UNIVERSITÉ DE REIMS CHAMPAGNE-ARDENNE

ÉCOLE DOCTORALE SCIENCES DU NUMERIQUE ET DE L'INGÉNIEUR

THÈSE EN CO-TUTELLE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE MOHAMMED PREMIER OUJDA

Discipline : ***PHYSIQUE ET INGENIERIE***

Spécialité : ***TRAITEMENT DU SIGNAL ET DE L'IMAGE***

Et

DOCTEUR DE L'UNIVERSITÉ DE REIMS CHAMPAGNE-ARDENNE

Discipline : ***AUTOMATIQUE, SIGNAL, PRODUCTIQUE, ROBOTIQUE***

Spécialité : TRAITEMENT DU SIGNAL ET DE L'IMAGE

Présentée et soutenue publiquement par

YOUSSEF ECH-CHOUDANY

Le 8 décembre 2018

ANALYSE DES SIGNAUX NON-STATIONNAIRES A L'AIDE D'UNE NOUVELLE DEMARCHE DE CLASSIFICATION : APPLICATION A L'IDENTIFICATION DE L'ENDOMMAGEMENT DE MATERIAUX COMPOSITES PAR EMISSION ACOUSTIQUE ET A LA DETECTION DE LA CRISE D'EPILEPSIE PAR EEG

Thèse dirigée par M. **Benaissa BELLACH,** Professeur des universités**, Université Mohammed Premier, Oujda, Maroc**

Et par M. **Frédéric MORAIN-NICOLIER,** Professeur des universités**, Université de Reims Champagne-Ardenne, France**

JURY

M. Mohammed EL HASSOUNI,	Professeur des universités,	Université Mohamed V, Rabat (Maroc),	**Président**
M. Pierre BEAUSEROY**,**	Professeur des universités,	Université de Technologie de Troyes (France),	**Rapporteur**
Mme Yamina KHLIFI,	Maître de Conférences HDR,	Université Mohammed Premier, Oujda (Maroc),	**Examinateur**
M. Hamid TAIRI,	Professeur des universités,	Université Sidi Mohamed Ben Abdellah, Fès (Maroc),	**Examinateur**
M. Benaissa BELLACH,	Professeur des universités,	Université Mohammed Premier, Oujda (Maroc),	**Co-Directeur**
M. Frédéric MORAIN-NICOLIER,	Professeur des universités,	Université de Reims-Champagne-Ardenne (France),	**Co-Directeur**
M. Mustapha ASSARAR,	Maître de Conférences HDR,	Université de Reims-Champagne-Ardenne (France),	**Co-Encadrant**
M. Daniel SCIDA,	Maître de Conférences,	Université de Reims Champagne-Ardenne (France),	**Co-Encadrant**

Thèse en co-tutelle

Analyse des signaux non-stationnaires à l'aide d'une nouvelle démarche de classification : application à l'identification de l'endommagement de matériaux composites par émission acoustique et à la détection de la crise d'épilepsie par EEG

Présentée par :	**Youssef ECH-CHOUDANY**[1,3]
Co-Directeur de thèse en France :	**Frédéric MORAIN-NICOLIER**[1]
Co-Directeur de thèse au Maroc :	**Benaissa BELLACH**[3]
Co-encadrants de thèse :	**Mustapha ASSARAR**[2] et **Daniel SCIDA**[2]

1. **C**entre de **R**echerche en **S**cience et **T**echnologie de l'**I**nformation et de la **C**ommunication (**CReSTIC**), Troyes, France.
2. **L**aboratoire d'**I**ngénierie et **S**ciences des **M**atériaux (**LISM**), Troyes, France.
3. **L**aboratoire des **S**ystèmes **É**lectroniques, **I**nformatiques et **I**mages (**LSE2I**), Oujda, Maroc.

Maroc

France

Remerciement

Tout d'abord, je tiens à remercier très sincèrement mes co-directeurs de thèse M. Frédéric MORAIN-NICOLIER professeur à l'Université de Reims Champagne-Ardenne (URCA) et M. Benaissa BELLACH professeur à l'Université Mohammed Premier (UMP) d'Oujda, pour avoir dirigé ce travail de recherche, pour m'avoir intégré dans leurs équipes de travail et m'avoir fait confiance tout au long de ces années.

Mes remerciements et ma sincère reconnaissance s'adressent à mes co-encadrants de thèse, M. Mustapha ASSARAR maître de conférences HDR à l'URCA et M. Daniel SCIDA maître de conférences à l'URCA, pour leurs précieux conseils et leurs aides durant toute la période du travail.

En témoignage de mon profond respect, je remercie M. Pierre BEAUSEROY professeur à l'Université de Technologie de Troyes, M. Mohammed EL HASSOUNI professeur à l'Université Mohammed V de Rabat et M. Hamid TAIRI professeur à l'Université Sidi Mohamed Ben Abdellah de Fès, de m'avoir fait l'honneur d'accepter de juger ce travail en tant que rapporteurs. Je remercier très vivement Mme. Yamina KHILIFI Maître de conférences HDR à l'UMP d'Oujda, d'avoir accepté d'être l'examinateur de cette thèse.

Je remercier très chaleureusement l'ex région française Champagne-Ardenne et le ministère de l'enseignement supérieur et de la recherche scientifique marocaine pour le co-financement de ces travaux de recherche.

J'ai effectué une partie de mon travail de thèse au sein du LISM de à l'IUT de Troyes où je remercie mes collègues Khouloud CHEOUR et El Hadi SAIDANE pour leurs aides précieuses dans la partie élaboration.

Bien sûr, toute ma reconnaissance va vers cinq grands amis que j'ai eu la chance de rencontrer au cours de la thèse : Abderrazak CHILALI, Mohamed Amine HARAOUBIA, Hoai Phuong NGUYEN, Cuong Hung TRAN et Duy An PHAM.

J'adresse également mes remerciements à mes amis de toujours, Moussa EL YAHYAOUI, Tarik ZARROUK, Mohammed amine AZZA et Hachim AZZAHHAFI pour leurs soutiens et leurs encouragements.

Je remercier grandement Mlle Kaoutar BAIBAI, pour l'aide qu'elle m'a apporté et pour ses encouragements.

Je remercie avec un grand plaisir et beaucoup de gratitude mes parents, mes sœurs, mes frères et toute ma famille, pour leur soutien constant et essentiel durant toutes mes années d'études, pour leur grande patience,

leur sacrifice, leur confiance absolue et tout leur amour.

Enfin, durant mon itinérance, entre l'université de Reims Champagne-Ardenne, l'université de technologie de Troyes et l'université Mohammed Premier d'Oujda, j'ai croisé beaucoup de gens et je ne saurai les citer ici. J'ai toujours été bien accueilli, je remercie l'ensemble de ces personnes.

Table des matières

Table des figures **XI**

Liste des tableaux **XV**

Introduction générale **1**

I Discrimination des mécanismes d'endommagement de matériaux composites à fibres végétales par une classification de signaux d'émissions acoustiques **7**

1 Introduction **9**

1.1 Généralité 9

1.2 Approche de classification 11

1.3 Objectif de la première partie 11

1.4 Organisation de la première partie 12

2 Matériaux composites à fibres végétales et émission acoustique : état de l'art et problématiques **17**

2.1 Introduction 18

2.2 Les matériaux composites à fibres naturelles 18

2.2.1 Les fibres naturelles 18

2.2.2 Les matériaux composites à fibres de lin 20

2.3 Mécanismes d'endommagement de matériaux composites fibreux 21

2.3.1 Les mécanismes d'endommagement intra-pli 22

2.3.2 Les mécanismes d'endommagements inter-pli 23

2.4 Identification de mécanismes d'endommagements à partir des signaux d'émission acoustique . 24

2.4.1 Présentation de la technique d'émission acoustique . . . 24
2.4.2 Identification des mécanismes d'endommagement par émission acoustique : état de l'art 26
2.5 Problématiques . . . 35
2.6 Conclusion . . . 37

3 Analyse de l'endommagement des composites par une classification supervisée 39
3.1 Introduction . . . 40
3.2 Matériaux et procédure expérimentale . . . 41
3.2.1 Constituants . . . 41
3.2.2 Mise en œuvre des matériaux composites . . . 42
3.2.3 Essai mécanique et procédure d'acquisition des signaux d'EA . . . 43
3.3 Démarche de classification . . . 46
3.3.1 Méthode incrémentale . . . 47
3.3.2 Support Vector Data Description . . . 50
3.3.3 Méthode de k-moyennes . . . 52
3.3.4 Méthode de k-PPV . . . 53
3.3.5 Mesures statistique pour évaluer les performances des classifieurs (Courbe ROC, Validation croisée) . . . 54
3.4 Résultats . . . 57
3.4.1 Sélection des descripteurs d'EA . . . 57
3.4.2 Construction de la base de données d'apprentissage . . . 60
3.4.3 Performance de la base de données d'apprentissage. . . . 65
3.4.4 Comparaison entre la méthode incrémentale et la méthode de k-moyennes . . . 66
3.4.5 Application de la classification aux composite hybrides lin-verre . . . 70
3.5 Conclusion . . . 73

II Approche de classification basée sur une mesure de dissimilarité : application aux signaux EEG d'épilepsie et aux images 75

4 Introduction 77
4.1 Généralité . . . 77
4.2 Représentation par dissimilarité . . . 79
4.3 Objectif de la deuxième partie . . . 80
4.4 Organisation de la deuxième partie . . . 81

5 Représentation et classification par dissimilarité **83**

5.1 Introduction . . . 83

5.2 Représentation des données . . . 84

5.3 Mesure de dissimilarité . . . 85

5.4 Représentation par dissimilarité . . . 86

5.4.1 Prolongement des matrices de dissimilarités . . . 87

5.4.2 Le classifieur k-PPV . . . 87

5.4.3 Espace de dissimilarités . . . 88

5.5 Conclusion . . . 89

6 Approche de classification basée sur une mesure de dissimilarité entre les RTF : application aux signaux EEG d'épilepsie **91**

6.1 Introduction . . . 92

6.2 Approche de classification basée sur une mesure de dissimilarité entre les RTF . . . 94

6.2.1 Classification dans le plan T-F . . . 94

6.2.2 Approche de classification basée sur une mesure de dissimilarité entre les RTF . . . 95

6.3 Analyse temps-fréquence . . . 97

6.3.1 Préambule . . . 97

6.3.2 Définition de la non-stationnarité . . . 102

6.3.3 Définition de la DSP . . . 102

6.3.4 Distribution de Wigner-Ville . . . 102

6.3.5 La classe de Cohen de RTF quadratique . . . 104

6.3.6 Synthèse . . . 107

6.4 Mesure de dissimilarité entre les RTF . . . 107

6.5 Application de l'approche sur une base de données de signaux EEG d'épilepsie . . . 109

6.5.1 Préambule . . . 109

6.5.2 Base de données des signaux EEG d'épilepsie . . . 111

6.5.3 Résultats . . . 112

6.6 Conclusion . . . 125

7 Approche de classification basée sur la carte de dissimilarité locale : application aux images **131**

7.1 Introduction . . . 132

7.2 Approche de classification basée sur une mesure de dissimilarité . . . 133

7.3 Carte de dissimilarité locale 134
7.3.1 CDL pour image binaire 134
7.3.2 CDL pour image en niveaux de gris 135
7.3.3 Seuils non régulièrement espacés pour la CDL 136
7.4 Base de données 137
7.5 Résultats et discussion 137
7.5.1 Performance de la CDL en fonction de nombre de coupes et les seuils 138
7.5.2 CDL contre la distance euclidienne et la SSIM 140
7.5.3 Classification dans l'espace de dissimilarités 145
7.5.4 Autre application : classification d'une base de données de lettrines 147
7.6 Conclusion 148

Conclusion générale **149**

Production scientifique **157**

Bibliographie **159**

Table des figures

1.1 Contraintes à la rupture et modules spécifiques des fibres naturelles comparés à ceux des fibres synthétiques [1] 10
1.2 Les étapes de classification de signaux d'EA par la méthode de k-moyennes. 13
1.3 Eprouvettes de matériaux composites unidirectionnelles à fibres de lin et à fibres de verre. . . 14

2.1 Classification des fibres naturelles en fonction de leur origine [2]. 19
2.2 Propagation de la rupture dans d'un composite à fibres unidirectionnelles. -a- rupture transverse, -b- rupture longitudinale, -c- décohésion fibre-matrice [3]. 22
2.3 Propagation de la rupture dans le cas d'une faible adhérence fibre-matrice [3]. 23
2.4 Rupture de fibre [3]. 23
2.5 Mécanismes d'endommagements observés dans les stratifiés [3]. 24
2.6 Principe de base de la technique d'EA. 26
2.7 La forme d'onde de signaux d'EA (Signaux de type A, B, C et D) des endommagments d'un matéraiu composite UD à fibre de verre sollicité en traction [4]. 29

3.1 Renforts UD utilisés pour l'élaboration des composites de l'étude : (a) lin et (b) verre. 41
3.2 Schéma d'explication (a) et montage expérimental (b) du procédé de l'infusion de résine. . . . 42
3.3 Machine de traction et dispositifs de mesure de l'activité acoustique. 44
3.4 Influence du PDT sur une salve typique d'EA pour définir le pic maximal de la salve. 45
3.5 Influence du HDT sur une salve typique d'EA pour définir la durée de la salve. 45
3.6 Forme typique d'une salve d'EA caractérisée par les principaux paramètres temporels. 46
3.7 Principe de la méthode incrémentale (MI) 48
3.8 Logigramme du regroupement par la Méthode Incrémentale (MI) 49
3.9 Illustration du principe de la méthode de k-PPV avec un problème de classification à deux classes. 54

