont calculés à partir du centre de la forme [Sonka 2007]. Pour une image discrète de taille  $M * N$  ou  $f(x, y)$  est la fonction décrivant le contenu de l'image en chacun de ses pixels de coordonnée  $x$  et  $y$ , le moment central  $m_{p,q}$  de l'ordre  $(p + q)$  est défini comme suit:

$$m_{p,q} = \sum_{p=0}^{M-1} \sum_{q=0}^{N-1} x^p y^q f(x, y) \quad (3.1)$$

Si une image est considérée comme une fonction continue  $p(x, y)$ , le moment central est défini par :

$$m_{p,q} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q p(x, y) dx dy \quad (3.2)$$

Le moment  $m_{0,0}$  d'ordre 0 représente l'air de la forme de l'objet (le nombre des pixels). Les deux moments  $m_{0,1}$ ,  $m_{1,0}$  d'ordre 1 permettent de déterminer les coordonnées du centre de gravité de l'image :

$$x_c = \frac{m_{0,1}}{m_{0,0}} \quad \text{et} \quad y_c = \frac{m_{1,0}}{m_{0,0}} \quad (3.3)$$

Ces moments de base sont d'utilités limitées puisque ils varient à chaque transformation géométrique tels que : la translation, la rotation, ainsi que le changement d'échelle [Huang 2010]. Un ensemble de moments invariants serait plus utile. Ceci peut être dérivé en calculant les moments centrés  $\mu_{p,q}$  par l'équation suivante :

$$\mu_{p,q} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - x_c)^p (y - y_c)^q f(x, y) dx dy \quad (3.4)$$

Les moments centrés normalisés  $\eta_{p,q}$  sont calculés en utilisant les moments centrés  $\mu_{p,q}$  :

$$\eta_{p,q} = \frac{\mu_{p,q}}{\mu_{0,0}^\gamma} \quad (3.5)$$

Ou,  $\gamma = (p + q)/2$  et  $p + q = 2, 3, \dots$ .

A partir des moments centrés normalisés, Hu a proposé un ensemble  $\phi_i$  de sept moments invariants [Zhang 2012]:

$$\phi_1 = \eta_{20} + \eta_{02} \quad (3.6)$$

$$\phi_2 = (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2 \quad (3.7)$$

$$\phi_3 = (\eta_{30} + \eta_{12})^2 + (3\eta_{21} - \mu_{03})^2 \quad (3.8)$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \quad (3.9)$$

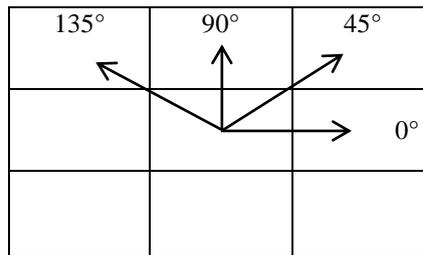
$$\begin{aligned} \phi_5 = & (\eta_{30} + 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ & + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \end{aligned} \quad (3.10)$$

$$\phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad (3.11)$$

$$\begin{aligned} \phi_7 = & (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - (\eta_{30} - \\ & 3\eta_{21})(\eta_{21} + \eta_{03}) [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \end{aligned} \quad (3.12)$$

### 3.2.2.2 La matrice de co-occurrence

La matrice de co-occurrence des niveaux de gris (*Gray Level Cooccurrence Matrix*, GLCM) est parmi les premières stratégies d'extraction des caractéristiques de texture [Haralick 1973]. Cette approche consiste à explorer les dépendances spatiales des textures en comptant les distributions des niveaux de gris co-occurents entre deux pixels séparés par un certain décalage. GLCM est une matrice carrée symétrique dans laquelle les lignes et les colonnes correspondent aux niveaux de gris dans l'image ou aux plages des niveaux de gris identifiées. L'élément de la matrice  $p_{d,\theta}(i, j)$  représente le nombre d'occurrences de niveau de gris correspondant à la ligne  $i$  et le niveau de gris correspondant à la colonne  $j$  pour un décalage spatial donné. Ce décalage est généralement décrit par le choix de la distance  $d$  et d'angle  $\theta$  de déplacement. La distance peut être combinée avec chacune des quatre directions (horizontale, verticale, diagonale gauche et droite) comme indiqué sur la figure 3.4. Les caractéristiques de texture de Haralick sont calculées pour chacune de ces directions d'adjacence [Sharma 2001 ; Zhang 2002].



**Figure 3.4.** Les quatre directions d'adjacence pour calculer les caractéristiques de texture de Haralick [Zemmal 2018].

Pour une image de taille  $M \times N$ , les caractéristiques de texture sont calculées en faisant la moyenne  $P_{d,\theta}(i, j)$  sur les quatre matrices de cooccurrence directionnelle :

$$P_{d,\theta}(i, j) = \frac{p_{d,\theta}(i, j)}{M \times N} \quad (3.13)$$

Les caractéristiques de Haralick utilisées dans notre approche sont les suivantes:

**Energie**

$$P_1 = \sum_i \sum_j P_{d,\theta}(i, j)^2 \quad (3.14)$$

$$\text{Contraste} \quad P_2 = \sum_i \sum_j ((i, j)^2 \times P_{d,\theta}(i, j)) \quad (3.15)$$

$$\text{Entropie} \quad P_3 = - \sum_i \sum_j (P_{d,\theta}(i, j) \times \log(i, j)) \quad (3.16)$$

$$\text{Variance} \quad P_4 = - \sum_i \sum_j ((i - \mu)^2 \times P_{d,\theta}(i, j)) \quad (3.17)$$

$$\text{Corrélation} \quad P_5 = \sum_i \sum_j (P_{d,\theta}(i, j) \frac{(i - \mu)(j - \mu)}{\sigma}) \quad (3.18)$$

$$\text{Moment inverse} \quad P_6 = \sum_i \sum_j (\frac{1}{1 + (i - j)^2} P_{d,\theta}(i, j)) \quad (3.19)$$

$$\text{moyenne des sommes} \quad P_7 = \sum_{i=0}^{2(N_g-1)} i \cdot P_{x+y}(i) \quad (3.20)$$

$$P_{x+y} = \sum_i \sum_j P_{d,\theta}(i, j), \quad k = i + j = 0, 1, 2, \dots, 2(N_g - 1) \quad (3.21)$$

$$\text{Variance des sommes} \quad P_8 = \sum_{i=0}^{2(N_g-1)} (i - P_7)^2 \cdot P_{x+y}(i) \quad (3.22)$$

$$\text{Entropie des sommes} \quad P_9 = - \sum_{i=0}^{2(N_g-1)} \log(P_{x+y}(i)) \cdot P_{x+y}(i) \quad (3.23)$$

$$\text{Variance de différence} \quad P_{10} = - \sum_{i=0}^{2(N_g-1)} (i - P'_{10}) \cdot P_{x-y}(i) \quad (3.24)$$

$$P^{x-y} = \sum_i \sum_j P_{d,\theta}(i, j), \quad k = |i - j| = 0, 1, 2, \dots, 2(N_g - 1) \quad (3.25)$$

$$P'_{10} = \sum_{i=0}^{2(N_g-1)} i \cdot P_{x-y}(i) \quad (3.26)$$

$$\text{Entropie de différence} \quad P_{11} = - \sum_{i=0}^{2(N_g-1)} \log(P^{x-y}(i)) \cdot P^{x-y}(i) \quad (3.27)$$

### 3.2.3 Approche basée sur les mesures de diversité pour la sélection d'ensemble de classifieurs.

De nombreux chercheurs ont démontré que la combinaison des classifieurs identiques ne présente aucun avantage au niveau de l'amélioration de la performance du système

[Zenobi 2001 ; Shipp 2002 ; Kuncheva 2003a]. La situation idéale est d'avoir un ensemble initial de classifieurs avec des erreurs non corrélées. La diversité a donc été reconnue comme une des caractéristiques principales de la conception d'un système multi-classifieurs performant. Plusieurs méthodes ont été suggérées pour générer un large groupe initial de classifieurs diversifiés comme les méthodes Bagging [Breiman 1996a], Boosting [Freund 1997] et Random Subspace [Ho 1998].

Le mécanisme de sélection de classifieurs consiste à choisir des meilleurs membres de l'ensemble initial, de sorte que le sous-ensemble sélectionné peut atteindre un taux de classification optimal [Kuncheva 2002a ; Didaci 2004 ; Ruta 2005]. Les méthodes de sélection de classifieurs sont divisés en deux groupes: statique et dynamique. A la sélection statique [Aksela 2003], le sous-ensemble des classifieurs le plus performant est choisi une fois pour toutes au cours de la phase d'apprentissage et il est utilisé pour reconnaître tous les échantillons de la base de test. Le choix d'un meilleur sous-ensemble de classifieurs durant la phase d'apprentissage dépend toujours de la connaissance a priori définie sur la base d'apprentissage ou de validation.

L'objectif principal de notre approche est de valider le rôle que joue la diversité entre les classifieurs sur l'amélioration de performance de classification aux images médicales. Dans cette proposition [Cheriguene 2016b ; Cheriguene 2016c], nous introduisons la diversité à deux niveaux durant les phases de construction et de sélection des classifieurs. Afin de renforcer la diversité entre les classifieurs de l'ensemble initial, notre système ensembliste (RSS-SCS) est basé d'abord sur la création d'un groupe initial de classifieurs en utilisant différents sous-espaces de caractéristiques, à l'aide de la technique de génération Random Subspace (RSS). D'autre part, notre système est basé sur la sélection statique de classifieurs (SCS) en utilisant différentes mesures de diversité, dans lequel les exemples de test sont classés ultérieurement par le même ensemble de configuration. Durant cette phase, la diversité est adoptée en tant que critère de sélection pour choisir le sous-ensemble des classifieurs les plus diversifiés.

#### ***3.2.3.1 Génération de classifieurs basé sur la méthode RSS***

La diversité peut être obtenue en utilisant le même algorithme d'apprentissage sur différents échantillons d'entraînement pour former des classifieurs différents [Cunningham 2000]. D'autres techniques permettent la construction d'un ensemble de classifieurs de même type qui sont différenciés au niveau des caractéristiques manipulées pendant

l'apprentissage. La méthode Random Subspace (RSS) [Tremblay 2004 ; Xia 2015 ; Hosseini 2016] est une de ces techniques qui travaille sur des données d'apprentissage modifiées en termes d'espace de caractéristiques afin d'obtenir des meilleurs classifieurs.

Ho [1998] a prouvé que la méthode RSS est capable d'améliorer l'erreur de généralisation. Le principe de base de l'algorithme RSS est d'entraîner chaque classifieur sur un sous-espace de caractéristiques différent généré aléatoirement. Chacun de ces sous-espaces aléatoires est de dimension  $D$ , avec  $D < F$  où  $F$  est la dimension de l'espace de description original. Le but de cette approche est de générer aléatoirement des sous-ensembles d'entités avec l'algorithme RSS de façon à les utiliser comme une entrée de chaque classifieur de base. La méthode RSS permettra de minimiser la redondance des informations, éliminer la perturbation causée par l'utilisation d'un grand nombre de caractéristiques et renforcer la diversité entre les sorties des classifieurs.

Trois classifieurs de base sont considérés dans notre étude : le perceptron multicouche (MLP) [Tang 2016], la machine à vecteurs de support (SVM) [Hearst 1998] et les arbres de décision (AD) [Quigley 1986]. Ces trois classifieurs sont très largement utilisés dans plusieurs domaines de reconnaissance de formes, notamment pour leurs bonnes performances et leur capacité à la généralisation. De plus, pour les classifieurs d'arbre de décision, nous avons utilisé l'algorithme C4.5 [Quinlan 1993]. L'ensemble initial homogène de classifieurs contient 30 classifieurs. Chaque classifieur a été formé en utilisant au hasard 50% des caractéristiques de la base de données d'apprentissage. Le pourcentage proposé a été déterminé par Ho [1998]. L'auteur a montré que les meilleurs résultats sont obtenus lorsque nous utilisons la moitié du jeu de composants.

#### ***3.2.3.2 Sélection statique de classifieurs***

L'utilisation d'une mesure de diversité appropriée durant la phase de sélection est importante pour contrôler la similarité et le degré d'indépendance entre les classifieurs [Kuncheva 2002b]. Plusieurs mesures comme la corrélation, l'entropie et d'autres ont été proposées pour estimer la diversité. Bien que ces mesures soient utilisées dans plusieurs travaux et appliquées sur différents domaines de la reconnaissance de formes, il n'est pas possible de distinguer la supériorité d'une mesure sur une autre [Kuncheva 2003a]. Ceci est fortement lié premièrement à la nature de la base de données utilisée, des échantillons, et aussi aux types des classifieurs de base adoptés. Pour cette raison, nous avons testé plusieurs mesures de diversité qui ont prouvé leurs efficacités à la littérature [Kuncheva

2003a]. Les mesures utilisées pour estimer la diversité des sous-ensembles sont des mesures par-paire comme : Q-statistique ( $Q$ ), Corrélacion ( $\rho$ ), mesure de désaccord ( $D$ ), Weighted Count of Errors and Correct results (WCEC) et des mesures globales, à savoir : l'entropie ( $Ent$ ), Difficulté ( $\theta$ ) et diversité généralisé ( $DG$ ).

Après la génération d'un ensemble de  $L$  classifieurs à l'étape précédente, nous avons créé tous les sous-ensembles possibles de taille  $L^*$ , ( $L^* < L$ ). Puis, la diversité a été calculée pour chaque sous-ensemble en utilisant une des mesures mentionnée au paragraphe précédent. Le sous-ensemble qui a la valeur maximal de la diversité est sélectionné [Cheriguene 2016b ; Cheriguene 2016c].

La dernière étape du système RSS-SCS consiste à choisir un moyen efficace pour combiner les sorties des classifieurs sélectionnés. Plusieurs règles d'agrégation peuvent être utilisées pour combiner les résultats des classifieurs. Dans notre étude, nous avons appliqué le vote majoritaire en tant que technique de fusion. Le concept du vote majoritaire est simple à implémenter et à appliquer. Étant donné une forme  $x$  non-étiquetée, chaque classifieur vote pour une classe spécifique. Le nombre de votes pour chacune des classes est compté et la sortie finale de l'ensemble est celle qui reçoit le plus grand nombre de votes [Lam 1994 ; Ruta 2001a ; Cheriguene 2015].

Le vote majoritaire est la règle de fusion employée dans notre système parce que (1) elle a prouvé son efficacité empiriquement et théoriquement dans plusieurs problèmes de classification; (2) c'est une règle non-paramétrique qui n'assume pas de connaissances préalables sur les classifieurs ; (3) et elle peut être utilisée avec n'importe quel type de classifieurs de base.

Malgré cela, nous avons également évalué d'autres règles de combinaison, telles que: moyenne, produit, médiane, minimum et maximum. Cependant, le vote à la majorité était légèrement meilleur par rapport à ces règles de combinaison. Ainsi, pour des raisons de clarté, nous avons décidé de ne rapporter que les résultats se basant sur la règle de fusion du vote majoritaire. Par ailleurs, les méthodes ensemblistes que nous avons choisies pour évaluer la performance de la méthode proposée ont également utilisé le vote majoritaire.

L'algorithme 3.1 explique les principales étapes du processus proposé.

---

**Algorithme 3.1** : Les étapes de l'algorithme proposé RSS-SCS

---

1. Pour  $j = 1, \dots, L$  répéter
    - a) Choisir parmi le vecteur de caractéristiques initial  $V_0$ , un sous-espace aléatoire  $V_j^D$ , de dimension  $D$ , ou  $D < F/2$  et  $F$  est la taille de l'espace de caractéristiques  $V_0$ .
    - b) Projeter toutes les données d'apprentissage dans ce nouveau sous-espace  $V_j^D$
    - c) Entraîner le classifieurs  $c_j$  en utilisant la base d'apprentissage  $B_j^D$  de  $V_j^D$
  - Fin pour
  2. Créer les sous-ensembles de classifieurs de taille  $L^*$  en appliquant toutes les combinaisons possibles.
  3. Quantifier la diversité de chaque sous-ensemble en utilisant une des mesures de diversité mentionnées au-dessus.
  4. Sélectionner le sous-ensemble  $C^*$  le plus diversifié de classifieurs.
  5. Combiner les résultats générés par les membres de  $C^*$  utilisant le vote majoritaire.
- 

### 3.3 Etude expérimentale

#### 3.3.1 Paramètres expérimentaux

##### 3.3.1.1 Base de données utilisées

Nos expérimentations ont été appliquées sur deux bases de données médicales différentes : DDSM et RIM-One. La sous base DDSM [Heath 1998, Heath 2000], considérée dans notre système de diagnostic de la maladie du cancer du sein, est constituée de 302 images dont 102 images sont malignes et 200 sont bénignes. Pour le diagnostic de la maladie du glaucome, nous avons utilisé toute la base de données RIM-ONE qui est constitué de 169 images [Fumero 2011].

Afin d'évaluer notre méthode, le protocole de validation croisée  $n$ -fold [Krogh1995] a été utilisé, dans lequel l'ensemble de données est divisé en plusieurs sous-ensembles de taille égale et mutuellement exclusifs. Dans nos expérimentations, les bases DDSM et RIM-One ont été divisées en 5 partitions de tailles à peu près égales. L'algorithme d'apprentissage a été ensuite exécuté 5 fois, en utilisant à chaque fois 3 partitions comme

ensemble d'apprentissage, une partition a été utilisée comme ensemble de validation et la partition restante comme ensemble de test. La précision finale sera la moyenne sur des 5 essais. Chaque algorithme a été exécuté 5 fois d'une manière indépendante et les moyennes des résultats sont présentées et discutées dans la section suivante.

### 3.3.1.2 Mesures de performance utilisées

Evaluer un système de diagnostic permet de mesurer ses performances vis-à-vis de ses objectifs. De ce fait, le choix des mesures à utiliser diffère selon les objectifs fixés. Plusieurs mesures d'évaluation de performance ont été utilisées pour déterminer l'efficacité du système ensembliste proposé, à savoir : la précision (taux de classification), la sensibilité et la spécificité. Les mesures que nous allons évoquer utilisent la matrice de confusion [Sahiner 2001, Li 2002], qui permet la différenciation des erreurs selon chaque classe en vue d'évaluer un classifieur (tableau 3.1).

**Tableau 3.1.** Matrice de confusion.

	Décision positive (+)	Décision négative (-)
Étiquette positive (+)	Vrai Positif (VP)	Faux Négatif (FN)
Étiquette négative (-)	Faux Positif (FP)	Vrai Négatif (VN)

- La **précision** (taux de classification) mesure la proportion de vrais positifs et de vrais négatifs dans tous les cas évalués. Cette précision est mesurée de la façon suivante:

$$\text{Précision} = \frac{VP + VN}{VP + VN + FP + FN} \quad (3.28)$$

- La **sensibilité** (taux de vrais positifs) mesure la proportion d'individus positifs qui sont effectivement bien détectés par le test par rapport à l'ensemble, donnée par :

$$\text{Sensibilité} = \frac{VP}{VP + FN} \quad (3.29)$$

- La **spécificité** (taux de vrais négatifs) mesure la proportion de vrais négatifs effectivement bien détectés par le test :

$$\text{Spécificité} = \frac{VN}{VN + FP} \quad (3.30)$$

La méthode RSS-SCS a été également évaluée sur la base de la courbe caractéristique opérationnelle du récepteur, connue par l'abréviation ROC (*Receiver Operating Characteristic*). La courbe ROC est une méthode de représentation graphique des performances d'un classifieur à deux classes [Berbaum 1989, Fawcett 2006]. Cette représentation est largement utilisée au domaine médical afin de discriminer entre la population des malades et des non-malades [Sahiner 2001, Li 2002]. La courbe ROC représente la sensibilité en ordonnée en fonction de la quantité (1-spécificité) en abscisse.

Afin d'obtenir une description quantitative de la performance de la classification, à partir de la représentation ROC, l'aire sous cette courbe notée généralement par AUC (Area Under Curve) a été utilisé durant nos expérimentation [Berbaum 1989]. La mesure de l'aire sous la courbe ROC peut être obtenue à partir du calcul de l'aire des rectangles juxtaposés occupant l'aire de la courbe. Etant donné que l'aire sous la courbe est une portion de l'aire d'un carré unitaire, sa valeur est toujours entre 0 et 1. La signification de l'aire sous la courbe ROC est évaluée de la façon suivante [Cheikhrouhou 2012]:

- Si  $AUC = 0.5$  : le résultat du diagnostic est d'apport nul.
- Si  $0.5 < AUC < 0.7$  : le résultat est peu informatif.
- Si  $0.7 < AUC < 0.9$  : le résultat est moyennement informatif.
- Si  $0.9 < AUC < 1$  : le résultat est très informatif.
- Si  $AUC = 1$  : le résultat est parfait.

### 3.3.2 Résultats expérimentaux

#### 3.3.2.1 Analyse de performance de RSS-SCS

Pour souligner l'importance et clarifier l'effet de l'approche proposée, nous avons effectué une première expérimentation pour observer le comportement de l'approche RSS-SCS et analyser sa performance en utilisant différentes paramètres (type de classifieurs individuels, critère de sélection, taille du sous-ensemble sélectionné  $L^*$ ). Plusieurs mesures de diversité ont été utilisées comme critère de sélection afin d'analyser le rôle de la notion de diversité dans les processus de construction et de sélection de classifieurs. Dans le but d'analyser l'influence de la taille du sous-ensemble à sélectionner sur la performance de notre système ensembliste RSS-SCS, l'algorithme d'apprentissage a été exécuté pour

différentes tailles de  $L^* = \{5, 10, 15\}$ . Les tableaux 3.2 et 3.3 montrent les résultats de taux de classification obtenus pour chaque base de données. Les figures 3.5, 3.6 et 3.7 illustrent les moyennes de précision obtenus par les trois classifieurs individuel (CI) MLP, SVM et AD et par la méthode RSS-SCS.

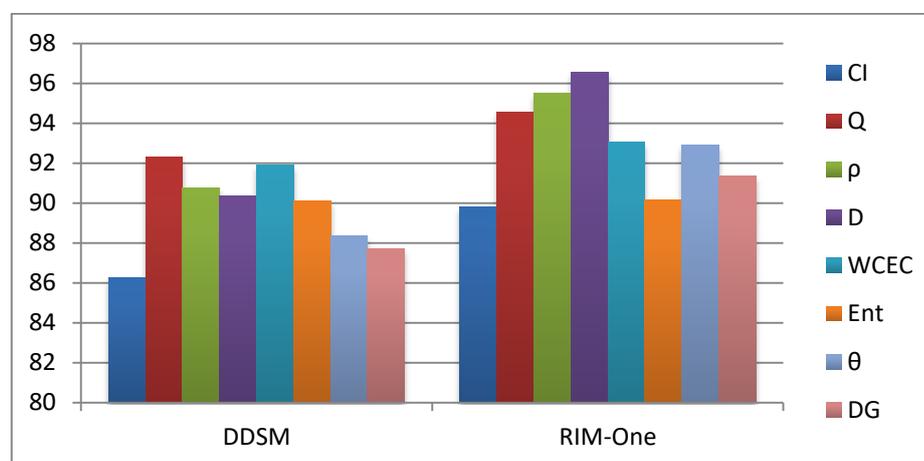
Les résultats indiqués dans les tableaux 3.2 et 3.3 ainsi que la figure 3.5 montrent que notre approche se comporte différemment selon les mesures de diversité. Les mesures Q-statistique ( $Q$ ), Désaccord ( $D$ ) et WCEC produisent des meilleures performances pour tous les ensembles de classifieurs de la base DDSM, tandis que les mesures  $Q$ ,  $\rho$  et  $D$  améliorent les résultats de la base RIM-One. Ainsi, les mesures par-paire ont des taux de classification supérieurs à ceux obtenus par les mesures globales. Nous constatons également de ces résultats que la sélection de classifieurs que ce soit par les mesures par-paire ou par les mesures globales produit une amélioration en taux de classification par rapport à l'utilisation de chacun des classifieurs individuellement.

**Tableau 3.2.** Précision du système proposé RSS-SCS sur la base de données DDSM.

Méthode	CI	$Q$	$\rho$	$D$	WCEC	Ent	$\theta$	DG	Moyenne
$L^* = 5$	MLP	90.73	90.20	90.07	<b>92.45</b>	91.52	86.49	86.62	89.72
	SVM	92.85	89.27	<b>93.64</b>	91.66	90.20	86.62	89.54	90.54
	AD	92.72	89.67	92.85	<b>92.98</b>	92.58	91.92	90.47	<b>91.88</b>
$L^* = 10$	MLP	91.26	92.45	87.42	<b>93.64</b>	92.05	88.08	86.49	90.20
	SVM	<b>91.66</b>	89.01	88.08	84.77	83.45	82.52	80.53	85.72
	AD	<b>94.84</b>	94.04	84.77	93.11	88.35	90.86	86.49	<b>90.35</b>
$L^* = 15$	MLP	90.86	92.45	<b>94.04</b>	92.85	92.32	91.26	87.55	91.62
	SVM	91.52	87.68	<b>93.78</b>	91.92	90.07	85.96	89.67	90.08
	AD	<b>94.31</b>	92.19	88.35	93.78	90.47	91.39	92.05	<b>91.79</b>
<b>Moyenne</b>		<b>92.30</b>	90.77	90.33	91.91	90.11	88.34	87.71	/

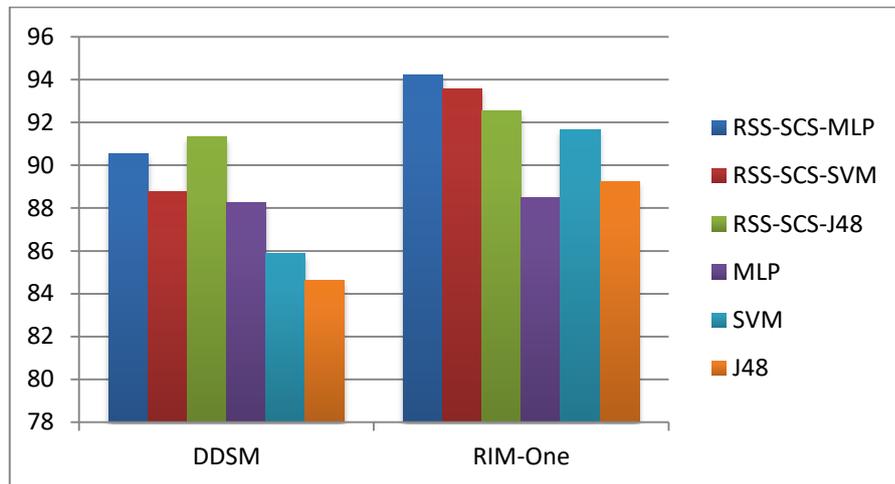
**Tableau 3.3.** Précision du système proposé RSS-SCS sur la base de données RIM-One.

Méthode	CI	$Q$	$\rho$	$D$	WCEC	Ent	$\theta$	DG	Moyenne
$L^* = 5$	MLP	<b>96.33</b>	95.50	96.09	89.59	87.57	93.49	91.12	<b>92.81</b>
	SVM	92.78	<b>93.25</b>	91.24	88.76	91.12	91.24	91.48	91.41
	AD	88.17	95.27	<b>96.92</b>	91.95	84.62	92.66	90.41	91.43
$L^* = 10$	MLP	<b>98.11</b>	95.62	97.28	93.61	90.77	88.76	92.31	<b>93.78</b>
	SVM	<b>96.57</b>	95.03	96.21	93.02	92.78	94.91	84.02	93.22
	AD	90.30	94.56	<b>96.80</b>	94.08	85.68	89.35	89.94	91.53
$L^* = 15$	MLP	97.04	97.16	<b>98.82</b>	95.86	93.61	95.74	93.85	96.01
	SVM	97.87	<b>97.99</b>	97.75	95.50	92.78	96.21	94.67	<b>96.11</b>
	AD	93.85	95.15	<b>97.87</b>	94.91	92.31	93.73	94.44	94.61
<b>Moyenne</b>		94.56	95.50	<b>96.55</b>	93.03	90.14	92.90	91.36	/


**Figure 3.5.** Précision moyenne du système proposé RSS-SCS en utilisant différentes mesures de diversité sur les bases de données DDSM et RIM-One.

Si on considère les résultats de l'approche RSS-SCS en utilisant différents types de classifieurs de base (figure 3.6), nous constatons que les résultats des trois ensembles de classifieurs RSS-SCS-MLP, RSS-SCS-SVM et RSS-SCS-AD surpassent les résultats des classifieurs individuels MLP, SVM et AD respectivement. La méthode RSS-SCS en

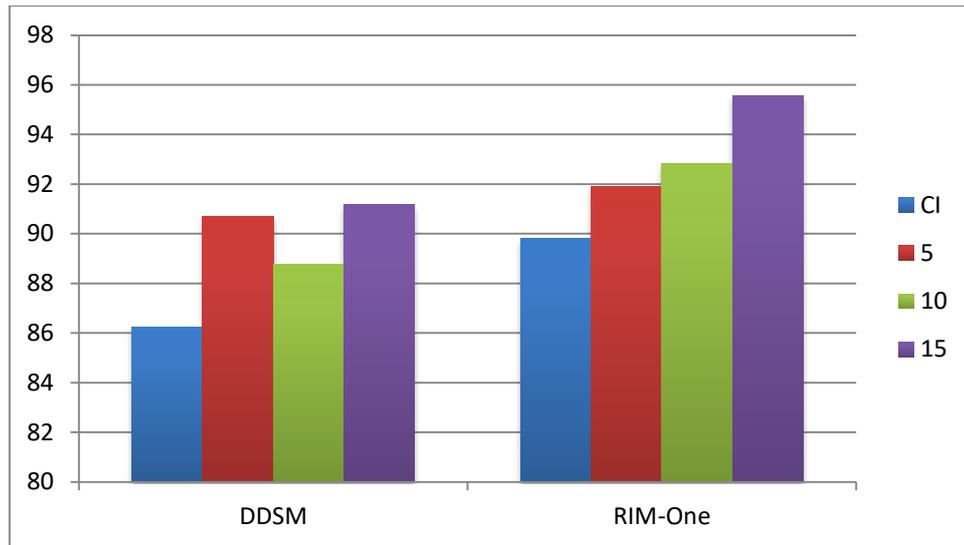
utilisant les ADs comme classifieurs de base nous permettant d'obtenir des meilleures performances pour la base DDSM et l'ensemble RSS-SCS-MLP génère des meilleurs taux de classification pour la base RIM-One. Nous pouvons constater à travers ces résultats que le système RSS-SCS génère des performances variables selon l'ensemble initial de classifieurs adoptés.