3.10 Courbe ROC pour les classifieurs : A, B , parfait et aléatoire. le classifieur A est meilleur par rapport à B. 56
3.11 Spectres fréquentiels de quatre signaux d'EA. 58
3.12 Dendrogramme obtenu à partir de la matrice de corrélation. 59
3.13 Performances en fonction du paramètre sigma de la base de données d'apprentissage de matériaux composites à : (a) fibres de verre et (b) fibres de lin. 61
3.14 Base de données d'apprentissage des composites à fibres de verre. Classification réalisée par la méthode incrémentale. Projection sur le plan (FC, Durée). -a- Les trois classes d'endommagements projeté au même endroit, -b- classe d'endommagement des fibres, -c- classe d'endommagement des interfaces, -d- classe d'endommagement de la matrice. 63
3.15 Base de données d'apprentissage des composites à fibres de lin. Classification réalisée par la méthode incrémentale. Projection effectuée sur le plan (FC, Durée). -a- Les trois classes d'endommagements projeté au même endroit, -b- classe d'endommagement des fibres, -c- classe d'endommagement des interfaces, -d- classe d'endommagement de la matrice. 63
3.16 Projection sur le plan (*FC, Durée*) pour chaque classe d'endommagement des signaux d'EA des deux bases de données d'apprentissage. (a) classe d'endommagement des fibres, (b) classe d'endommagement des interfaces et (c) classe d'endommagement de la matrice. 64
3.17 Projection sur le plan (*FC, Durée*) des clusters obtenus par la méthode de k-moyennes sur une base de données d'un composite UD0 de lin. La méthode de 3-moyennes est appliquée plusieurs fois (changement des conditions initiales à chaque application) sur la même base de données : (a) première, (b) deuxième, (c) troisième et (d) quatrième application. 68
3.18 Projection sur le plan (FC, Durée) des trois clusters obtenus par : (a) Méthode incrémentale et (b) Méthode de k-moyennes. Base de données d'EA obtenu par essai mécanique sur un matériau composite UD0 de lin. 69
3.19 Evolution du nombre de salves d'EA en fonction du temps dans les trois cluster obtenus par : (a) la méthode incrémentale, (b) la méthode de k-moyennes. Base de données d'EA obtenu par essai mécanique sur un matériau composite UD0 de lin. 70
3.20 (a) Projection sur le plan (FC, durée) et (b) évolution du nombre de salve des signaux d'EA des classes d'endommagement du matériaux composite hybride $[V_2/L_2]_s$. Les signaux d'EA sont classifiés par la méthode de 1-PPV en s'appuient sur la base d'apprentissage. 71
3.21 (a) Projection sur le plan (FC, durée) et (b) évolution du nombre de salve des signaux d'EA des classes d'endommagement du matériaux composite hybride $[L_2/V_2]_s$. Les signaux d'EA sont classifiés par la méthode de 1-PPV en s'appuient sur la base d'apprentissage. 71

4.1 Différence entre l'approche basée sur les descripteurs et l'approche basée sur la dissimilarité. . 78
4.2 Espace vectoriel de dissimilarités 80

6.1 Approche typique d'analyse et classification de signal. 93
6.2 Approche de classification basée sur une mesure de dissimilarité entre les RTF. 96
6.3 Procédure pour choisir le couple (Mesure de dissimilarité, RTF) utiliser pour l'approche de classification basée sur une mesure de dissimilarité entre les RTF. 98
6.4 deux signaux contiennent trois sinusoïdes et leurs transformées de Fourier. 99
6.5 RTF idéal d'un signal. 100
6.6 Echantillonnage du plan T-F : (a) TFCT et (b) Ondelette. 101
6.7 WVD d'un signal avec deux composantes sinusoïdale. Illustration des termes d'interférences. 103
6.8 Représentations (temps, fréquence, RTF idéalisée) d'un signal composé de quatre chrip 105
6.9 Exemples de RTF quadratiques de la classe de Cohen du signal à quatre chirp 106
6.10 Exemples de signaux EEG de la base de données épileptique [5]. 112
6.11 Le temps d'exécution de la première (calcule des RTF) et la deuxième (calcule de la matrice d'apprentissage) étapes en fonction du nombre de signaux dans la base de données d'apprentissage 119
6.12 Le temps d'exécution de la troisième étape en fonction du nombre de signaux dans la base de données d'apprentissage. L'approche est combinée avec les classifieurs : RNA, SVM, ADL et k-PPV. 119
6.13 Le temps d'exécution de l'algorithme complet de l'approche proposée en fonction du nombre de signaux dans la base de données d'apprentissage. L'approche est combinée avec les classifieurs : RNA, SVM, ADL et k-PPV. 120
6.14 Espace mémoire nécessaire pour l'exécution de l'algorithme de l'approche proposée en fonction de la taille de la base de données d'apprentissage (nombre de signaux). 121
6.15 Précision de la classification en fonction du SNR. Les classifieurs sont évalués en utilisant : (a) le problème de classification à 2 classes, (b) le problème de classification à 3 classes et (c) le problème de classification à 5 classes. 123

7.1 Processus des méthodes de classification basées sur une mesure de dissimilarité. 134
7.2 CDL obtenues à partir de trois paires d'images en niveaux de gris. 136
7.3 Exemple de chiffres manuscrits de la base de données MNIST. 138
7.4 Précision et durée de la simulation en fonction de nombre de coupes de la CDL. Seuils régulièrement espacée. Performances évaluées par le classifieur k-PPV. 139
7.5 Histogrammes de 4 images de chiffres manuscrits. 139

7.6 Précision et durée de la simulation en fonction de nombre de coupes de la CDL. Seuils non régulièrement espacée. Performances évaluées par le classifieur k-PPV. 141
7.7 Précision en fonction de nombre de classes obtenus par les trois distances (CDL, Euclidienne et SSIM). Performances évaluées par le classifieur k-PPV. 142
7.8 Durée de la simulation en fonction de nombre de classes obtenus par les trois distances (CDL, Euclidienne et SSIM). Performances évaluées par le classifieur k-PPV. 143
7.9 Précision en fonction de nombre d'individus obtenus par les trois distances (CDL, Euclidienne et SSIM). Performances évaluées par le classifieur k-PPV. 144
7.10 Durée de la simulation en fonction de nombre d'individus obtenus par les trois distances (CDL, Euclidienne et SSIM). Performances évaluées par le classifieur kPP. 144
7.11 Précision en fonction de nombre d'individus obtenus par les trois distances (CDL, Euclidienne et SSIM). Performances évaluées par le classifieur RNA basée sur la dissimilarité comme descripteur. 146
7.12 Durée de la simulation en fonction de nombre d'individus obtenus par les trois distances (CDL, Euclidienne et SSIM). Performances évaluées par le classifieur RNA basée sur la dissimilarité comme descripteur. 146
7.13 Quelques exemples de lettrines de la base de données 147

Liste des tableaux

2.1 Propriétés mécaniques d'une liste non exhaustive de fibres végétales et de verre [6, 7]. 19

2.2 Valeurs d'amplitude (en dB) de différents mécanismes d'endommagement pour différents matériaux composites à fibres synthétiques. 28

2.3 Caractéristiques des signaux de type A, B, C et D des endommagements d'un composite UD à fibre de verre sollicité en traction [4]. 28

2.4 Valeurs de fréquences (en kHz) des principaux mécanismes d'endommagement pour différents matériaux composites UD à fibres synthétiques. Les signaux ont été analysés par la FFT. . . 30

2.5 Valeurs de fréquences (en kHz) des principaux mécanismes d'endommagement pour différents matériaux composites à fibres synthétiques. Les signaux ont été analysés par la TOC, TOD, THH ou TFCT. 30

2.6 Différentes approches utilisées pour la classification de signuax d'EA et l'identification des mécanismes d'endommagement de matériaux composites à fibres synthétiques. 32

2.7 Valeurs d'amplitude (en dB) de différents mécanismes d'endommagement pour différents matériaux composites à fibres végétales. 33

2.8 Différentes approches utilisées pour la classification de signaux d'EA et l'identification des mécanismes d'endommagement de matériaux composites à fibres végétales. 33

3.1 Propriétés de la résine époxyde (données du fabricant) [8]. 41

3.2 Propriétés mécaniques des fibres de lin et de verre. 41

3.3 Nombre de plis, épaisseur et fraction volumique des fibres dans les matériaux composites UD non hybrides. 43

3.4 Nombre de plis, séquence d'empilement, épaisseur et fraction volumique des fibres dans les matériaux composites UD hybrides. 43

3.5 Valeur des paramètres du logiciel d'acquisition recommandées pour le test des matériaux composites. 46

3.6 Les paramètres temporels et fréquentiels définis sur une salve d'EA. 47
3.7 Matériaux composites UD et ses mécanismes d'endommagements 49
3.8 Matrice de confusion pour un problème de classification à deux classes. 55
3.9 Matrice de corrélation entre les descripteurs de signaux d'EA. 59
3.10 AUC moyenne en fonction de la valeur de sigma. Performances obtenus par le classifieur *1*-PPV sur différentes base de données des composites à fibres de verre et à fibres de lin. 61
3.11 Performances obtenues par le classifieur k-PPV sur les bases de données de test. 66
3.12 Nombre d'évènement d'EA en pourcentage produite par chaque mode d'endommagement. Classification des bases de données des matériaux composites hybrides ($[L_2/V_2]_s$,$[V_2/L_2]_s$). . 72

6.1 Quelques RTF de la classe de Cohen et leurs fonctions noyaux. 104
6.2 f-divergences couramment utilisées dans le traitement du signal pour la classification des RTFs. 109
6.3 Les instances des trois problèmes de classification des signaux EEG. 112
6.4 Précisions obtenues par le classifieur k-PPV en utilisant différentes mesures de dissimilarité. Évaluation de la performance en utilisant le problème de classification à 5 classes et la validation croisée *10-fold*. 114
6.5 Précisions obtenues par le classifieur k-PPV en utilisant les trois catégories de RTF. La distance d_{DSL_1} est utilisée comme mesure de dissimilarité. Évaluation de la performance en utilisant le problème de classification à 5 classes et la validation croisée *10-fold*. 114
6.6 Précisions obtenues par le k-PPV basée sur les mesures de dissimilarités entre les RTFs comme distances. La distribution SPWVD avec différentes fenêtres de lissage est utilisée. La distance d_{DSL_1} est exploitée comme mesure de dissimilarité. Évaluation des performances en utilisant le problème de classification à 5 classes (instances : A, B, C, D, E) et la validation croisée *10-flod*. 115
6.7 Les Valeurs numériques des paramètres de chaque classifieurs. 115
6.8 Classification d'une base de données de signuax EEG d'épélepsie. Les performances obtenues par le classifieurs RNA, SVM et l'ADN implémentés dans l'espace de dissimilarité et le classifieur k-PPV. La RTF de SPWVD avec une fenêtre de lissage Hanning et la distance d_{DSL_1} ont été utilisées. Évaluation des performances à l'aide du problème de classification à 5 classes et la validation croisée *10-flod*. 116

6.9 Classification d'une base de données de signuax EEG d'épélepsie. Les performances obtenues par le classifieurs RNA, SVM et l'ADN implémentés dans l'espace de dissimilarité et le classifieur k-PPV. La RTF de SPWVD avec une fenêtre de lissage Hanning et la distance d_{DSL_1} ont été utilisées. Évaluation en utilisant les instances du problème de classification à 2 classes. Évaluation de la performance en utilisant la précision et la validation croisée *10-flod*. 117

6.10 Classification d'une base de données de signuax EEG d'épélepsie. Les performances obtenues par le classifieurs RNA, SVM et l'ADN implémentés dans l'espace de dissimilarité et le classifieur k-PPV. La RTF de SPWVD avec une fenêtre de lissage Hanning et la distance d_{DSL_1} ont été utilisées. Evaluation en utilisant les instances du problème de classification à 3 classes. Évaluation de la performance en utilisant la précision et la validation croisée *10-flod*. 117

6.11 La notation Big-O de la complexité en temps des trois étapes de l'algorithme de l'approche de classification basée sur une mesure de dissimilarité. 120

6.12 Comparaison de la précision (en %) obtenue par différentes méthodes de détection des crises d'épilepsie appliquées aux données de [39]. Comparaison des résultats précédemment rapportés pour un problème de classification à 2 classes. Instances de classification définies à partir de (A, B, C, D, E). 124

6.13 Comparaison de la précision (en %) obtenue par différentes méthodes de détection des crises d'épilepsie appliquées aux données de [39]. Comparaison des résultats précédemment rapportés pour un problème de classification à 3 classes. Instances de classification définies à partir de (A, B, C, D, E). 124

6.14 Comparaison de la précision obtenue par différentes méthodes de détection des crises d'épilepsie appliquées aux données de [39]. Comparaison des résultats précédemment rapportés pour un problème de classification à 5 classes (instance A-B-C-D-E). 124

6.15 Diverses approches utilisées pour la détection de crise épileptique appliquée à la base de données [5]. Travaux publiés entre 2004 et 2018. 126

7.1 Précisons et durées de simulation obtenues par le RNA et le k-PPV. Application sur la base de données de chiffre manuscrit. 147

7.2 Précisons et durées de simulation obtenues par le RNA et le k-PPV. Application sur la base de données de lettrines. 148

Acronymes

k-PPV k-Plus Proches Voisins. 14

ADL Analyse Discriminante Linéaire. 92

ADQ Analyse Discriminante Quadratique. 110

ATF Analyse Temps-Fréquence. 29

CDL Carte de Dissimilarité Locale. 80, 81

CND Contrôle Non Destructive. 24

EA Emission Acoustique. 24

EEG Electroencéphalographie. 109

FC Fréquence Centroïde. 47

FI Fréquence d'Initiation. 47

FM Fréquence Moyenne. 47

FR Fréquence de Réverbération. 47

PWVD Pseudo Wigner-Ville Distribution. 103

RNA Réseau de Neurones artificiels. 92

RTF Représentations Temps-Fréquence. 94

RVDT Real Valued Distance Transform. 135

SPWVD Smoothed Pseudo Wigner-Ville Distribution. 103

SSIM Structural SIMilarity. 137

SV Support Vectors. 51

SVDD Support Vector Data Description. 50

SVM Séparateurs à Vastes Marges. 52

T-F Temps-Fréquence. 10

TFCT Transformée de Fourier à Court Terme. 29

THH Transformée de Hilbert Huang. 29

TO Transformée en Ondelettes. 110

TOC Transformée en Ondelettes Continue. 29

TOD Transformée en Ondelettes Discrète. 29

UD UniDirectionnels. 22

WVD Wigner-Ville Distribution. 97

Introduction générale

Les travaux de recherche développés dans le cadre de cette thèse ont fait l'objet d'une cotutelle internationale entre l'Université de Reims Champagne-Ardenne (URCA) et l'Université Mohammed Premier (UMP) d'Oujda. Cette thèse a également fédéré les compétences de trois laboratoires, à savoir le LISM1 [1] (URCA) pour son expertise dans le domaine de la caractérisation des matériaux et le CReSTIC [2] (URCA) et le LSE2I [3] (UMP) pour leurs compétences en matière du traitement du signal et de l'image.