**Figure 3.6.** Précision moyenne du système proposé RSS-SCS sur les bases de données DDSM et RIM-One en utilisant différents types de classifieurs de base.

Si on regarde les résultats de la figure 3.7, nous observons que le système RSS-SCS génère des résultats différents selon la taille du sous-ensemble sélectionné  $L^*$ . Les résultats obtenus sur la base RIM-One indiquent que RSS-SCS offrent les meilleures performances pour une valeur de  $L^* = 15$ . Dans le cas des images mammographiques (DDSM), le RSS-SCS obtient des meilleurs résultats pour une valeur de  $L^* = 10$ .

Après la validation de notre approche, nous pouvons conclure que le système proposé sur la base DDSM, en utilisant la mesure Q-statistique comme critère de sélection, avec un ensemble initial des ADs et un nombre de classifieurs à sélectionner  $L^* = 10$ , obtient le meilleur résultat dont la précision atteint 94.84%. Dans le cas de la base RIM-One, le RSS-SCS en utilisant la mesure de désaccord ( $D$ ), avec un ensemble initial de MLPs et une taille  $L^* = 15$ , offre les meilleurs résultats dont la performance atteint 98.92%. Les résultats de cette expérience indiquent que le comportement des méthodes de sélection varie en fonction des bases de données appliquées, la mesure de diversité utilisée, le type de classifieur adopté et le nombre de classifieurs à sélectionner.



**Figure 3.7.** Précision moyenne du système proposé RSS-SCS sur les bases de données DDSM et RIM-One en utilisant différentes tailles de  $C^*$ .

### 3.3.2.2 Etude comparative du système proposé

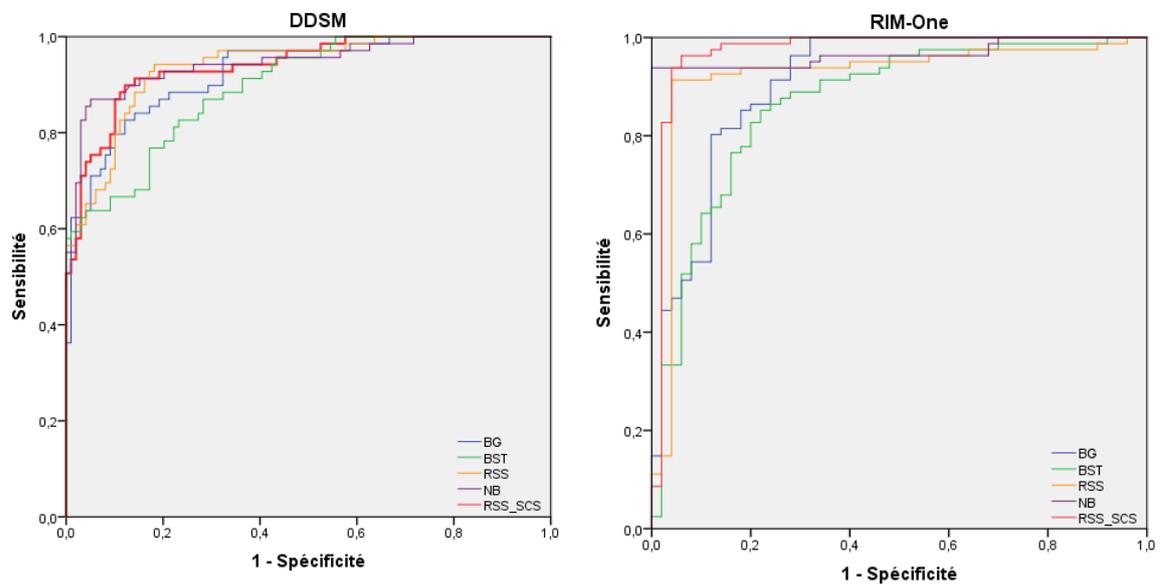
Dans le but d'examiner les meilleures performances obtenus par l'approche RSS-SCS pour chaque base de données, nous avons testé d'autres techniques de construction ensemblistes les plus utilisées dans la littérature, à savoir : le Bagging (BG) [Breiman 1996a], le Boosting (BST) [Freund 1997] et le Random Subspace (RSS) [Ho 1998]. Ainsi, RSS-SCS a été comparé avec d'autres méthodes de sélection comme N-Best (NB) [Ruta 2005] qui sélectionne les N meilleurs classifieurs en se basant sur leurs performances. Le tableau 3.4 résume les taux de précision (PRE), sensibilité (SEN), spécificité (SPE) et AUC obtenus pour chaque base de données par les différents EdC. Les courbes ROC des bases DDSM et RIM-One sont illustrés à la figure 3.8.

D'après ces résultats, nous observons que la méthode RSS-SCS donne toujours des bons résultats par rapport aux méthodes BG, BST et RSS. Ceci démontre que l'utilisation d'une phase de sélection de classifieurs durant le processus de construction d'un système multi-classifieurs est plus efficace que la combinaison de tous les membres de l'ensemble initial. Cependant, la technique de sélection NB obtient des résultats mieux que ceux donnés par la méthode RSS-SCS pour la base DDSM ; ceci peut être expliqué par le fait que le RSS-SCS basé sur les mesures de diversité peut sélectionner des membres faibles en termes de performance. Ce phénomène nous montre que l'utilisation unique des mesures

de diversité lors de la sélection des classifieurs ne peut être meilleure que l'utilisation de la précision pour trouver les sous-ensembles de classifieurs les plus compétents.

**Tableau 3.4.** Résultats obtenus sur les bases de données DDSM et RIM-One par différents EdC.

DDSM					RIM-One			
EdC	PRE	SEN	SPE	AUC	PRE	SEN	SPE	AUC
<b>Bagging</b>	92.05	87.25	94.50	92.60	92.30	90.54	93.68	90.80
<b>Boosting</b>	90.06	83.33	93.50	89.70	89.34	87.83	90.52	86.8
<b>RSS</b>	93.04	89.21	95.00	93.70	93.49	90.54	95.78	92.2
<b>NB</b>	<b>95.43</b>	<b>94.31</b>	<b>96.00</b>	<b>94.50</b>	97.63	95.94	98.94	96.6
<b>RSS-SCS</b>	94.84	<b>94.31</b>	95.10	94.00	<b>98.81</b>	<b>97.29</b>	<b>100</b>	<b>97.3</b>



**Figure 3.8.** Courbes ROC des bases de données DDSM et RIM-One.

Parmi les divers travaux proposés dans la littérature, seulement ceux qui utilisent les mêmes bases de données que la nôtre ont été choisis dans le but de comparer les résultats de notre méthode RSS-SCS. Les comparaisons de notre système avec certaines méthodes

proposées pour la détection du cancer du sein (DDSM) et du glaucome (RIM-One) sont illustrées par le tableau 3.5 et 3.6.

Ces résultats indiquent que l'approche RSS-SCS peut être considérée comme une technique d'ensemble de classification optimale pour la détection de la maladie du glaucome avec une précision de 98.81%. De ces tableaux, on peut également dire que RSS-SCS a prouvé son efficacité dans le domaine du diagnostic des images mammographies et surtout dans le cas d'images malignes avec un taux de sensibilité de 94.31%.

**Tableau 3.5.** Différents systèmes de classification du cancer du sein.

Travail	Classification	Pre	Sen	Spe
[Nazaré-Silva 2015]	SVM	83.50	92.30	82.20
[Singh 2016]	Ada-DEWNN	87.20	82.50	90.30
[Abubacker 2018]	RNA	<b>95.10</b>	92.20	96.30
[Zemmal 2018]	TSVM (semi-supervisée)	94.80	94.20	96.30
<b>Notre approche</b>	RSS-SCS	94.84	<b>94.31</b>	95.10

**Tableau 3.6.** Différents systèmes de classification pour le diagnostic du glaucome.

Travail	Classification	Pre	Sen	Spe
[Raja 2015]	SVM	98.00	97.50	96.30
[Ghosh 2015]	RNA	87.47	87.50	87.45
[Haleem 2016]	SVM	94.40	92.30	95.30
[Bechar 2018]	Co-forest	90.85	/	/
[Zemmal 2018]	TSVM	98.00	99.10	97.70
<b>Notre approche</b>	RSS-SCS	<b>98.81</b>	97.29	<b>100</b>

### 3.4 Conclusion

Dans ce chapitre, nous avons présenté un système de classification des images médicales basé sur le paradigme de surproduction et sélection. Pour représenter les images

source sous forme de vecteurs de primitives, ce système a été précédé par une phase d'extraction de caractéristiques qui consiste à intégrer différentes familles de caractéristiques de texture et de forme. Le système multi-classifieurs proposé est basé sur la combinaison de la méthode Random Subspace pour la génération de classifieurs et la sélection statique en utilisant différentes mesures de diversité. Notre objectif principal est d'étudier plus en détail le rôle de la notion de diversité dans les processus de construction et de sélection des ensembles de classifieurs.

Plusieurs expérimentations ont été menées tout au long de ce chapitre ; elles ont montré la supériorité de notre méthode par rapport aux méthodes Bagging, Boosting et Random Subspace. Notre système a prouvé également son efficacité par rapport à des systèmes basés sur un seul classifieur avec une précision de 94.84% sur la base DDSM et 98.81% sur la base RIM-One. Les résultats obtenus ont indiqué également que la performance de la phase de sélection est liée au type de classifieurs membres, la mesure de diversité utilisée et le nombre de classifieurs à sélectionner.

L'étude comparative établie par rapport à d'autres paradigmes de sélection confirme l'observation faite dans les travaux précédents [Kuncheva 2003a ; Ruta 2005]. L'utilisation unique de la diversité comme critère de sélection expose inévitablement à la construction d'un ensemble constitué de classifieurs peu précis. Les ensembles de classifieurs qui ont une grande valeur de diversité ne sont pas nécessairement ceux qui produisent le taux de classification le plus élevé. La précision de classifieurs sélectionnés est un facteur important qui a prouvé son rôle dans l'amélioration de la performance du processus de sélection. Il semble y avoir un compromis entre la diversité et la précision individuelle.

Pour cette raison, nous allons maintenant tenter de chercher à approfondir l'étude et la compréhension de ce compromis au cours du chapitre suivant, en proposant un paradigme de sélection basé sur la précision et la diversité durant le processus de création d'un ensemble de classifieurs.

# Chapitre 4

## Sélection Statique de Classifieurs

### Combinant la Diversité et la Méthode mRMR

#### 4.1 Introduction

La sélection d'un sous-ensemble de classifieurs devient une étape impérative dans le processus de construction d'un système ensembliste. Elle vise à rechercher le sous-ensemble optimal de classifieurs qui améliore la performance et diminue également la complexité d'un système multi-classifieurs. Le critère de sélection est l'un des éléments de base du processus de sélection de classifieurs [Aksela 2006 ; Álvarez 2015 ; Verbiest 2016]. En effet, la plupart des travaux dans la littérature sont basés sur l'utilisation de la diversité ou de la précision comme critère de sélection. Les chercheurs ont montré que la sélection des sous-ensembles de classifieurs était plus performante lorsque les candidats de l'ensemble sont caractérisés par une précision élevée et une grande diversité en même temps [Rodriguez 2006 ; Mendialdua 2015]. Par conséquent, il est nécessaire d'équilibrer ces deux facteurs par un compromis entre eux.

La sélection de classifieurs peut être vue comme la sélection de caractéristiques mais à un niveau d'abstraction plus élevé. Dans [Vincent 2011], les auteurs ont défini la sélection de caractéristiques comme un processus de recherche permettant de trouver un sous-ensemble "pertinent" de caractéristiques parmi celles de l'ensemble de départ, en utilisant un ou plusieurs critères de sélection. En effet, la sélection de caractéristiques utilise des méthodes qui peuvent facilement être adaptées à la sélection de classifieurs. Plusieurs

méthodes de sélection de caractéristiques ont été proposées dont le critère diffère en corrélation, ou relevance entre caractéristiques. La particularité de la méthode mRMR (*minimum Redundancy Maximum Relevance*) consiste à sélectionner un sous-ensemble optimal de caractéristiques qui sont dissemblables au maximum les uns aux autres (peu redondantes) et ont aussi la plus haute relevance par rapport à la classe cible [Peng 2005]. La méthode mRMR a tenté de déterminer un compromis entre la relevance et la redondance des caractéristiques en éliminant les membres similaires tout en préservant les plus pertinents [El Akadi 2011 ; Li 2012a].

Dans le chapitre précédent, un modèle de sélection de classifieurs basé sur les mesures de diversité a été proposé. Les résultats expérimentaux ont établi l'impact de la diversité dans le processus de sélection de classifieurs. Néanmoins, nous avons remarqué que l'utilisation de la diversité seule lors de la phase de sélection est insuffisante pour améliorer la performance de l'ensemble. La précision est donc un facteur primordial prouvant son efficacité dans le processus de sélection [Kuncheva 2003a ; Ruta 2005].

Dans ce chapitre, nous cherchons à clarifier le lien existant entre la précision et la diversité dans un ensemble de classifieurs, en abordant sa construction selon deux approches de sélection différentes. La première méthode proposée est basée sur un nouveau compromis « diversité-précision » intégrant la méthode mRMR et les mesures de diversité. Notre objectif est de sélectionner les classifieurs précis tout en conservant un haut niveau de diversité entre eux. Cette proposition vise à réduire le coût de calcul de l'étape de sélection en appliquant l'algorithme d'ordonnancement et choisissant un nombre spécifique de classifieurs [Cheriguene 2016a].

La limite principale de cette approche concerne la question du nombre de classifieurs à sélectionner pour une application donnée. Généralement, la taille du meilleur sous-ensemble de classifieurs n'est pas connue a priori, ce qui nous mène à chercher d'autres algorithmes que l'algorithme d'ordonnancement. De plus, chaque classifieur a été évalué individuellement et la performance globale de l'ensemble à sélectionner n'était pas prise en compte. Pour remédier à ces limites, nous avons proposé d'améliorer notre approche en utilisant un algorithme de recherche glouton basé sur un nouveau compromis diversité-performance prenant en considération la performance globale du sous-ensemble candidat afin d'obtenir un ensemble optimal de classifieurs [Cheriguene 2018].

Un ensemble de 100 classifieurs hétérogènes générés à l'aide de la méthode Random Subspace a été utilisé comme pool initial de classifieurs. Ainsi, la règle de vote pondéré a été utilisée pour agréger les prédictions individuelles des prédicteurs sélectionnés et prendre la décision finale.

Dans ce chapitre, nous commençons par la présentation du principe de la méthode de sélection de caractéristiques mRMR. Ensuite, nous détaillons notre contribution consistant à mettre en place deux approches de sélection statique de classifieurs basées sur l'hybridation de la méthode mRMR et des mesures de diversité. Nous évaluons également les approches proposées sur différents bases de données de référence. Enfin, nous concluons sur quelques extensions qui seront proposées.

### 4.2 La méthode mRMR

Le problème de la sélection d'un sous-ensemble de caractéristiques peut être vu comme une recherche dans un espace d'hypothèses ou un ensemble de solutions possibles [Blum 1997]. La sélection de caractéristiques a le même objectif de la sélection de classifieurs, qui est la réduction du coût de calcul et de nombre de caractéristiques tout en conservant la qualité de performance de reconnaissance et/ou de classification. Différents critères ont été développés pour évaluer les sous-ensembles candidats, à savoir : *l'information mutuelle*.

Dans la théorie de l'information et la théorie des probabilités, l'information mutuelle de est une quantité mesurant la dépendance statistique entre deux variables aléatoires [Cover 1994 ; Paninski 2003]. L'information mutuelle est une mesure symétrique de l'information commune de deux variables. Elle mesure la quantité d'information qu'apporte une variable aléatoire sur une autre. C'est la diminution de l'incertitude associée à une variable aléatoire grâce à la connaissance de l'autre variable aléatoire.

Supposons que nous avons une variable aléatoire  $Z$  qui prend des valeurs dans l'ensemble  $\{z_1, z_2, \dots, z_n\}$  avec les probabilités correspondantes  $(p(z_1), p(z_2), \dots, p(z_n))$ . L'information mutuelle ( $IM$ ) entre deux variables discrètes aléatoires  $Z$  et  $K$ , est donné par la formule suivante:

$$IM(Z; K) = \sum_i \sum_j p(z_i, k_j) \log_2 \frac{p(z_i, k_j)}{p(z_i) p(k_j)} \quad (4.1)$$

Où,  $p(z_i, k_j)$  est la densité de probabilité jointe entre les deux variables  $Z$  et  $K$ . Si la valeur d'information mutuelle entre deux variables  $Z$  et  $K$  est grande, cela signifie que  $Z$  et  $K$  sont dépendantes.

En utilisant le concept d'information mutuelle, la méthode mRMR sélectionne le sous-ensemble de caractéristiques qui sont peu redondantes et ont aussi la plus haute relevance par rapport à la classe cible [Peng 2005]. La méthode du maximum de relevance (*Maximum Relevance*, MR) est défini comme la valeur maximale de l'information mutuelle entre les entités individuelles  $X_i$  et les étiquettes de classe  $\Omega$ . Étant donné un ensemble de  $F$  caractéristiques, la relevance est définie par:

$$MR = \text{Max}_{X_i \in F} IM(X_i, \Omega) \quad (4.2)$$

Le critère de la redondance minimale (minimum Redundancy, mR) est défini comme la valeur minimale de l'information mutuelle entre une paire de caractéristiques  $X_i$  et  $X_j$ :

$$mR = \text{Min}_S \frac{1}{|F|} \sum_{X_i, X_j \in F} IM(X_i, X_j) \quad (4.3)$$

Le critère de minimisation de la redondance et de maximisation de la relevance (mRMR) fusionne ces deux critères [Peng 2005]. Afin de sélectionner le sous-ensemble optimal de caractéristiques les plus pertinentes à la cible  $\Omega$  et les moins redondantes entre eux, deux méthodes différentes, à savoir: le *MID* (*Mutual Information Difference criterion*) et *MIQ* (*Mutual Information Quotient criterion*), ont été proposées pour combiner les deux critères des équations (4.2) et (4.3) en une seule fonction de score et représente le critère mRMR:

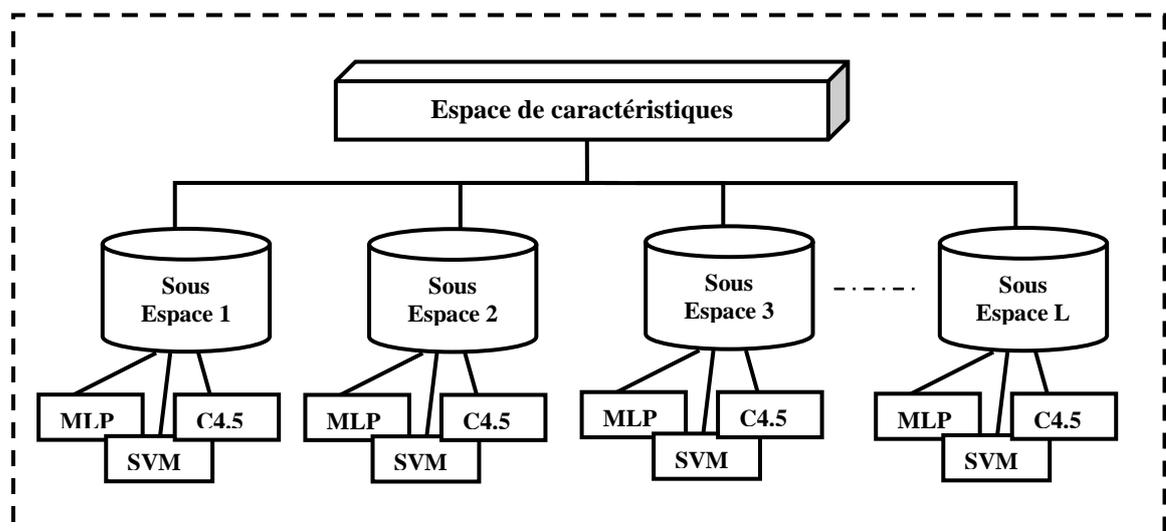
$$MID = \text{Max}_F (IM(X_i, \Omega) - \frac{1}{|F|} \sum_{X_i, X_j \in F} IM(X_i, X_j)) \quad (4.4)$$

$$MIQ = \text{Max}_F (IM(X_i, \Omega) / \frac{1}{|F|} \sum_{X_i, X_j \in F} IM(X_i, X_j)) \quad (4.5)$$

## 4.3 Sélection statique de classifieurs basé sur la méthode mRMR et les mesures de diversité

### 4.3.1 Génération des classifieurs

La méthode Random Subspace (RSS) est une des méthodes de génération de la diversité qui a prouvé son efficacité dans de nombreux problèmes de classification [Ho 1998 ; Hosseini 2016]. La technique RSS consiste à former un ensemble de divers classifieurs, en utilisant différents sous-espaces de caractéristiques. Dans ce travail, nous avons utilisé la méthode RSS à la base de trois modèles de classification, tels que: MLP, SVM et C4.5, afin de générer un pool de 100 classifieurs hétérogènes [Cheriguene 2018]. Le principe de base est d'entraîner, à chaque itération, ces trois classifieurs simultanément sur un sous-espace aléatoire de l'espace de description. La figure 4.1 illustre le schéma de création de l'ensemble initial de classifieurs.



**Figure 4.1.** Schéma de génération d'ensemble préliminaire des classifieurs [Cheriguene 2018].

### 4.3.2 Sélection de classifieurs

La sélection de classifieurs peut être définie comme un processus de recherche permettant de choisir un sous-ensemble optimal de classifieurs à partir d'un ensemble plus large en fonction de certains critères [Álvarez 2015 ; Choi 2016]. En pratique, la performance des classifieurs et la diversité entre eux sont les deux facteurs qui doivent être prises en considération pour construire un ensemble performant. La sélection de

classifieurs les plus divers ou les plus précis ne permettrait pas toujours d'obtenir de meilleures performances [Mendialdua 2015 ; Visentini 2016]. Par conséquent, il est nécessaire d'équilibrer ces deux facteurs en utilisant un compromis entre eux. Dans ce chapitre, nous proposons un nouveau schéma de sélection de prédicteurs nommé : *Maximum Relevance Maximum Diversity (MRMD)* [Cheriguene 2016a] qui combine à la fois la relevance (se basant sur l'information mutuelle) et les mesures de diversité pour sélectionner les  $L^*$  meilleurs candidats parmi l'ensemble initial de classifieurs.

#### 4.3.2.1 L'approche proposée MRMD

L'approche MRMD tente d'intégrer l'utilisation de la précision des classifieurs individuels et des mesures de diversité comme critère de sélection [Cheriguene 2016a]. Le sous-ensemble de classifieurs fortement corrélés avec les étiquettes de classe sera pris en considération durant la sélection. La relevance a été donc utilisée pour estimer la performance de chaque membre de l'ensemble  $C$ . Tout d'abord, les sorties des classifieurs ont été représentées par une matrice de taille  $N * L$ , où  $N$  est le nombre d'objets de la base de validation et  $L$  est le nombre de classifieurs. Chaque colonne représente les sorties  $X_{c_j}$  du classifieur  $c_j$  pour tous les  $N$  exemples,  $X_{c_j} = [X_{c_j,1}, X_{c_j,2}, \dots, X_{c_j,N}]^T$  avec  $X_{c_j,i}$  est la sortie du classifieur  $c_j$  pour la forme  $x_i$ . Nous définissons la relevance (*REV*) d'un classifieur  $c_j$  comme la valeur de l'information mutuelle entre les sorties de ce classifieur et le vecteur des classes  $X_\Omega$  des  $N$  exemples:

$$REV(c_j) = IM(X_{c_j}, X_\Omega) \quad (4.6)$$

D'autre part, les membres sélectionnés doivent être différents les uns des autres pour assurer un certain niveau de diversité. Cette dernière est donc nécessaire pour éliminer les classifieurs redondants. De plus, Les auteurs de [Li 2012b] ont prouvé qu'une seule mesure de diversité n'était pas suffisante pour capturer toutes les diversités pertinentes dans l'ensemble. Pour Cela, nous avons proposé une combinaison des mesures de diversité pour quantifier la diversité. Dans notre cas, nous avons choisi les deux mesures Q-statistique ( $Q$ ) et Désaccord ( $D$ ), ce choix sera montré expérimentalement. La diversité combinée  $Div_{com}$  entre deux classifieurs  $c_j$  et  $c_h$  est proposé comme suit :

$$Div_{com}(c_j, c_h) = \frac{1}{2} \cdot \left( \frac{1 - Q(c_j, c_h)}{2} + D(c_j, c_h) \right) \quad (4.7)$$

La diversité globale d'un classifieur  $c_j$  est déterminée par la moyenne des valeurs de diversité combinée  $Div_{com}(c_j, c_h)$  du classifieur  $c_j$  par rapport aux autres membres  $c_h$  de l'ensemble  $E_t$ , qui est définie par :

$$Div(c_j) = \frac{1}{|C| - 1} \sum_{h \neq j} Div_{com}(c_j, c_h) \quad (4.8)$$

Le critère de sélection MRMD est déterminé comme suit :

$$MRMD(c_j) = \frac{Div(c_j)}{REV(c_j)} \quad (4.9)$$

La sélection d'un sous-ensemble optimal de classifieurs nécessite une recherche exhaustive parmi  $2^L - 1$  combinaisons de classifieurs possibles. La méthode MRMD utilise l'algorithme de recherche par ordonnancement pour réduire l'espace de recherche et diminuer la complexité de calcul [Cheriguene 2016a]. En premier lieu, les classifieurs de l'ensemble initial  $C = \{c_1, c_2, \dots, c_L\}$ , sont triés par ordre décroissant selon le critère  $MRMD$ . En second lieu, les  $L^*$  meilleurs classifieurs sont sélectionnés pour former l'ensemble final  $C^*$ . Le choix de la valeur de  $L^*$  reste un problème à résoudre où elle varie selon l'application visée [Cheriguene 2016a].

#### 4.3.2.2 *L'approche MRMD améliorée utilisant la recherche séquentielle et la performance globale (MRMD-II)*

##### a) *Critère de sélection*

Dans l'approche MRMD, chaque classifieur a été évalué individuellement sans tenir en compte de la performance de l'ensemble ainsi que sa diversité. En effet, ces deux dernières sont des critères importants durant la sélection d'un ensemble optimal de prédicteurs [Zhou 2004, Gacquer 2009]. Dans le but d'améliorer notre précédente proposition, nous avons intégré les deux critères de la performance ensemble et la diversité globale dans la mesure MRMD [Cheriguene 2018].

De la même manière de l'approche MRMD, les sorties d'un sous ensemble de classifieurs  $E_t$  sont présentées par un vecteur  $X_{E_t} = [X_{E_t,1}, X_{E_t,2}, \dots, X_{E_t,N}]$  de taille  $N$ , où  $N$  est le nombre des exemples de la base de validation, et  $X_{E_t,i}$  est la réponse de l'ensemble  $E_t$  pour l'exemple  $x_i$ . Ces réponses sont obtenues en combinant les décisions des membres de  $E_t$  par la règle de vote majoritaire. Dans la version révisée qu'est MRMD-II, la relevance a été utilisée pour estimer la performance d'un ensemble de classifieurs ainsi que celles de ses membres. La relevance d'un ensemble  $E_t$  est donné par :

$$REV(E_t) = IM(X_{E_t}, X_\Omega) \quad (4.10)$$

Le critère  $MRMD - II$  est basé donc sur une combinaison de la performance globale de l'ensemble et la moyenne des performances individuelles de ses membres. Le paramètre  $MRMD - II$  fusionne les critères de relevance et de diversité en utilisant un compromis entre eux. Le  $MRMD - II$  d'un ensemble  $E_t$  est obtenu à l'aide de la formule suivante:

$$MRMD - II(E_t) = \frac{REV(E_t)}{Div(E_t)} + \frac{1}{|E_t|} * \sum_{c_j \in E_t} \frac{REV(c_j)}{Div(c_j)} \quad (4.11)$$

, où  $Div(c)$  est la diversité moyenne du classifieur  $c_j$  par rapport aux autres membres  $c_h$  de l'ensemble  $E_t$ , qui est définie à l'équation 4.8.  $Div(E_t)$  est la diversité globale de l'ensemble  $E_t$  qui est déterminée par:

$$Div(E_t) = \frac{2}{|E_t|(|E_t| - 1)} \sum_{j=1}^{|E_t|-1} \sum_{h=j+1}^{|E_t|} Div_{com}(c_j, c_h) \quad (4.12)$$

**b) Algorithme de recherche**

Bien que l'approche MRMD conduit à réduire la complexité de la phase de sélection, il n'y a aucune garantie que la solution optimale sera trouvée. En pratique, on ne connaît pas toujours le nombre de classifieurs pertinents et donc l'application d'un autre algorithme de recherche est imposée. Pour cette raison, la méthode améliorée (MRMD-II) a été renforcé par un algorithme de recherche glouton qui ne nécessite pas de déterminer à priori la taille des classifieurs à sélectionner [Cheriguene 2018]. À partir d'un ensemble contenant le classifieur le plus pertinent, les autres classifieurs sont sélectionnés séquentiellement un par un de telle sorte que chaque classifieur sélectionné permettant d'obtenir des meilleures performances lors de la combinaison de celle-ci avec notre sous-ensemble déjà sélectionné.

En d'autres termes, à chaque itération, un classifieur est ajouté pour fournir des meilleures performances avec le sous ensemble qui a déjà été sélectionné.

Soit  $C = \{c_1, c_2, \dots, c_L\}$  un ensemble initial de  $L$  classifieurs de base et  $C^*$  est le sous-ensemble optimal final. Supposons que  $C_t^*$  est le sous-ensemble optimal sélectionné à l'itération  $t$ , avec  $C_0^* = \{c_{best}\}$  et  $c_{best}$  est le classifieur le plus pertinent de l'ensemble initial  $C$ , avec  $MDMR - II(E_t)$  est la mesure de compétence d'un ensemble  $E_t$ .

Le processus de sélection MRMD-II fonctionne comme suit: à chaque itération  $t$ , tous les sous-ensembles possibles  $E_{t,j} = C_{t-1}^* \cup \{c_j\}$  sont créés en utilisant l'ensemble des classifieurs restants  $c_j \in C - C_{t-1}^*$ . Le meilleur ensemble  $E_{t,j}$  qui maximise la mesure  $MRMD - II$  par rapport aux autres ensembles  $E_{t,h}$  est sélectionné, c'est-à-dire  $MDMR - II(E_{t,j}) > MDMR - II(E_{t,h})$  pour tout  $j \neq h$ . Si la mesure  $MDMR - II$  du meilleur ensemble  $E_{t,j}$  est meilleure que l'ensemble optimal  $C_{t-1}^*$  de l'itération précédente  $t - 1$ , c'est à dire  $MDMR - II(E_{t,j}) > MDMR - II(C_{t-1}^*)$ , alors  $E_{t,j}$  sera l'ensemble optimal de l'itération  $t$  où,  $C_t^* = E_{t,j}$ . Sinon, l'algorithme s'arrête et renvoie l'ensemble  $C_{t-1}^*$  comme l'ensemble optimal final  $C^*$  où,  $C^* = C_{t-1}^*$ .

L'algorithme 4.1 explique les principales étapes du processus de sélection de classifieurs proposé MRMD-II.

**Algorithme 4.1:** Processus de sélection statique de classifieurs MRMD-II

---

**Entrées:**

base de données de validation étiquetée  $\mathbf{VD}$  avec  $M$  Classes.

Ensemble initial de  $L$  classifieurs  $C^* = \{c_1, c_2, \dots, c_L\}$ .

**Sorties:**

L'ensemble final des classifieurs sélectionnés  $C^*$ .

---

1. **Calculer** la relevance  $REV$  de chaque classifieur  $c_j$  sur  $\mathbf{VD}$  en utilisant l'équation (4.6).

2. **Si**  $(REV(c_j) > REV(c_h))$  pour tous  $h \neq j$ , alors  $C_0^* = \{c_j\}$ .

3. **Pour**  $t=1, \dots, L-1$

**Construire** l'ensemble de classifieurs  $E_{t,j} = C_{t-1}^* \cup \{c_j\}$ , pour tous  $c_j \in C - C_{t-1}^*$ .

**Calculer** la mesure de performance  $MRMD - II$  sur  $\mathbf{VD}$  pour chaque ensemble de classifieurs  $E_{t,j}$  en utilisant l'équation (4.11).

**Si**  $MDMR - II(E_{t,j}) > MDMR - II(E_{t,h})$  pour tous  $j \neq h$  et  $MDMR - II(E_{t,j}) > MDMR - II(C_{t-1}^*)$ , alors  $C_t^* = E_{t,j}$ .

**Sinon**  $C_t^* = C_{t-1}^*$

**Fin Si**

**Fin pour**

4. **Retourner**  $C^*$  comme ensemble final.

---

### 4.3.3 Combinaison de classifieurs sélectionnés

Après la phase de sélection, les prédictions individuelles des apprenants sélectionnés seront combinées pour prédire la décision finale. Dans ce chapitre, la méthode de vote pondérée a été utilisée pour fusionner les décisions des classifieurs sélectionnés.

Soit  $C^* = \{c_1, c_2, \dots, c_{L^*}\}$  l'ensemble de  $L^*$  classifieurs sélectionnés,  $x$  est l'exemple d'entrée et  $c_{i,j}$  est la sortie du  $j^{\text{ème}}$  classifieur pour la  $i^{\text{ème}}$  classe. Considérons  $w_j$  le poids du  $j^{\text{ème}}$  classifieur, la prédiction finale utilisant le vote majoritaire pondéré peut être définie comme suit:

$$VMP(x) = \max_{i \in \Omega} \sum_{i=1}^{L^*} w_j c_{i,j} \text{ et } \sum_{i=1}^L w_j = 1 \quad (4.13)$$

Plusieurs schémas ont été proposés pour estimer les poids des classifieurs [Moreno-Seco 2006]. En général, ces poids sont estimés en utilisant la précision individuelle de chaque classifieur. Soit  $p_i$  et  $p_j$  sont les précisions des  $i^{\text{ème}}$  et  $j^{\text{ème}}$  classifieurs sur la base de validation. Le poids  $w_j$  est calculé par:

$$w_j = \frac{p_j}{\sum_i p_i} \quad (4.14)$$

Dans ce travail, la règle du vote pondéré meilleur-pire (Best-Worst Weighted Vote) a été adoptée pour quantifier les poids. L'idée de base de ce schéma de fusion est d'identifier le meilleur et le pire membre de l'ensemble sélectionné. Dans ce cas, les valeurs  $p_i$  sont déterminées à l'aide de l'expression suivante:

$$p_i = \frac{e_k - e_b}{e_w - e_b} + \varepsilon \quad (4.15)$$

, où,  $e_w$  indique l'erreur maximale parmi les classifieurs et  $e_b$  est l'erreur minimale. La valeur  $p_i$  varie dans l'intervalle  $[0,1]$ .

## 4.4 Etude expérimental

### 4.4.1 Paramètres expérimentaux

Les performances des méthodes présentées précédemment ont été évaluées sur deux bases de données de Kuncheva (*Kuncheva collection for real datasets*, LKC) [Kuncheva 2004] et 18 benchmarks réelles issues de l'UCI (*Machine Learning Repository*) [Asuncion 2007]. Les bases que nous avons utilisées pour nos expérimentations sont issues de différents domaines de reconnaissance de formes et représentent des problèmes de classification supervisée fréquemment utilisés dans la littérature. De plus, la dimension des données est dans l'intervalle  $[7 : 61]$ , le nombre de classes d'origine dans  $[2 : 10]$  et le nombre de données disponibles sont dans l'intervalle  $[98 : 5000]$ . Les différents jeux de données utilisés sont décrits dans le tableau 4.1.

**Tableau 4.1.** Caractéristiques des bases de données utilisées.

Base de données	#Instances	#Caractéristiques	#Classes
Breast-w	699	9	2
Credit-a	690	43	2
Credi-g	1000	61	2
Colic	368	22	2
Contraction	98	27	2
Diabetes	768	8	2
Echocardiogram	132	12	2
Ecoli	336	7	8
Glass	214	9	6
Heart-statlog	123	25	2
Hepatitis	155	19	2
Ionosphere	351	34	2
Landast	6435	36	6
Laryngeal-2	213	16	2
Parkinson's	195	22	2
Segment	2310	19	7
Vote	435	16	2
Waveform	5000	21	3
Yeast	1486	8	10
Zoo	101	16	7

Afin de réduire l'impact de variabilité dans les exemples d'apprentissage, le protocole de validation croisée en  $k$  blocs [Krogh 1995] a été appliqué sur ces jeux de données. Chaque base de données est divisée au hasard en  $k$  partitions de tailles à peu près égales. Tour à tour, une partition est utilisée pour former la base de test, une autre partition est utilisée comme base de validation et les partitions restantes sont utilisées pour constituer la base d'apprentissage. Un problème peut également se poser en ce qui concerne le nombre de partitions à créer. Les auteurs de [García 2015] ont déclaré que 5 ou 10 blocs peuvent être un bon choix avec des bases de données de tailles différentes. En outre, les répétitions du processus étaient également souhaitable afin d'assurer la commutation entre les données d'entraînement et les données de test autant que possible et d'éviter les écarts élevés.

Par conséquent, une validation croisée en 5 blocs a été adoptée pour des bases qui ont moins de 300 instances) (respectivement, 10 blocs pour des bases qui ont plus de 300 instances), afin de parvenir à des conclusions fiables et solides concernant la performance du modèle proposé. Dans ce chapitre, une validation croisée de  $5 \times 5$  blocs (respectivement,  $5 \times 10$  blocs) a été appliquée sur chaque base de données, et le processus a été répété 5 fois (respectivement, 10 fois) pour chacune, donnant un total de 25 résultats (respectivement, 50) de test. Les moyennes des résultats sont présentées et discutées dans les sections suivantes.

### 4.4.2 Évaluation de la méthode MRMD

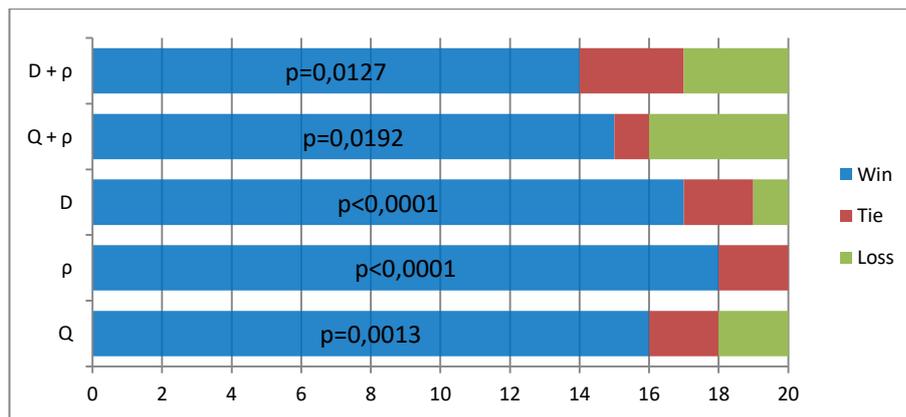
#### 4.4.2.1 Analyse des mesures de diversité

Comme l'était déjà mentionné, une seule mesure de diversité n'était pas suffisante pour atteindre des taux de reconnaissance élevés. Dans ce cas, deux questions se posent: devons-nous combiner différentes mesures de diversité? Et quelles mesures devrais être choisir Afin de répondre à ces questions, nous avons évalué une version modifiée de l'approche proposée MRMD dans laquelle une seule mesure de diversité a été utilisée à la fois. Dans cette version modifiée, l'équation 4.8 ne combine pas les mesures de diversité et le critère  $Div_{com}(c_j, c_h)$  est remplacé par une seule mesure de diversité. Les mesures adoptées sont des mesures par-paire qui ont prouvé son efficacité dans la littérature [Kuncheva 2003a], à savoir : Q-statique ( $Q$ ), corrélation ( $\rho$ ) et Désaccord ( $D$ ). D'autre part, les différentes combinaisons de ces mesures ( $Q + \rho$ ;  $D + \rho$  et  $Q + D$ ) ont été évaluées. Pour effectuer cette expérimentation, nous avons utilisé différents nombre de classifieurs à sélectionner  $L^* = \{10, 25\}$ . Dans ce cas, l'objectif est d'évaluer l'impact de variation de la taille  $L^*$  sur notre méthode de sélection MRMD.

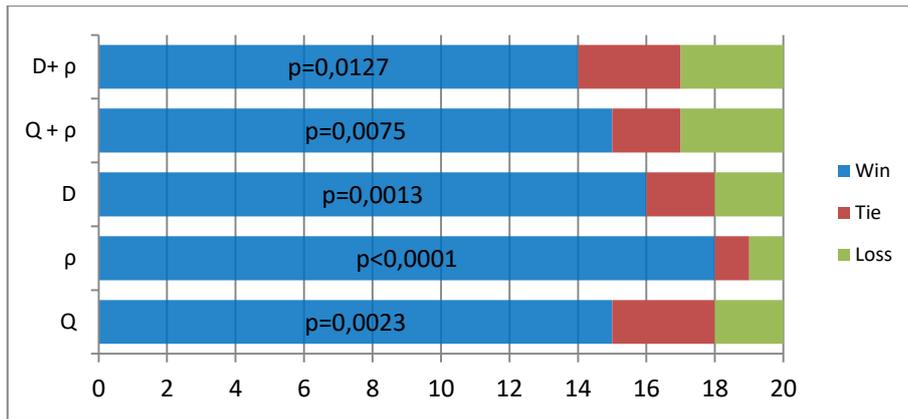
Comme nous avons plusieurs méthodes à comparer sur différentes bases de données, nous avons utilisé les tests des signes (*sign test*), nommé aussi les tests statistiques, «win/tie/loss» et le «two-tailed p-value» [Yang 2001] permettant de détecter les différences statistiques entre deux stratégies. Le test de signe win/tie/loss présente le nombre de fois pour lesquelles un classifieur  $A$  a produit des résultats de précision meilleurs/neutres/inférieurs qu'une autre méthode de classification  $B$ . Le p-value est la probabilité d'obtenir une corrélation aussi grande que la valeur observée par hasard lorsque la corrélation réelle est nulle.

Selon ce test de signe [Yang 2001], pour  $N$  ensembles de données, une méthode  $A$  est meilleure que  $B$ , avec un niveau de confiance de 5%, si le nombre de victoires plus la moitié des égalités est au moins égal à  $N / 2 + 1,96 \sqrt{N} / 2$ . Pour  $N = 20$  base de données, le nombre de victoires nécessaires est 15. D'autre part, si la valeur de p-value est significativement faible (p-value  $< 0,05$ ), il est raisonnable de conclure qu'il est peu probable que le résultat a été obtenu par hasard et donc que le test de signe win/tie/loss représente un avantage significatif de la méthode  $A$  par rapport à la méthode  $B$ . Les figures 4.2 et 4.3 illustrent les tests des signes entre les résultats de précision de l'approche MRMD utilisant la mesure de diversité ( $Q + D$ ) et les résultats de MRMD basée sur différentes estimations de diversité.

Les résultats des tests des signes win/tie/loss montrent une supériorité significative de critère ( $Q + D$ ) par rapport aux mesures ( $Q + \rho$ ),  $Q$ ,  $\rho$  et  $D$ . Toutefois,  $Q + D$  suggère un avantage marginal par rapport au critère restant ( $D + \rho$ ), notez, cependant, que leur p-value pour le test de signe win/tie/loss par rapport à  $Q + D$  est de 0,0127, ce qui serait accepté comme significatif. En outre, nous remarquons aussi que les résultats obtenus par la combinaison des mesures de diversité sont significativement meilleure que ceux donnés par l'utilisation d'une seule mesure de diversité. Nous pouvons constater à travers ces résultats que l'ajout du critère de  $Div_{com}$  a nettement amélioré la performance de notre méthode MRMD.



**Figure 4.2.** Comparaison entre les résultats obtenus par MRMD utilisant différentes mesures de diversité, avec  $L^* = 10$ .



**Figure 4.3.** Comparaison entre les résultats obtenus par MRMD utilisant différentes mesures de diversité, avec  $L^* = 25$ .

#### 4.4.2.2 Etude comparative de la méthode proposée MRMD

Dans le but d'examiner les meilleures performances obtenus par la méthode proposée MRMD, nous avons testé d'autres méthodes de sélection de classifieurs en utilisant le même algorithme de recherche de la méthode MRMD tout en changeant le critère de sélection. Les critères utilisés à cette étude comparative sont :  $Div_{com}$ ,  $REV$  et mRMR. Pour effectuer cette expérimentation, nous avons utilisé deux nombres de classifieurs à sélectionner  $L^* = \{10, 25\}$ . Les résultats obtenus sont présentés au tableau 4.2. La moyenne de précision, le test win/tie/loss et le p-value correspond pour chaque méthode sont également présentés en bas de ce tableau.

Ce tableau montre que les pourcentages de précision de notre approche sont meilleurs sur la plus part des jeux de données. Les résultats de la précision de MRMD sont meilleurs que ceux obtenues par les autres approches de sélection de classifieurs pour 17 cas sur 20 (avec  $L^* = 10$  ou  $L^* = 25$ ). En ce qui concerne les win/tie/loss (p-value) tests, on voit clairement que le critère de sélection MRMD utilisant une taille  $L^* = 10$  est meilleur que les autres critères  $Div_{com}$ ,  $REV$  et mRMR, avec des scores d'évaluation de 18/0/2 (0.0004), 17/0/3 (0.0026) et 20/0/0 (<0.0001) respectivement. Pour une taille  $L^* = 25$ , le critère MRMD obtient des signes de test win/tie/loss statistiquement significatifs par rapport aux critères  $Div_{com}$ ,  $REV$  et mRMR, avec des scores d'évaluation respectivement de 17/2/1 (0.0001), 18/1/1 (<0.0001) et 19/0/1 (<0.0001). Ces résultats nous permettent de distinguer les avantages de combinaison de la relevance et des mesures de diversité comme critère de sélection.

**Tableau 4.2.** Taux de classification des différentes méthodes de sélection, en utilisant différentes valeurs de  $L^*$ .

Base de données	$L^* = 10$				$L^* = 25$			
	$Div_{com}$	$REV$	$mRMR$	$MRMD$	$Div_{com}$	$REV$	$mRMR$	$MRMD$
Breast-w	92.64	93.27	91.34	<b>95.87</b>	95.63	94.28	93.64	<b>97.52</b>
Credit-a	84.97	82.53	81.12	<b>86.72</b>	86.14	87.43	86.23	<b>87.97</b>
Credi-g	75.18	72.64	75.39	<b>77.05</b>	75.39	75.94	74.19	<b>77.23</b>
Colic	80.87	81.45	79.38	<b>83.64</b>	<b>85.23</b>	81.78	82.63	<b>85.23</b>
Conraction	90.52	89.10	90.52	<b>92.38</b>	91.21	<b>91.88</b>	89.71	91.45
Diabetes	73.42	73.06	72.08	<b>75.36</b>	74.29	74.74	73.28	<b>75.12</b>
Echocardiogram	<b>83.41</b>	<b>83.41</b>	81.38	83.27	85.47	85.47	84.69	<b>86.85</b>
Ecoli	85.54	84.87	84.60	<b>85.80</b>	86.03	85.94	85.30	<b>86.14</b>
Glass	75.67	<b>77.23</b>	76.45	76.63	78.23	77.56	78.59	<b>78.94</b>
Heart-statlog	79.34	80.69	80.25	<b>82.79</b>	<b>82.64</b>	81.34	81.27	<b>82.64</b>
Hepatitis	85.93	85.41	85.41	<b>87.36</b>	87.12	88.67	<b>90.53</b>	89.37
Ionosphere	90.78	91.62	92.44	<b>93.55</b>	92.33	<b>93.55</b>	92.71	<b>93.55</b>
Landsat	89.23	89.01	87.53	<b>92.43</b>	92.47	91.36	90.88	<b>93.68</b>
Laryngeal-2	94.25	95.84	93.77	<b>96.09</b>	95.19	95.42	94.12	<b>96.67</b>
Parkinson's	87.34	88.20	85.36	<b>90.16</b>	87.15	89.14	88.34	<b>90.33</b>
Segment	92.15	94.71	93.91	<b>95.27</b>	93.75	92.73	92.34	<b>94.71</b>
Vote	96.23	96.27	94.55	<b>96.57</b>	95.18	96.43	94.37	<b>96.57</b>
Waveform	84.74	85.19	85.02	<b>87.51</b>	<b>89.50</b>	86.12	86.34	87.68
Yeast	<b>79.12</b>	79.09	78.42	78.83	80.68	80.43	79.84	<b>81.82</b>
Zoo	92.83	91.74	91.69	<b>96.67</b>	93.65	92.10	93.75	<b>94.18</b>
<b>Moyenne</b>	85.72	85.77	85.03	<b>87.70</b>	87.36	87.12	86.64	<b>88.38</b>
<b>Win/tie/loss</b>	18/0/2	17/0/3	20/0/0	/	17/2/1	18/1/1	19/0/1	/
<b>P-value</b>	0.0004	0.0026	<0.0001	/	0.0001	<0.0001	<0.0001	/

A partir de ces résultats, nous observons aussi que la méthode de sélection MRMD génère des résultats différents selon la taille de l'ensemble sélectionné  $L^*$ . Pour notre méthode MRMD, les sous-ensembles de classifieurs sélectionnés de taille  $L^* = 25$  représentent les résultats les plus encourageants. Cependant, l'utilisation d'un nombre fixe des classifieurs à sélectionner ne permet pas toujours de trouver une solution optimale.

Pour cela, nous avons utilisé, à la méthode improuvée (MRMD-II), un algorithme de recherche séquentiel qui ne nécessite pas de déterminer à priori la taille des classifieurs à sélectionner.

### 4.4.3 Evaluation de la méthode MRMD-II

Plusieurs expérimentations ont été menées afin d'évaluer et déterminer l'efficacité de l'approche proposée MRMD-II. Nous les décrivons dans les sections suivantes les suivies des résultats obtenus.

#### 4.4.3.1 Etude Comparative

La méthode proposée MRMD-II a été comparée avec d'autres approches ensemblistes les plus utilisées dans la littérature, à savoir: Bagging (BGG), Boosting (BST) et Random Subspace (RSS) en utilisant trois différents apprenants de base: MLP, SVM et C4.5. L'objectif de cette comparaison est de vérifier et démontrer l'impact de la méthode de sélection proposée sur le processus de construction d'un système multi-classifieurs. La taille du pool initial des classifieurs est initialisée à 100 pour toutes les expérimentations.

Les tableaux 4.3, 4.4 et 4.5 montrent les résultats de précision obtenus par les méthodes MRMD-II, BGG, BST, RSS et CI (classifieur individuel) sur chaque base de données en utilisant différents classifieurs de base. Les tests des signes ont été aussi utilisés pour vérifier la signification de l'amélioration de notre méthode proposée. La moyenne de précision, le test win/tie/loss et leur p-value de chaque méthode sont également présentés en bas de chaque tableau.

Les tableaux 4.3 à 4.5 montrent que la performance de l'approche proposée (MRMD-II) est supérieure à celles des autres méthodes en terme de précision. Dans 17 cas sur 20, le MRMD-II utilisant MLP produit des meilleures performances que les autres ensembles de classifieurs. En ce qui concerne les résultats des tests win/tie/loss (p-value), la méthode MRMD-II est meilleure que les autres méthodes CI, BGG, BST et RSS avec des scores d'évaluation de 20/0/0 ( $<0.0001$ ), 18/0/2 (0.0004), 19/0/1 ( $<0.0001$ ) et 19/0/1 ( $<0.0001$ ) respectivement.

**Tableau 4.3.** Comparaison entre l'utilisation d'un classifieur unique et différents méthodes ensembliste utilisant le classifieur MLP.

Base de données	CI-MLP	BGG-MLP	BST-MLP	RSS-MLP	MRMD-II-MLP
<b>Breast-w</b>	96.71	96.57	95.57	96.71	<b>97.85</b>
<b>Credit-a</b>	81.45	84.06	81.16	83.48	<b>85.66</b>
<b>Credi-g</b>	71.50	74.90	75.80	75.10	<b>76.00</b>
<b>Colic</b>	83.15	<b>85.33</b>	83.15	83.15	83.96
<b>Contraction</b>	82.65	82.65	85.71	83.67	<b>88.79</b>
<b>Diabetes</b>	75.21	76.69	75.13	76.56	<b>77.34</b>
<b>Echocardiogram</b>	86.36	87.88	84.09	89.19	<b>90.15</b>
<b>Ecoli</b>	85.12	84.61	84.23	85.20	<b>87.47</b>
<b>Glass</b>	63.55	<b>67.76</b>	66.23	65.42	66.03
<b>Heart-statlog</b>	82.59	83.70	83.74	83.65	<b>84.07</b>
<b>Hepatitis</b>	75.48	79.35	82.54	81.29	<b>85.89</b>
<b>Ionosphere</b>	86.61	86.61	91.45	91.17	<b>92.00</b>
<b>Landsat</b>	88.86	91.07	91.39	89.76	<b>92.43</b>
<b>Laryngeal-2</b>	90.46	90.75	91.18	90.75	<b>96.67</b>
<b>Parkinson's</b>	76.41	81.03	79.49	81.54	<b>88.77</b>
<b>Segment</b>	90.94	93.35	93.35	93.12	<b>94.08</b>
<b>Vote</b>	95.36	95.09	95.63	94.94	<b>96.10</b>
<b>Waveform</b>	85.22	86.32	86.57	86.17	<b>87.63</b>
<b>Yeast</b>	79.33	80.71	80.44	<b>82.53</b>	81.09
<b>Zoo</b>	89.11	89.11	90.10	91.09	<b>96.00</b>
<b>Moyenne</b>	83.30	84.88	84.85	85.22	<b>87.40</b>
<b>Win /Tie/ loss</b>	20/0/0	18/0/2	19/0/1	19/0/1	/
<b>p-value</b>	< 0.0001	0.0004	< 0.0001	< 0.0001	/

En utilisant des ensembles de SVM, le MRMD-II produit des taux de reconnaissance élevés sur 15 bases de données. Les scores win/tie/loss (p-value) de l'approche MRMD-II utilisant SVM, montrent un avantage significatif par rapport aux autres méthodes CI, BGG, BST et RSS avec des résultats de 20/0/0 (<0,0001), 18/0/2 (0,0004), 19/0/1 (<0,0001) et 16/1/3 (0,0044) respectivement.

Pour des ensembles de C4.5, MRMD-II améliore la performance de 16 jeux de données. Les scores win/tie/loss (p-value) par rapport aux méthodes CI, BGG, BST et RSS sont respectivement de 20/0/0 (<0.0001), 17/1/2 (0.0007), 17/1/2 (0.0007) et 17/0/3

(0,0026). En outre, les performances moyennes de MRMD-II utilisant différents classifieurs de base sont plus élevées que celles obtenues par les autres approches de classification, avec 87,40% en utilisant MLP, 87,63% avec SVM et 87,17% avec C4.5.

**Tableau 4.4.** Comparaison entre l'utilisation d'un classifieur unique et différents méthodes ensembliste utilisant le classifieur SVM.

Base de données	CI-SVM	BGG-SVM	BST-SVM	RSS-SVM	MRMD-II-SVM
<b>Breast-w</b>	95.28	96.14	95.14	96.28	<b>97.13</b>
<b>Credit-a</b>	82.48	82.75	79.86	<b>84.35</b>	83.38
<b>Credi-g</b>	71.90	75.70	74.90	75.81	<b>76.30</b>
<b>Colic</b>	84.24	85.87	80.71	<b>86.41</b>	<b>86.41</b>
<b>Contraction</b>	89.80	87.76	84.69	87.76	<b>90.87</b>
<b>Diabetes</b>	74.61	<b>77.08</b>	74.83	76.02	75.13
<b>Echocardiogram</b>	87.88	89.39	85.61	87.12	<b>90.15</b>
<b>Ecoli</b>	81.25	85.12	85.42	84.41	<b>86.60</b>
<b>Glass</b>	69.63	75.70	74.47	<b>76.64</b>	76.20
<b>Heart-statlog</b>	80.37	82.19	81.26	82.46	<b>82.59</b>
<b>Hepatitis</b>	81.91	81.94	81.94	82.58	<b>83.97</b>
<b>Ionosphere</b>	90.88	92.31	90.88	92.02	<b>93.57</b>
<b>Landsat</b>	89.52	91.09	91.24	90.48	<b>92.48</b>
<b>Laryngeal-2</b>	90.03	90.75	90.46	90.90	<b>96.82</b>
<b>Parkinson's</b>	68.72	74.36	74.87	75.36	<b>86.73</b>
<b>Segment</b>	90.14	<b>92.74</b>	91.22	91.79	92.66
<b>Vote</b>	95.86	96.10	95.48	95.63	<b>96.78</b>
<b>Waveform</b>	85.22	86.55	87.32	86.41	<b>87.45</b>
<b>Yeast</b>	78.29	79.48	<b>82.68</b>	81.62	82.47
<b>Zoo</b>	89.11	91.09	91.19	93.07	<b>95.00</b>
<b>Moyenne</b>	83.86	85.71	84.71	85.86	<b>87.63</b>
<b>Win /Tie/ loss</b>	20/0/0	18/0/2	19/0/1	16/1/3	/
<b>p-value</b>	< 0.0001	0.0004	< 0.0001	0.0044	/

**Tableau 4.5.** Comparaison entre l'utilisation d'un classifieur unique et différents méthodes ensembliste utilisant le classifieur C4.5.

Base de données	CI-C4.5	BGG-C4.5	BST-C4.5	RSS-C4.5	MRMD-II-C4.5
Breast-w	94.13	95.57	95.14	95.14	<b>95.57</b>
Credit-a	80.33	85.07	82.31	<b>84.20</b>	83.08
Credi-g	70.60	75.60	74.20	74.50	<b>76.00</b>
Colic	85.27	85.32	83.42	85.33	<b>85.60</b>
Contraction	85.71	88.78	87.67	86.73	<b>89.75</b>
Diabetes	73.18	<b>77.21</b>	75.52	73.96	74.86
Echocardiogram	84.09	87.88	84.85	86.36	<b>90.90</b>
Ecoli	82.44	85.42	83.23	83.63	<b>86.01</b>
Glass	68.69	74.30	74.78	74.77	<b>75.36</b>
Heart-statlog	80.37	81.48	81.63	83.58	<b>84.81</b>
Hepatitis	78.06	78.06	78.71	80.65	<b>86.53</b>
Ionosphere	85.47	90.03	92.02	90.88	<b>94.29</b>
Landsat	88.15	90.42	<b>91.08</b>	90.47	<b>91.08</b>
Laryngeal-2	90.61	90.75	90.61	91.04	<b>97.83</b>
Parkinson's	73.33	74.87	77.44	75.38	<b>83.67</b>
Segment	88.47	91.16	91.38	<b>92.05</b>	91.89
Vote	96.48	96.09	95.17	95.63	<b>96.78</b>
Waveform	83.62	85.35	84.12	84.97	<b>85.74</b>
Yeast	79.07	79.54	80.89	<b>80.23</b>	80.15
Zoo	92.01	90.10	92.06	90.10	<b>93.42</b>
<b>Moyenne</b>	83.00	85.15	84.81	84.98	<b>87.17</b>
<b>Win /Tie/ loss</b>	20/0/0	17/1/2	17/1/2	17/0/3	/
<b>p-value</b>	< 0.0001	0.0007	0.0007	0.0026	/

Par conséquent, puisque chaque ensemble de classifieurs homogènes améliore les performances de la méthode MRMD-II sur différents jeux de données, nous avons proposé d'utiliser dans les expérimentations suivantes un ensemble hétérogène de classifieurs composé des MLP, SVM et C.4 ainsi que l'algorithme Random Subspace dans le but de renforcer la diversité préliminaire.