Les méthodes de classification sont de plus en plus utilisées dans divers domaines afin d'apporter des informations supplémentaires concernant les objets ou les sources à classifiés. De ce fait, plusieurs travaux de recherche ont tenté d'améliorer ces méthodes ou d'introduire des indicateurs pour tester leur fiabilité. C'est dans ce contexte que le présent travail se situe. Par souci de concision et d'une certaine cohérence, ce mémoire est scindé en deux parties distinctes. Nous avons développé dans la première partie (du chapitre 1 à 3) une nouvelle méthode de classification non-supervisée pour séparer les signaux acoustiques collectés à partir d'essais mécaniques effectués sur divers matériaux composites. Cette nouvelle méthode de classification a permis d'identifier les mécanismes d'endommagement, engendrant la rupture des matériaux composites étudiés, à partir de leur signature acoustique. La deuxième partie de ce mémoire (du chapitre 4 à 7) a pour objectif la proposition d'une nouvelle approche de classification sans emploi de descripteurs, appliquée aux signaux électroencéphalogramme (EEG) d'épilepsie et aux images. Concernant les signaux non-stationnaires (chapitre 6), l'idée originale de cette approche consiste à combiner l'Analyse Temps-Fréquence (ATF) avec l'approche de classification basée sur une mesure de dissimilarité. L'objectif de cette démarche est d'améliorer et de simplifier la décision dans un environnement non-stationnaire. Ce qui concerne les images (chapitre 7), l'objectif est de proposer une mesure de dissimilarité afin d'améliorer les performances de l'approche de classification basée sur une mesure de dissimilarité.

Pour montrer davantage l'intérêt des travaux de recherche développés dans ce manuscrit, le contenu ces deux parties est précisé en quelques lignes ci-après.

1. Laboratoire d'Ingénierie et Sciences des Matériaux.
2. Centre de Recherche en Science et Technologie de l'Information et de la Communication.
3. Laboratoire des Systèmes Electroniques, Informatiques et Images.

Première partie : discrimination des mécanismes d'endommagement de matériaux composites à fibres végétales par une classification de signaux d'émissions acoustiques.

De par leur constitution, les matériaux composites sont des matériaux hétérogènes puisqu'ils associent au moins deux constituants différents : un renfort (généralement sous forme de fibres) et une matrice (souvent en résine). Pendant leur utilisation, les matériaux composites et plus particulièrement ceux à fibres végétales subissent des agressions dues à une large variété d'endommagement : exposition à l'humidité, à des cycles de température, à la fatigue mécanique ou à des impacts. Ces événements entraînent principalement de la fissuration de matrice, la décohésion de l'interface fibre/matrice et la rupture des fibres. Ces dommages peuvent servir de précurseurs à un endommagement qui va croître avec la répétition des sollicitations imposées à la structure en service, et éventuellement conduire à sa ruine totale. Il est donc indispensable de recourir au Contrôle Non Destructif (CND) de ces matériaux pour évaluer leur état de santé. Le travail proposé dans cette partie consiste à développer une méthode de Contrôle Non Destructif (CND) par Emissions Acoustiques (EA). Cette technique permet d'analyser et d'identifier plus en détails ces mécanismes d'endommagement intervenant au sein des matériaux à l'aide d'une classification des signaux acoustiques.

Cette partie est présentée en trois premiers chapitres. Après une introduction (chapitre 1), nous avons mené, dans le chapitre 2, une étude bibliographique relative à l'identification des sources d'EA au sein d'un matériau composite, après avoir décrit les endommagements rencontrés.

Le chapitre 3 est consacré à la démarche expérimentale menée sur les différents matériaux composites de l'étude, la démarche de classification et illustre aussi les résultats obtenus. Ainsi, pour identifier les principaux mécanismes d'endommagement provoquant la ruine du matériau et construire une base de données de signaux d'EA, des éprouvettes modèles, de résine pure et d'autres de type unidirectionnel (UD), ont été réalisées puis testées lors d'essais mécaniques spécifiques. L'essai de traction a été retenu puisque des tests sur résine et à différentes orientations sur UD, entre la direction de traction et celle des fibres, sont associés à certains types d'endommagement. Ensuite, une nouvelle méthode de classification de signaux d'EA est proposée, appelée méthode incrémentale. Contrairement aux méthodes actuelles, cette méthode ne nécessite pas une étape supplémentaire d'identification de groupes. Chaque groupe de signaux obtenus par cette méthode sera directement corrélé avec un et un seul mécanisme d'endommagement. Enfin, la méthode de classification que nous avons développée est appliquée, dans un premier temps, à des matériaux composites dont les modes d'endommagement sont connus (de type UD), ce qui a permis de valider la méthode. De plus, cette méthode de classification des évènements d'EA est menée sur des composites dont la structure est plus complexe (stratifiés à renforts hybrides de fibres de lin et de verre).

Deuxième partie : Approche de classification basée sur une mesure de dissimilarité : application aux signaux EEG épileptiques et aux images.

L'objectif de la deuxième partie de ce manuscrit est de proposer une nouvelle approche pour la décision en environnement non-stationnaire, par une fertilisation croisée des domaines de l'analyse temps-fréquence et la reconnaissance des formes par une mesure de dissimilarité.

La classification des signaux non-stationnaires est un problème classique du diagnostic médical assisté par ordinateur, du contrôle non destructif, de la reconnaissance vocale, etc. Pour résoudre un problème de classification, un espace de représentation doit être d'abord choisi. Ensuite, un classifieur peut être implémenté dans cet espace. La représentation la plus populaire est basée sur un espace euclidien de caractéristiques. Celles-ci peuvent être défini à partir de la représentation temporelle, fréquentielle ou encore temps-fréquence du signal. Ces caractéristiques devraient idéalement caractériser les signaux et être également pertinentes pour permettre au classifieur de distinguer les classes. De telles caractéristiques doivent être définies par des experts exploitant leurs connaissances du domaine d'application.

Un inconvénient de l'utilisation des caractéristiques est que différents signaux peuvent avoir la même représentation, car ils diffèrent par des propriétés qui n'ont pas été exprimées dans l'ensemble de caractéristiques choisi. Cela se traduit par un chevauchement entre les classes. Dans certaines zones de l'espace de représentation, des signaux appartenant à différentes classes sont représentés par les mêmes vecteurs de caractéristiques. Par conséquent, les signaux dans la zone de chevauchement ne peuvent pas être distingués par le classifieur, ce qui conduit à une erreur de classification intrinsèque, généralement appelé erreur de Bayes. De plus, la représentation basée sur les caractéristique nécessite la sélection de ceux-ci, ce qui est parfois difficile et dépendant du domaine d'application.

Une autre manière de représenter les objets consiste à les décrire en utilisant une mesure de dissimilarité. L'idée est que la notion de proximité est plus fondamentale que celle d'une caractéristique ou d'une classe. La représentation par dissimilarité offre la possibilité d'unifier les deux représentations, statistique et structurelle. Pour une telle représentation, le chevauchement entre les classes n'existe pas si les objets sont bien étiquetés. Ceci doit être en accord avec des applications du monde réel puisqu'il n'y a pas d'objet appartenant à plus d'une classe si les classes sont bien définies. Ainsi, avec une dissimilarité définie comme distance, seuls les objets identiques ont une distance nulle et doivent avoir la même étiquette, car ils sont identiques.

La représentation par dissimilarité a été principalement conçue pour résoudre des problèmes de classification. Dans cette approche, les objets sont représentés par un vecteur de leurs dissimilarités par rapport à un ensemble d'objets prototypes généralement appelé ensemble de représentation. Ainsi, les classifieurs peuvent être construits dans l'espace de dissimilarité où chaque dimension correspond à la dissimilarité par rapport à un objet de l'ensemble de représentation. Ensuite, les classifieurs peuvent être utilisés pour identifier un nouvel objet représenté de la même manière. Par conséquent, la réussite d'un classifieur est déterminée par la mesure de dissimilarité définie par l'utilisateur, dans laquelle les informations sur le domaine d'application

peuvent être exprimées. Même les « mauvaises » mesures de dissimilarité peuvent être utilisées (au prix d'un grand ensemble d'apprentissage), à condition que seuls les objets identiques présentent une dissimilarité nulle. Il est important de noter que tout classifieur basé sur l'espace de caractéristique peut également être utilisé dans l'espace de dissimilarité.

Les signaux non-stationnaires ont des caractéristiques spectrales qui varient dans le temps. Ces signaux ne peuvent être correctement traités par les méthodes usuelles d'analyse spectrale (transformation de Fourier) : en effet, la transformation de Fourier permet de représenter un signal temporel dans le domaine fréquentiel en le décomposant en une somme de signaux tonals. Ainsi, la transformation de Fourier n'exhibe pas clairement de ce fait les informations sur le séquencement temporel du signal. Cette décomposition nécessite par essence que le signal soit stationnaire au sens où ses propriétés spectrales n'évoluent pas au cours du temps. Pour les besoins du traitement du signal, on a cherché à représenter simultanément les informations concernant les structures temporelles et fréquentielles contenues dans le signal. Ces transformées sont appelées représentations temps-fréquence (RTF). Elles permettent de voir la répartition de l'énergie du signal dans le plan temps-fréquence. Il convient de considérer le passage par RTF au domaine temps-fréquence, comme une distribution de l'information contenue dans le signal analysé de façon à en faciliter l'interprétation. Les RTF présentent l'avantage majeur de mettre en évidence les comportements non-stationnaires du signal. Une grande diversité de RTF s'offre à présent à l'utilisateur, par exemple les décompositions linéaires telles que les transformées en ondelettes et de Fourier à court-terme, ou les distributions d'énergie telles que le spectrogramme et la distribution de Wigner-Ville.

Ainsi, nous avons proposé une nouvelle approche de classification de signaux non-stationnaires en combinant l'espace de dissimilarité avec l'ATF. Afin de mettre en évidence la pertinence de cette approche, des applications à des signaux réels sont considérées. Les résultats expérimentaux montrent que l'approche que nous proposons conduit à une amélioration notable des performances d'un classifieur comparée à l'approche basée sur les caractéristiques. De plus, ces résultats montrent également que l'approche est robuste aux bruits. Cependant, cette approche augmente la complexité en temps, en fonction de la taille de la base de données d'apprentissage, d'un classifieur. Ce qui peut restreindre l'applicabilité de l'approche à des problèmes de classification avec une base d'apprentissage de petite taille. Pour résoudre ce problème, nous proposons d'adopter la méthode de sélection de prototypes et de choisir une mesure de dissimilarité permettant au classifieur de converger le plus rapidement possible vers sa précision maximale. Ainsi, cette idée a été développée en s'intéressant à des mesures de dissimilarité utilisée pour la comparaison des images. Il s'agit précisément de la carte de dissimilarité locale (CDL), qui a été combinée avec l'approche de classification basée sur une mesure de dissimilarité. Pour valider l'efficacité de la CDL, des applications à des bases de données constituées d'images sont considérées. Les résultats expérimentaux montrent que la CDL permet au classifieur de converger très

rapidement vers de bonne précision.

Cette partie s'articule autour de trois chapitres (chapitres 5 à 7) après une introduction (chapitre 4). Une revue bibliographique sur les techniques de classification basées sur une représentation par dissimilarité permettant de situer les travaux de la deuxième partie est présentée dans le chapitre 5. Ainsi, nous évoquerons la méthode de prolongement de dissimilarités et le classifieur k-PPV comme étant la méthode la plus simple et la plus directe dans ce contexte. Également nous décrivons le principe de construction d'un classifieur directement dans un espace de dissimilarités comme étant l'approche la plus globale.

Dans le chapitre 6, le principe de l'approche de classification basée sur une mesure de dissimilarité entre les RTF de signaux sera d'abord présenté. Il s'agira aussi de montrer que l'analyse la plus adaptée aux signaux non-stationnaires est l'analyse temps-fréquence. Nous introduirons également une liste non exhaustive de mesures de dissimilarité utilisées pour comparer les RTF. En dernier lieu, les avantages de l'approche proposée en l'appliquant à un problème de classification de signaux EEG d'épilepsie seront mis en évidence.

Le chapitre 7 soulignera les capacités de la CDL pour la comparaison des images. Ainsi, la CDL sera combinée dans un premier lieu avec l'approche de classification basée sur une mesure de dissimilarité. Ensuite, nous présenterons le principe de la CDL pour la comparaison des images binaires et en niveaux de gris. En dernier lieu, les performances en termes de classification de cette approche sur deux bases de données constituées d'images seront évaluées.

Enfin, les résultats obtenus dans les deux parties de cette thèse sont synthétisés dans la conclusion générale, qui propose aussi un certain nombre de perspectives.

Première partie

Discrimination des mécanismes d'endommagement de matériaux composites à fibres végétales par une classification de signaux d'émissions acoustiques

Chapitre 1

Introduction

Sommaire

1.1 Généralité 9

1.2 Approche de classification 11

1.3 Objectif de la première partie 11

1.4 Organisation de la première partie 12

1.1 Généralité

Les matériaux composites à fibres naturelles offrent une alternative intéressante à certains composites à renforts usuels comme le verre. En effet, les fibres naturelles présentent des propriétés mécaniques intéressantes et de nombreux atouts si l'on s'intéresse aux impacts environnementaux. Il s'agit de ressources renouvelables, biodégradables, neutres quant aux émissions de CO2 dans l'atmosphère et requérant peu d'énergie pour être produites. De ce fait, plusieurs travaux de recherche ont été menés pour mieux comprendre ces fibres. Le potentiel des fibres naturelles, comme le lin et le chanvre, tient au fait que leurs propriétés spécifiques (propriétés mécaniques ramenées à la masse volumique) sont comparables voire supérieures à celles des fibres de verre (Figure 1.1). Toutefois, le développement des composites à fibres naturelles dans des applications industrielles nécessite une information supplémentaire sur leur processus d'endommagement afin de mieux prédire leur durée de vie. En effet, lorsqu'un mécanisme d'endommagement se produit dans un matériau composite, une onde transitoire, résultant de la libération d'énergie stockée, se propage de la source d'endommagement vers la surface du matériau. Cette onde peut être enregistrée par des capteurs fixés sur le matériau. Ce phénomène est connu sous le nom d'Emission Acoustique (EA). Dans l'EA, on fait l'hypothèse que chaque signal reçu lors de l'endommagement d'un matériau composite correspond à un événement dont

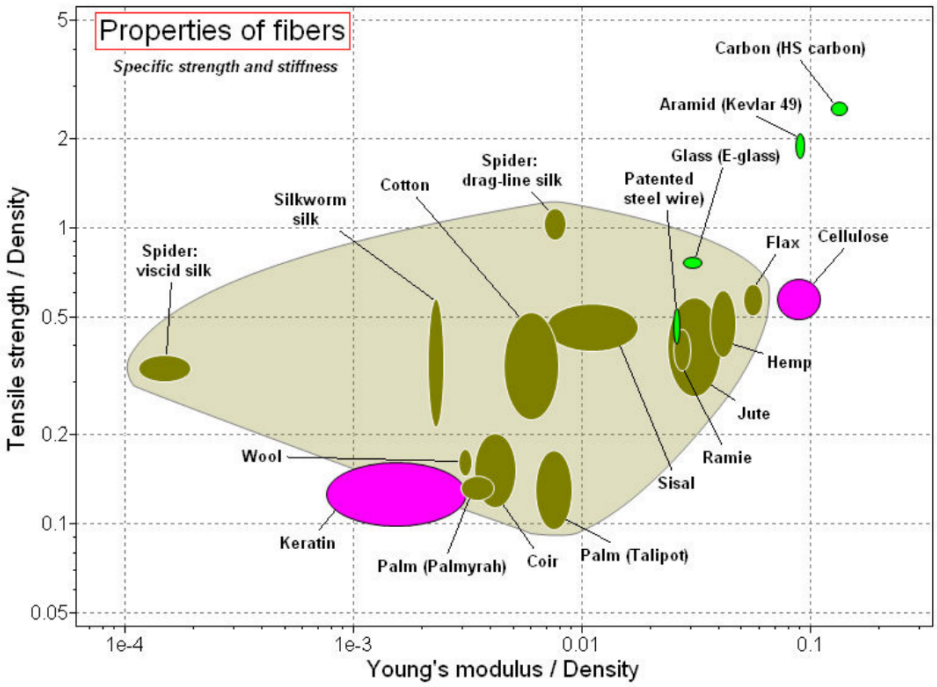


FIGURE 1.1 – Contraintes à la rupture et modules spécifiques des fibres naturelles comparés à ceux des fibres synthétiques [1]

la forme est directement liée aux caractéristiques de cet événement. Le défi consiste alors à établir une bonne corrélation entre les mécanismes d'endommagement qui se sont produits au sein du matériau composite et les signaux d'EA résultants.