Afin de vérifier l'impact du compromis diversité-performance proposé sur le processus de sélection, les performances de la méthode MRMD-II ont été comparées avec celles de différents systèmes de sélection et de combinaison de classifieurs, à savoir:

- RSS\_HET: cet ensemble génère un ensemble de classifieurs en utilisant la méthode RSS basée sur des classifieurs hétérogènes ;
- SB (Single Best): cet algorithme de sélection choisit le meilleur classifieur de l'ensemble en fonction de sa précision ;
- DIV (Diversité): ce schéma de sélection définit la compétence d'un sous-ensemble de classifieurs en fonction des mesures de diversité et sélectionne les membres les plus diversifiés;
- PER (Pertinence): ce système sélectionne le sous-ensemble de classifieurs les plus pertinents selon la mesure de relevance (pertinence) du critère MRMD-II;
- MRMD: Maximum Relevance Maximum Diversity est le schéma de sélection présenté précédemment, avec une taille  $L^* = 25$ .

Le tableau 4.6 illustre les précisions de différents schémas de sélection sur chaque ensemble de données, ainsi que les tests des signes win/tie/loss (p-value). Les résultats de ce tableau montrent que, contrairement à la méthode SB, tous les schémas de sélection de classifieurs sont capables de surpasser l'ensemble RSS hétérogène (RSS-HET) sur tous les jeux de données.

Ces résultats prouvent aussi l'efficacité de notre approche MRMD-II dans la plupart des cas. Les résultats obtenus par la méthode MRMD-II sont plus encourageants que l'approche MRMD sur 16 bases de données. Néanmoins, pour les bases de données Credit-g et Glass, le MRMD-II est légèrement moins performant que l'approche MRMD. Ces résultats peuvent être expliqués par le fait que la performance globale, dans ces cas, ne peut pas présenter la compétence de l'ensemble plus correctement que la performance moyenne de ses membres. Nous constatons également que l'utilisation de la performance globale de l'ensemble avec un algorithme de recherche séquentiel est généralement plus efficace que l'utilisation des performances individuelles de ces membres.

En outre, le tableau 4.6 montre que la combinaison des mesures de diversité et de précision pour le processus de sélection est meilleure que l'utilisation de chacun d'eux séparément. Les performances de MRMD-II sont meilleures que celles produites par les approches DIV et PER dans la plupart des cas. En ce qui concerne les tests win/tie/loss (p-value), les résultats de la méthode MRMD-II sont meilleures que celle de RSS-HET, SB,

DIV, PER et MRMD avec des scores d'évaluation de 20/0/0 (<0.0001), 20/0/0 (<0,0001), 19/1/0 (<0,0001), 19/1/1 (<0,0001) et 16/2/2 (0,0013) respectivement.

**Tableau 4.6.** Comparaison entre la méthode RSS et différentes méthodes de sélection utilisant un ensemble de classifieurs hétérogènes.

Base de données	RSS-HET	SB	DIV	PER	MRMD	MRMD-II
<b>Breast-w</b>	97.14	97.70	97.12	97.27	97.52	<b>98.00</b>
<b>Credit-a</b>	83.20	85.36	85.97	87.69	<b>87.97</b>	<b>87.97</b>
<b>Credi-g</b>	74.69	73.30	75.60	74.90	<b>77.23</b>	76.40
<b>Colic</b>	82.88	84.23	84.78	86.14	85.23	<b>86.41</b>
<b>Contraction</b>	90.75	90.87	90.87	86.70	91.45	<b>92.87</b>
<b>Diabetes</b>	73.95	72.52	76.17	76.43	75.12	<b>78.95</b>
<b>Echocardiogram</b>	85.55	78.03	85.60	84.84	86.85	<b>87.88</b>
<b>Ecoli</b>	85.11	86.01	86.01	85.14	86.14	<b>87.90</b>
<b>Glass</b>	76.86	76.67	77.67	76.86	<b>78.94</b>	78.56
<b>Heart-statlog</b>	77.35	78.83	79.85	80.23	82.64	<b>85.58</b>
<b>Hepatitis</b>	83.16	87.82	89.74	84.61	89.37	<b>90.38</b>
<b>Ionosphere</b>	92.61	93.58	93.97	93.55	93.55	<b>94.59</b>
<b>Landsat</b>	90.36	91.64	91.56	92.74	93.68	<b>93.75</b>
<b>Laryngeal-2</b>	90.60	95.67	90.17	<b>96.67</b>	<b>96.67</b>	<b>96.67</b>
<b>Parkinson's</b>	78.87	85.71	82.65	88.26	90.33	<b>91.32</b>
<b>Segment</b>	90.15	92.69	93.73	92.38	94.71	<b>94.82</b>
<b>Vote</b>	95.62	95.64	95.18	96.10	96.57	<b>96.78</b>
<b>Waveform</b>	85.31	87.36	87.39	85.47	87.68	<b>88.00</b>
<b>Yeast</b>	80.02	81.22	81.34	80.47	81.82	<b>82.00</b>
<b>Zoo</b>	90.03	95.07	<b>96.00</b>	94.00	94.18	<b>96.00</b>
<b>Moyenne</b>	85.21	86.50	87.07	87.02	88.38	<b>89.24</b>
<b>Win/tie/loss</b>	20/0/0	20/0/0	19/1/0	19/1/0	16/2/2	/
<b>p-value</b>	<0.0001	<0.0001	<0.0001	<0.0001	0.0013	/

Les tests de signes win/tie/loss et le p-value sont également utilisés pour analyser les résultats expérimentaux de différentes versions de schéma de sélection proposé. Le tableau 4.7 résume les scores obtenus entre les approches de sélection MRMD-II-MLP, MRMD-II-SVM, MRMD-II-C4.5 et MRMD-II-HET.

**Tableau 4.7.** Comparaison des résultats de précision de différentes versions de schéma de sélection MRMD-II.

		MRMD-II- HET	MRMD-II- MLP	MRMD-II- SVM	MRMD-II- C4.5
<b>MRMD-II- HET</b>	win/tie/loss	/	17/2/1	15/2/3	17/2/1
	p-value	/	0.0001	0.0075	0.0001
<b>MRMD-II- MLP</b>	win/tie/loss	/	/	10/1/9	10/1/9
	p-value	/	/	1.0000	1.0000
<b>MRMD-II- SVM</b>	win/tie/loss	/	/	/	14/1/5
	p-value	/	/	/	0.0636

Le tableau 4.7 montre que le schéma de sélection proposé utilisant un ensemble de classifieurs hétérogènes produit des meilleurs résultats dans la plupart des cas. L'ensemble MRMD-II-HET obtient des résultats win/tie/loss statistiquement significatifs par rapport aux méthodes MRMD-II-MLP, MRMD-II-SVM et MRMD-II-J48.

#### 4.4.3.2 Problème des données déséquilibrées

Quatre jeux de données, à savoir : Diabetes, Ecoli, Glass et Ionosphere ont été utilisés pour analyser le comportement de l'approche proposée MRMD-II sur des bases de données déséquilibrées. La précision ou le taux d'erreur n'est pas une mesure utile pour les données déséquilibrées. Par exemple, supposons que nous avons un ratio de 1:100 pour la classe minoritaire et la classe majoritaire, un classifieur qui assigne toutes les instances à la classe majoritaire aura une précision de 99%. En outre, d'autres paramètres d'évaluation ont été pris en compte, à savoir : la sensibilité et la spécificité, pour évaluer la performance de notre approche sur ces bases de données.

Afin d'appliquer ces métriques sur un problème de classification multi-classes, il est nécessaire de le décomposer en un ensemble de sous problèmes binaire. Dans cette expérimentation, nous avons utilisé la méthode de décomposition *un contre tous* (*one versus All*) consistant à opposer une classe particulière au reste des autres classes. Cette

stratégie est appliquée sur les jeux de données Ecoli et Glass. Les jeux de données Diabetes et Ionosphere sont déséquilibrés de manière inhérente. Les détails de ces jeux de données sont résumés au tableau 4.8. Le tableau 4.9 résume les taux de sensibilité (SEN), spécificité et (SPE) et obtenus pour chaque base de données par l'approche MRMD-II en utilisant différents pool de classifieurs.

**Tableau 4.8.** Description des bases de données déséquilibrées

Base de données	#minoritaire	#majoritaire	Ratio de déséquilibre
<b>Diabetes</b>	268	500	1.86
<b>Ecoli-1</b>	77	259	3.36
<b>Ecoli-3</b>	35	301	8.60
<b>Glass-0</b>	70	144	2.06
<b>Glass-4</b>	13	201	15.46
<b>Ionosphere</b>	126	225	2.02

**Tableau 4.9.** Résultats de Sensibilité (SEN) et de Spécificité (SPE) obtenus sur des jeux de données déséquilibrées en utilisant différents ensembles de classifieurs.

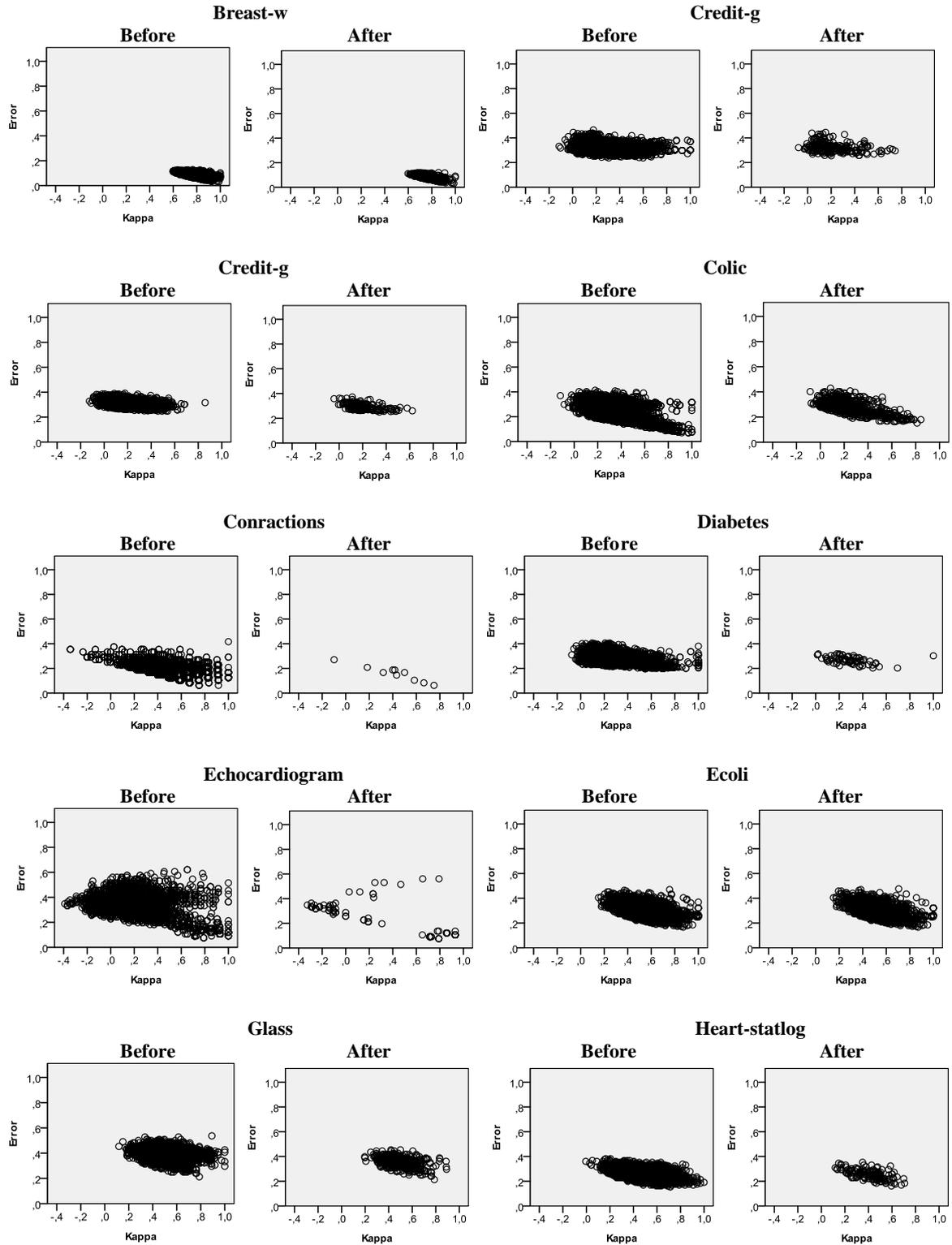
Base de données		MRMD	MRMD-II- HET	MRMD-II- MLP	MRMD-II- SVM	MRMD-II- C4.5
<b>Diabetes</b>	SEN	58.58	<b>61.19</b>	58.20	60.44	59.70
	SPE	84.00	<b>88.00</b>	87.64	83.23	83.14
<b>Ecoli-1</b>	SEN	74.02	<b>77.92</b>	76.62	74.06	75.32
	SPE	89.98	<b>90.78</b>	90.75	90.35	89.21
<b>Ecoli-3</b>	SEN	71.43	<b>74.34</b>	71.43	71.37	68.51
	SPE	88.03	<b>89.38</b>	<b>89.38</b>	88.37	88.09
<b>Glass-0</b>	SEN	<b>75.71</b>	72.86	57.19	64.31	61.42
	SPE	80.55	81.25	70.13	<b>82.63</b>	81.94
<b>Glass-4</b>	SEN	53.86	<b>61.54</b>	38.45	46.18	46.15
	SPE	<b>80.59</b>	79.60	67.66	78.61	77.11
<b>Ionosphere</b>	SEN	88.88	<b>90.47</b>	84.92	88.09	89.68
	SPE	96.00	<b>96.88</b>	96.00	<b>96.88</b>	96.82
<b>Moyenne</b>	SEN	71.40	<b>73.94</b>	66.25	70.31	67.06
	SPE	85.40	<b>86.26</b>	83.08	85.60	84.75

Selon les résultats du tableau 4.9, la méthode MRMD-II-HET surpasse les autres méthodes dans la plupart des cas. Selon la métrique de sensibilité, la méthode MRMD-II-HET obtient les meilleurs résultats avec une moyenne de 73.94%. Nous observons aussi de ces résultats que la méthode MRMD-II-HET atteint la sensibilité la plus élevée sur la plus part des jeux de données avec une moyenne de 86.26%. Malgré les résultats remarquable, la méthode MRMD-II n'était pas capable d'améliorer la sensibilité des données déséquilibrées.

### ***4.4.3.3 Analyse du compromis diversité-performance utilisant diagramme Kappa-Erreur***

Afin d'étudier et analyser la relation entre la diversité et la précision des membres de l'ensemble sélectionné, nous avons utilisé le diagramme kappa-erreur. Ce diagramme [Margineantu 1997] est représenté par un nuage de points, où chaque point présente une paire de classifieurs  $(c_i, c_j)$ . La coordonnée  $x$  indique la valeur de la mesure de diversité Kappa ( $k$ ) qui estime le degré d'accord entre deux classifieurs, et la coordonnée  $y$  indique l'erreur moyenne de ces deux classifieurs. Ainsi, un ensemble comportant  $L$  classifieurs est représenté par un nuage de  $\frac{L(L-1)}{2}$  points correspondants à toutes les paires qui le constituent. Les diagrammes kappa-erreur avant et après le processus de sélection MRMD-II pour les 20 bases de données sont illustrées dans les figures 4.4 et 4.5.

Ces figures montrent que le nuage des points sélectionnés est situé à la partie gauche, la partie inférieure ou à la partie inférieure gauche du nuage initial (avant la sélection). Cela vérifie l'efficacité de la mesure de sélection proposée pour contrôler l'équilibre entre la diversité et la performance en choisissant les classifieurs précis, divers et les membres précis-divers en même temps. Dans les diagrammes des bases de données Echocardiogram, Landsat et Diabetes, peu de points ont été situés sur la partie haute droite, ce qui indique une précision et une diversité plus faibles de certaines paires de classifieurs. Cependant, la méthode d'agrégation utilisée surmonte cette lacune en donnant des poids faibles à ces membres. Les résultats précédents établissent l'efficacité de la méthode proposée.



**Figure 4.4.** Diagrammes Kappa-Erreur des bases : Breast-w, Credit-a, Credit-g, Colic, Contractions, Diabetes, Echocardiogram, Ecoli, Glass et Heart-statlog.

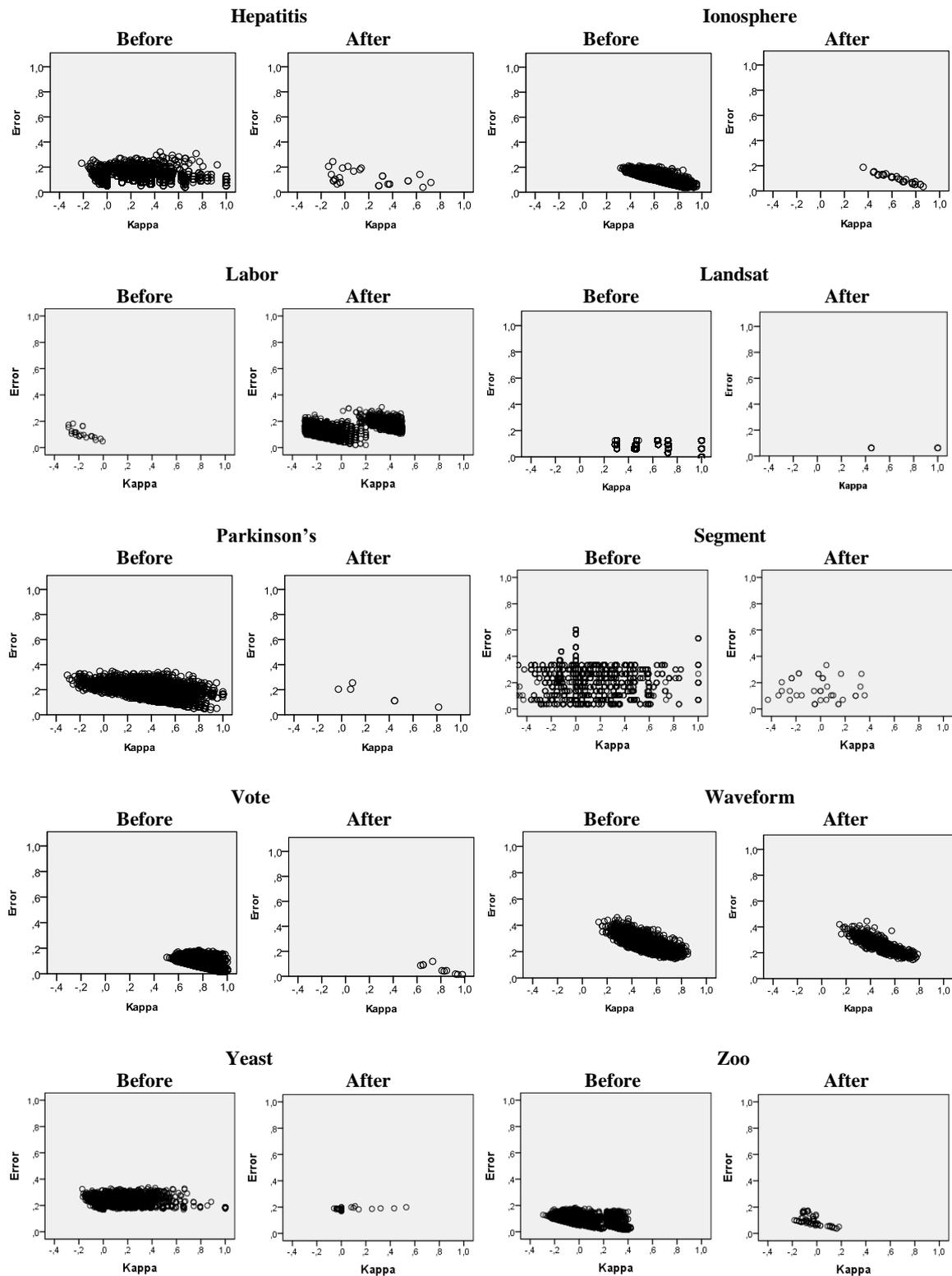


Figure 4.5. Diagrammes Kappa-Erreur des bases: Glass, Heart-statlog, Hepatitis, Ionosphere, Landsat, Parkinson's, Vote, Waveform, Yeast et Zoo.

## 4.5 Conclusion

Dans ce chapitre, nous avons proposé une nouvelle méthode de sélection de classifieurs (MRMD) qui utilise un algorithme de recherche par ordonnancement basé sur un nouveau critère de sélection. Ce critère intègre le critère de sélection de caractéristiques mRMR et les mesures de diversité dans le but de contrôler simultanément la diversité et les performances individuelles des classifieurs.

Les résultats obtenus sur 20 bases de données de références issues de UCI et LKC ont montré l'efficacité de l'approche de sélection MRMD et son impact sur la construction d'un système multi-classifieurs robuste. De plus, ces résultats nous ont permis de distinguer les avantages de combinaison de la relevance et des mesures de diversité comme critère de sélection de classifieurs. Malgré les résultats encourageants obtenus par cette proposition, nous avons rencontré une difficulté pratique concernant la taille du sous-ensemble final de classifieurs, nécessaire pour résoudre adéquatement le problème posé. Le nombre des classifieurs à sélectionner de l'approche MRMD doit être déterminé a priori.

Pour surmonter ce problème, nous avons proposé une autre méthode de sélection (MRMD-II) basée sur la recherche gloutonne permettant de contrôler le nombre de classifieurs dans le sous-ensemble solution. La méthode MRMD-II permet de sélectionner les classifieurs ayant la plus grande relevance tout en conservant un niveau de diversité élevé entre ces membres. La relevance a été utilisée au critère MRMD-II pour déterminer la performance globale d'un sous ensemble candidat ainsi que les performances individuels de ses membres.

Plusieurs expérimentations ont été menées pour évaluer la performance de l'approche MRMD-II. L'étude comparative faite par rapport à d'autres paradigmes de sélection a prouvé la supériorité de l'approche proposée. De plus, Le MRMD-II utilisant un ensemble hétérogène de classifieurs a été capable de surpasser plusieurs méthodes de sélection sur la plupart des jeux de données. Les résultats des diagrammes d'erreur kappa ont montré aussi l'efficacité du critère proposé (Maximum Relevance Maximum Diversity-II) pour contrôler l'équilibre entre la diversité des classifieurs et leurs performances.

Finalement, nous avons observé que ces approches de sélection ont choisi un seul ensemble de classifieurs pour classer les échantillons en phase de test ou en phase de généralisation. Cependant, comme mentionné dans le chapitre 2, il n'y a aucune garantie

que le sous-ensemble sélectionné est la solution la plus susceptible pour classer chaque échantillon de test individuellement. De plus, notre approche MRMD-II n'était pas capable d'améliorer la sensibilité des données déséquilibrées. Dans le chapitre suivant de cette thèse, nous proposons une approche de sélection dynamique de classifieurs pour les bases de données déséquilibrées afin de déterminer si le gain de performance peut être augmenté ou non.

# Chapitre 5

## Sélection Dynamique de Classifieurs Combinant la Diversité et la Méthode mRMR

### 5.1 Introduction

Dans les chapitres 3 et 4, nous avons présenté deux stratégies de sélection statique de classifieurs. Nous avons étudié différents algorithmes de recherche basés sur différents critères pour guider le processus de construction et sélection de classifieurs. Nous avons également présenté un nouveau compromis «diversité-performance» hybridant la technique mRMR avec les mesures de diversité. L'objectif était de trouver le sous-ensemble de classifieurs les plus précis et les plus divers simultanément au cours de la phase de sélection, et de l'utiliser pour prédire la classe de tous les échantillons de la base de données de test. Cependant, il n'y a aucune garantie que le sous-ensemble sélectionné soit optimal en phase de généralisation, car chaque classifieur est un expert dans une région locale différente de l'espace de caractéristiques.

Afin de pallier cette limite, dans ce chapitre, nous proposons une extension dynamique de la méthode de sélection statique MRMD-II (*Maximum Relevance Maximum Diversity*), présentée au chapitre précédent. Dans les systèmes de sélection dynamique, les membres sélectionnés sont définis au cours du processus de classification en tenant compte des caractéristiques de l'échantillon à traiter. Seuls les classifieurs les plus compétents dans la région de compétence sont choisis dynamiquement pour classer chaque échantillon de test ou inconnu.

Outre la sélection dynamique de classifieurs, ce chapitre traite le problème de données déséquilibrées qui a attiré une attention considérable de la communauté de fouille de données. Le problème des données déséquilibrées se produit lorsque le nombre d'instances d'une classe est nettement plus petit que celui des autres classes. En particulier, dans un problème de classification à deux classes, la classe contenant moins d'instances est appelée classe minoritaire, tandis que l'autre est appelée classe majoritaire.

Généralement, dans ce type de cas, les algorithmes de classification ont des difficultés au cours de la phase d'apprentissage liées aux déséquilibres des classes. Les systèmes d'apprentissage auront toujours tendance à négliger la classe minoritaire, qui est dans la majorité des cas la plus intéressante, à cause de cette disproportion. Pour cette raison, diverses stratégies d'échantillonnage ont été utilisées pour modifier la distribution des données d'apprentissage et donc diminuer la fréquence de la classe majoritaire, à savoir : le sous-échantillonnage aléatoire, l'échantillonnage ciblé, le sous-échantillonnage centré, le sur-échantillonnage basé sur la génération synthétique, et plusieurs combinaisons de ces techniques [Galar 2013].

Une situation déséquilibrée se produit souvent dans des applications réelles telles que la classification de documents, la recherche d'images par le contenu [Akbari 2004], l'analyse des risques financiers [Thammasiri 2014], et plus particulièrement dans la classification des bases de données médicales [Chawla 2002]. Lorsqu'on aborde des problèmes de données déséquilibrés, les gens ont tendance à se préoccuper davantage de la classe minoritaire, car le coût de la classification erroné des échantillons minoritaires est beaucoup plus élevé que celui de la classe majoritaire [Galar 2013 ; Maldonado 2014 ; Menardi 2014]. Dans le domaine de la santé et tout particulièrement dans le diagnostic du cancer, annoncer à un patient qu'il n'a pas de cancer alors qu'il en porte un, pourrait avoir des conséquences désastreuses sur sa santé et pourrait entraîner son décès [Ritschard 2007]. Puisque ce problème est présent dans de nombreuses situations médicales réelles, nous sommes particulièrement intéressés par l'analyse des performances de notre approche sur des bases de données médicales.

Par conséquent, dans ce chapitre, nous présentons un nouveau système multi-classifieurs pour la classification des données médicales déséquilibrées. L'approche proposée est basée sur la combinaison des techniques d'échantillonnage et une nouvelle méthode de sélection dynamique des ensembles de classifieurs. Avant l'étape de génération

des classifieurs initiaux, une approche d'échantillonnage est considérée pour rééquilibrer l'ensemble de données d'apprentissage. Ensuite, la méthode Random Subspace est employée en se basant sur ces données pour créer un pool initial de classifieurs à l'aide d'un ensemble hétérogène de modèles de base. Le choix d'utilisation d'une approche de prétraitement est fait en fonction de leur nature aléatoire qui pourra améliorer la diversité du pool de classifieurs [Kuncheva 2003a]. D'autre part, puisque la sélection dynamique exploite la compétence locale des classifieurs de base, elle peut surmonter les difficultés en raison de leur qualité variable [Krawczyk 2016 ; Roy 2018]. Ces avantages nous ont donné des raisons de croire qu'un système multi-classifieurs qui utilise l'échantillonnage à la phase de génération et applique une sélection dynamique sur les classifieurs générés produira des résultats meilleurs par rapport à une méthode d'ensemble statique.

Dans ce chapitre, nous présentons dans un premier temps les principales stratégies proposées pour traiter les problèmes de l'asymétrie des classes. Ensuite, nous détaillons notre contribution qui consiste à mettre en place une approche de sélection dynamique de classifieurs basée sur un nouveau compromis «diversité-performance» proposé pour les données équilibrées et déséquilibrées ainsi que quelques expérimentations.

## 5.2 Apprentissage sensible au déséquilibre des classes

Le déséquilibre des classes [He 2008] se réfère aux problèmes de classification dans lesquels le nombre d'observations appartenant à une classe, dépasse fortement celui des observations dans les autres classes. En particulier, dans un scénario à deux classes, les données d'apprentissage de la classe majoritaire sont largement supérieures en nombre à celles de la classe minoritaire. Lorsqu'une base de données est déséquilibrée, les classifieurs classiques privilégient généralement la classe majoritaire, et donc échouent de classer correctement les instances de la classe minoritaire, ce qui dégrade la performance de classification [Prati 2015]. Diverses techniques ont été développées dans la littérature pour résoudre ce problème [Galar 2012]. Selon les auteurs de [Galar 2013], ces approches peuvent être réparties en trois catégories :

### 5.2.1 Stratégies algorithmiques

Ces approches visent à réduire le biais envers la classe majoritaire en adaptant les méthodes d'apprentissage existantes pour traiter les données déséquilibrées. Certains chercheurs ont proposé d'introduire un biais dans les algorithmes d'apprentissage comme

les algorithmes kPPV et SVM [Verpoulos 1991 ; Batista 2004]. D'autres travaux ont proposé de modifier le seuil de décision en fournissant une probabilité pour chaque individu d'appartenir à une classe. La décision est donc prise en fixant un seuil sur cette probabilité. C'est le cas du bayésien naif ou de certains réseaux de neurones [Hamdi 2012].

### 5.2.2 Stratégies d'échantillonnage

Ce type de stratégies inclut des algorithmes de prétraitement qui rééquilibrent la distribution des classes en ré-échantillonnant l'espace de données, et donc évitant des modifications de l'algorithme d'apprentissage [García 2012]. Pour cette raison, ce sont les approches les plus couramment utilisées pour manipuler les données déséquilibrées [Douzas 2017 ; Mathew 2018]. Ces approches permettent de redresser les échantillons d'apprentissage de façon à diminuer la fréquence de la classe majoritaire ou augmenter celle de la classe minoritaire. Branco et al. [Branco 2016] ont divisé ces méthodes en trois catégories : le sous-échantillonnage et le sur-échantillonnage, (qui se basent sur la notion de voisinage) et la combinaison des deux techniques précédentes. Le sous-échantillonnage consiste à rééquilibrer les jeux de données en supprimant de la base d'apprentissage des échantillons appartenant à la classe majoritaire. À l'opposé, le sur-échantillonnage permet d'équilibrer les données en constituant des échantillons de la classe minoritaire de manière à encourager les algorithmes d'apprentissage à converger vers un type de modèle spécifique. Dans ce chapitre, nous nous intéressons aux techniques d'échantillonnage de bases de données déséquilibrées et plus précisément aux techniques de sur-échantillonnage.

Cependant, les stratégies d'échantillonnage ont montré différents inconvénients. Le sur-échantillonnage peut augmenter artificiellement le risque de sur-apprentissage, tandis que le sous-échantillonnage présente le risque de perdre des informations potentiellement utiles. Pour pallier ces inconvénients, plusieurs techniques avancées de ré-échantillonnage ont été proposées. Le sous-échantillonnage aléatoire (*Random Under-Sampling*, RUS) [García 2012] est basé sur la suppression aléatoire des exemples de la classe majoritaire pour équilibrer les données. Wilson [1972] a proposé d'utiliser la règle des plus proches voisins pour supprimer des individus de la classe majoritaire "*Neighborhood Cleaning Rule*". Chawla et al. [Chawla 2002] ont proposé une technique de sur-échantillonnage appelée SMOTE (*Synthetic Minority Oversampling Technique*) qui permet de générer des individus artificiels dans la classe minoritaire. Basé sur le principe de l'algorithme

SMOTE, plusieurs méthodes ont été proposées dans la littérature, telles que *Borderline-SMOTE* [Han 2005], *Random Balance* (RB) [Dos Santos 2011] et *Random Minority Oversampling Technique* (RAMO) [Chen 2010].

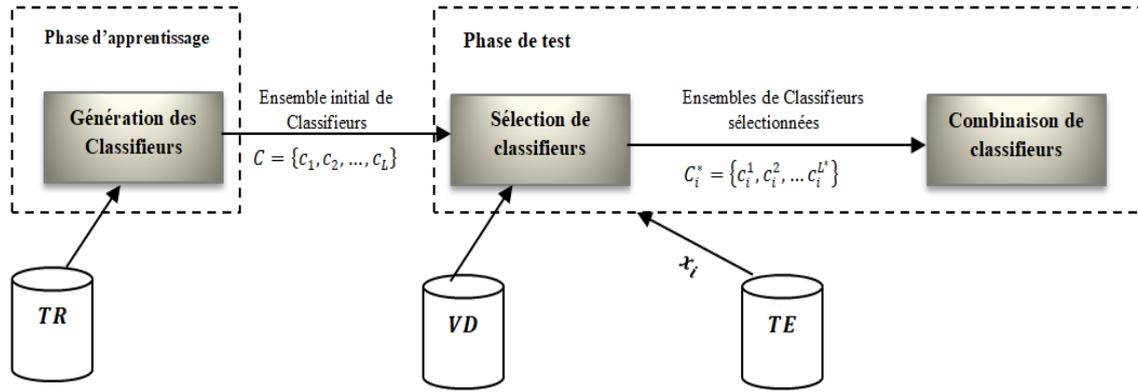
### 5.2.3 Stratégies ensemblistes

Plus récemment, les approches d'ensemble de classifieurs sont devenues une solution populaire pour résoudre le problème de l'asymétrie des classes. Généralement, ces approches ont été combinées avec des méthodes d'échantillonnage. Tout d'abord, une technique de ré-échantillonnage est appliquée pour équilibrer la distribution des classes, puis un ensemble de modèles d'apprentissage différents est formé en se basant sur l'ensemble de données équilibrées.

La méthode ensembliste la plus utilisée est l'algorithme Boosting [Schapire 1998], qui a été appliqué à de nombreux algorithmes d'ensemble connus tels que SMOTEBoost [Chawla 2003], RUSBoost [Seiffert 2010], EasyEnsemble [Liu 2009] et EUSboost [Díez-Pastor 2015]. Dans [Nanni 2015], les auteurs ont indiqué qu'un processus de création d'un système multi-classifieurs typique contient généralement 3 procédés, qui sont, le ré-échantillonnage, la génération des classifieurs de base, et la règle de fusion. De plus, l'algorithme Bagging a été aussi appliqué dans de nombreux algorithmes d'ensemble tels que SMOTE-Bagging [Wang 2009], UnderBagging [Galar 2012].

## 5.3 Sélection dynamique de classifieurs basée sur la méthode mRMR et les mesures de diversité

Le système multi-classifieurs proposé dans cette thèse comprend trois phases principales: la génération de classifieurs pour créer un ensemble initial de classifieurs, la sélection dynamique pour choisir le sous-ensemble optimal de chaque exemple de test  $x_i$  et l'étape de combinaison qui consiste à fusionner les décisions des classifieurs sélectionnés pour prédire la classe de  $x_i$ . La phase de génération des classifieurs initiaux est exécutée pendant la phase d'apprentissage, tandis que les étapes de sélection et d'agrégation sont effectuées pendant la phase de test. La figure 5.1 présente le schéma général du système proposé. Dans les sections suivantes, chaque étape du système proposé sera largement discutée.



**Figure 5.1.** Un organigramme du système multi-classifieurs proposé.

### 5.3.1 Génération d'ensemble de classifieurs

Avant la création d'un pool de départ de classifieurs, une approche de sur-échantillonnage est appliquée si la base de données d'apprentissage (**TR**) est déséquilibrée. Les techniques de sur-échantillonnage permettent de réduire le degré de distribution déséquilibrée des données d'apprentissage en augmentant le nombre d'individus de la classe minoritaire. Plus précisément, la technique de sur-échantillonnage de la minorité synthétique (SMOTE) [Chawla 2002] est utilisée dans notre approche. Ce choix est justifié en fonction de leur efficacité prouvée dans différents problèmes de données déséquilibrées binaires [Wang 2017 ; Fernández 2018 ; Roy 2018]. De plus, en raison de leur nature aléatoire, le sur-échantillonnage peut améliorer la diversité du pool de classifieurs [Kuncheva 2007]. Après l'échantillonnage de la base d'apprentissage, la méthode Random Subspace est utilisée pour créer un pool initial de classifieurs à l'aide d'un ensemble hétérogène de modèles de base.

#### 5.3.1.1 Approche de sur-échantillonnage SMOTE

L'approche SMOTE [Chawla 2002] est une technique de sur-échantillonnage permettant la création des individus artificiels dans la classe minoritaire. Le processus SMOTE fonctionne comme suit:

- soit  $x_i$  une instance de la classe minoritaire ; pour créer une instance artificielle à partir de  $x_i$  ; SMOTE isole d'abord les  $k$ -plus proches voisins de  $x_i$ , de la classe minoritaire.

- Ensuite, il sélectionne aléatoirement certain nombre de voisins et génère aléatoirement des individus synthétiques le long de la ligne entre l'individu de la classe minoritaire  $x_i$  et ses voisins sélectionnés.

Une fois la réalisation de la phase de sur-échantillonnage est achevée, nous procédons à la génération du pool initial des classifieurs par l'approche Random Subspace.

### 5.3.2 Sélection dynamique de classifieurs

Le modèle proposé de sélection est une extension de la méthode statique MRMD-II proposée au chapitre 4. La méthode proposée, appelée: DES-MRMD-III est divisée en trois étapes:

- **Définition de la région de compétence** : à cette étape, nous utilisons l'algorithme k-plus proche voisins (kPPV) pour définir la région locale  $\theta_i$  entourant l'exemple de test  $x_i$ .
- **Détermination du critère de sélection MRMD-III**: le critère de sélection proposé est défini à cette étape pour estimer le niveau de compétence des classifieurs de base en fonction de la région de compétence  $\theta_i$  en utilisant une version modifiée (MRMD-III) du critère MRMD-II proposé au chapitre précédent.
- **Schéma de sélection dynamique de classifieurs**: notre méthode exploite les avantages de l'approche de recherche séquentielle en avant (*forward*) pour limiter l'espace de recherche et fournir un sous-ensemble optimal de classifieurs  $C_i^*$  ( $C_i^* \subset C$ ). La combinaison des membres de  $C_i^*$  est ensuite utilisé pour prédire la classe de l'échantillon de test  $x_i$ ,

#### 5.3.2.1 Définition de la région de compétence

Dans cette phase, la région de compétence  $\theta_i$  d'un exemple de test  $x_i$  est pré-estimée lors de la phase de test en utilisant les k-plus proches voisins [Cruz 2017] de  $x_i$  dans la base de données de la sélection dynamique *DSEL*. La même approche de sur-échantillonnage (SMOTE) est appliquée sur la base *DSEL* afin de rééquilibrer les données de  $\theta_i$ . Le principe de l'algorithme des k-plus proches voisins (kPPV) est très simple et intuitif. Le kPPV est basé sur une comparaison directe en calculant les distances entre le vecteur de caractéristiques représentant l'exemple à classer  $x_i$  et les vecteurs

caractéristiques des exemples de référence de l'ensemble de validation. Dans ce chapitre, la distance euclidienne est utilisée pour calculer les distances entre ces caractéristiques.

Soit  $X_i = (x_{i1}, x_{i2}, \dots, x_{iF})$  le vecteur caractéristique de l'exemple à classer  $x_i$  et  $F$  le nombre de caractéristiques. La distance euclidienne entre  $X_i$  et  $X_q$  (avec  $X_q = (x_{q1}, x_{q2}, \dots, x_{qN})$ ) peut être définie comme suit:

$$D(X_i, X_q) = \sqrt{\sum_{k=1}^N (x_{ik} - x_{qk})^2} \quad (5.1)$$

### 5.3.2.2 Détermination du critère de sélection MRMD-III

L'identification d'un sous-ensemble optimal de classifieurs nécessite un choix judicieux des critères de sélection pour obtenir un taux de classification élevé [Aksela 2003]. La diversité entre les classifieurs est un facteur important pour assurer l'efficacité de l'ensemble. Cependant, la sélection d'un ensemble avec la plus grande diversité peut apporter des classifieurs faibles. Par conséquent, il est préférable de sélectionner un ensemble de modèles performants et de préserver la diversité entre eux en même temps.

Dans l'approche précédente MRMD-II, nous avons utilisé un nouveau compromis «diversité-performance» basé sur la relevance et la combinaison des mesures de diversité pour équilibrer la performance de l'ensemble et leur accord de classification. Dans cette approche, une version modifiée de critère de sélection proposé *Maximum relevance Maximum Diversité* (MRMD-III) est utilisée. Dans le but d'estimer la dissimilarité des classifieurs à sélectionner, nous utilisons la même combinaison de mesures de diversité  $Div_{com}$  (équation 4.7) qui a prouvé son efficacité sur différents jeux de données. La corrélation d'un sous-ensemble de classifieur  $E_i$  est calculée en se basant seulement sur la diversité globale de  $E_i$ . La diversité globale  $Div(E_i)$  est définie par :

$$Div(E_i) = \frac{2}{|E_i|(|E_i| - 1)} \sum_{j=1}^{|E_i|-1} \sum_{h=j+1}^{|E_i|} Div_{com}(c_j, c_h) \quad (5.2)$$

La mesure proposée est inspirée du contexte de la méthode de sélection mRMR en utilisant la relevance ( $REV$ ) pour estimer les précisions des ensembles candidats. La mesure  $REV$  est déterminée par la combinaison de la relevance d'un ensemble de classifieurs  $E_i$  et la relevance moyenne de ses membres. La relevance d'un classifieur  $c_j$  ou

d'un sous-ensemble  $E_i$  est définie par la valeur d'information mutuelle entre les sorties de ce classifieur (ou de l'ensemble) et le vecteur des classes  $X_\Omega$  des exemples (équations 4.6 et 4.10). La relevance totale d'un sous-ensemble  $E_i$  est définie par :

$$\overline{REV}(E_i) = MI(V_{E_i}, \Omega) + \frac{1}{|E_i|} \sum_{c_j \in E_i} MI(V_{c_j}, \Omega) \quad (5.3)$$

$$\text{et; } MI(V_{E_i}, \Omega) = \int_{V_{E_i}} \int_{\Omega} p(V_C, \Omega) \log_2 \frac{p(V_{E_i}, \Omega)}{p(V_{E_i}) p(\Omega)} \quad (5.4)$$

Cependant, dans le cas des données déséquilibrées, la mesure de relevance  $REV$  ne peut pas être considérée comme un critère de sélection, car elle peut introduire un biais vers la classe majoritaire, ce qui peut conduire à une mauvaise estimation de la compétence du classifieur et par conséquent, à la sélection des classifieurs incompetents. Pour cette raison, nous proposons une autre mesure de relevance équilibrant les proportions des classes. La nouvelle mesure de relevance  $\overline{REV}_{IMB}(E_i)$  permet de calculer la relevance des classifieurs de chaque classe d'une manière indépendante, en divisant la région de compétence  $\theta_i$  en deux sous-ensembles:  $\theta_i^{min}$  pour la classe minoritaire et  $\theta_i^{maj}$  pour la classe majoritaire. Une telle décomposition nous permet de sélectionner un sous-ensemble de classifieurs compétents dans les deux régions de compétence minoritaires et majoritaires. La relevance totale  $\overline{REV}_{IMB}$  des données déséquilibrées est déterminée par :

$$\overline{REV}_{IMB}(E_i) = (1 - \alpha) \cdot \overline{REV}_{min}(E_i) + \alpha \cdot \overline{REV}_{maj}(E_i) \quad (5.5)$$

Où,  $\overline{REV}_{min}(E_i)$  et  $\overline{REV}_{maj}(E_i)$  sont les mesures de relevance d'un sous-ensemble  $E_i$  dans les classes minoritaire et majoritaire, respectivement. Le paramètre  $\alpha$  permet de contrôler l'équilibre entre la relevance des exemples minoritaires et majoritaires. Soit  $|\theta_i^{min}|$  et  $|\theta_i^{maj}|$  les nombres des individus des régions de compétences  $\theta_i^{min}$  et  $\theta_i^{maj}$ , le rapport de déséquilibre  $\alpha$  peut être calculé comme suit:

$$\alpha = \frac{|\theta_i^{min}|}{|\theta_i^{maj}| + |\theta_i^{min}|} \quad (5.6)$$

Le critère de sélection MRMD-III proposé pour les données déséquilibrées sera donc défini comme suit:

$$MRMD - III(E_i) = \frac{\overline{REV}_{IMB}(E_i)}{Div(E_i)} \quad (5.7)$$

### 5.3.2.3 Schéma de sélection dynamique de classifieurs

Cette étape effectue la sélection d'un sous-ensemble de classifieurs compétents localement à partir de l'ensemble initial  $C$ . La qualité d'apprentissage d'un modèle de sélection peut être affectée par le choix du nombre de classifieurs dans le sous-ensemble solution [Álvarez 2015]. En effet, l'utilisation d'un nombre spécifique de classifieurs est inefficace et la recherche de la solution optimale dans toutes les combinaisons possibles est un problème NP-difficile. Pour ces raisons, un algorithme de recherche séquentielle en avant est appliqué lors du processus de sélection. Le critère proposé (MRMD-III) est utilisé comme mesure de sélection pour contrôler l'équilibre entre la précision et la diversité de chaque sous-ensemble de  $C$ . Dans la phase d'agrégation, la décision finale du système proposé est fournie en combinant les sorties des classifieurs sélectionnés par la règle de vote pondérée proposée au chapitre 4.

L'algorithme 5.1 résume les principales étapes du processus proposé de sélection dynamique de classifieurs DES-MRMD-III.

---

**Algorithme 5.1:** Processus de sélection dynamique de classifieurs DES-MRMD-III

---

**Entrées:**

base de données de validation étiquetée  $VD$  avec  $M$  Classes.

Ensemble initial de  $L$  classifieurs  $C = \{C_1, C_2, \dots, C_L\}$ .

Ensemble de test  $TE$  constitué de  $K$  exemples;  $TE = \{x_1, x_2, \dots, x_{N_{TE}}\}$ .

**Sorties:**

L'ensemble final des classifieurs sélectionnés  $E_i^*$ .

---

**Pour**  $i = 1 \dots N_{TE}$ .

1. **Déterminer** la région de compétence  $\theta_i$  à partir de  $VD$  en utilisant les k-PPV de  $x_i$ .
  2. **Sélectionner** le classifieur le plus compétent de l'ensemble  $C$ , avec  $E_0^* = \{C_{best}\}$  en utilisant la relevance  $REV$  de l'équation (4.6).
  3. **Exécuter** l'étape 3 de l'algorithme 4.1.
  4. **Retourner**  $C_i^*$  comme ensemble final de  $x_i$ .
-

## 5.4 Expérimentation

### 5.4.1 Les bases de données utilisées

Pour évaluer les performances de notre méthode proposée, nous avons utilisé douze jeux de données médicales à deux classes qui ont été utilisés dans de nombreux travaux concernant la classification des données médicales. Sept jeux de données sont issus de l'UCI (*Machine Learning Repository*) [Asuncion 2007], trois bases provenant de la collection de Kuncheva [Kuncheva 2004] et les deux jeux de données DDSM et RIM-One qui ont été utilisées au chapitre 3. Afin de diminuer l'influence de la variabilité dans les exemples d'apprentissage, nous avons utilisé le même protocole de validation croisée en 5 blocs adopté au chapitre 4. Le tableau 5.1 résume les caractéristiques de ces bases de données.

**Tableau 5.1.** Caractéristiques des bases de données utilisées.

Base de données	#Exemples	#Positif	#Négatif	#Caractéristiques
<b>Appendicitis</b>	106	21	85	8
<b>DDSM</b>	1000	500	500	28
<b>Diabetes</b>	768	268	500	8
<b>Echocardiogram</b>	131	43	88	12
<b>RIM-One</b>	166	74	92	28
<b>Heart disease</b>	270	120	150	13
<b>Hepatitis</b>	155	32	123	19
<b>Laryngeal1</b>	692	53	639	16
<b>Parkinson's</b>	195	48	147	22
<b>Respiratory</b>	85	40	45	17
<b>Weaning</b>	302	151	151	17
<b>Wisconsin breast cancer (WBC)</b>	699	241	458	9

### 5.4.2 Résultats expérimentaux

#### 5.4.2.1 Etude expérimentale de la méthode DES-MRMD-II

En premier lieu, nous avons étudié les performances de notre approche de sélection dynamique en utilisant le critère MRMD-II présenté au chapitre 4. Notre méthode DES-MRMD-II a été comparée avec d'autres méthodes de sélection de classifieurs, à savoir :

- DES-LA : Dans cette méthode, les classifieurs présentant les précisions les plus élevées dans la région de compétence  $\theta_i$  sont sélectionnés pour prédire la classe de  $x_i$ .
- DES-kPPV : cette méthode sélectionne dynamiquement les classifieurs en se basant sur leurs diversités et performances [Soares 2006 ; de Souto 2008].
- KNORA-E : explore les propriétés du concept d'oracle pour sélectionner l'ensemble de classifieurs le plus appropriés pour classer  $x_i$  [Ko 2008].
- MRMD-II : c'est la méthode de sélection statique présentée au chapitre 4 [Cheriguene 2018].

Les objectifs de ces expérimentations effectuées sur les douze jeux de données médicales sont d'une part de tester l'effet de la sélection dynamique de classifieurs sur l'amélioration du taux de classification et d'autre part de montrer les performances de l'approche que nous avons proposée. Les tableaux 5.2, 5.3 et 5.4 présentent les résultats de précision, sensibilité et spécificité, respectivement, des ensembles DES-LA, DES-kPPV, KNORA-E, MRMD-II, et DES-MRMD-II.

Les résultats expérimentaux obtenus ici montrent que la technique proposée DES-MRMD-II réalise des meilleures performances par rapport aux autres méthodes de sélection dynamique. L'analyse de ces tableaux montre également que la sélection dynamique de classifieurs améliore le taux de classification par rapport à la sélection statique (MRMD-II).

À partir du tableau 5.2, on remarque que l'algorithme DES-MRMD-II offre des taux de classification qui dépassent les résultats des autres méthodes sur la plupart des cas. Le DES-MRMD-II obtient les meilleurs résultats pour 10 jeux de données avec une précision moyenne de 90,07% et il a augmenté la précision par 4.14%, 3.55%, 1.32% et 2.06% par rapport aux méthodes DES-LA, DES-kPPV, KNORA-E, MRMD-II, respectivement. La plus grande performance prédictive pour l'ensemble de données Hepatitis est obtenue par la méthode statique MRMD-II, tandis que la plus grande précision pour la base de données Weaning est obtenue par l'approche KNORA-E.

**Tableau 5.2.** Résultats de précision de l'approche DES-MRMD-II et différentes méthodes de sélection.

Base de données	DES-LA	DES-kPPV	KNORA-E	MRMD-II	DES-MRMD-II
Appendicitis	85.84	84.90	87.73	87.73	<b>88.61</b>
DDSM	92.11	93.47	94.56	93.75	<b>97.28</b>
Diabetes	75.72	76.57	78.25	78.95	<b>80.60</b>
Echocardiogram	89.31	90.08	90.82	87.88	<b>91.65</b>
Heart disease	83.33	84.81	87.03	85.58	<b>88.51</b>
Hepatitis	87.09	85.80	87.09	<b>90.38</b>	87.09
Laryngeal1	74.30	75.69	78.91	78.00	<b>82.51</b>
Parkinson's	89.74	90.76	93.84	91.32	<b>95.38</b>
Respiratory	89.41	90.58	92.94	90.58	<b>94.11</b>
RIM-One	91.59	92.60	94.05	93.91	<b>95.07</b>
Weaning	76.52	72.83	<b>82.14</b>	80.10	81.85
WBC	96.28	96.85	97.71	98.00	<b>98.42</b>
<b>Moyenne</b>	85.93	86.52	88.75	88.01	<b>90.07</b>

Comme indiqué aux tableaux 5.3 et 5.4, notre méthode donne les meilleurs résultats de sensibilité et de spécificité sur la plus part des jeux de données. Nous observons que DES-MRMD-II présente majoritairement des résultats de sensibilité meilleurs ou égaux que ceux obtenus par la méthode de sélection statique MRMD-II. De plus, les résultats de sensibilité de DES-MRMD-II surpassent ceux obtenus par les méthodes dynamiques pour 10 jeux de données médicales. En termes de spécificité, la méthode de sélection proposée génère des résultats équivalents ou meilleurs que ceux obtenus par les méthodes de sélection DES-LA et DES-kPPV. Les résultats de spécificité obtenus montrent aussi que l'algorithme DES-MRMD-II réalise une meilleure performance dans 10 cas sur 12, par rapport aux méthodes KNORA-E et MRMD-II.

De façon générale, la sensibilité moyenne obtenue par DES-MRMD-II (83,78%) est supérieure à celles des algorithmes DES-LA, DES-kPPV, KNORA-E et MRMD-II par des différences de 5,21%, 4,44%, 1,61% et 1,91% respectivement. En outre, DES-MRMD-II donne une amélioration de spécificité de 1,13% à 4,52% par rapport aux autres schémas de sélection.

**Tableau 5.3.** Résultats de sensibilité de différentes méthodes de sélection.

Base de données	DES-LA	DES-kPPV	KNORA-E	MRMD-II	DES-MRMD-II
Appendicitis	61.90	61.90	66.67	66.67	<b>71.42</b>
DDSM	52.23	56.71	56.71	61.19	<b>63.05</b>
Diabetes	95.68	92.24	93.10	93.10	<b>97.41</b>
Echocardiogram	79.06	83.72	<b>86.04</b>	79.06	<b>86.04</b>
Heart disease	71.66	81.66	80.83	78.33	<b>82.49</b>
Hepatitis	<b>81.25</b>	78.12	78.12	78.12	<b>81.25</b>
Laryngeal1	59.25	66.66	<b>76.54</b>	74.07	72.83
Parkinson's	83.33	83.78	85.41	85.41	<b>91.66</b>
Respiratory	<b>95.00</b>	87.50	92.5	90.00	92.50
RIM-One	91.85	90.87	93.15	93.15	<b>94.46</b>
Weaning	76.51	73.50	79.47	<b>85.43</b>	73.50
WBC	95.02	95.43	97.51	97.92	<b>98.75</b>
Moyenne	78.56	79.34	82.17	81.87	<b>83.78</b>

**Tableau 5.4.** Résultats de spécificité de différentes méthodes de sélection.

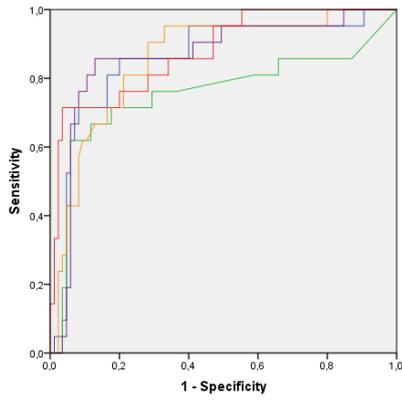
Base de données	DES-LA	DES-kPPV	KNORA-E	MRMD-II	DES-MRMD-II
Appendicitis	91.76	90.58	<b>92.94</b>	<b>92.94</b>	<b>92.94</b>
DDSM	87.00	87.60	89.8	87.00	<b>90.00</b>
Diabetes	86.02	95.58	<b>97.05</b>	94.85	<b>97.05</b>
Echocardiogram	<b>94.31</b>	93.18	93.18	92.04	<b>94.31</b>
Heart disease	92.66	87.33	<b>92.00</b>	<b>92.00</b>	93.33
Hepatitis	88.61	87.80	89.43	92.68	88.61
Laryngeal1	<b>83.33</b>	81.06	80.30	80.30	<b>83.33</b>
Parkinson's	93.19	92.37	<b>96.59</b>	93.87	<b>96.59</b>
Respiratory	86.66	91.12	93.33	91.12	<b>95.55</b>
RIM-One	91.38	93.99	94.77	94.51	<b>95.56</b>
Weaning	69.53	79.47	84.76	74.83	<b>90.00</b>
WBC	96.94	97.59	97.87	<b>98.25</b>	<b>98.25</b>
Moyenne	88.44	89.91	91.83	90.36	<b>92.96</b>

En outre, les résultats de la méthode DES-MRMD-II ont été comparés aux autres modèles de sélection en se basant sur le critère de l'aire sous la courbe caractéristique opérationnelle du récepteur (ROC). Le tableau 5.5 et les figures 5.2 et 5.3 illustrent les résultats des courbes ROC de différentes méthodes de sélection sur les douze bases de données.

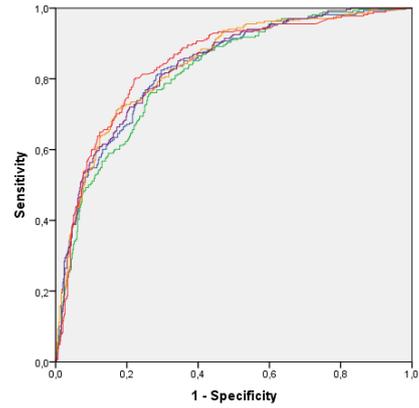
**Tableau 5.5.** Résultats d'AUC de différentes méthodes de sélection.

Base de données	DES-LA	DES-kPPV	KNORA-E	MRMD-II	DES-MRMD-II
Appendicitis	85.5	75.5	86.2	85.6	<b>87.7</b>
DDSM	81.6	83.0	83.5	83.8	<b>84.4</b>
Diabetes	94.8	96.2	97.4	97.0	<b>98.5</b>
Echocardiogram	91.4	92.2	93.4	93.3	<b>93.8</b>
Heart disease	91.4	91.5	92.5	92.6	<b>93.0</b>
Hepatitis	88.0	86.4	90.5	<b>90.9</b>	90.2
Laryngeal1	74.6	77.8	82.6	81.4	<b>82.9</b>
Parkinson's	94.0	95.0	<b>98.5</b>	96.4	<b>98.5</b>
Respiratory	95.3	95.9	96.9	96.4	<b>97.6</b>
RIM-One	95.1	95.8	96.1	96.1	<b>96.6</b>
Weaning	83.7	84.5	90.0	85.9	<b>90.5</b>
WBC	98.9	99.0	99.1	98.8	<b>99.3</b>
<b>Moyenne</b>	89.52	89.40	92.22	91.51	<b>92.75</b>

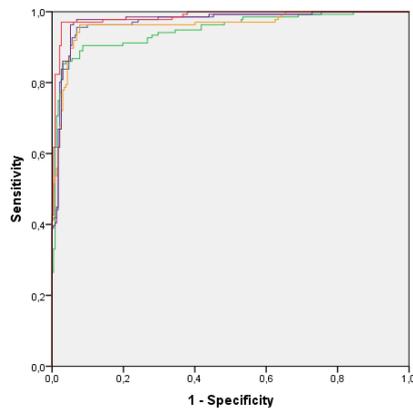
D'après les différentes courbes de ces figures, on remarque que la courbe ROC du DES-MRMD-II représente la plus grande surface sous la courbe ce qui revient à la plus grande valeur de AUC. Le tableau 5.3 confirme ce résultat étant donné que, pour les différentes stratégies de sélection testées, les plus grandes valeurs d'AUC sont fournies par la méthode DES-MRMD-II. Pour la base de données Hepatitis, la valeur d'AUC de la méthode DES-MRMD-II est légèrement inférieure à celle des méthodes MRMD-II et KNORA-E. Les différences entre l'AUC moyen de la méthode DES-MRMD-II et les AUC moyens des autres méthodes de sélection sont les suivantes: 3.23% (DES-LA), 3.35% (DES-kPPV), 0.53% (KNORA-E) et 1.24% (MRMD-II).