En effet, durant le processus de dégradation des matériaux composites (sollicitations mécaniques, vieillissement), plusieurs mécanismes d'endommagement à l'échelle microscopique peuvent intervenir selon la nature du composite et de ses constituants (fibres et résine). L'EA permet d'analyser et d'identifier plus en détails ces mécanismes d'endommagement, à l'aide d'une classification basée sur un ou plusieurs paramètres temporels, fréquentiels ou Temps-Fréquence (T-F). Plusieurs méthodes de classification ont été employées pour identifier ces mécanismes : les méthodes à un paramètre temporel [9–15] et les méthodes à plusieurs paramètres comme amplitude, durée, énergie, etc [16–31, 31–37]. L'utilisation des méthodes mono-paramétriques peut s'avérer insuffisante pour la discrimination des modes d'endommagement, notamment pour des matériaux plus complexes. Les classifications à plusieurs paramètres sont plus pertinentes mais nécessitent une connaissance préalable du processus d'endommagement et présentent certaines limites. Par exemple, pour les agro-composites à renfort hybride (constitués de deux types de fibres par exemple), ces analyses ne permettent pas de distinguer les endommagements liés à chaque fibre. Il est donc nécessaire de proposer une nouvelle méthode de classification qui permettra de tenir compte de ces particularités et ainsi d'améliorer l'identification des types d'endommagement.

1.2 Approche de classification

Deux méthodes de classification peuvent être distinguées, supervisée et non-supervisée. Une méthode de classification supervisée classe un objet en s'appuyant sur un apprentissage préalable. Elle se base sur un jeu de données connues à priori (une base d'apprentissage) à la suite d'une expérience par l'utilisateur pour construire le modèle. Une méthode de classification non-supervisée permet de diviser un ensemble d'objets en différents sous-ensembles homogènes. Les objets de chaque sous-ensemble partagent des caractéristiques communes. Les indices de qualité permettent de fournir des indications sur le degré d'homogénéité d'un point de vue mathématique. Ensuite, la tâche de l'utilisateur est d'identifier les sous-ensembles obtenus, c'est-à-dire l'extraction du phènomène physique que chaque sous-ensemble caractérise. Les objets à classifier sont souvent représentés par leurs vecteurs de caractéristiques, que l'on nomme descripteurs.

La résolution d'un problème de classification se base sur les descripteurs. La réussite d'une méthode de classification est directement liée à la pertinence des descripteurs. La présence de descripteurs redondants ou non pertinents peut réduire considérablement cette performance. Une définition de la pertinence d'un descripteur a été donnée dans [38]. Selon cette définition, un descripteur peut être très pertinent, peu pertinent ou bien non pertinent. Un descripteur est dit très pertinent si son absence entraîne une détérioration significative de la performance du système de classification. Il devient peu pertinent si sa présence ou absence entraine respectivement une légère amélioration ou détérioration de la performance du système de classification. Le descripteur non pertinent, qu'il soit absent ou présent, n'a aucun effet sur la performance du système de classification.

1.3 Objectif de la première partie

Ces dernières décennies, diverses méthodes ont été employées pour classifier les signaux d'EA et discriminer les mécanismes d'endommagement des matériaux composites. Le k-moyennes [39] est l'une des méthodes les plus utilisées dans la littérature pour classifier les signaux d'EA [16–26, 28–36]. Le principe de cette méthode consiste à regrouper un ensemble de signaux d'EA à partir de leurs vecteurs de descripteurs en un nombre de groupe donné (figure 1.2). Ensuite, l'utilisateur doit corréler les groupes obtenus aux mécanismes d'endommagement correspondants. Cependant, lorsque de nombreux mécanismes d'endommagement se produisent dans le même matériau composite, il est plutôt difficile d'associer un type d'endommagement à un groupe donné de signaux d'EA. En effet, pour être sûr d'identifier correctement les endommagements, il est nécessaire de combiner l'algorithme de k-moyennes à d'autres informations trouvées dans la littérature ou obtenues via des observations microscopiques sur les éprouvettes. Il est également important de noter que la

nature non stationnaire des signaux d'EA rend également difficile la comparaison entre les différents groupes de signaux.

Le but de ce travail est de remédier à ce problème en construisant une base d'apprentissage de signaux d'EA où chaque groupe est associé à un mode d'endommagement bien identifié. Pour ce faire, nous avons développé une nouvelle méthode de classification de signaux d'EA dite incrémentale. L'avantage de cette méthode par rapport à celle de k-moyennes se situe dans le fait que nous n'avons pas besoin d'une étape d'identification des groupes obtenus. Chaque groupe obtenu par la méthode incrémentale doit être en correspondance avec un et un seul endommagement précis. Le principe de la méthode incrémentale consiste à combiner deux bases de données de signaux d'EA dont l'une d'entre elles contient un type d'endommagement supplémentaire par rapport à l'autre. L'objectif sera de séparer les signaux correspondant à l'endommagement supplémentaire et ceux liés aux autres endommagements. Pour construire la base de données d'apprentissage, les signaux d'EA sont d'abord collectés à partir d'essais mécaniques spécifiques sur des éprouvettes spécialement conçues (figure 1.3) pour favoriser un ou plusieurs types d'endommagement. Ils sont ensuite classifiés à l'aide de la méthode incrémentale à partir de leurs descipteurs. Les résultats obtenus sont ensuite comparés à ceux obtenus à l'aide de la méthode des k-moyennes. Enfin, les signaux identifiés lors de l'apprentissage sur des échantillons témoins sont utilisés comme référence pour séparer les mécanismes d'endommagement présents dans des matériaux composites plus complexe.

1.4 Organisation de la première partie

Plusieurs travaux de recherche menés sur les composites à fibres naturelles s'inspirent fortement de ceux effectués sur les composites traditionnels (à fibres synthétique). Par ailleurs, les constituants des fibres végétales et synthétiques sont totalement différents, et cette différence n'a pas été prise en compte. Cette partie est divisée en deux chapitres. Avant de présenter les démarches de la nouvelle approche de classification et les résultats obtenus (chapitre 3), une étude bibliographique sur les différentes études d'EA menées sur des matériaux composites fibreux est d'abord introduite dans le chapitre 2.

Le principal objectif de ce dernier est de montrer les limites des approches actuelles de classification et d'identification de mécanismes d'endommagement de matériaux composites fibreux à partir de signaux d'EA. En préambule, une brève présentation des avantages de la fibre de lin est donnée. Puis, les mécanismes d'endommagement qui interviennent dans les matériaux composites fibreux sont répertoriés. Ensuite, une étude bibliographique s'est focalisée sur les différentes travaux de recherche qui ont été menés jusqu'à présent afin d'identifier les mécanismes d'endommagement. Enfin, à partir de l'étude bibliographique, différentes problématiques sont soulignés et des solutions sont formulées.

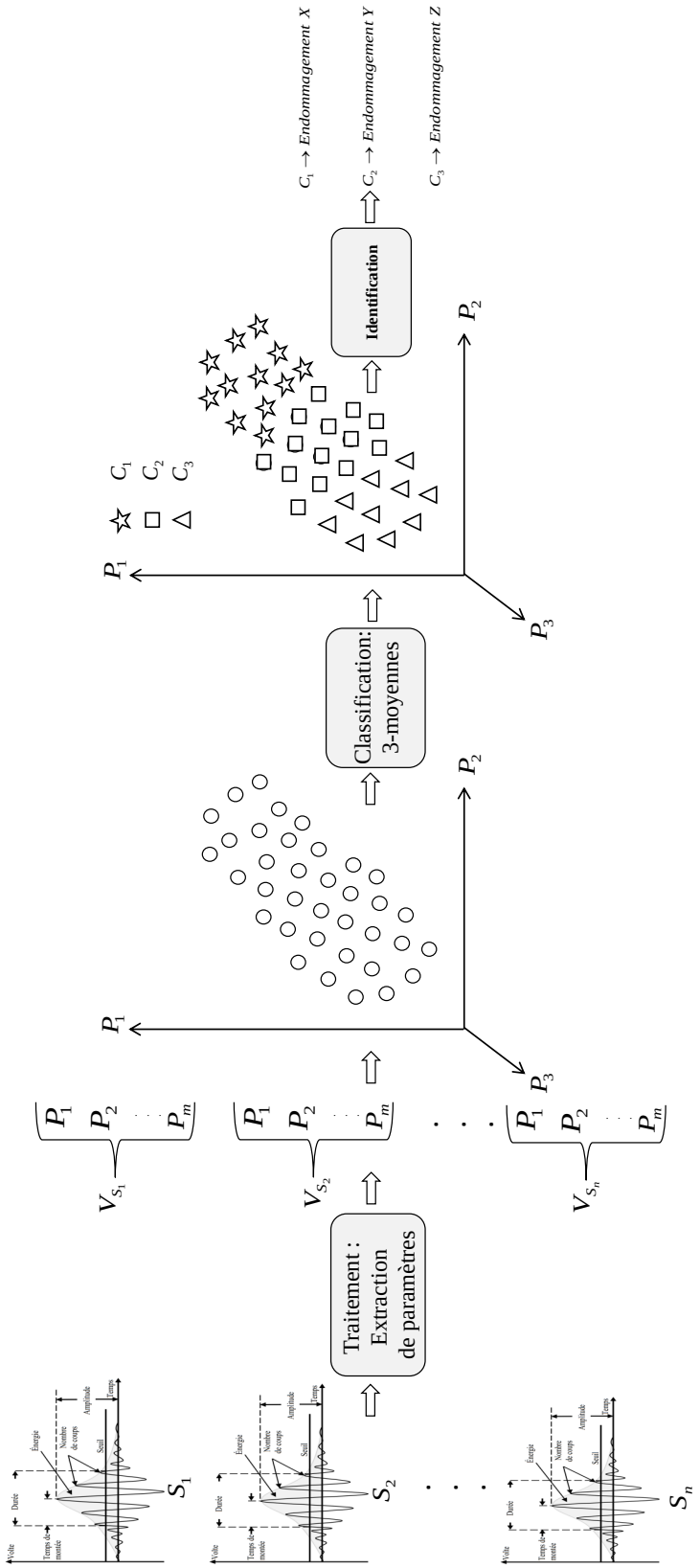


FIGURE 1.2 – Les étapes de classification de signaux d'EA par la méthode de k-moyennes.

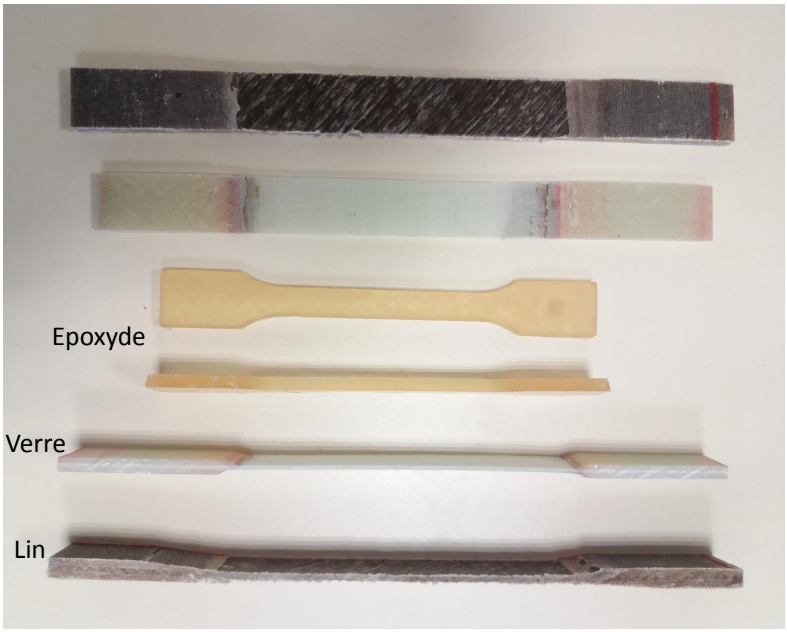


FIGURE 1.3 – Eprouvettes de matériaux composites unidirectionnelles à fibres de lin et à fibres de verre.

Le chapitre 3 comporte trois volets. Le premier s'intéresse à l'étude expérimentale qui a permis de construire une base de données de signaux d'EA à partir d'échantillons témoins sollicités en traction. Le deuxième volet aborde notre approche de classification, la méthode incrémentale. Enfin, les résultats sont présentés et analysés dans le troisième volet.

Le premier volet intitulé "étude expérimentale" décrit les différentes étapes afin de construire une base de données de signaux d'EA. Dans un premier temps, les constituants (fibre, résine) des matériaux composites de l'étude et leurs propriétés mécaniques distinctives sont présentés. Dans un second temps la mise en ouvre des matériaux de l'étude et le protocole expérimentale sont décrits.

Dans le deuxième volet intitulé "démarche de classification", nous présentons dans un premier temps le principe de la méthode incrémentale. Ensuite, nous introduisons le principe de la méthode de k-moyennes et celui de la méthode de k-Plus Proches Voisins (k-PPV). Enfin, nous présentons quelques mesures statistiques pour évaluer les performances d'une méthode de classification supervisée.

Le troisième volet de ce chapitre est réservé aux résultats. Nous présentons d'abord une démarche pour choisir les descripteurs les plus pertinents afin de caractériser les signaux d'EA. Nous utilisons ensuite la méthode incrémentale pour construire deux bases de données d'apprentissage de signaux d'EA. Une base de données correspond aux matériaux composites à fibres de lin et l'autre à ceux à fibres de verre. Une comparaison entre les signaux d'EA des endommagements de matériaux composites à fibres de lin et ceux à fibres de verre est aussi réalisée. Puis, les performances du classifieur k-PPV sont évaluées en utilisant les deux bases de données pour l'apprentissage afin d'estimer leurs qualités. Nous présentons ensuite une comparaison

entre les résultats obtenus par la méthode incrémentale avec ceux obtenus par la méthode de k-moyennes. Nous combinons enfin les deux bases de données d'apprentissages avec le classifieur k-PPV afin d'identifier les signaux d'EA d'un matériau composite plus complexe.

Chapitre 2

Matériaux composites à fibres végétales et émission acoustique : état de l'art et problématiques

Sommaire

2.1 Introduction 18

2.2 Les matériaux composites à fibres naturelles 18

2.2.1 Les fibres naturelles 18

2.2.2 Les matériaux composites à fibres de lin 20

2.3 Mécanismes d'endommagement de matériaux composites fibreux 21

2.3.1 Les mécanismes d'endommagement intra-pli 22

2.3.2 Les mécanismes d'endommagements inter-pli 23

2.4 Identification de mécanismes d'endommagements à partir des signaux d'émission acoustique 24

2.4.1 Présentation de la technique d'émission acoustique 24

2.4.2 Identification des mécanismes d'endommagement par émission acoustique : état de l'art 26

2.5 Problématiques 35

2.6 Conclusion 37

2.1 Introduction

Ce chapitre dresse une synthèse bibliographique sur les matériaux composites à fibres végétales et les approches actuelles de classification de signaux d'EA. Cette synthèse permettra la compréhension et l'interprétation des résultats obtenus au chapitre 3. Le présent chapitre est ainsi organisé en trois parties :

— La première partie décrit le potentiel des matériaux composites à fibres naturelles et spécialement ceux à fibres végétales.

— La deuxième partie présente les principaux mécanismes d'endommagement qui interviennent au sein des matériaux composites fibreux.

— La troisième partie est consacrée à une étude bibliographique sur les approches actuelles de discrimination de mécanismes d'endommagement par la technique de l'EA. Après une description du principe de la technique de l'EA, une liste non exhaustive des différentes approches qui ont été utilisées pour la classification de signaux d'EA est présentée. A l'issue de l'étude bibliographique, une liste des problématiques est soulignée et des solutions sont formulées.

2.2 Les matériaux composites à fibres naturelles

De manière générale, un matériau composite est constitué de l'assemblage de deux matériaux de nature différentes, se complétant et permettant d'aboutir à un matériau dont l'ensemble des performances est supérieur à celui des composants pris séparément. Cette assemblage souvent de résine et de fibres offre aux concepteurs la possibilité de façonner un "matériau à la carte", ce qui permet d'élaborer des structures avec un design novateur tout en garantissant les différentes contraintes qui leur sont imposées par le cahier des charges. Par ailleurs, le développement de ces matériaux dans diverses applications industrielles soulève de plus en plus des questions concernant leur recyclabilité. De ce fait, de nombreux travaux de recherche se sont intéressés à des matériaux composites à constituants naturels. Car ces derniers sont souvent faciles à recycler et demandent peu d'énergie pour être produits. Dans ce qui suit, nous présentons les principaux constituants d'origine naturelle qui sont utilisés pour élaborer les biomatériaux composites.

2.2.1 Les fibres naturelles

Les renforts se présentent dans un matériau composite sous forme de fibres continues ou discontinues. Le rôle du renfort est d'assurer la fonction de résistance mécanique aux efforts. La matrice assure quant à elle la

cohésion entre les renforts de manière à répartir les sollicitations mécaniques. L'arrangement des fibres, leur orientation permettent de renforcer les propriétés mécaniques de la structure. Dans le cadre de cette thèse, nous étudions plus particulièrement les matériaux composites dont le renfort est constitué de fibres longues continues d'origine naturelle.

Les fibres naturelles peuvent être d'origine animale, minérale ou végétale comme l'indique la figure 2.1. Ces fibres, notamment celles d'origine végétale, constituent actuellement une alternative intéressante à certaines fibres usuelles comme les fibres de verre. En effet, les fibres naturelles présentent des propriétés mécaniques spécifiques comparables voire supérieures à celles de certaines catégories de fibres de verre. A titre d'exemple, on peut constater à partir du tableau 2.1 que la fibre de lin présente les meilleures propriétés mécaniques des fibres naturelles.

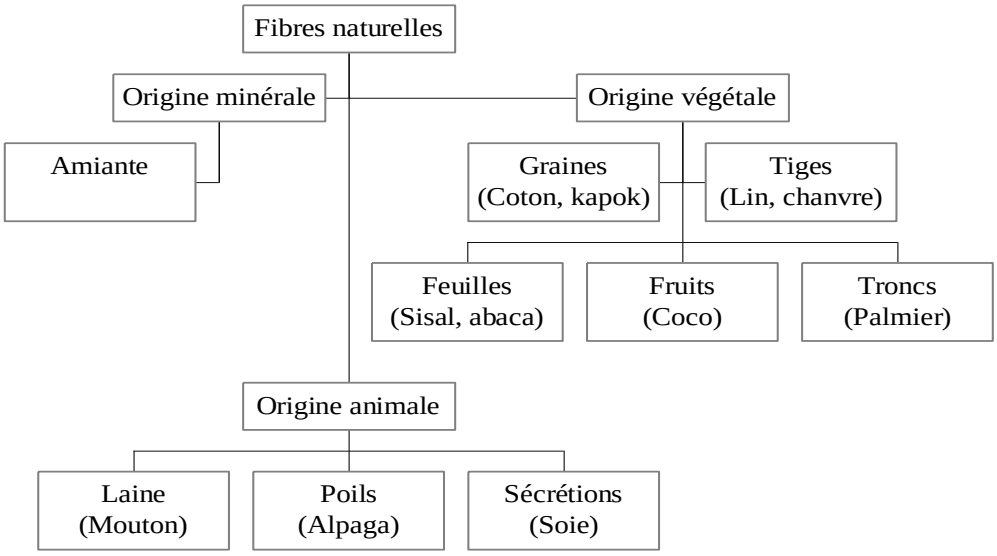


FIGURE 2.1 – Classification des fibres naturelles en fonction de leur origine [2].

Tableau 2.1 – Propriétés mécaniques d'une liste non exhaustive de fibres végétales et de verre [6, 7].

Fibre	Diamètre (µm)	Densité (g/cm^3)	Module d'Young (GPa)	Contrainte à la rupture (MPa)	Déformation à la rupture (%)
Lin	12-24	1,5	23-68,2	345-1454	1,6-3,6
Sisal	100-300	1,5	9,4-25	390-635	2-9
Abaca	17-21	1,5	12	400	3-10
Chanvre	16-31,2	1,48	19-44,5	270-788	0,8-2,5
Ramie	25-30	1,5	24,5	560	2,5
Jute	160-185	1,3	14,5-30	393-800	1,8-5
Coco	17-21	1,2	4-6	175	30
Verre E	16	2,6	74	2500	3,5

2.2.2 Les matériaux composites à fibres de lin

La fibre de lin est l'une des fibres végétales la plus utilisée comme renforts pour les matériaux composites destinés à l'industrie automobile. Ceci justifie l'augmentation ces dernières décennies du nombre de travaux de recherche sur les composites à fibre de lin. Ces travaux de recherche peuvent être divisés en deux catégories. La première, concerne les études sur la fibre élémentaire de lin. Ces études ont été menées sur la microstructure de la fibre de lin à différentes échelles, sa composition chimique [40, 41] et son comportement mécanique [42, 43]. La deuxième catégorie englobe les travaux de recherche sur les fibres de lin utiliseés comme renfort de matériaux composites. La plupart des études sont axées sur les comportements mécaniques des matériaux composites à fibres de lin avec différentes matrices thermodurcissables [33, 44], thermoplastiques [45, 46], biodégradables [47], polymère [46, 48] ou biopolymères [8]. D'autres études ont portées sur l'analyse des paramètres influençant les propriétés mécaniques, comme la forme, le taux et l'arrangement des fibres [48, 49], les traitements chimiques des fibres et modifications de la matrice [50, 51], les procédés de fabrication [52, 53] et la température d'essai [33]. De plus, la cinétique de l'absorption d'eau et le vieillissement des composites à fibres de lin font partie des problématiques actuelles [8, 33, 37].

D'autre part, l'hybridation de deux types de fibres, végétale et synthétique, permet de profiter des avantages qu'offrent les fibres végétales tout en gardant les propriétés mécaniques intéressantes des fibres synthétiques. L'avantage de l'hybridation du renfort d'un composite est qu'un type de fibre peut compléter ce qui manque à l'autre. Par exemple, en proposant une solution composite hybride, les performances mécaniques de certaines fibres synthétiques peuvent être combinées avec les avantages environnementaux et les gains de masse que permettent les fibres végétales. Plusieurs travaux de recherche ont étudié expérimentalement l'effet de l'hybridation sur les propriétés mécaniques de différents composites hybrides fibres végétales-fibres de verre [54–60]. Tous ces travaux de recherche ont montré que l'hybridation a considérablement modifié le comportement mécanique des fibres végétales. En effet, le composite hybride présente généralement des propriétés intermédiaires entre les composites à fibres synthétiques et les composites à fibres végétales. Cependant, peu de travaux ont étudiés les mécanismes d'endommagement des matériaux composites hybrides.

Malgré les efforts entrepris par les chercheurs sur les matériaux composites à fibres végétales et spécialement à fibres de lin, de nombreuses questions restent posées quant à la possibilité d'utiliser ce type de composites à une plus grande échelle. Certaines problématiques liées à leurs mécanismes d'endommagement doivent être résolues. En effet, l'étude des phénomènes d'endommagement est un enjeu important pour évaluer la durabilité des matériaux composites. La bonne analyse et l'interprétation des variables d'endommagement permettent d'établir des critères de rupture ou des modèles de prédiction de la durée de vie des matériaux composites. Par conséquent, plusieurs études se sont intéressées aux mécanismes d'endommagement de ma-

tériaux composites à fibres végétales. Ces études ont été beaucoup influencées par les techniques qui ont été utilisées pour identifier les mécanismes d'endommagement de matériaux composites à fibres usuelles. En plus de la difficulté liée à la complexité des mécanismes d'endommagement, les constituants à l'échelle microscopique des fibres végétales et synthétiques sont totalement différents, et ceci n'a pas été pris en compte dans ces études.

Au regard de ce bref contexte de l'état de l'art sur les matériaux composites à fibres végétales, il s'avère nécessaire de se pencher sur l'étude des mécanismes d'endommagement de ce type de matériaux afin d'évaluer leurs durabilités. Comparées à d'autres fibres végétales, en termes de propriétés mécaniques, la fibre de lin est un candidat idéal. De surcroît, le lin est une ressource disponible localement, et dont la France est le premier producteur européen. En effet, la France représente 75% des surfaces européennes implantées de lin [61]. En effet, il s'agit de valoriser une ressource locale, renouvelable et justifier son utilisation pour mettre en œuvre un matériau avec un faible impact sur l'environnement. Ainsi, cette partie de la présente étude doctorale porte sur l'étude des mécanismes d'endommagement au cours des essais mécaniques de deux matériaux composites non hybrides (à fibres verre et à fibres lin) et un matériau composite hybride lin-verre. Afin d'accomplir cet objectif, les parties suivantes du présent chapitre dressent une étude bibliographique non exhaustive des différents travaux de recherche qui ont étudié les mécanismes d'endommagement des matériaux composites fibreux.

2.3 Mécanismes d'endommagement de matériaux composites fibreux

Cette partie est consacrée aux principaux mécanismes d'endommagement provoquant la ruine totale d'un matériau composite. Ce dernier est constituée de couches successives de fibres unidirectionnelles ou bidirectionnelles imprégnés de résine. Les couches sont également nommées plis. La compréhension des mécanismes d'endommagement d'un stratifié est un élément clé dans le développement des matériaux composites, notamment ceux à fibres naturelles.

Dans les matériaux composites stratifiés, plusieurs mécanismes d'endommagement peuvent intervenir. Il s'agit notamment des mécanismes liés à la rupture des fibres, à la fissuration de la matrice et ceux couplés aux fibres et à la matrice. Ces mécanismes d'endommagement sont les plus couramment identifiés par les chercheurs et sont présentés ci-dessous en détails. Nous distinguons les mécanismes d'endommagement intervenant à l'intérieur d'un pli du stratifié nommés intra-pli (section 2.6.1) et ceux entre plis du stratifié, nommés inter-pli(section 2.6.2).

2.3.1 Les mécanismes d'endommagement intra-pli

La rupture finale d'un composite unidirectionnel est le résultat de l'accumulation de divers mécanismes d'endommagement élémentaires :

— La rupture des fibres,

— La rupture transverse de la matrice,

— La rupture longitudinale de la matrice,

— La rupture de l'interface fibre-matrice (décohésion).

Généralement, un mécanisme n'est pas isolé, mais divers mécanismes coexistent. Ces mécanismes se développent suivant la nature des matériaux et les conditions de sollicitations mécaniques imposées.

a-Fissuration de la matrice et de l'interface fibre-matrice : dans un matériau composite UniDirectionnels (UD), la fissuration de la matrice peut se produire, soit par fissuration transverse (figure 2.2-a) lorsque la contrainte en traction dans la matrice atteint la contrainte à la rupture de la matrice, soit par fissuration longitudinale (figure 2.2-b) lorsque la contrainte de cisaillement dans la matrice atteint la contrainte en cisaillement à la rupture. Ce dernier mode de rupture se produit lorsque la contrainte de décohésion est supérieure à la contrainte en cisaillement à la rupture de la matrice. Dans le cas contraire, il se produit la rupture par décohésion de l'interface fibre-matrice (figure 2.2-c).

Dans le cas d'un composite UD soumis à une traction longitudinale, l'initiation de la rupture se produit généralement soit par rupture des fibres lorsque la déformation à la rupture des fibres est plus faible que celle de la matrice, soit par rupture transverse de la matrice dans le cas contraire. La propagation de la fissuration (figure 2.3) peut produire deux mécanismes d'endommagement, l'extraction des fibres et la décohésion fibre-matrice.