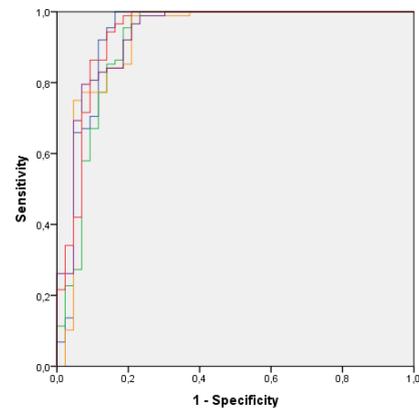
**Appendicitis**



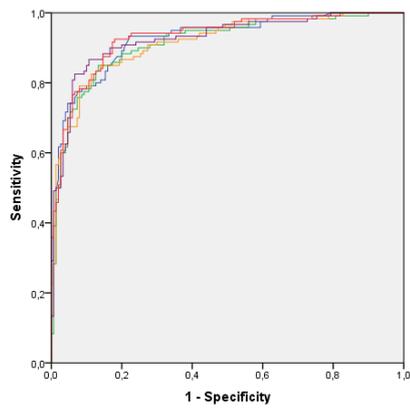
**Diabetes**



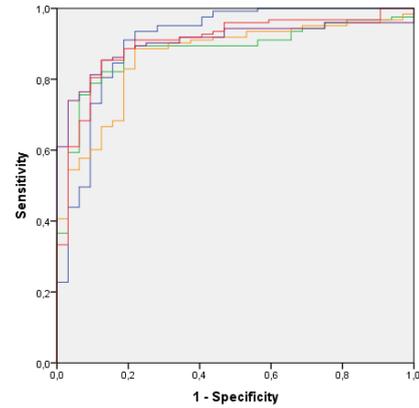
**DDSM**



**Echocardiogram**

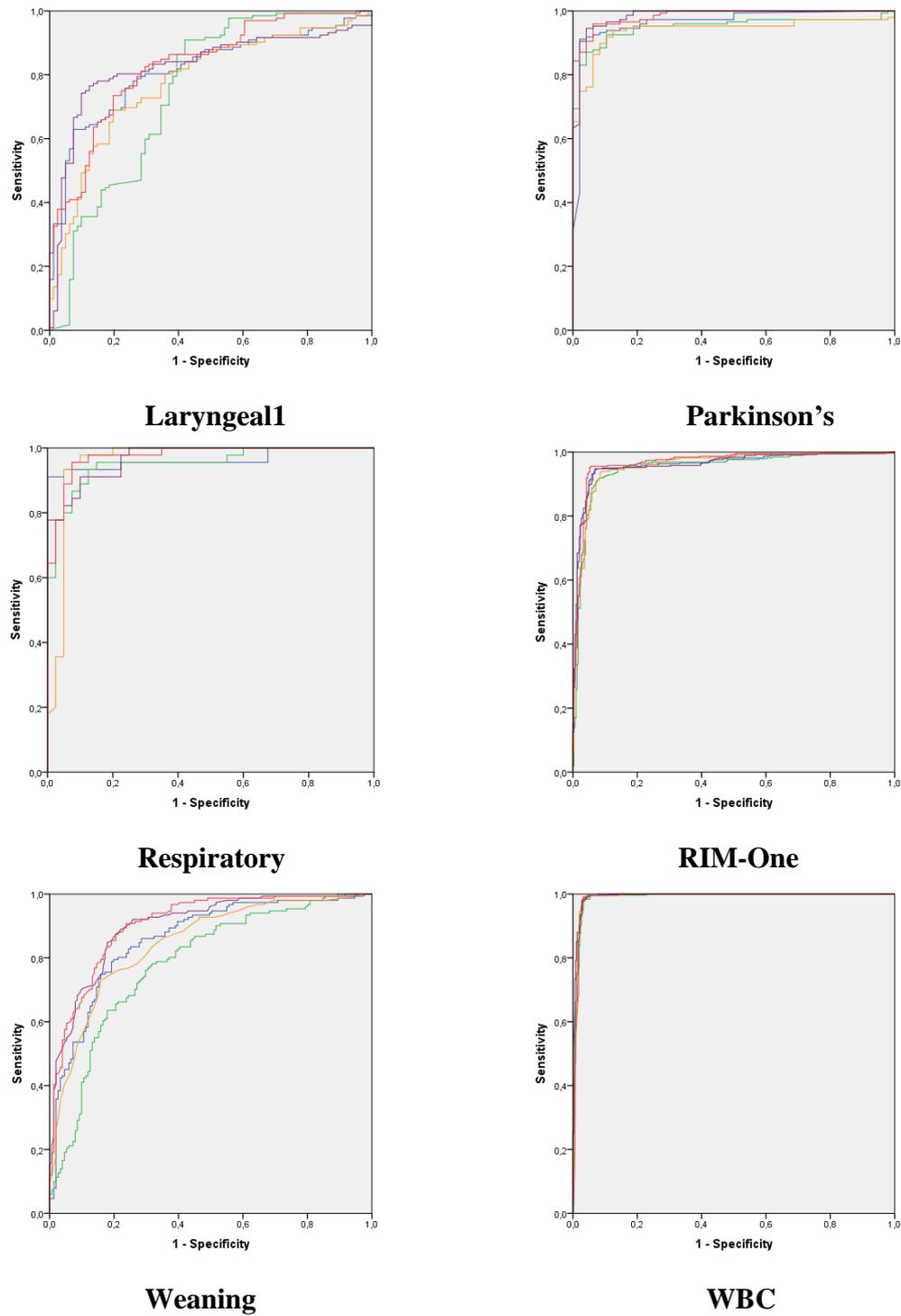


**Heart-disease**



**Hepatitis**

**Figure 5.2.** Les courbes ROC des bases Appendicitis, Diabetes, DDSM, Echocardiogram, Heart-disease et Hepatitis.



**Figure 5.3.** Les courbes ROC des bases Laryngeal1, Parkinson's, Respiratory, RIM-One, Weaning et WBC.

En conclusion, à partir des résultats des tableaux 5.2 à 5.5, nous constatons que l'ajout d'un compromis diversité-performance améliore la performance de la sélection dynamique de classifieurs.

#### 5.4.2.2 Problèmes des données déséquilibrées

Afin d'évaluer la performance de notre approche de sélection dynamique sur des problèmes liés aux données déséquilibrées, nous avons sélectionné six jeux de données utilisés dans l'expérimentation précédente. Le tableau 5.6 présente pour chaque base, le nombre d'individus de chaque classe (minoritaire et majoritaire) et la distribution des classes minoritaires et majoritaires (Ratio de déséquilibre).

**Tableau 5.6.** Description des bases de données déséquilibrées

Base de données	#Minoritaire	#Majoritaire	Ratio de déséquilibre
<b>Appendicitis</b>	21	85	4.05
<b>Diabetes</b>	268	500	1.87
<b>Echocardiogram</b>	43	88	2.05
<b>Hepatitis</b>	32	123	3.84
<b>Laryngeal1</b>	53	639	12.06
<b>Parkinson's</b>	48	147	3.06

Dans cette expérimentation, nous avons comparé la méthode DES-MRMD-III utilisant la mesure de sélection  $\overline{REV}_{IMB}$  et la technique SMOTE avec la méthode DES-MRMD-II pour les cas équilibrés et avec deux méthodes variantes de l'algorithme Boosting : SMOTEBoost [Chawla 2003] et RAMOBoost [Chen 2010]. L'approche DES-MRMD-II a été comparée également avec deux autres versions de notre méthode: (1) DES-MRMD-II en utilisant seulement la mesure de relevance  $\overline{REV}_{IMB}$  pour les données déséquilibrées (DES-MRMD-II+ $\overline{REV}_{IMB}$ ) et (2) DES-MRMD-II combiné avec la technique de prétraitement SMOTE (DES-MRMD-II+SMOTE). Les résultats de précision, sensibilité et spécificité sur ces jeux de données sont illustrés par les tableaux 5.7 et 5.8.

**Tableau 5.7.** Taux de classification de différents ensembles de classifieurs sur des bases de données déséquilibrées.

Base de données	DES-MRMD-II	SMOTE-Boost	RAMO-Boost	DES-MRMD-III+REV <sub>IMB</sub>	DES-MRMD-II+SMOTE	DES-MRMD-III
Appendicitis	88.61	86.79	87.73	89.62	<b>90.56</b>	91.50
Diabetes	80.60	83.59	83.72	85.80	84.76	<b>87.23</b>
Echocardiogram	91.65	91.60	92.36	93.89	92.36	<b>94.65</b>
Hepatitis	87.09	87.74	87.74	90.32	89.03	<b>92.90</b>
Laryngeal1	82.51	83.09	83.23	83.52	83.52	<b>83.81</b>
Parkinson's	95.38	95.38	95.38	95.89	95.89	<b>96.41</b>
Moyenne	87.64	88.03	88.36	89.84	89.35	<b>91.08</b>

**Tableau 5.8.** Résultats de sensibilité (SEN) et spécificité (SPE) de différents ensembles de classifieurs sur des bases de données déséquilibrées.

Base de données		DES-MRMD-III	SMOTE-Boost	RAMO-Boost	DES-MRMD-III+REV <sub>IMB</sub>	DES-MRMD-III+SMOTE	DES-MRMD-III <sub>IMB</sub>
Appendicitis	SEN	71.42	66.66	71.42	85.71	80.95	<b>90.47</b>
	SPE	92.94	91.76	91.76	90.58	92.94	<b>94.11</b>
Diabetes	SEN	63.05	72.37	72.01	78.35	74.62	<b>80.97</b>
	SPE	90.00	89.60	90.00	89.80	90.20	<b>90.60</b>
Echocardiogram	SEN	86.04	88.37	88.37	<b>93.02</b>	88.37	<b>93.02</b>
	SPE	<b>94.31</b>	93.18	<b>94.31</b>	<b>94.31</b>	93.18	<b>94.31</b>
Hepatitis	SEN	81.25	84.37	84.37	90.62	87.50	<b>93.75</b>
	SPE	88.61	88.61	88.61	90.24	89.43	<b>92.68</b>
Laryngeal1	SEN	72.83	77.36	81.13	<b>88.68</b>	83.01	86.79
	SPE	83.33	<b>83.57</b>	83.41	83.09	<b>83.57</b>	<b>83.57</b>
Parkinson's	SEN	91.66	89.58	91.66	<b>95.83</b>	93.75	<b>95.83</b>
	SPE	96.59	<b>97.27</b>	96.59	95.91	96.59	96.59
Moyenne	SEN	77.70	79.78	81.49	88.70	84.70	<b>90.13</b>
	SPE	90.96	90.66	90.78	90.66	90.98	<b>91.97</b>

Le tableau 5.7 montre clairement que les résultats obtenus par les trois versions de notre méthode proposée pour les données déséquilibrées sont meilleurs que ceux des

algorithmes SMOTEBoost, RAMOBoost et DES-MRMD-II sur tous les jeux de données. Nous observons de ces résultats que la méthode DES-MRMD-II+ $\overline{REV}_{IMB}$  fournit des taux de reconnaissance élevés en le comparant à ceux de DES-MRMD-II+SMOTE sur les bases de Diabetes, Echocardiogram et Hepatitis. Pour les jeux de données Laryngeal1 et Parkinson's, DES-MRMD-II+SMOTE et DES-MRMD-II+ $\overline{REV}_{IMB}$  obtiennent les mêmes résultats de précision, tandis que pour la base Appendicitis, le résultat de DES-MRMD-II+SMOTE dépassent celui de DES-MRMD-II+ $\overline{REV}_{IMB}$ .

Par ailleurs, ces résultats montrent aussi que la combinaison (appelée DES-MRMD-III) de la méthode SMOTE et de la sélection dynamique en utilisant la mesure  $\overline{REV}_{IMB}$  génère de meilleures performances plutôt que d'utiliser chacune séparément. La méthode proposée DES-MRMD-III fournit les meilleurs taux de classification sur tous les ensembles de données avec une moyenne de 91,08%.

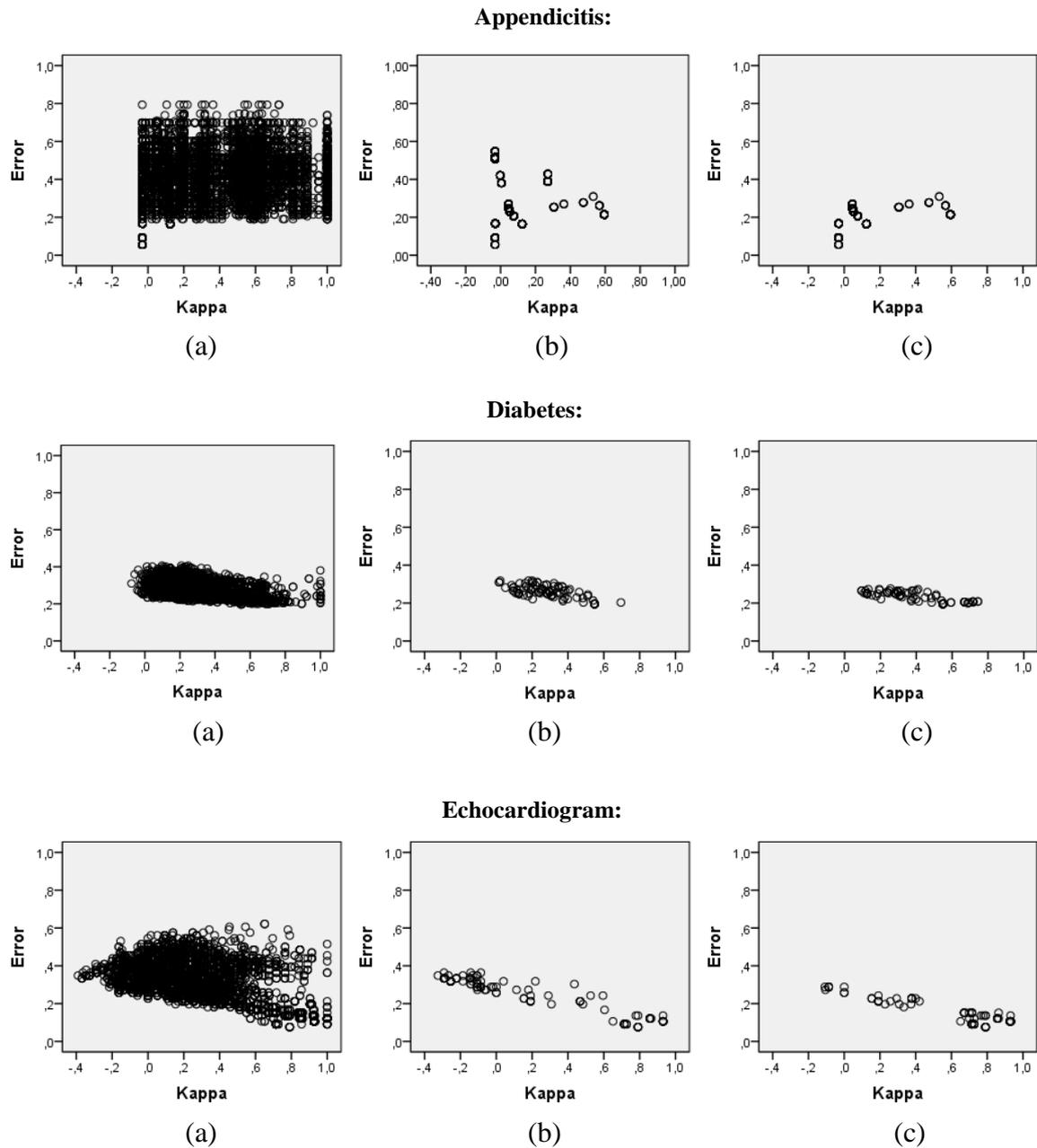
En termes de sensibilité et spécificité, le tableau 5.8 montre également que les résultats de la méthode DES-MRMD-III sont meilleurs que les autres ensembles de classifieurs sur tous les jeux de données avec des moyennes de 90,13% pour la sensibilité et de 91,97% pour la spécificité.

### 5.4.2.3 Relation entre la diversité et la précision

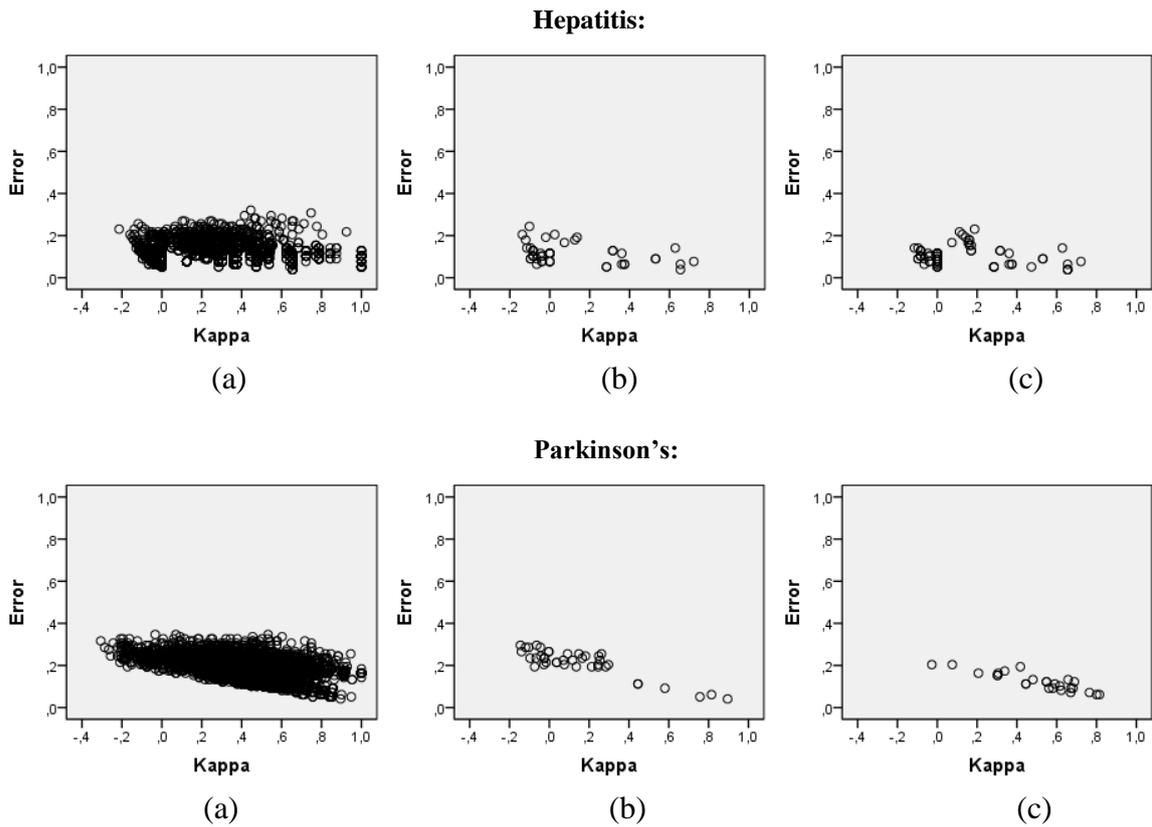
Dans la conception d'un système multi-classifieurs, le compromis entre la diversité et la précision est un facteur important pour garantir le succès de la phase de sélection. Le diagramme kappa-erreur [Margineantu 1997] est un modèle populaire utilisé pour montrer l'influence de la relation entre la diversité et la précision sur la création d'un ensemble de classifieurs. Dans ce chapitre, le diagramme kappa-erreur est utilisé pour examiner le comportement de nos critères proposés de sélection et pour analyser aussi la corrélation entre les membres de l'ensemble sélectionné sur les bases de données déséquilibrées. Les figures 5.4 et 5.5 présentent les diagrammes kappa-erreur des méthodes Random Subspace, DES-MRMD-II et DES-MRMD-III pour les bases de données Appendicitis, diabetes, Echocardiogram, Hepatitis and Parkinson's. Le diagramme kappa-erreur de l'ensemble Laryngeal1 n'est pas présenté car un seul classifieur est sélectionné dans les deux schémas dynamiques.

Ces figures montrent que les diagrammes de dispersion des ensembles sélectionnés sont généralement situés dans la partie gauche, la partie inférieure ou à la partie inférieure

gauche du diagramme kappa-erreur de la méthode Random Subspace. Cela confirme l'existence d'un équilibre entre la précision et la diversité dans les ensembles de classifieurs. De plus, la méthode DES-MRMD-III sélectionne des classifieurs généralement moins divers mais plus précis par rapport à la technique DES-MRMD-II. Cependant, la métrique de pertinence proposée pour les ensembles de données déséquilibrés permet de réduire le biais vers la classe majoritaire et d'atteindre ainsi un meilleur taux de précision.



**Figure 5.4.** Les diagrammes kappa-erreur des méthodes : (a) Random Subpace, (b) DES-MRMD-II et (c) DES-MRMD-III sur les bases Appendicitis, Diabetes et Echocardiogram.



**Figure 5.5.** Les diagrammes kappa-erreur des méthodes : (a) Random Subspace, (b) DES-MRMD-II et (c) DES-MRMD-III sur les bases Hepatitis et Parkinson's.

## 5.5 Conclusion

Dans ce chapitre, nous avons présenté un système multi-classifieurs basé sur l'échantillonnage des données et le concept de la sélection dynamique de classifieurs, permettant d'apporter une solution au problème des classes déséquilibrées. L'approche de sur-échantillonnage SMOTE a été considérée dans ce système pour rééquilibrer les données d'apprentissage et améliorer la diversité entre les classifieurs de l'ensemble initial. Le processus de sélection proposé est une extension de l'approche statique MRMD-II proposée au chapitre précédent. Cette méthode de sélection dynamique (DES-MRMD-III) est basée sur la recherche séquentielle en avant (*forward*) pour limiter l'espace de recherche et fournir un sous-ensemble optimal de classifieurs. Afin de mieux guider le traitement de déséquilibre des classes, l'utilisation d'un critère de sélection basé sur la distribution des classes nous paraît une solution efficace pour choisir les classifieurs les plus compétents dans une région locale de caractéristiques en tenant compte de leurs diversités.

Les résultats expérimentaux menés sur les douze bases de données médicales ont montré que l'extension dynamique (DES-MRMD-II) de l'approche MRMD-II pour les cas équilibrés a été capable de surpasser plusieurs méthodes de sélection statique et dynamique sur différents jeux de données. L'amélioration des résultats par rapport aux schémas dynamiques peut être justifié à notre avis par l'algorithme de recherche glouton utilisé et le critère de sélection proposé qui nous a permis d'établir un équilibre entre la précision et la diversité des classifieurs. En outre, la méthode DES-MRMD-II a choisis différents sous-ensembles de classifieurs pour classer chaque échantillon de test ce qui justifie l'amélioration des résultats statiques.

Par ailleurs, plusieurs expérimentations ont été menées pour évaluer la performance de l'approche DES-MRMD-III sur des cas déséquilibrés. Les résultats expérimentaux obtenus sur six bases de données médicales ont montré que les résultats obtenus par DES-MRMD-III sont plus encourageant que l'approche précédente DES-MRMD-II. De plus, ces résultats ont montré l'efficacité du schéma de sélection dynamique DES-MRMD-III par rapport à d'autres méthodes ensemblistes proposées pour des problèmes de données déséquilibrées. Ces résultats peuvent s'expliquer par le fait que le traitement du problème de l'asymétrie des classes est appliqué en deux étapes: premièrement, la méthode de sur-échantillonnage SMOTE a été utilisée dans l'étape de génération de l'ensemble afin de créer un pool de classifieurs diversifié. Ensuite, Un poids basé sur la distribution des classes a été proposé pour équilibrer l'estimation des compétences des classifieurs.

## Conclusion et Perspectives

Ensemble de classifieurs ou système multi-classifieurs est un domaine de recherche qui a donné lieu à de nombreuses études et à de nouvelles approches. Une telle approche permet de réduire la variance des erreurs de classification des différents classifieurs élémentaires, ce qui a pour conséquence de stabiliser la décision finale en présence de changements. Un système multi-classifieurs est généralement composé de trois phases principales : la génération des classifieurs initiaux, la sélection des meilleurs candidats et la fusion des décisions des classifieurs sélectionnés. Toutefois, de nombreuses études des différentes références confirment que la sélection de classifieurs est généralement meilleure que la combinaison de tous les membres de l'ensemble initial. Ceci nous a donc amené à choisir le concept de sélection de classifieurs comme support de nos travaux. Les différents travaux réalisés durant cette thèse apportent plusieurs contributions concernant la sélection de classifieurs pour des problèmes de classification supervisée.

La première contribution proposée dans cette thèse concerne l'utilisation d'ensembles de classifieurs pour la classification des images médicales. Nous avons proposé un nouveau système multi-classifieurs (RSS-SCS) qui combine l'algorithme Random Subspace (RSS) et la sélection statique de classifieurs (SCS) selon le paradigme de surproduction et choix. Cette approche consiste à construire un pool initial de classifieurs en utilisant différents sous-espaces de caractéristiques dans le but de renforcer et créer un certain niveau de diversité entre les membres de l'ensemble. La phase de sélection consiste à choisir le sous-ensemble de classifieurs adéquats en utilisant les mesures de diversité. Ce choix a été effectué pour pouvoir clarifier et étudier le rôle de la diversité au cours de la phase de sélection des ensembles de classifieurs.

Nous avons réalisé de nombreuses expérimentations pour évaluer le système ensembliste proposé en utilisant trois classifieurs de base (MLP, SVM, AD), différentes tailles du sous-ensemble sélectionné et deux bases de données médicales (DDSM, RIM-One) pour la validation. En outre, plusieurs mesures de diversité ont été utilisées comme critère de sélection afin d'analyser l'influence de la diversité sur le processus de sélection de classifieurs. Les résultats obtenus ont prouvé que la performance du processus de sélection peut être liée au type de classifieur de base adopté, à la mesure de diversité utilisée ainsi qu'à la taille du sous-ensemble de classifieurs à sélectionner. Les résultats d'expérimentaux obtenus lors de la comparaison de l'approche RSS-SCS avec d'autres méthodes ensemblistes, à savoir : Bagging, Boosting et Random Subspace, ont démontré que notre approche est plus performante en termes de précision, sensibilité, spécificité et AUC sur les deux bases médicales DDSM et RIM-One. Ceci démontre que la sélection de classifieurs est plus efficace que la combinaison de tous les membres de l'ensemble initial. Cependant, la comparaison de cette approche avec l'algorithme de sélection N-Best nous a montré que l'utilisation unique des mesures de diversité comme critère de sélection est insuffisante pour trouver les sous-ensembles de classifieurs les plus compétents. La performance de membres candidats est un facteur important qui doit être prise en considération lors du processus de sélection de classifieurs.

La deuxième contribution de cette thèse correspond à l'introduction d'une nouvelle approche de sélection de classifieurs permettant d'offrir de meilleures performances dans différents domaines de la reconnaissance de formes. Afin de contrôler le compromis entre la performance et la diversité entre les membres sélectionnés, l'approche proposée (MRMD) est basée sur l'intégration de la méthode mRMR et les mesures de diversité en utilisant un algorithme de recherche par ordonnancement. La relevance est utilisé dans le critère MRMD pour estimer la performance de chaque classifieur candidat. Les résultats expérimentaux réalisés sur vingt bases de données ont prouvé que la combinaison de la relevance et des mesures de diversité comme critère de sélection est meilleure que l'utilisation de chacun d'eux séparément. Ces résultats ont montré également que notre approche MRMD fournit des résultats différents selon le nombre de classifieurs à sélectionner. Cependant, l'utilisation d'un nombre fixe des classifieurs à sélectionner ne permet pas toujours de trouver une solution optimale.

Pour cette raison, une deuxième méthode MRMD-II a été proposée afin d'améliorer l'algorithme MRMD pour de meilleures performances en utilisant un algorithme de recherche gloutonne qui permet de contrôler la taille du sous-ensemble sélectionné. Cette méthode intègre aussi la relevance pour prendre en considération les performances individuelles des classifieurs sélectionnés ainsi que la performance globale de leur combinaison. Les résultats que nous avons obtenus sur différents benchmarks de l'UCI et LKC ont montré que cette méthode est efficace et robuste en termes de taux de classification. Cette étude a prouvé que les résultats obtenus par la méthode MRMD-II sont meilleurs que ceux de l'approche précédente MRMD. En outre, l'étude comparative faite par rapport aux autres méthodes d'ensemble de classifieurs confirme encore une fois que la sélection d'un sous-ensemble de classifieurs est meilleure que la combinaison de tous les membres de l'ensemble initial. D'autre part, les diagrammes kappa-erreur ont été utilisés durant nos expérimentations pour pouvoir étudier le compromis existant entre la précision et la diversité des classifieurs sélectionnés. Les résultats de ces diagrammes ont montré l'efficacité du critère proposé (Maximum Relevance Maximum Diversity-II) pour contrôler l'équilibre entre la diversité des classifieurs et leurs performances.