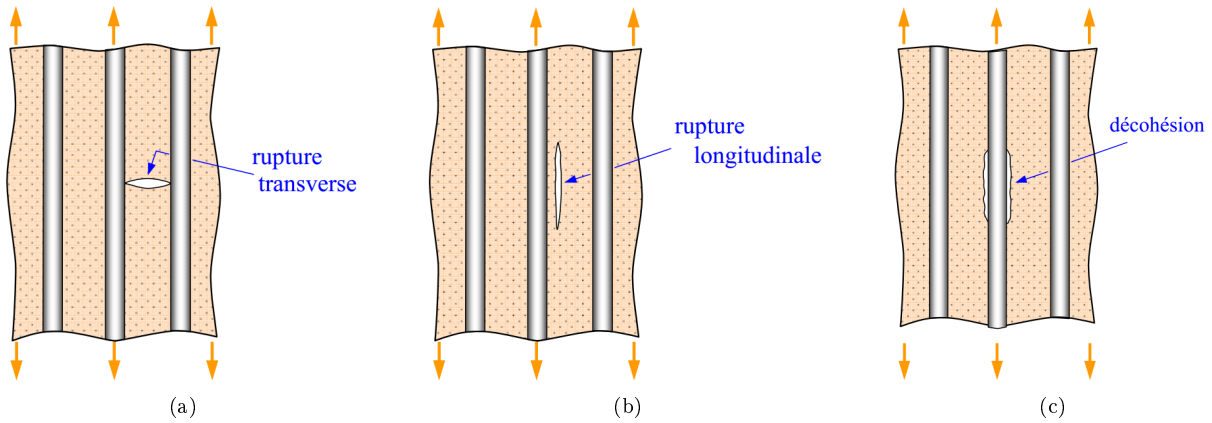


FIGURE 2.2 – Propagation de la rupture dans d'un composite à fibres unidirectionnelles. -a- rupture transverse, -b- rupture longitudinale, -c- décohésion fibre-matrice [3].

b-Rupture de fibres : dans un matériau composite UD soumis à des sollicitations mécaniques, la rupture des fibres intervient lorsque la contrainte de traction axiale dans les fibres dépasse la contrainte maximale admissible. Ce type de rupture est illustré schématiquement dans la figure 2.4.

La rupture finale d'un matériau composite UD est le résultat de l'accumulation des mécanismes cités ci-dessus.

2.3.2 Les mécanismes d'endommagements inter-pli

Aux mécanismes élémentaires décrits précédemment (rupture de fibres, rupture longitudinale de la matrice, rupture transverse de la matrice, décohésion fibre-matrice), s'ajoute (figure 2.5) un mécanisme d'endommagement entre les plis, appelé rupture par délaminage, dans le cas des stratifiés. Le délaminage entre les couches est une conséquence du développement de la fissuration de la matrice à l'interface fibre/matrice. En effet, la densité de fissure augmente dans le pli avec l'augmentation de la sollicitation. Les plis adjacents subissent alors des endommagements inter-laminaire avec une propagation de la fissure qui se développe d'un pli à l'autre, provoquant le délaminage entre deux plis. Les mécanismes de rupture induits dépendent de la nature des constituants, de l'architecture des couches et du mode de sollicitation imposé.

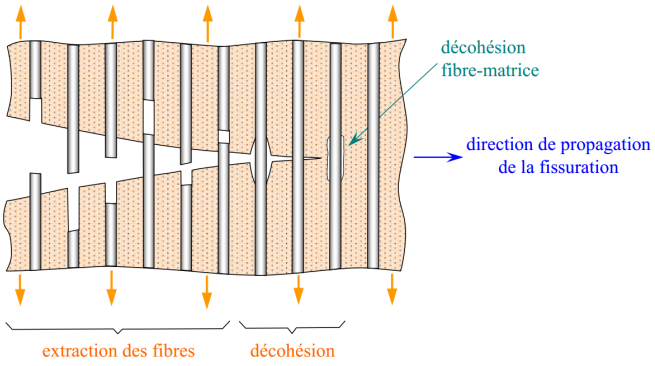


FIGURE 2.3 – Propagation de la rupture dans le cas d'une faible adhérence fibre-matrice [3].

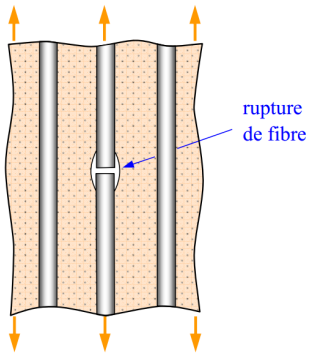


FIGURE 2.4 – Rupture de fibre [3].

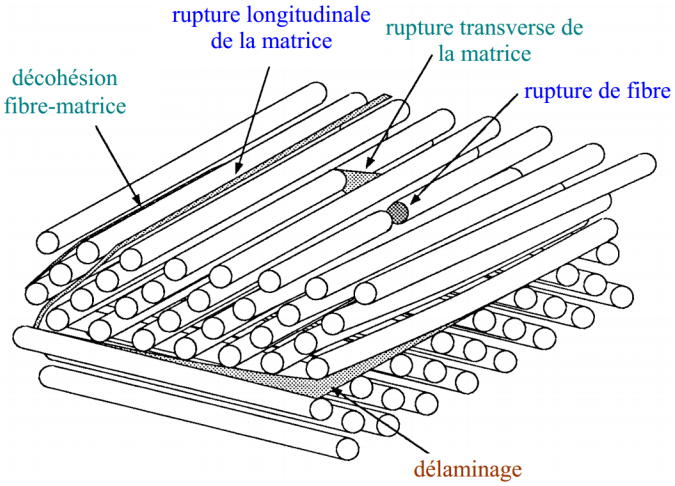


FIGURE 2.5 – Mécanismes d'endommagements observés dans les stratifiés [3].

Cette introduction sur les modes de ruptures des composites à fibres UD a permis de référencer les principaux mécanismes d'endommagement qui peuvent se produire lors d'une sollicitation mécanique. On constate qu'il y a des mécanismes d'endommagement qui touchent les fibres (rupture), la matrice (fissurations longitudinale et transversale) et l'interface fibres/matrice (décohésion). Il y a aussi des mécanismes couplés aux fibres et à la matrice. Dans ces derniers, lorsqu'un mécanisme d'endommagement a lieu au niveau des fibres, des mécanismes liés à la matrice se produisent, et vice versa. Finalement, il y a le mécanisme d'endommagement qui touche les plis d'un matériau composite (délaminage).

Ces mécanismes sont divisés en 3 catégories :

— Tous les mécanismes d'endommagement qui touchent spécialement les fibres sont nommés 'Endommagement des fibres'.

— Tous les mécanismes d'endommagement qui touchent individuellement la matrice sont nommés 'Endommagement de la matrice'.

— Tous les mécanismes d'endommagement qui touchent les interfaces fibres/matrice sont nommés 'Endommagement des interfaces fibres/matrice'.

2.4 Identification de mécanismes d'endommagements à partir des signaux d'émission acoustique

2.4.1 Présentation de la technique d'émission acoustique

L'Emission Acoustique (EA) est une technique de Contrôle Non Destructive (CND). Cette technique permet de déceler en temps-réel l'évolution de l'endommagement dans un matériau. C'est une technique

non destructive, ne nécessitant pas de signal excitateur. Généralement, l'EA est définie comme des ondes élastiques transitoires produites par une redistribution soudaine de la contrainte dans un matériau [62–64] . Lorsqu'une structure est soumise à une contrainte externe, des modifications internes ont lieu comme l'amorçage et la propagation de fissures, la déformation plastique locale, la corrosion, les transformations de phase, les frottements, etc. Ces sources, appelées aussi événement, libèrent de l'énergie sous forme d'ondes de contrainte dans le matériau, qui se propage comme une onde élastique à travers le matériau. Pendant la propagation, l'onde d'EA pourra subir des effets d'atténuation. Comme discuté dans [65], parmi les éléments qui peuvent éventuellement contribuer à ces effets sont l'architecture du composite et l'orientation des fibres.

Deux principaux types de signaux d'EA sont généralement observés, les signaux transitoires et continus. Les signaux transitoires sont des événements micro-mécaniques individuels bien séparés dans le temps. Ils peuvent être clairement distingués du bruit. Ces salves sont caractérisées par une courte durée et une amplitude plus élevée que le bruit de fond. Les sources de ces signaux peuvent être les ruptures micro et macroscopiques, les frottements et les impacts mécaniques. Les signaux d'EA continus quant à eux sont étroitement espacés et caractérisés par une faible amplitude. Ces signaux sont chevauchés dans le temps et il n'est pas possible de les séparés en salves. Ils sont principalement observés dans les matériaux métalliques (déformation plastique).

Dans la technique de CND par EA, les ondes générées par une source sont détectées au moyen de capteurs piézoélectriques [66] en contact direct avec la surface du composite qui convertissent les mouvements de surface en signaux électriques. Le capteur convertit l'énergie libérée en signal électrique, qui est transféré à un équipement électronique pour un traitement ultérieur. La figure 2.6 illustre le principe de la technique d'EA.

Les trois principales utilisations de la technique d'EA sont les suivantes : 1) la localisation des sources, 2) la caractérisation des matériaux (performance mécanique et évaluation des mécanismes d'endommagement) et 3) la surveillance de l'état de santé des structures et des machines. Cette partie du présent travail s'inscrit dans la troisième utilisation de la technique d'EA.

Les autres techniques de CND (radiographie, ultrasons, courants de Foucault, etc.) détectent des discontinuités géométriques en introduisant une certaine forme d'énergie dans la structure étudiée (techniques actives). En revanche, pour la technique par EA, les défauts sont auto-générateurs des ondes de contrainte (technique passive). Ces ondes peuvent être détectées par les capteurs piézoélectriques placés sur la surface du matériau. Les caractéristiques des signaux acquis sont utiles pour identifier les mécanismes d'endommagement qui se produisent dans le matériau.

Depuis la première application de la technique d'EA, il a été démontré que l'activité acoustique est

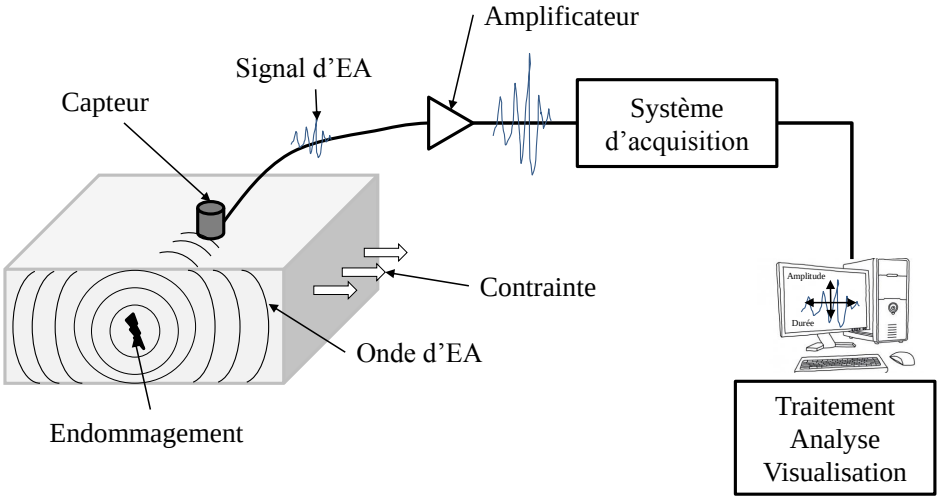


FIGURE 2.6 – Principe de base de la technique d'EA.

directement liée aux mécanismes d'endommagement. Par conséquent, de nombreuses études ont été réalisées dans le but de trouver la meilleure méthode qui fournit l'information pour aboutir à une identification et à une classification précises des mécanismes d'endommagement.

2.4.2 Identification des mécanismes d'endommagement par émission acoustique : état de l'art

De nos jours, de nombreuses méthodes d'analyse et de traitement des signaux d'EA ont été étudiées pour identifier les mécanismes d'endommagement des matériaux composites à renforts fibreux (fibres usuelles et/ou naturelles). D'une manière générale, le principe de ces méthodes peut être scindé en deux étapes, à savoir l'extraction des caractéristiques et la classification. L'extraction de caractéristiques vise à identifier les paramètres distinctifs des signaux d'EA. Cette étape a une incidence directe sur la pertinence de la classification finale. D'autre part, la classification vise à regrouper les signaux en se basant sur leurs caractéristiques. Les signaux du même groupe doivent être homogènes et représentent un seul mécanisme d'endommagement. Cependant, le défi est de savoir comment affecter un groupe à un mécanisme d'endommagement. L'extraction de caractéristiques à partir de signaux d'EA est une tâche complexe et difficile en raison de l'atténuation et la modification des signaux pendant la propagation dans des structures anisotropes telles que les matériaux composites. Différentes approches basées sur les signaux d'EA ont été développées dans le but de détecter, identifier et classifier les différents mécanismes d'endommagement de matériaux composites à renfort fibreux. Dans cette section, une revue de la littérature sur les travaux de recherche les plus importants sur l'identification des mécanismes d'endommagement est présentée. Celle-ci est divisée en trois, d'abord sur les composites

à fibres usuelles, ensuite sur les composites à fibres végétales et enfin sur les composites hybrides.

a) Matériaux composites à fibres conventionnelles

a.1) Analyse temporelle

De nombreuses études ont montré que les mécanismes sources des signaux d'EA peuvent être identifiés en utilisant les paramètres temporels des signaux enregistrés [9–15]. Ensuite, ces signaux sont généralement analysés dans le domaine temporel en considérant un ou plusieurs descripteurs. Enfin, chaque mécanisme d'endommagement est associé à un intervalle donné du descripteur en question. Plusieurs travaux de recherche utilisant cette démarche ont montré que les mécanismes d'endommagement de certains matériaux composites peuvent être classés par amplitude croissante : fissurations matricielles, décohésions, délaminages et ruptures de fibres (tableau 2.2). Par contre, leurs résultats étaient différents du fait que l'amplitude dépend de la morphologie des fibres (courtes ou longues), l'orientation des fibres, de la géométrie de l'échantillon et de l'emplacement des capteurs. Par conséquence, il est également intéressant de considérer d'une manière simultanée d'autres descripteurs pour améliorer la caractérisation des signaux d'EA. Ainsi, d'autres travaux de recherche se sont intéressés à l'analyse simultanée de plusieurs paramètres temporels pour associer aux signaux d'EA les mécanismes d'endommagement correspondants [27, 67]. A titre d'exemple, Huguet et al. [27] et Godin et al. [68] ont procédé à une analyse conventionnelle simultanée de plusieurs paramètres temporels (amplitude, temps de montée, durée et énergie) lors d'un essai de traction sur des composites UD à matrice polyester et à matrice époxyde renforcées par des fibres de verre. Ces auteurs ont identifié 4 catégories dénommées types A, B, C et D. Les principales caractéristiques de ces signaux sont présentées dans le tableau 2.3 et leurs formes d'ondes sont données sur la Figure 2.7. Les signaux de type A présentent des amplitudes comprises entre 50 et 70dB. Ces signaux sont caractérisés également par des formes d'onde de faible durée, un temps de montée assez lent et une énergie relativement faible. S'agissant des signaux de type B, ils ont des amplitudes comprises entre 70 et 100dB et leurs formes d'onde présentent des temps de montée assez courts et des énergies plus élevées. Concernant les signaux de type C, ils sont très proches de ceux de type B avec des amplitudes et énergies moins importantes. Enfin, les signaux de type D ont des durées très longues, des temps de montée lents, ils sont assez énergétiques avec une gamme d'amplitude assez large.

Tableau 2.2 – Valeurs d'amplitude (en dB) de différents mécanismes d'endommagement pour différents matériaux composites à fibres synthétiques.

Auteur	Matériau composite (fibre/matrice)	Type de renfort	Fissurations matricielles	Décohésion	Délaminage	Déchaussement des fibres	Rupture des fibres
Kotsikos et al. [9]	Verre/Polyester	UD	40-50	-	55-70	-	>80
Kim et al. [10]	Carbone/Epoxy	tissé uni	40-70	-	-	-	60-100
Chen et al [11]	Carbone/Verre	fibre courte	60-80	-	-	-	70-90
Ceysson et al. [12]	Carbone/Epoxy	UD	50	-	62	-	-
Barré et al. [13]	Verre/Polypropylène	fibre courte	40-55	60-65	-	65-85	85-95
Zhuang et al. [14]	Polyéthylène/Polyéthylène	UD	30-60	30-45	60-85	60-80	80-97

Tableau 2.3 – Caractéristiques des signaux de type A, B, C et D des endommagements d'un composite UD à fibre de verre sollicité en traction [4].

Endommagement	Temps de montée	Durée du signal	Décroissance du signal	Type du signal
Fissurations matricielles (A)	Moyen	Moyenne	Lente	Résonnant
Décohésion (B)	Court	Moyenne	Courte	Impulsionnel
Rupture des fibres (C)	Court	Courte	Courte	Impulsionnel
Délaminage (D)	Lent	Très longue	Lente	Résonnant

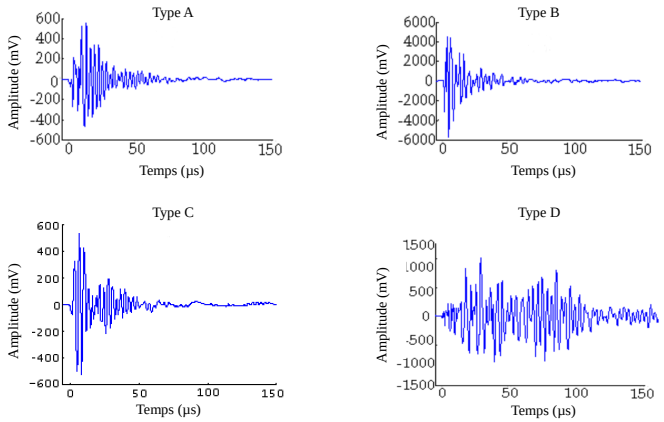


FIGURE 2.7 – La forme d'onde de signaux d'EA (Signaux de type A, B, C et D) des endommagments d'un matéraiu composite UD à fibre de verre sollicité en traction [4].

a.2) Analyse fréquentielle et temps-fréquence

Les parametres temporelles ne sont pas toujours pertinents pour caractériser les signaux d'EA, notamment pour des composites de nature complexe. Ainsi, d'autres travaux ont été menés dans le domaine fréquentiel [69–75] afin de caractériser les signaux d'EA par leurs contenus fréquentiels. Ces études ont montré que chaque mécanisme d'endommagement peut être caractérisé par un pic fréquentiel spécifique (Tableau 2.4) différent, obtenu par l'analyse de Fourier. Par contre, le pic de fréquence est toujours déterminé à partir d'un point unique dans le spectre de fréquence (le point de grandeur maximale). Ce pic donnerait une description appropriée du contenu fréquentiel uniquement si le signal contient une seule composante de fréquence, ce qui est rarement le cas pour les signaux d'EA [76].

Puisque les signaux d'EA sont de natures non stationnaires, une simple transformée de Fourier ne peut pas fournir des informations sur l'évolution des composantes fréquentielles en fonction du temps de telles signaux.

Pour surmonter cette limitation, les techniques d'Analyse Temps-Fréquence (ATF) sont utilisées pour obtenir des informations supplémentaires sur le comportement d'un signal. Pour les signaux d'EA, les techniques typiques utilisées pour étudier ce comportement sont les suivantes : la Transformée de Fourier à Court Terme (TFCT), la Transformée en Ondelettes Discrète (TOD), la Transformée en Ondelettes Continue (TOC) et la Transformée de Hilbert Huang (THH) [77–82]. D'un côté, ces techniques sont utilisées afin d'identifier les intervalles de fréquences de signaux d'EA des principaux mécanismes d'endommagement des composites à fibres usuelles (tableau 2.5). D'un autre coté, ces techniques sont utilisées dans le but d'extraire des descripteurs T-F. Ces derniers sont combinés avec les descripteurs temporels afin de classifier les signaux par les techniques d'analyse statistique multi-variable.

Tableau 2.4 – Valeurs de fréquences (en kHz) des principaux mécanismes d'endommagement pour différents matériaux composites UD à fibres synthétiques. Les signaux ont été analysés par la FFT.

Auteur	Matériau composite (fibre/matrice)	Fissurations matricielles	Décohésion	Délaminage	Déchaussement des fibres	Rupture des fibres
DeGroot et al. [69]	Carbone/Epoxy	90-180	240-310	-	-	>300
Giordano et al. [70]	Carbone/Polyester	-	-	-	-	500
Arumugam et al. [71]	Verre/Epoxy	90-120	220-245	260-295	130-200	300-450
Ramirez et al. [72]	Verre/Polypropylène	100	200-300	200-300	-	>400
Gutkin et al. [73]	Carbone/Polymère	50-100	200-300	50-150	500-600	400-500
Eaton [74]	Carbone/Epoxy	<150	-	-	-	>400
Sause et al. [75]	Carbone/Epoxy	0-250	250-450	-	-	450-800

Tableau 2.5 – Valeurs de fréquences (en kHz) des principaux mécanismes d'endommagement pour différents matériaux composites à fibres synthétiques. Les signaux ont été analysés par la TOC, TOD, THH ou TFCT.

Auteur	Matériau composite (fibre/matrice)	Type de renfort	Fissurations matricielles	Décohésion	Délaminage	Rupture des fibres
Qi et al. [77]	Carbone/Polymère	UD	50-150kHz	150-250	-	250-310
Oskouei et al. [78]	Verre/Polyester	UD	125-200	250-310	-	375-440
Han et al. [80]	Carbon/Epoxy	Tissé	105-145	150-240	-	275-404
Ni et al. [81]	Carbone/Epoxy	UD	<100	200-300	-	400-450
Hamdi et al. [82]	Verre/Polymère	UD	150	180-280	30-80	340-400

a.3) Analyse statistique multi-variable

Afin d'améliorer l'identification des mécanismes d'endommagement, plusieurs études [16–29] ont mis l'accent sur la nécessité d'utiliser des méthodes statistiques multi-variables pour analyser les signaux d'EA. Ces méthodes permettent de segmenter les vecteurs de descripteurs en groupes homogènes. Les éléments de chaque groupe doivent correspondre à un et un seul mécanisme d'endommagement. Ensuite, chaque groupe doit être corrélé avec un mécanisme d'endommagement. Le tableau 2.6 résume les différentes méthodes statistiques multi-variables qui ont été utilisées pour identifier les mécanismes d'endommagement des composites à fibres usuelles.

b) Matériaux composites à fibres végétales

Influencées par les études menées sur les composites à fibres usuelles, les études menées sur les matériaux composites à fibres végétales ont suivi la même direction. Ainsi, les approches employées sur les composites à fibres usuelles ont été aussi utilisées dans le cas des composites à fibres végétales. Le tableau 2.7 présente les intervalles d'amplitude de signaux qui ont été attribués à chaque mécanisme d'endommagement. À partir de ce tableau, on remarque que les amplitudes des signaux d'EA sont toujours classées par ordre croissant. Les signaux de faibles amplitudes sont toujours corrélés avec les mécanismes d'endommagement de la matrice, à savoir la fissuration matricielle et frottement matrice/matrice. Les signaux de moyennes amplitudes quant à eux sont toujours attribués aux mécanismes d'endommagement de l'interface fibre/matrice comme la décohésion et les frottements matrice/fibre. Les signaux de grandes amplitudes sont attribués aux mécanismes d'endommagement liés à la fibre comme la déchaussement et la rupture des fibres.

Pourtant, l'étude menée par Romhany et al. [83] sur les mécanismes d'endommagement des fibres élémentaires de lin prouve que l'amplitude seul n'est pas un paramètre pertinent pour caractériser un signal d'EA. Dans cette étude, les auteurs ont identifié trois mécanismes d'endommagement, (1) les fissurations longitudinales des fibres, (2) les fissurations transversales des fibres et (3) la rupture des fibres. Aussi, ils ont proposés une gamme d'amplitude à chaque mécanisme d'endommagement (1) : <35dB, (2) : 35-60dB et (3) >60dB. A partir de cette étude, on constate que les mécanismes d'endommagement des fibres peuvent être source de toutes gammes d'amplitudes (faibles, moyennes et grandes). Ainsi, utiliser uniquement les gammes d'amplitudes, afin de corréler les signaux d'EA aux mécanismes d'endommagement, n'est pas forcément un choix judicieux.

Tableau 2.6 – Différentes approches utilisées pour la classification de signaux d'EA et l'identification des mécanismes d'endommagement de matériaux composites à fibres synthétiques.

Auteur	Matériau composite (fibre/matrice)	Type de renfort	Descripteur	Méthode
Pappas et al. [16]	Carbone/Carbone	UD	Amplitude, Temps de montée, énergie, nombre de coups et durée	k-moyenne
Pashmforoush et al. [29]	Verre/Polyester	UD	Temps de montée, énergie, nombre de coups, durée et fréquence moyenne	k-moyenne, ACP et Harmony search
Kempf et al. [28]	Verre/Epoxy	UD	pic de fréquence, pic de fréquence pondérée et trois puissances partielles	k-moyennes et ACP
Huguet et al. [27]	Verre/Polyester	UD	Amplitude, Temps de montée, énergie, nombre de coups, durée le nombre de coups pendant la montée	CAO de Kohonen
Al-Jumaili et al. [26]	Carbone/Polymère	UD	Amplitude, Temps de montée, énergie, nombre de coups, durée, pic de fréquence, fréquence moyenne et Fréquence Centroid	k-moyennes, c-moyenne floue, Hiérarchique regroupement et ACP
Oliveira et al. [25]	Verre/Polyester	UD	Amplitude, Temps de montée, énergie et Centre de gravité du spectre	CAO de Kohonen et k-moyennes
Oskouei et al. [24]	Verre/Polyester	UD, Tissé	Pic de fréquence, temps de montée et durée	C-moyennes floue et ACP
Fotouhi et al. [23]	Verre/Epoxy	UD, Tissé	Descripteurs basés sur la décomposition en paquet d'ondelette	C-moyennes floue et ACP
Mohammadi et al. [22]	Verre/Epoxy	UD	fréquence moyenne, le nombre de coups, le pic de fréquence, amplitude, énergie et le temps de montée	C-moyennes floue, analyse par élément finie et la transformée d'odelette
Masmoudi et al. [21]	Verre/Epoxy	UD	Amplitude, Temps de montée, énergie, nombre de coups et durée	k-moyennes
Sause et al. [20]	Carbone/plastique	UD	Pic de fréquence, pic de fréquence pondérée et puissance partielle	k-moyennes
Wang et al [19]	Polyéthylène/Polyéthylène	UD	Amplitude, pic de fréquence et durée	Hiérarchique regroupement, k-moyennes et ANN
Li et al. [18]	Verre/Epoxy	Tissé	Amplitude et pic de fréquence	k-moyennes++ et ACP
Heidary et al. [17]	Verre/Epoxy	Tissé	Amplitude, énergie, nombre de coups, durée, temps de montée et fréquence moyenne	c-moyenne floue et ACP
Boussetta et al. [84]	Verre/Polyester	UD	Energie, Temps de montée et durée	k-moyennes, COA de Kohonen

Tableau 2.7 – Valeurs d'amplitude (en dB) de différents mécanismes d'endommagement pour différents matériaux composites à fibres végétales.

Auteur	Bravo et al. [85]	El Mahi et al. [30]	Czigány et al. [86]	Elouaer et al. [87]	Aslan [88]	Romhany et al. [83]	Dogossy et al. [89]
Matériau composite (fibre/matrice)	Bouleau/ Polyéthylène	Lin/ Epoxy	Basalt/ Polypropylène	Chanvre/ Polypropylène	Lin/ LPET	Lin/ -	Maïs/ Polyéthylène
Type de renfort	Fibre courte	UD	Fibre longue	fibre courte	fibre longue	fibre élémentaire	fibre longue
Fissuration matricielle	35-45	42-60	21-30	40-60	40-50	-	<25
Frottement matrice/matrice	40-55	-	35-45	-	40-55	-	-
Décohésion	45-60	60-70	45-60	60-65	45-60	-	-
Frottement matrice/fibre	55-85	-	-	65-85	-	-	-
Déchaussement des fibres	-	-	-	64-85	60-80	-	26-40
Rupture des fibres	-	70-100	60-100	85-95	80-96	>60	>45
Fissuration longitudinale des fibres	-	-	-	-	-	>35	-
Fissuration Transversale des fibres	-	-	-	-	-	35-60	-

Tableau 2.8 – Différentes approches utilisées pour la classification de signaux d'EA et l'identification des mécanismes d'endommagement de matériaux composites à fibres végétales.