La dernière contribution proposée dans cette thèse aborde le problème des données déséquilibrées qui se produit lorsque le nombre des individus d'une classe est nettement plus petit que celui d'autres classes. Cette contribution présente une approche ensembliste combinant la sélection dynamique d'ensemble de classifieurs et la méthode de sur-échantillonnage des données SMOTE pour traiter la classification dans un environnement déséquilibré. L'approche SMOTE a été considérée dans ce système multi-classifieurs pour rééquilibrer les données d'apprentissage et améliorer la diversité entre les classifieurs de l'ensemble initial. Le système de sélection proposé présente une extension dynamique de la méthode de sélection statique MRMD-II pour permettre la sélection du sous-ensemble de classifieurs le plus confiant pour étiqueter chaque échantillon individuellement. De cette manière, au lieu de choisir une seule solution en phase de généralisation, tous les ensembles candidats issus de la population générée par un algorithme de recherche séquentielle sont considérés pour sélectionner la solution la plus compétente pour chaque échantillon de test. De plus, nous avons proposé de mener notre sélection dynamique en utilisant un nouveau compromis diversité-précision (MRMD-III) pour les données déséquilibrées qui se base sur les mesures de diversité et la relevance pondérée par la distribution des classes.

L'évaluation des performances de l'approche proposée a été d'abord effectuée sur deux bases de données médicales en appliquant une sélection dynamique basée sur le critère MRMD-II des données équilibrées. Les résultats expérimentaux ont démontré que les sous-ensembles de classifieurs sélectionnés dynamiquement par le critère MRMD-II présentent des meilleures performances par rapport aux taux obtenus par d'autres méthodes de sélection dynamique ainsi que la méthode statique proposée au chapitre 4. Par ailleurs, d'autres expérimentations ont été menées sur six bases de données déséquilibrées afin d'examiner la performance de l'approche DES-MRMD-III. Les résultats expérimentaux obtenus démontrent que l'approche DES-MRMD-III est meilleure que l'approche précédente DES-MRMD-II. De plus, l'approche DES-MRMD-III a prouvé son efficacité par rapport à d'autres approches ensemblistes proposées pour les cas déséquilibrés.

### Perspectives

De nombreuses perspectives peuvent être envisagées suite à cette thèse, notamment :

- Le critère de sélection utilisé dans les approches proposées (statique et dynamique) permet d'équilibrer le compromis entre la précision et la diversité de l'ensemble en utilisant le même taux d'importance à chacun de ces deux facteurs. Une poursuite des travaux est de faire modifier et varier l'importance donnée à la précision et/ou à la diversité lors de la recherche des classifieurs compétents, et d'observer son influence sur la qualité globale des résultats obtenus.
- L'utilisation d'un algorithme de recherche est une autre perspective à venir. Afin de diminuer le coût de calcul de notre système de sélection dynamique, d'autres méthodes de recherche éventuelle comme les algorithmes génétiques peuvent être appliqués.
- Il serait également important d'examiner l'impact de notre stratégie dynamique de classifieurs dans des problèmes multi-classes de données déséquilibrées.
- L'étude et la comparaison de différentes stratégies pour la génération de l'ensemble initial de classifieurs, ce qui présente l'entrée de la phase de sélection dynamique de notre méthode, constitue une autre perspective de nos travaux.

# Liste des Publications

## Conférences et colloques nationaux et internationaux

- Cheriguene, S., & Azizi, N. (2015, April). A Classifier Ensemble Model based on Random Subspace and Static Classifiers Selection : Application on classification problems. *2ème Conférence sur la Vision Artificielle*.
- Cheriguene, S., Azizi, N., Dey, N., Ashour, A. S., Mnerie, C. A., Olariu, T., & Shi, F. (2016, August). Classifier Ensemble Selection Based on mRMR Algorithm and Diversity Measures: An Application of Medical Data Classification. In *International Workshop Soft Computing Applications* (pp. 375-384). Springer, Cham.
- Cheriguene, S., Azizi, A., Farah, N., & Ziani, A. Medical Data Classification System based on bayesian networks Selection Ensemble. (2016, October). In *the 3rd International Conference on Bayesian Networks and Applications*.
- Cheriguene, S., Azizi, A., & Ziani, A. (2016, December). A two stage Classifier Selection Ensemble based on mRMR Algorithm and Diversity Measures. In the *2nd Conference on Computing Systems and Applications*, Algiers, Algeria.

## Publications nationales et internationales

- Cheriguene, S., Azizi, N., Dey, N., Ashour, A. S., & Ziani, A. (2018). A new hybrid classifier selection model based on mRMR method and diversity measures. *International Journal of Machine Learning and Cybernetics*, 1-16.

## Chapitres d'ouvrages scientifiques

- Cheriguene, S., Azizi, N., Zemmal, N., Dey, N., Djellali, H., & Farah, N. (2016). Optimized tumor breast cancer classification using combining random subspace and static classifiers selection paradigms. In *Applications of intelligent optimization in biology and medicine* (pp. 289-307). Springer, Cham.
- Cheriguene, S., Azizi, N., & Dey, N. (2016). Ensemble classifiers construction using diversity measures and random subspace algorithm combination: application to glaucoma diagnosis. In *Medical Imaging in Clinical Applications* (pp. 131-152). Springer, Cham.

# Bibliographie

- [Abubacker 2018] Abubacker, N. F., Azman, A., Doraisamy, S., & Murad, M. A. A. (2018). An integrated method of associative classification and neuro-fuzzy approach for effective mammographic classification. *Neural Computing and Applications*, 28(12), 3967-3980.
- [Ahmed 2017] Ahmed, M., Rasool, A. G., Afzal, H., & Siddiqi, I. (2017). Improving handwriting based gender classification using ensemble classifiers. *Expert Systems with Applications*, 85, 158-168.
- [Akbari 2004] Akbari, R., Kwek, S., & Japkowicz, N. (2004, September). Applying support vector machines to imbalanced datasets. In *European conference on machine learning* (pp. 39-50). Springer, Berlin, Heidelberg.
- [Aksela 2003] Aksela, M. (2003, June). Comparison of classifier selection methods for improving committee performance. In *International Workshop on Multiple Classifier Systems* (pp. 84-93). Springer, Berlin, Heidelberg.
- [Aksela 2006] Aksela, M., & Laaksonen, J. (2006). Using diversity of errors for selecting members of a committee classifier. *Pattern Recognition*, 39(4), 608-623.
- [Álvarez 2015] Álvarez, A., Sierra, B., Arruti, A., López-Gil, J. M., & Garay-Vitoria, N. (2015). Classifier subset selection for the stacked generalization method applied to emotion recognition in speech. *Sensors*, 16(1), 21.
- [Asuncion 2007] Asuncion, A., & Newman, D. (2007). UCI machine learning repository.
- [Azizi 2010] Azizi, N., Farah, N., Sellami, M., & Ennaji, A. (2010, April). Using diversity in classifier set selection for arabic handwritten recognition. In *International Workshop on Multiple Classifier Systems* (pp. 235-244). Springer, Berlin, Heidelberg.

- [Azizi 2013] Azizi, N., Tlili-Guiassa, Y., & Zemmal, N. (2013). A computer-aided diagnosis system for breast cancer combining features complementarily and new scheme of SVM classifiers fusion. *International Journal Of Multimedia and Ubiquitous Engineering*, 8(4), 45-58.
- [Bakker 2003] Bakker, B., & Heskes, T. (2003). Clustering ensembles of neural network models. *Neural networks*, 16(2), 261-269.
- [Baluja 1994] Baluja, S. (1994). Population-based incremental learning. a method for integrating genetic search based function optimization and competitive learning (No. CMU-CS-94-163). Carnegie-Mellon Univ Pittsburgh Pa Dept Of Computer Science.
- [Banfield 2003] Banfield, R. E., Hall, L. O., Bowyer, K. W., & Kegelmeyer, W. P. (2003, June). A new ensemble diversity measure applied to thinning ensembles. In *International Workshop on Multiple Classifier Systems* (pp. 306-316). Springer, Berlin, Heidelberg.
- [Bashbaghi 2017] Bashbaghi, S., Granger, E., Sabourin, R., & Bilodeau, G. A. (2017). Dynamic ensembles of exemplar-svms for still-to-video face recognition. *Pattern recognition*, 69, 61-81.
- [Batista 2004] Batista, G. E., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD explorations newsletter*, 6(1), 20-29.
- [Bauer 1999] Bauer, E., & Kohavi, R. (1999). An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine learning*, 36(1-2), 105-139.
- [Bechar 2018] Bechar, M. E. A., Settouti, N., Barra, V., & Chikh, M. A. (2018). Semi-supervised superpixel classification for medical images segmentation: application to detection of glaucoma disease. *Multidimensional Systems and Signal Processing*, 29(3), 979-998.
- [Belacel 1999] Belacel, N. (1999). Méthodes de classification multicritère, méthodologie et applications à l'aide au diagnostic médicale. (Thèse de doctorat), Université Libre de Bruxelles, Belgique.
- [Berbaum 1989] Berbaum, K. S., Dorfman, D. D., & Franken Jr, E. A. (1989). Measuring observer performance by ROC analysis: indications and complications. *Investigative Radiology*, 24(3), 228-233.

- [Bernard 2009] Bernard, S. (2009). *Forets Aléatoires : De l'Analyse des Mécanismes de Fonctionnement à la Construction Dynamique*. Human-Computer Interaction. (Thèse de doctorat), Université de Rouen, France.
- [Bilmes 2000] Bilmes, J. A., & Kirchoff, K. (2000). Directed graphical models of classifier combination: Application to phone recognition. In *Sixth International Conference on Spoken Language Processing*.
- [Bishop 2006] Bishop, C.M. (2006). *Pattern recognition and machine learning (Information science and statistics)*. Springer-Verlag New York.
- [Blum 1997] Blum, A. L., & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial intelligence*, 97(1-2), 245-271.
- [Boser 1992] Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992, July). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory* (pp. 144-152). ACM
- [Branco 2016] Branco, P., Torgo, L., & Ribeiro, R. P. (2016). A survey of predictive modeling on imbalanced domains. *ACM Computing Surveys (CSUR)*, 49(2), 31.
- [Breiman 1996a] Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2), 123-140.
- [Breiman 1996b] Breiman, L. (1996). Heuristics of instability and stabilization in model selection. *The annals of statistics*, 24(6), 2350-2383.
- [Breiman 2000] Breiman, L. (2000). Randomizing outputs to increase prediction accuracy. *Machine Learning*, 40(3), 229-242.
- [Britto 2014] Britto Jr, A. S., Sabourin, R., & Oliveira, L. E. (2014). Dynamic selection of classifiers—a comprehensive review. *Pattern Recognition*, 47(11), 3665-3680.
- [Brown 2009] Brown, G. (2009). Ensemble learning. In *Encyclopedia of Machine Learning* (pp. 312-320). Springer, Boston, MA.
- [Brown 2010] Brown, G., & Kuncheva, L. I. (2010, April). “Good” and “bad” diversity in majority vote ensembles. In *International workshop on multiple classifier systems* (pp. 124-133). Springer, Berlin, Heidelberg.
- [Canuto 2007] Canuto, A. M., Abreu, M. C., de Melo Oliveira, L., Xavier Jr, J. C., & Santos, A. D. M. (2007). Investigating the influence of the choice of the ensemble members in accuracy and diversity of selection-based and fusion-

- based methods for ensembles. *Pattern recognition letters*, 28(4), 472-486.
- [Cavalin 2012] Cavalin, P. R., Sabourin, R., & Suen, C. Y. (2012). LoGID: An adaptive framework combining local and global incremental learning for dynamic selection of ensembles of HMMs. *Pattern recognition*, 45(9), 3544-3556.
- [Cavalin 2013a] Cavalin, P. R., Sabourin, R., & Suen, C. Y. (2013). Dynamic selection approaches for multiple classifier systems. *Neural Computing and Applications*, 22(3-4), 673-688.
- [Chabat 2000] Chabat, F., Hansell, D. M., & Yang, G. Z. (2000). Computerized decision support in medical imaging. *IEEE Engineering in Medicine and Biology Magazine*, 19(5), 89-96.
- [Chawla 2002] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357.
- [Chawla 2003] Chawla, N. V., Lazarevic, A., Hall, L. O., & Bowyer, K. W. (2003, September). SMOTEBoost: Improving prediction of the minority class in boosting. In *European conference on principles of data mining and knowledge discovery* (pp. 107-119). Springer, Berlin, Heidelberg.
- [Cheikhrouhou 2012] Cheikhrouhou, I. (2012). Description et classification des masses mammaires pour le diagnostic du cancer du sein. (Thèse de Doctorat), Université d'Evry-Val d'Essonne, France.
- [Chen 2010] Chen, S., He, H., & Garcia, E. A. (2010). RAMOBoost: ranked minority oversampling in boosting. *IEEE Transactions on Neural Networks*, 21(10), 1624-1642.
- [Cheng 2006] Cheng, H. D., Shi, X. J., Min, R., Hu, L. M., Cai, X. P., & Du, H. N. (2006). Approaches for automated detection and classification of masses in mammograms. *Pattern recognition*, 39(4), 646-668.
- [Cheriguene 2015] Cheriguene, S., & Azizi, N. (2015, April). A Classifier Ensemble Model based on Random Subspace and Static Classifier Selection : Application on classification problems. 2ème *Conférence sur la Vision Artificielle*.
- [Cheriguene 2016a] Cheriguene, S., Azizi, N., Dey, N., Ashour, A. S., Mnerie, C. A., Olariu, T., & Shi, F. (2016, August). Classifier Ensemble Selection Based on mRMR Algorithm and Diversity Measures: An Application of Medical Data Classification. In *International Workshop Soft Computing Applications* (pp.

- 375-384). Springer, Cham.
- [Cheriguene 2016b] Cheriguene, S., Azizi, N., Zemmal, N., Dey, N., Djellali, H., & Farah, N. (2016). Optimized tumor breast cancer classification using combining random subspace and static classifiers selection paradigms. In *Applications of intelligent optimization in biology and medicine* (pp. 289-307). Springer, Cham.
- [Cheriguene 2016c] Cheriguene, S., Azizi, N., & Dey, N. (2016). Ensemble classifiers construction using diversity measures and random subspace algorithm combination: application to glaucoma diagnosis. In *Medical Imaging in Clinical Applications* (pp. 131-152). Springer, Cham.
- [Cheriguene 2016d] Cheriguene, S., Azizi, A., Farah, N., & Ziani, A. Medical Data Classification System based on bayesian networks Selection Ensemble. (2016, October). In *the 3rd International Conference on Bayesian Networks and Applications*.
- [Cheriguene 2016e] Cheriguene, S., Azizi, A., & Ziani, A. (2016, December). A two stage Classifier Selection Ensemble based on mRMR Algorithm and Diversity Measures. In *the 2nd Conference on Computing Systems and Applications*, Algiers, Algeria.
- [Cohen 1960] Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1), 37-46.
- [Çomak 2007] Comak, E., Polat, K., Güneş, S., & Arslan, A. (2007). A new medical decision making system: least square support vector machine (LSSVM) with fuzzy weighting pre-processing. *Expert Systems with Applications*, 32(2), 409-414.
- [Cormack 1971] Cormack, R. M. (1971). A review of classification. *Journal of the Royal Statistical Society. Series A (General)*, 321-367.
- [Cover 1994] Cover, T. M., Thomas, J. A., & Kieffer, J. (1994). Elements of information theory. *SIAM Review*, 36(3), 509-510.
- [Cruz 2011] Cruz, R. M., Cavalcanti, G. D., & Ren, T. I. (2011, July). A method for dynamic ensemble selection based on a filter and an adaptive distance to improve the quality of the regions of competence. In *The 2011 International Joint Conference on Neural Networks* (pp. 1126-1133). IEEE.
- [Cruz 2013] Cruz, R. M., Cavalcanti, G. D., Tsang, R., & Sabourin, R. (2013). Feature representation selection based on classifier projection space and oracle

- analysis. *Expert Systems with Applications*, 40(9), 3813-3827.
- [Cruz 2014] Cruz, R. M., Sabourin, R., & Cavalcanti, G. D. (2014, August). On meta-learning for dynamic ensemble selection. In *Pattern Recognition (ICPR), 2014 22nd International Conference on* (pp. 1230-1235). IEEE.
- [Cruz 2015] Cruz, R. M., Sabourin, R., & Cavalcanti, G. D. (2015). A DEEP analysis of the META-DES framework for dynamic selection of ensemble of classifiers. *arXiv preprint arXiv:1509.00825*.
- [Cruz 2017] Cruz, R. M., Sabourin, R., & Cavalcanti, G. D. (2017, May). Analyzing different prototype selection techniques for dynamic classifier and ensemble selection. In *International Joint Conference on Neural Networks* (pp. 3959-3966).
- [Cunningham 2000] Cunningham, P., & Carney, J. (2000, May). Diversity versus quality in classification ensembles based on feature selection. In *European Conference on Machine Learning* (pp. 109-116). Springer, Berlin, Heidelberg.
- [Davis 1991] Davis, L. (1991). Handbook of genetic algorithms. Van Nostrand Reinhold New York.
- [Dempster 1988] Dempster, A. P. (1988). Probability, evidence, and judgment. *Decision Making: Descriptive, normative, and prescriptive interactions*, 284-298.
- [de Souto 2008] de Souto, M. C., Soares, R. G., Santana, A., & Canuto, A. M. (2008, June). Empirical comparison of dynamic classifier selection methods based on diversity and accuracy for building ensembles. In *Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on* (pp. 1480-1487). IEEE.
- [Didaci 2004] Didaci, L., & Giacinto, G. (2004, June). Dynamic classifier selection by adaptive k-nearest-neighbourhood rule. In *International Workshop on Multiple Classifier Systems* (pp. 174-183). Springer, Berlin, Heidelberg.
- [Dietterich 1994] Dietterich, T. G., & Bakiri, G. (1994). Solving multiclass learning problems via error-correcting output codes. *Journal of artificial intelligence research*, 2, 263-286.
- [Dietterich 2000a] Dietterich, T. G. (2000). Ensemble methods in machine learning. In *International workshop on multiple classifier systems* (pp. 1-15). Springer, Berlin, Heidelberg.

- [Dietterich 2000b] Dietterich, T. G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine learning*, 40(2), 139-157.
- [Díez-Pastor 2015] Díez-Pastor, J. F., Rodríguez, J. J., García-Osorio, C., & Kuncheva, L. I. (2015). Random balance: ensembles of variable priors classifiers for imbalanced data. *Knowledge-Based Systems*, 85, 96-111.
- [Dominguez 2009] Domínguez, A. R., & Nandi, A. K. (2009). Toward breast cancer diagnosis based on automated segmentation of masses in mammograms. *Pattern Recognition*, 42(6), 1138-1148.
- [Dos Santos 2006] Dos Santos, E. M., Sabourin, R., & Maupin, P. (2006, July). Single and multi-objective genetic algorithms for the selection of ensemble of classifiers. In *Neural Networks, 2006. IJCNN'06. International Joint Conference on* (pp. 3070-3077). IEEE.
- [Dos Santos 2008] Dos Santos, E. M., Sabourin, R., & Maupin, P. (2008). A dynamic overproduce-and-choose strategy for the selection of classifier ensembles. *Pattern recognition*, 41(10), 2993-3009.
- [Dos Santos 2009] Dos Santos, E. M., Sabourin, R., & Maupin, P. (2009). Overfitting cautious selection of classifier ensembles with genetic algorithms. *Information Fusion*, 10(2), 150-162.
- [Dos Santos 2011] Dos Santos, E. M., & Sabourin, R. (2011, June). Classifier ensembles optimization guided by population oracle. In *Evolutionary Computation (CEC), 2011 IEEE Congress on* (pp. 693-698). IEEE.
- [Douzas 2017] Douzas, G., & Bacao, F. (2017). Self-Organizing Map Oversampling (SOMO) for imbalanced data set learning. *Expert systems with Applications*, 82, 40-52.
- [Duda 1973] Duda, R. O., & Hart, P. E. (1973). Pattern classification and scene analysis. *A Wiley-Interscience Publication, New York: Wiley, 1973*.
- [Duin 2000] Duin, R. P., & Tax, D. M. (2000, June). Experiments with classifier combining rules. In *International Workshop on Multiple Classifier Systems* (pp. 16-29). Springer, Berlin, Heidelberg.
- [Efron 1994] Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

- [El Akadi 2011] El Akadi, A., Amine, A., El Ouardighi, A., & Aboutajdine, D. (2011). A two-stage gene selection scheme utilizing MRMR filter and GA wrapper. *Knowledge and Information Systems*, 26(3), 487-500.
- [El-Atlas 2014] El Atlas, N., El Aroussi, M., & Wahbi, M. (2014, November). Computer-aided breast cancer detection using mammograms: A review. In *Complex Systems (WCCS), 2014 Second World Conference on* (pp. 626-631). IEEE.
- [Fawcett 2006] Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, 27(8), 861-874.
- [Fernández 2018] Fernández, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary. *Journal of Artificial Intelligence Research*, 61, 863-905.
- [Freund 1997] Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1), 119-139.
- [Fuchs 2015] Fuchs, K., Gertheiss, J., & Tutz, G. (2015). Nearest neighbor ensembles for functional data with interpretable feature selection. *Chemometrics and Intelligent Laboratory Systems*, 146, 186-197.
- [Fumero 2011] Fumero, F., Alayón, S., Sanchez, J. L., Sigut, J., & Gonzalez-Hernandez, M. (2011, June). RIM-ONE: An open retinal image database for optic nerve evaluation. In *2011 24th international symposium on computer-based medical systems (CBMS)* (pp. 1-6). IEEE.
- [Gacquer 2008a] Gacquer, D. (2008). Sur l'utilisation active de la diversité dans la construction d'ensembles de classifieurs. Application à la détection de fumées nocives sur site industriel. (Thèse de Doctorat), Université de Valenciennes et du Hainaut-Cambresis, France.
- [Gacquer 2008b] Gacquer, D., Delmotte, F., Delcroix, V., & Piechowiak, S. (2008). A genetic approach for training diverse classifier ensembles. In *Proceedings of IPMU* (Vol. 8, p. 799).
- [Galán 2002] Galán, S. F., Aguado, F., Díez, F. J., & Mira, J. (2002). NasoNet, modeling the spread of nasopharyngeal cancer with networks of probabilistic events in discrete time. *Artificial Intelligence in Medicine*, 25(3), 247-264.
- [Galar 2012] Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., & Herrera, F. (2012). A review on ensembles for the class imbalance problem: bagging-,

- boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(4), 463-484.
- [Galar 2013] Galar, M., Fernández, A., Barrenechea, E., & Herrera, F. (2013). EUSBoost: Enhancing ensembles for highly imbalanced data-sets by evolutionary undersampling. *Pattern Recognition*, 46(12), 3460-3471.
- [García 2012] García, V., Sánchez, J. S., & Mollineda, R. A. (2012). On the effectiveness of preprocessing methods when dealing with different levels of class imbalance. *Knowledge-Based Systems*, 25(1), 13-21.
- [Ghosh 2015] Ghosh, A., Sarkar, A., Ashour, A. S., Balas-Timar, D., Dey, N., & Balas, V. E. (2015). Grid color moment features in glaucoma classification. *Int J Adv Comput Sci Appl*, 6(9), 1-14.
- [García 2015] García, V., Marqués, A. I., & Sánchez, J. S. (2015). An insight into the experimental design for credit risk and corporate bankruptcy prediction systems. *Journal of Intelligent Information Systems*, 44(1), 159-189.
- [Giacinto 1999] Giacinto, G., & Roli, F. (1999). Methods for dynamic classifier selection. In *Image Analysis and Processing, 1999. Proceedings. International Conference on* (pp. 659-664). IEEE.
- [Giacinto 2000] Giacinto, G., Roli, F., & Fumera, G. (2000). Design of effective multiple classifier systems by clustering of classifiers. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on* (Vol. 2, pp. 160-163). IEEE.
- [Giacinto 2001a] Giacinto, G., & Roli, F. (2001). Design of effective neural network ensembles for image classification purposes. *Image and Vision Computing*, 19(9-10), 699-707.
- [Giacinto 2001b] Giacinto, G., & Roli, F. (2001). Dynamic classifier selection based on multiple classifier behaviour. *Pattern Recognition*, 34(9), 1879-1881.
- [Glover 1998] Glover, F., & Laguna, M. (1998). Tabu search. In *Handbook of combinatorial optimization* (pp. 2093-2229). Springer, Boston, MA.
- [Golestani 2007] Golestani, A., Amiri, K. A., & Motlagh, M. J. (2007, July). A novel adaptive-boost-based strategy for combining classifiers using diversity concept. In *Computer and Information Science, 2007. ICIS 2007. 6th IEEE/ACIS International Conference on* (pp. 128-134). IEEE.
- [Gosselin 1997] Gosselin, B. (1997). Cooperation of multilayer perceptron classifiers. *Procs*

- of the 8th Works Circuits Systems and Signal Processing*, 187-190.
- [Gunes 2001] Gunes, V. (2001). Reconnaissance des formes évolutives par combinaison, coopération et sélection de classifieurs. (Thèse de Doctorat), Université de La Rochelle, France.
- [Günter 2004] Günter, S., & Bunke, H. (2004). Optimization of weights in a multiple classifier handwritten word recognition system using a genetic algorithm. *ELCVIA: electronic letters on computer vision and image analysis*, 3(1), 025-44.
- [Guo 2004] Guo, P., & Lyu, M. R. (2004). A pseudoinverse learning algorithm for feedforward neural networks with stacked generalization applications to software reliability growth data. *Neurocomputing*, 56, 101-121.
- [Haleem 2016] Haleem, M. S., Han, L., Van Hemert, J., Fleming, A., Pasquale, L. R., Silva, P. S., & Aiello, L. P. (2016). Regional image features model for automatic classification between normal and glaucoma in fundus and scanning laser ophthalmoscopy (slo) images. *Journal of medical systems*, 40(6), 132.
- [Hamdi 2012] Hamdi, F. (2012). Apprentissage en distributions déséquilibrées (Thèse de Doctorat), Université de Paris 13, Paris, France.
- [Han 2005] Han, H., Wang, W. Y., & Mao, B. H. (2005, August). Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In *International Conference on Intelligent Computing* (pp. 878-887). Springer, Berlin, Heidelberg.
- [Hansen 1990] Hansen, L. K., & Salamon, P. (1990). Neural network ensembles. *IEEE transactions on pattern analysis and machine intelligence*, 12(10), 993-1001.
- [Hansen 1997] Hansen, P., & Jaumard, B. (1997). Cluster analysis and mathematical programming. *Mathematical programming*, 79(1-3), 191-215.
- [He 2008] He, H., & Garcia, E. A. (2008). Learning from imbalanced data. *IEEE Transactions on Knowledge & Data Engineering*, (9), 1263-1284.
- [Hearst 1998] Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4), 18-28.
- [Heath 1998] Heath, M., Bowyer, K., Kopans, D., Kegelmeyer, P., Moore, R., Chang, K., & Munishkumar, S. (1998). Current status of the digital database for screening mammography. In *Digital mammography* (pp. 457-460). Springer,

- Dordrecht.
- [Heath 2000] Heath, M., Bowyer, K., Kopans, D., Moore, R., & Kegelmeyer, W. P. (2000, June). The digital database for screening mammography. In *Proceedings of the 5th international workshop on digital mammography* (pp. 212-218). Medical Physics Publishing.
- [Henriet 2000] Henriët, L. (2000). Système d'évaluation et de classification multicritères pour l'aide à la décision, construction de modèles et procédures d'affectation. (Thèse de doctorat), Université Paris Dauphine, Paris, France.
- [Heutte 1994] Heutte, L. (1994). Reconnaissance de caractères manuscrits : application à la lecture automatique des chèques et des enveloppes postales. (Thèse de Doctorat), Université de Rouen, France.
- [Ho 1992] Ho, T. K. (1992). Theory of Multiple Classifier Systems and Its Application to Visual Word Recognition.
- [Ho 1998] Ho, T. K. (1992). The random subspace method for constructing decision forests. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(8):832–844.
- [Hosseini 2016] Hosseini, M. P., Hajisami, A., & Pompili, D. (2016, July). Real-time epileptic seizure detection from eeg signals via random subspace ensemble learning. In *Autonomic Computing (ICAC), 2016 IEEE International Conference on* (pp. 209-218). IEEE.
- [Hu 1962] Hu, M. K. (1962). Visual pattern recognition by moment invariants. *IRE transactions on information theory*, 8(2), 179-187.
- [Huang 1995] Huang, Y. S., & Suen, C. Y. (1995). A method of combining multiple experts for the recognition of unconstrained handwritten numerals. *IEEE transactions on pattern analysis and machine intelligence*, 17(1), 90-94.
- [Huang 2010] Huang, Z., & Leng, J. (2010, April). Analysis of Hu's moment invariants on image scaling and rotation. In *Computer Engineering and Technology (ICCET), 2010 2nd International Conference on* (Vol. 7, pp. V7-476). IEEE.
- [Huang 2017] Huang, M. W., Chen, C. W., Lin, W. C., Ke, S. W., & Tsai, C. F. (2017). SVM and SVM ensembles in breast cancer prediction. *PLoS one*, 12(1), e0161501.
- [Hull 1983] Hull, J. J., Srihari, S. N., & Choudhari, R. (1983). An integrated algorithm

- for text recognition: comparison with a cascaded algorithm. *IEEE transactions on pattern analysis and machine intelligence*, (4), 384-395.
- [Hull 1988] Hull, J. J. (1988). *A computational theory of visual word recognition* (Doctoral dissertation, State University of New York at Buffalo).
- [Jain 2000] Jain, A. K., Duin, R. P., & Mao, J. (2000). Statistical pattern recognition: A review. *IEEE Transactions on pattern analysis and machine intelligence*, 22(1), 4-37.
- [Kahramanli 2009] Kahramanli, H., & Allahverdi, N. (2009). Extracting rules for classification problems: AIS based approach. *Expert Systems with Applications*, 36(7), 10494-10502.
- [Kittler 1997a] Kittler, J., Matas, J., Jonsson, K., & Sánchez, M. R. (1997). Combining evidence in personal identity verification systems. *Pattern Recognition Letters*, 18(9), 845-852.
- [Kittler 1997b] Kittler, J., Hojjatoleslami, S. A., & Windeatt, T. (1997, September). Weighting Factors in Multiple Expert Fusion. In *BMVC*.
- [Kittler 1998] Kittler, J. (1998). Combining classifiers: A theoretical framework. *Pattern analysis and Applications*, 1(1), 18-27.
- [Ko 2008] Ko, A. H., Sabourin, R., & Britto Jr, A. S. (2008). From dynamic classifier selection to dynamic ensemble selection. *Pattern Recognition*, 41(5), 1718-1731.
- [Kolen 1991] Kolen, J. F., & Pollack, J. B. (1991). Back propagation is sensitive to initial conditions. In *Advances in neural information processing systems* (pp. 860-867).
- [Krawczyk 2016] Krawczyk, B. (2016). Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence*, 5(4), 221-232.
- [Krogh 1995] Krogh, A., & Vedelsby, J. (1995). Neural network ensembles, cross validation, and active learning. In *Advances in neural information processing systems* (pp. 231-238).
- [Kuncheva 2000] Kuncheva, L. I. (2000). Clustering-and-selection model for classifier combination. In *Knowledge-Based Intelligent Engineering Systems and Allied Technologies, 2000. Proceedings. Fourth International Conference on* (Vol. 1, pp. 185-188). IEEE.

- [Kuncheva 2001] Kuncheva, L. I., & Whitaker, C. J. (2001, February). Ten measures of diversity in classifier ensembles: limits for two classifiers. In *Intelligent Sensor Processing (Ref. No. 2001/050), A DERA/IEE Workshop on* (pp. 10-1). IET.
- [Kuncheva 2002a] Kuncheva, L. I. (2002). Switching between selection and fusion in combining classifiers: An experiment. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 32(2), 146-156.
- [Kuncheva 2002b] Kuncheva, L. I., & Kountchev, R. K. (2002). Generating classifier outputs of fixed accuracy and diversity. *Pattern recognition letters*, 23(5), 593-600.
- [Kuncheva 2002c] Kuncheva, L. I. (2002). A theoretical study on six classifier fusion strategies. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (2), 281-286.
- [Kuncheva 2003a] Kuncheva, L. I., & Whitaker, C. J. (2003). Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine learning*, 51(2), 181-207.
- [Kuncheva 2003b] Kuncheva, L. I. (2003, June). That elusive diversity in classifier ensembles. In *Iberian conference on pattern recognition and image analysis* (pp. 1126-1138). Springer, Berlin, Heidelberg.
- [Kuncheva 2004] Kuncheva, L. (2004). Ludmila Kuncheva Collection. URL< <http://www.bangor.ac.uk/~mas00a/activities/patrec1.html>.
- [Kuncheva 2007] Kuncheva, L. I. (2004). Combining pattern classifiers: methods and algorithms. John Wiley & Sons.
- [Kurzynski 2010] Kurzynski, M., Woloszynski, T., & Lysiak, R. (2010, October). On two measures of classifier competence for dynamic ensemble selection-experimental comparative analysis. In *Communications and Information Technologies (ISCIT), 2010 International Symposium on* (pp. 1108-1113). IEEE.
- [Kwok 1998] Kwok, S. W., & Carter, C. (1990). Multiple decision trees. In *Machine Intelligence and Pattern Recognition* (Vol. 9, pp. 327-335). North-Holland.
- [Lam 1994] Lam, L., & Suen, C. Y. (1994, October). A theoretical analysis of the application of majority voting to pattern recognition. In *Pattern Recognition, 1994. Vol. 2-Conference B: Computer Vision & Image Processing*,

- Proceedings of the 12th IAPR International. Conference on* (Vol. 2, pp. 418-420). IEEE.
- [Lam 1997] Lam, L., & Suen, S. Y. (1997). Application of majority voting to pattern recognition: an analysis of its behavior and performance. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 27(5), 553-568.
- [Lanyi 1985] Lanyi, M. (1985). Morphologic analysis of microcalcifications. In *Early Breast Cancer* (pp. 113-135). Springer, Berlin, Heidelberg.
- [Lazarevic 2001] Lazarevic, A., & Obradovic, Z. (2001). Effective pruning of neural network classifier ensembles. In *Neural Networks, 2001. Proceedings. IJCNN'01. International Joint Conference on* (Vol. 2, pp. 796-801). IEEE.
- [Lee 1995] Lee, D. S., & Srihari, S. N. (1995, August). A theory of classifier combination: the neural network approach. In *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on* (Vol. 1, pp. 42-45). IEEE.
- [Li 2002] Li, L., Clark, R. A., & Thomas, J. A. (2002). Computer-aided diagnosis of masses with full-field digital mammography. *Academic Radiology*, 9(1), 4-12.
- [Li 2012a] Li, B. Q., Hu, L. L., Niu, S., Cai, Y. D., & Chou, K. C. (2012). Predict and analyze S-nitrosylation modification sites with the mRMR and IFS approaches. *Journal of Proteomics*, 75(5), 1654-1665.
- [Li 2012b] Li, N., Yu, Y., & Zhou, Z. H. (2012, September). Diversity regularized ensemble pruning. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 330-345). Springer, Berlin, Heidelberg.
- [Lin 2003] Lin, X., Yacoub, S., Burns, J., & Simske, S. (2003). Performance analysis of pattern classifier combination by plurality voting. *Pattern Recognition Letters*, 24(12), 1959-1969.
- [Lin 2014] Lin, C., Chen, W., Qiu, C., Wu, Y., Krishnan, S., & Zou, Q. (2014). LibD3C: ensemble classifiers with a clustering and dynamic selection strategy. *Neurocomputing*, 123, 424-435.
- [Liu 2001] Lui, H., & Motoda, H. (2001). Instance selection and construction for data mining. *Kluwer Academic Publishers*.

- [Liu 2008] Liu, F. T., Ting, K. M., Yu, Y., & Zhou, Z. H. (2008). Spectrum of variable-random trees. *Journal of Artificial Intelligence Research*, 32, 355-384.
- [Liu 2009] Liu, X. Y., Wu, J., & Zhou, Z. H. (2009). Exploratory undersampling for class-imbalance learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(2), 539-550.
- [Liu 2013] Liu, G. H., & Yang, J. Y. (2013). Content-based image retrieval using color difference histogram. *Pattern recognition*, 46(1), 188-198.
- [Loughrey 2004] Loughrey, J., & Cunningham, P. (2005). Overfitting in wrapper-based feature subset selection: The harder you try the worse it gets. In *Research and development in intelligent systems XXI* (pp. 33-43). Springer, London.
- [Luukka 2011] Luukka, P. (2011). Feature selection using fuzzy entropy measures with similarity classifier. *Expert Systems with Applications*, 38(4), 4600-4607.
- [Maldonado 2014] Maldonado, S., Weber, R., & Famili, F. (2014). Feature selection for high-dimensional class-imbalanced data sets using Support Vector Machines. *Information Sciences*, 286, 228-246.
- [Mao 2011] Mao, S., Jiao, L. C., Xiong, L., & Gou, S. (2011). Greedy optimization classifiers ensemble based on diversity. *Pattern Recognition*, 44(6), 1245-1261.
- [Margineantu 1997] Margineantu, D. D., & Dietterich, T. G. (1997, July). Pruning adaptive boosting. In *ICML* (Vol. 97, pp. 211-218).
- [Markowska-Kaczmar 2006] Markowska-Kaczmar, U., & Matkowski, R. (2006, July). Experimental study of evolutionary based method of rule extraction from neural networks in medical data. In *Industrial Conference on Data Mining* (pp. 76-90). Springer, Berlin, Heidelberg.
- [Menardi 2014] Menardi, G., & Torelli, N. (2014). Training and assessing classification rules with imbalanced data. *Data Mining and Knowledge Discovery*, 28(1), 92-122.
- [Mendialdua 2015] Mendialdua, I., Arruti, A., Jauregi, E., Lazkano, E., & Sierra, B. (2015). Classifier subset selection to construct multi-classifiers by means of estimation of distribution algorithms. *Neurocomputing*, 157, 46-60.
- [Mitchell 1997] Mitchell, T. M. (1999). Machine learning and data mining. *Communications of the ACM*, 42(11), 30-36.

- [Mookiah 2012] Mookiah, M. R. K., Acharya, U. R., Lim, C. M., Petznick, A., & Suri, J. S. (2012). Data mining technique for automated diagnosis of glaucoma using higher order spectra and wavelet energy features. *Knowledge-Based Systems*, 33, 73-82.
- [Moreno-Seco 2006] Moreno-Seco, F., Inesta, J. M., De León, P. J. P., & Micó, L. (2006, August). Comparison of classifier fusion methods for classification in pattern recognition tasks. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)* (pp. 705-713). Springer, Berlin, Heidelberg.
- [Nanni 2015] Nanni, L., Fantozzi, C., & Lazzarini, N. (2015). Coupling different methods for overcoming the class imbalance problem. *Neurocomputing*, 158, 48-61.
- [Nazabal 2016] Nazabal, A., Garcia-Moreno, P., Artes-Rodriguez, A., & Ghahramani, Z. (2016). Human activity recognition by combining a small number of classifiers. *IEEE journal of biomedical and health informatics*, 20(5), 1342-1351.
- [Nazaré-Silva 2015] de Nazaré Silva, J., de Carvalho Filho, A. O., Silva, A. C., De Paiva, A. C., & Gattass, M. (2015). Automatic detection of masses in mammograms using quality threshold clustering, correlogram function, and SVM. *Journal of digital imaging*, 28(3), 323-337.
- [Oliveira 2005] Oliveira, L. S., Morita, M., Sabourin, R., & Bortolozzi, F. (2005, March). Multi-objective genetic algorithms to create ensemble of classifiers. In *International Conference on Evolutionary Multi-Criterion Optimization* (pp. 592-606). Springer, Berlin, Heidelberg.
- [Oliveira 2017] Oliveira, D. V., Cavalcanti, G. D., & Sabourin, R. (2017). Online pruning of base classifiers for dynamic ensemble selection. *Pattern Recognition*, 72, 44-58.
- [Oltean 2009] Oltean, M., & Dioşan, L. (2009). An autonomous GP-based system for regression and classification problems. *Applied Soft Computing*, 9(1), 49-60.
- [Opitz 1996] Opitz, D. W., & Shavlik, J. W. (1996). Actively searching for an effective neural network ensemble. *Connection Science*, 8(3-4), 337-354.
- [Paninski 2003] Paninski, L. (2003). Estimation of entropy and mutual information. *Neural computation*, 15(6), 1191-1253.

- [Partalas 2008] Partalas, I., Tsoumakas, G., Hatzikos, E. V., & Vlahavas, I. (2008). Greedy regression ensemble selection: Theory and an application to water quality prediction. *Information Sciences*, 178(20), 3867-3879.
- [Partridge 1996] Partridge, D., & Yates, W. B. (1996). Engineering multiversion neural-net systems. *Neural Computation*, 8(4), 869-893.
- [Partridge 1997] Partridge, D., & Krzanowski, W. (1997). Software diversity: practical statistics for its measurement and exploitation. *Information and software technology*, 39(10), 707-717.
- [Peng 2005] Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*, 27(8), 1226-1238.
- [Pham 2017] Pham, B. T., Bui, D. T., Prakash, I., & Dholakia, M. B. (2017). Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *Catena*, 149, 52-63.
- [Philipose 2018] Philipose, M., Shen, H., Wolman, A., & Agarwal, S. (2018). *U.S. Patent Application No. 15/958,943*.
- [Prabusankarlal 2015] Prabusankarlal, K. M., Thirumoorthy, P., & Manavalan, R. (2015). Assessment of combined textural and morphological features for diagnosis of breast masses in ultrasound. *Human-centric Computing and Information Sciences*, 5(1), 12.
- [Prati 2015] Prati, R. C., Batista, G. E., & Silva, D. F. (2015). Class imbalance revisited: a new experimental setup to assess the performance of treatment methods. *Knowledge and Information Systems*, 45(1), 247-270.
- [Quigley 1986] Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1), 81-106.
- [Quigley 2006] Quigley, H. A., & Broman, A. T. (2006). The number of people with glaucoma worldwide in 2010 and 2020. *British journal of ophthalmology*, 90(3), 262-267.
- [Quinlan 1993] Quinlan, J. R. (2014). C4. 5: programs for machine learning. Elsevier.
- [Radová 1997] Radová, V., & Psutka, J. (1997, April). An approach to speaker identification using multiple classifiers. In *Acoustics, Speech, and Signal*

- Processing, 1997. ICASSP-97., 1997 IEEE International Conference on* (Vol. 2, pp. 1135-1138). IEEE.
- [Rahman 2003] Rahman, A. F. R., & Fairhurst, M. C. (2003). Multiple classifier decision combination strategies for character recognition: A review. *Document Analysis and Recognition*, 5(4), 166-194.
- [Raja 2015] Raja, C., & Gangatharan, N. (2015). A Hybrid Swarm Algorithm for optimizing glaucoma diagnosis. *Computers in biology and medicine*, 63, 196-207.
- [Rao 2012] Rao, V. S. H., & Kumar, M. N. (2012). A new intelligence-based approach for computer-aided diagnosis of dengue fever. *IEEE transactions on information technology in biomedicine*, 16(1), 112-118.
- [Rastrigin 1981] Rastrigin, L. A., & Erenstein, R. H. (1981). Method of collective recognition. *Energoizdat, Moscow*, 595.
- [Reunanen 2003] Reunanen, J. (2003). Overfitting in making comparisons between variable selection methods. *Journal of Machine Learning Research*, 3(Mar), 1371-1382.
- [Ripley 2007] Ripley, B. D. (2007). *Pattern recognition and neural networks*. Cambridge university press.
- [Ritschard 2007] Ritschard, G., Zighed, D. A., & Marcellin, S. (2007). Données déséquilibrées, entropie décentrée et indice d'implication.
- [Rodriguez 2006] Rodriguez, J. J., Kuncheva, L. I., & Alonso, C. J. (2006). Rotation forest: A new classifier ensemble method. *IEEE transactions on pattern analysis and machine intelligence*, 28(10), 1619-1630.
- [Roli 2001] Roli, F., Giacinto, G., & Verzasca, G. (2001, July). Methods for designing multiple classifier systems. In *International Workshop on Multiple Classifier Systems* (pp. 78-87). Springer, Berlin, Heidelberg.
- [Roller 2013] Roller, R., & Stevenson, M. (2013). Identification of genia events using multiple classifiers. In *Proceedings of the BioNLP Shared Task 2013 Workshop* (pp. 125-129).
- [Roy 2018] Roy, A., Cruz, R. M., Sabourin, R., & Cavalcanti, G. D. (2018). A study on combining dynamic selection and data preprocessing for imbalance learning. *Neurocomputing*, 286, 179-192.

- [Ruck 1990] Ruck, D. W., Rogers, S. K., Kabrisky, M., Oxley, M. E., & Suter, B. W. (1990). The multilayer perceptron as an approximation to a Bayes optimal discriminant function. *IEEE Transactions on Neural Networks*, 1(4), 296-298.
- [Ruta 2000] Ruta, D., & Gabrys, B. (2000). An overview of classifier fusion methods. *Computing and Information systems*, 7(1), 1-10.
- [Ruta 2001a] Ruta, D., & Gabrys, B. (2001). Analysis of the correlation between majority voting error and the diversity measures in multiple classifier systems.
- [Ruta 2001b] Ruta, D., & Gabrys, B. (2001). Application of the evolutionary algorithms for classifier selection in multiple classifier systems with majority voting. In *International Workshop on Multiple Classifier Systems* (pp. 399-408). Springer, Berlin, Heidelberg.
- [Ruta 2005] Ruta, D., & Gabrys, B. (2005). Classifier selection for majority voting. *Information fusion*, 6(1), 63-81.
- [Sabourin 1993] Sabourin, M., Mitiche, A., Thomas, D., & Nagy, G. (1993, October). Classifier combination for hand-printed digit recognition. In *Document Analysis and Recognition, 1993., Proceedings of the Second International Conference on* (pp. 163-166). IEEE.
- [Sahiner 2001] Sahiner, B., Chan, H. P., Petrick, N., Helvie, M. A., & Hadjiiski, L. M. (2001). Improvement of mammographic mass characterization using spiculation measures and morphological features. *Medical physics*, 28(7), 1455-1465.
- [Santos 2006] Dos Santos, E. M., Sabourin, R., & Maupin, P. (2006, July). Single and multi-objective genetic algorithms for the selection of ensemble of classifiers. In *Neural Networks, 2006. IJCNN'06. International Joint Conference on* (pp. 3070-3077). IEEE.
- [Seiffert 2010] Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J., & Napolitano, A. (2010). RUSBoost: A hybrid approach to alleviating class imbalance. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 40(1), 185-197.
- [Sekar 2012] Sekar, B. D., Chui Dong, M., Shi, J., & Hu, X. Y. (2012). Fused hierarchical neural networks for cardiovascular disease diagnosis. *IEEE Sensors Journal*, 12(3), 644-650.

- [Shapire 1998] Schapire, R. E., Singer, Y., & Singhal, A. (1998, August). Boosting and Rocchio applied to text filtering. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 215-223). ACM.
- [Sharkey 1997] Sharkey, A. J., & Sharkey, N. E. (1997). Combining diverse neural nets. *The Knowledge Engineering Review*, 12(3), 231-247.
- [Sharkey 2000] Sharkey, A. J., Sharkey, N. E., Gerecke, U., & Chandroth, G. O. (2000, June). The “test and select” approach to ensemble combination. In *International Workshop on Multiple Classifier Systems* (pp. 30-44). Springer, Berlin, Heidelberg.
- [Sharma 2001] Sharma, M., & Singh, S. (2001, November). Evaluation of texture methods for image analysis. In *Intelligent Information Systems Conference, The Seventh Australian and New Zealand 2001* (pp. 117-121). IEEE.
- [Shipp 2002] Shipp, C. A., & Kuncheva, L. I. (2002). Relationships between combination methods and measures of diversity in combining classifiers. *Information fusion*, 3(2), 135-148.
- [Singh 2006] Singh, R., Vatsa, M., Noore, A., & Singh, S. K. (2006). Dempster-Shafer theory based classifier fusion for improved fingerprint verification performance. In *Computer vision, graphics and image processing* (pp. 941-949). Springer, Berlin, Heidelberg.
- [Singh 2016] Singh, S. P., & Urooj, S. (2016). An improved CAD system for breast cancer diagnosis based on generalized pseudo-Zernike moment and Ada-DEWNN classifier. *Journal of medical systems*, 40(4), 105.
- [Skalak 1996] Skalak, D. B. (1996, August). The sources of increased accuracy for two proposed boosting algorithms. In *Proc. American Association for Artificial Intelligence, AAAI-96, Integrating Multiple Learned Models Workshop* (Vol. 1129, p. 1133).
- [Smits 2002] Smits, P. C. (2002). Multiple classifier systems for supervised remote sensing image classification based on dynamic classifier selection. *IEEE Transactions on Geoscience and Remote Sensing*, 40(4), 801-813.
- [Soares 2006] Soares, R. G., Santana, A., Canuto, A. M., & de Souto, M. C. P. (2006, July). Using accuracy and diversity to select classifiers to build ensembles. In *Neural Networks, 2006. IJCNN'06. International Joint Conference on* (pp.

- 1310-1316). IEEE.
- [Sonka 1999] Sonka, M., Hlavac, V., & Boyle, R. (2014). *Image processing, analysis, and machine vision*. Cengage Learning.
- [Souza 2018] Souza, M. A., Cavalcanti, G. D., Cruz, R. M., & Sabourin, R. (2018). Online local pool generation for dynamic classifier selection: an extended version. *arXiv preprint arXiv:1809.01628*.
- [Srihari 1982] Srihari, S. N. (1982). Reliability analysis of majority vote systems. *Information Sciences*, 26(3), 243-256.
- [Stefano 2002] Stefano, C., Della Cioppa, A., & Marcelli, A. (2002, August). An adaptive weighted majority vote rule for combining multiple classifiers. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on* (Vol. 2, pp. 192-195). IEEE.
- [Stoean 2013] Stoean, R., & Stoean, C. (2013). Modeling medical decision making by support vector machines, explaining by rules of evolutionary algorithms with feature selection. *Expert Systems with Applications*, 40(7), 2677-2686.
- [Stork 2015] Stork, J., Ramos, R., Koch, P., & Konen, W. (2015). SVM ensembles are better when different kernel types are combined. In *Data Science, Learning by Latent Structures, and Knowledge Discovery* (pp. 191-201). Springer, Berlin, Heidelberg.
- [Suen 2000] Suen, C. Y., & Lam, L. (2000, June). Multiple classifier combination methodologies for different output levels. In *International workshop on multiple classifier systems* (pp. 52-66). Springer, Berlin, Heidelberg.
- [Sun 2017] Sun, X., Liu, L., Wang, Z., Miao, J., Wang, Y., Luo, Z., & Li, G. (2017). An optimized multi-classifiers ensemble learning for identification of ginsengs based on electronic nose. *Sensors and Actuators A: Physical*, 266, 135-144.
- [Sundararaj 2014] Sundararaj, G. K., & Balamurugan, V. (2014, November). An expert system based on texture features and decision tree classifier for diagnosis of tumor in brain MR images. In *Contemporary Computing and Informatics (IC3I), 2014 International Conference on* (pp. 1340-1344). IEEE.
- [Tax 2000] Tax, D. M., Van Breukelen, M., Duin, R. P., & Kittler, J. (2000). Combining multiple classifiers by averaging or by multiplying?. *Pattern recognition*, 33(9), 1475-1485.

- [Teoh 2004] Teoh, A., Samad, S. A., & Hussain, A. (2004). Nearest neighbourhood classifiers in a bimodal biometric verification system fusion decision scheme. *Journal of Research and Practice in Information Technology*, 36(1), 47-62.
- [Thammasiri 2014] Thammasiri, D., Delen, D., Meesad, P., & Kasap, N. (2014). A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition. *Expert Systems with Applications*, 41(2), 321-330.
- [Tremblay 2004] Tremblay, G., Sabourin, R., & Maupin, P. (2004, August). Optimizing nearest neighbour in random subspaces using a multi-objective genetic algorithm. In *null* (p. 208). IEEE.
- [Tsymbal 2005] Tsymbal, A., Pechenizkiy, M., & Cunningham, P. (2005). Diversity in search strategies for ensemble feature selection. *Information fusion*, 6(1), 83-98.
- [Tulyakov 2008] Tulyakov, S., Jaeger, S., Govindaraju, V., & Doermann, D. (2008). Review of classifier combination methods. In *Machine learning in document analysis and recognition* (pp. 361-386). Springer, Berlin, Heidelberg.
- [Vapnik 1998] Vapnik, V. (1998). *Statistical learning theory. 1998* (Vol. 3). Wiley, New York.
- [Verbiest 2016] Verbiest, N., Vluymans, S., Cornelis, C., García-Pedrajas, N., & Saeys, Y. (2016). Improving nearest neighbor classification using ensembles of evolutionary generated prototype subsets. *Applied Soft Computing*, 44, 75-88.
- [Veropoulos 1991] Veropoulos, K., Campbell, C., & Cristianini, N. (1999, July). Controlling the sensitivity of support vector machines. In *Proceedings of the international joint conference on AI* (Vol. 55, p. 60).
- [Vincent 2003] Vincent, P. (2003). Modèles à noyaux à structure locale. (Thèse de Doctorat), Université de Montréal, Canada
- [Vincent 2011] Vincent, P. N. & Cloppet, F. (2011). Sélection de caractéristiques : méthodes et applications. (Thèse de doctorat), université de Paris Descartes, Paris, France.
- [Visentini 2016] Visentini, I., Snidaro, L., & Foresti, G. L. (2016). Diversity-aware classifier ensemble selection via f-score. *Information Fusion*, 28, 24-43.

- [Wang 2006] Wang, X., & Wang, H. (2006). Classification by evolutionary ensembles. *Pattern Recognition*, 39(4), 595-607.
- [Wang 2009] Wang, S., & Yao, X. (2009, March). Diversity analysis on imbalanced data sets by using ensemble models. In *Computational Intelligence and Data Mining, 2009. CIDM'09. IEEE Symposium on* (pp. 324-331). IEEE.
- [Wang 2017] Wang, Q., Luo, Z., Huang, J., Feng, Y., & Liu, Z. (2017). A novel ensemble method for imbalanced data learning: bagging of extrapolation-SMOTE SVM. *Computational intelligence and neuroscience*, 2017.
- [Wilson 1972] Wilson, D. L. (1972). Asymptotic properties of nearest neighbor rules using edited data. *IEEE Transactions on Systems, Man, and Cybernetics*, (3), 408-421.
- [Woloszynski 2009] Woloszynski, T., & Kurzynski, M. (2009, September). On a new measure of classifier competence applied to the design of multiclassifier systems. In *International Conference on Image Analysis and Processing* (pp. 995-1004). Springer, Berlin, Heidelberg.
- [Woloszynski 2011] Woloszynski, T., & Kurzynski, M. (2011). A probabilistic model of classifier competence for dynamic ensemble selection. *Pattern Recognition*, 44(10-11), 2656-2668.
- [Woloszynski 2012] Woloszynski, T., Kurzynski, M., Podsiadlo, P., & Stachowiak, G. W. (2012). A measure of competence based on random classification for dynamic ensemble selection. *Information Fusion*, 13(3), 207-213.
- [Wolpert 1992] Wolpert, D. H. (1992). Stacked generalization. *Neural networks*, 5(2), 241-259.
- [Woods 1997] Woods, K., Kegelmeyer, W. P., & Bowyer, K. (1997). Combination of multiple classifiers using local accuracy estimates. *IEEE transactions on pattern analysis and machine intelligence*, 19(4), 405-410.
- [Xia 2015] Xia, J., Dalla Mura, M., Chanussot, J., Du, P., & He, X. (2015). Random subspace ensembles for hyperspectral image classification with extended morphological attribute profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 53(9), 4768-4786.
- [Xiao 2012] Xiao, J., Xie, L., He, C., & Jiang, X. (2012). Dynamic classifier ensemble model for customer classification with imbalanced class distribution. *Expert Systems with Applications*, 39(3), 3668-3675.

- [Xu 1992] Xu, L., Krzyzak, A., & Suen, C. Y. (1992). Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE transactions on systems, man, and cybernetics*, 22(3), 418-435.
- [Yang 2001] Yang, Y., & Webb, G. I. (2001, September). Proportional k-interval discretization for naive-Bayes classifiers. In *European Conference on Machine Learning* (pp. 564-575). Springer, Berlin, Heidelberg.
- [Yang 2011] Yang, L. (2011). Classifiers selection for ensemble learning based on accuracy and diversity. *Procedia Engineering*, 15, 4266-4270.
- [Yule 1900] Yule, G. U. (1900). VII. On the association of attributes in statistics: with illustrations from the material of the childhood society, &c. *Phil. Trans. R. Soc. Lond. A*, 194(252-261), 257-319.
- [Zemmal 2018] Zemmal, N. (2018). Techniques d'apprentissage pour la sélection d'attributs : Application à la reconnaissance des formes. (Thèse de Doctorat), Université Badji Mokhtar, Annaba, Algérie.
- [Zenobi 2001] Zenobi, G., & Cunningham, P. (2001, September). Using diversity in preparing ensembles of classifiers based on different feature subsets to minimize generalization error. In *European Conference on Machine Learning* (pp. 576-587). Springer, Berlin, Heidelberg.
- [Zhang 2001] Zhang, J., & Tan, T. (2002). Brief review of invariant texture analysis methods. *Pattern recognition*, 35(3), 735-747.
- [Zhang 2003] Zhang, S., Zhang, C., & Yang, Q. (2003). Data preparation for data mining. *Applied artificial intelligence*, 17(5-6), 375-381.
- [Zhang 2012] Zhang, L., Xiang, F., Pu, J., & Zhang, Z. (2012, August). Application of improved HU moments in object recognition. In *Automation and Logistics (ICAL), 2012 IEEE International Conference on* (pp. 554-558). IEEE.
- [Zhang 2015] Zhang, Y., Lo, D., Xia, X., & Sun, J. (2015, July). An empirical study of classifier combination for cross-project defect prediction. In *Computer Software and Applications Conference (COMPSAC), 2015 IEEE 39th Annual* (Vol. 2, pp. 264-269). IEEE.
- [Zhou 2002] Zhou, Z. H., Wu, J., & Tang, W. (2002). Ensembling neural networks: many could be better than all. *Artificial intelligence*, 137(1-2), 239-263.
- [Zhou 2018] Zhou, S., Qian, S., Chang, W., Xiao, Y., & Cheng, Y. (2018). A novel bearing multi-fault diagnosis approach based on weighted permutation

- entropy and an improved SVM ensemble classifier. *Sensors*, 18(6), 1934.
- [Zhu 2004] Zhu, X., Wu, X., & Yang, Y. (2004, November). Dynamic classifier selection for effective mining from noisy data streams. In *Data Mining, 2004. ICDM'04. Fourth IEEE International Conference on* (pp. 305-312). IEEE.
- [Zilly 2017] Zilly, J., Buhmann, J. M., & Mahapatra, D. (2017). Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation. *Computerized Medical Imaging and Graphics*, 55, 28-41.
- [Zouari 2004] Zouari, H. K. (2004). Contribution à l'évaluation des méthodes de combinaison parallèle de classifieurs par simulation. (Thèse de Doctorat), Université de Rouen, France.