Auteur	Matériau composite (fibre/matrice)	Type de renfort	Descripteur	Méthode
Assarar et al. [36]	chanvre/ Polypropylène	UD	amplitude, durée, temps de montée et nombre de coups	k-moyennes et ACP
Monti et al. [35]	Lin/ Thermoplastique	UD	Amplitude, temps de montée, durée, énergie et nombre de coups	k-moyennes
Saidane et al.[34]	verre-lin/ époxyde	UD	Amplitude, temps de montée, durée et énergie	k-moyennes
Assarar et al. [33]	Lin/ époxyde	UD	Amplitude, énergie, durée, temps de montée et nombre de coups	c-moyenne floue et ACP
Habibi et al. [32]	Lin/ époxyde	fibre courte et UD	Amplitude, durée et nombre de coups	k-moyennes
Chilali et al. [31]	Lin/ thermoplastique	Tissé	Amplitude, durée, temps de montée, nombre de coups et énergie	k-moyennes et ACP
El Mahi et al. [30]	Lin/époxyde	UD	Amplitude, énergie, nombre de coups, durée et temps de montée	k-moyennes

D'autres études ont exploité l'analyse statistique multi-variable pour classifier les signaux d'EA des matériaux composites à fibres végétales. Le tableau 2.8 donne une liste non exhaustive de ces études. Ce tableau montre que la méthode de k-moyenne est largement utilisée pour classifier les signaux d'EA et elle est souvent associée aux paramètres temporels comme descripteurs. D'autre par, la corrélation entre les groupes identifiés, via la méthode des k-moyennes, et les mécanismes d'endommagement correspondants est souvent réalisée en s'appuyant sur les résultats de la littérature. Cette corrélation a été faite spécialement par une comparaison entre les gammes d'amplitude ou de fréquence de signaux d'EA de chaque groupe, avec ceux de signaux typiques de quelques mécanismes d'endommagement. Cette procédure adoptée dans la plupart des études est discutable du fait que l'amplitude la fréquence de pic ne sont pas les seuls descripteurs pertinents pour caractériser les signaux d'EA.

c) Matériaux composites hybrides

La plupart des études sur les matériaux composites hybrides ont portée sur l'effet de l'hybridation du renfort fibreux sur les propriétés mécaniques. Cependant, ces composites hybrides peuvent présenter plusieurs mécanismes d'endommagement lorsqu'ils sont soumis à des sollicitations mécaniques. C'est dans ce contexte que quelques travaux de recherche ont utilisé la technique d'EA afin de caractériser ces mécanismes intervenant dans des composites hybrides renforcés de fibres naturelles [34, 90, 91]. Ces travaux se sont basés sur les mêmes techniques appliquées sur les composites à fibres synthétiques et végétales. À titre d'exemple, Saidane et al. [34] ont étudié les mécanismes d'endommagement d'un matériau composite hybride lin-verre associé à une matrice époxyde avant et après le vieillissement. Les essais mécaniques ont été couplés avec la technique de l'EA. La méthode de k-moyennes, basée sur les descripteurs temporels, a été utilisée pour regrouper les signaux d'EA. Une approche basée sur la minimisation du taux de chevauchement entre les groupes a été utilisée pour optimiser le nombre de paramètres et justifier le choix du nombre de groupes. Des observations microscopiques à balayage électronique ont été effectuée sur les échantillons testés afin d'identifier les mécanismes d'endommagement. Pour les matériaux non vieillis et vieillis, le taux de chevauchement a été minimal, pour trois et quatre classes, respectivement. En se basant sur les résultats de la littérature, à savoir la distribution de l'amplitude, les classes ont été corrélées avec les fissurations matricielles, la décohésion, la rupture des fibres et le délaminage. Ce dernier a été détecté seulement dans le matériau vieilli. Dans leur conclusion, les auteurs ont mis l'accent sur la limite de la méthode de classification adoptée. Cette méthode permet d'identifier les mécanismes d'endommagement typique du matériau composite hybride étudié. Cependant, elle ne permet pas de distinguer séparément les mécanismes d'endommagement liés aux fibres de lin fibres de verre.

2.5 Problématiques

Cette synthèse des travaux de la littérature montre que l'identification des mécanismes d'endommagement à partir de leur signature acoustique se fait en quatre étapes :

1. Les essais mécaniques couplés avec la technique de l'EA pour collecter les signaux d'EA.
2. L'extraction de vecteurs de caractéristiques significatifs : temporelles, fréquentielles et T-F qui permettent de séparer les signaux d'EA.
3. Le regroupement des signaux collectés en groupe, via une méthode de classification non-supervisée à partir des descripteurs identifiés dans l'étape 2.
4. La corrélation entre les groupes identifiés et les mécanismes d'endommagement correspondants.

À l'issue de cette étude bibliographique non exhaustive sur différentes approches, l'utilisation de la technique de l'EA pour identifier les mécanismes sources d'endommagement des matériaux composites fibreux (synthétiques, végétales et hybrides), nous a permis de discerner huit problèmes :

1. Du fait que les signaux d'EA sont de nature non-stationnaire et multi-composantes, le pic de fréquence ne permet pas de décrire leurs contenus fréquentiels de manière significatif.
2. L'amplitude seule n'est pas un descripteur pertinent pour classifier les signaux d'EA. En effet, l'utilisation de l'amplitude comme seul descripteur engendre parfois des résultats qui présentent des recouvrements. En effet, l'amplitude dépend de plusieurs paramètres comme la géométrie des éprouvettes, la localisation des capteurs et l'orientation des fibres. En plus, les études qui se sont focalisées sur les fibres élémentaires de lin ont montré que les mécanismes d'endommagement des fibres peuvent être associés à toutes les gammes d'amplitude.
3. Pour améliorer la classification des signaux d'EA, différents descripteurs temporels, fréquentiels et T-F ont été utilisés. Par contre, une mesure quantitative de la pertinence de chaque descripteur n'a pas été effectuée. La pertinence des descripteurs a été mesurée dans la plupart des cas de manière qualitative. En revanche, la réussite d'une méthode de classification est fortement liée aux pertinences des descripteurs.
4. Les méthodes de classification non-supervisées sont les plus utilisées pour la classification de signaux d'EA. La qualité de ces méthodes a été mesurée dans la plupart des cas qualitativement (observation microscopique, évolution du nombre de signaux de chaque classe en fonction du temps de l'essai).
5. La corrélation entre les groupes obtenus par la méthode de classification et les mécanismes d'endommagement a été obtenue dans la plupart des études en se basant sur les résultats de la littérature, à

savoir la distribution de l'amplitude.

6. Il n'y a pas de base de données d'apprentissage de signaux d'EA. En revanche, une base d'apprentissage s'avère primordiale pour mesurer la pertinence des descripteurs ou pour évaluer les performances d'une méthode de classification.
7. Toutes les approches utilisées pour l'identification des mécanismes d'endommagement des matériaux composites à fibres synthétiques ont été directement appliquées sur les matériaux composites à fibres naturelles. En revanche, la structure et les compositions chimiques des deux types de fibres sont par définition différentes, ainsi que leurs propriétés mécaniques. Ces différences n'ont pas été prises en compte.
8. Les approches actuelles de classification de signaux d'EA ne permettent pas de discriminer séparément les mécanismes d'endommagement liées à chaque fibre dans les matériaux composites hybrides à fibres végétales-fibres synthétiques.

À noter que les problèmes de 1 à 6 sont aussi valables pour les matériaux composites à fibres végétales et les matériaux composites hybrides.

Ainsi, les limites des approches actuelles de classification de signaux d'EA peuvent être résumées en trois points :

1. L'utilisation des méthodes non-supervisées.
2. L'étape d'identification des classes obtenues.
3. La pertinence des descripteurs.

Dans le but d'améliorer la classification de signaux d'EA, l'objectif de la première partie de cette thèse est de proposer des solutions adéquates pour les trois limites cités ci-dessus. La solution intuitive pour les limites 1 et 2 est l'utilisation des méthodes de classification supervisée. Par ailleurs, les méthodes supervisées nécessitent une base de données pour l'apprentissage. Or, à notre connaissance, aucune base de données d'apprentissage de signaux d'EA fiable n'est proposée dans la littérature. Donc, le premier objectif de notre travail est de construire cette base de données. Pour ce faire, nous avons développé une méthode de regroupement qui s'appelle la méthode incrémentale. Le principe de cette méthode et les résultats de son application sur les signaux d'EA sont présentés dans le chapitre suivant.

Une solution adéquate pour lever la limite 2 est l'utilisation d'une méthode de classification sans extraction de caractéristiques. Il s'agit d'avoir une approche de classification basée sur une mesure de dissimilarité. Cette solution a été développée et validée sur une base de données de signaux EEG d'épilepsie dans la deuxième partie de cette thèse. Son application sur les signaux d'EA, qui n'a pas pu être réalisée, constitue une des perspectives de ces travaux de thèse.

2.6 Conclusion

La revue bibliographique non exhaustive de ce chapitre a révélé les avantages que les fibres végétales, en particulier celles du lin, offrent par rapport aux fibres synthétiques. Ces fibres possèdent des propriétés mécaniques spécifiques comparables à celles du verre. C'est la raison pour laquelle plusieurs travaux de recherche ont émergés dans le développement des matériaux composites à fibres de lin. La plupart des études se sont intéressées au développement d'une filière de composites à fibres de lin en s'intéressant à leur comportement mécanique. D'autres se sont intéressés à l'analyse des mécanismes d'endommagement de ces matériaux composites par la technique d'EA. En effet, combiné la technique d'EA et l'analyse statistique multivariable permet d'analyser et identifier les mécanismes d'endommagement à l'aide d'une méthode classification basée sur les descripteurs d'un signal d'EA. Cependant, l'état de l'art sur les méthodes de classification et les descripteurs nous a permis de souligner plusieurs problématiques. Les méthodes de classification les plus utilisées pour classifier les signaux d'EA sont des méthodes non-supervisées. Ces dernières nécessitent une étape supplémentaire afin de corréler les groupes obtenus avec les mécanismes d'endommagement réels. Cette corrélation se fait la plupart du temps en se basant sur les résultats de la littérature, en particulier sur les gammes d'amplitude. La fréquence de pic et l'amplitude sont les plus utilisés pour caractériser un signal d'EA et pourtant ces descripteurs pris individuellement ne sont pas les seuls pertinents. En revanche, la réussite d'une méthode de classification est directement liée à la pertinence des descripteurs. Les méthodes de classification actuelles ne permettent pas d'identifier distinctivement les mécanismes d'endommagement d'un matériau composite hybride.

Pour toutes ces raisons, nous avons aussi choisi de contribuer à l'étude des mécanismes d'endommagement détectés par EA des matériaux bio-composites. Notre but est de proposer des solutions aux problématiques liées aux méthodes actuelles. Pour ce faire, nous avons choisi d'étudier trois types de matériaux composites, deux non hybrides (à fibres de verre et à fibres de lin) et un troisième hybride lin-verre. Dans une première démarche, nous avons proposé une méthode de regroupement afin de construire une vérité terrain de signaux d'EA. Dans la seconde démarche, la vérité terrain est combinée avec un classifieur supervisé afin de classifier les signaux d'EA des matériaux composites hybrides. Ces deux démarches se sont proposées afin de combler les problématiques liées à l'utilisation d'une méthode de classification non-supervisée et l'étape d'identification des groupes. Elles sont détaillées dans le chapitre suivant.

Chapitre 3

Analyse de l'endommagement des composites par une classification supervisée

Sommaire

3.1 Introduction **40**

3.2 Matériaux et procédure expérimentale **41**

3.2.1 Constituants 41

3.2.2 Mise en œuvre des matériaux composites 42

3.2.3 Essai mécanique et procédure d'acquisition des signaux d'EA 43

3.3 Démarche de classification **46**

3.3.1 Méthode incrémentale 47

3.3.2 Support Vector Data Description 50

3.3.3 Méthode de k-moyennes 52

3.3.4 Méthode de k-PPV 53

3.3.5 Mesures statistique pour évaluer les performances des classifieurs (Courbe ROC, Validation croisée) 54

3.4 Résultats **57**

3.4.1 Sélection des descripteurs d'EA 57

3.4.2 Construction de la base de données d'apprentissage 60

3.4.3 Performance de la base de données d'apprentissage. 65

3.4.4 Comparaison entre la méthode incrémentale et la méthode de k-moyennes 66
3.4.5 Application de la classification aux composite hybrides lin-verre 70
3.5 Conclusion 73

3.1 Introduction

Dans le chapitre précédent, l'étude bibliographique sur les matériaux composites à fibres naturelles a mis en évidence leurs avantages par rapport aux matériaux composites classiques. Cet état de l'art nous a permis aussi de choisir les matériaux de l'étude. Il s'agit des composites à fibres UD de lin, de verre et hybrides lin-verre. D'autre part, la deuxième partie de l'étude bibliographique sur les approches de classification de signaux d'EA, a révélé trois problématiques liées à ces approches. La première problématique est relative à la pertinence des descripteurs qui ont été utilisés pour caractériser les signaux d'EA. La deuxième et la troisième problématiques sont liées à l'utilisation des méthodes de classification non-supervisées et à l'étape d'identification des groupes obtenus par ces méthodes, respectivement. Ainsi, le but de ce chapitre est de proposer une solution adéquate à chacune de ces problématiques. Pour ce faire et dans un premier temps, les caractéristiques des matériaux composites de l'étude et les procédures expérimentales sont d'abord présentés. Les procédures expérimentales englobent, la méthode de mise en œuvre des composites, les tests mécaniques et les dispositifs de suivi de l'activité acoustique. Dans un second temps, nous proposerons une nouvelle méthode de classification de signaux d'EA afin de surmonter les limites des approches actuelles. Cette méthode a comme objectif la construction d'une vérité terrain de signaux d'EA. Cette vérité terrain représente des bases de données d'apprentissage de signaux d'EA associées aux différents types d'endommagement rencontrés dans les matériaux composites de l'étude. Les bases d'apprentissage seront combinées avec une méthode de classification supervisée afin d'établir un rôle de décision. Ce dernier sera utilisé afin de classifier des signaux d'EA venant d'un matériau composite où les mécanismes d'endommagement sont inconnus. Dans un dernier temps, dans la partie consacrée aux résultats, les démarches pour sélectionner les descripteurs d'EA les plus pertinents sont présentées. Ensuite, deux bases de données d'apprentissage sont construites par notre approche de classification et leurs performances sont évaluées. Après, une comparaison entre les résultats obtenus par notre méthode et la méthode de k-moyennes est donnée afin de mettre en évidence à la fois les avantages de notre approche et les inconvénients des approches actuelles. Enfin, la base d'apprentissage que nous avons élaborée est utilisée afin de discriminer les mécanismes d'endommagement de matériaux complexes à savoir les composites hybride lin-verre.

3.2 Matériaux et procédure expérimentale

3.2.1 Constituants

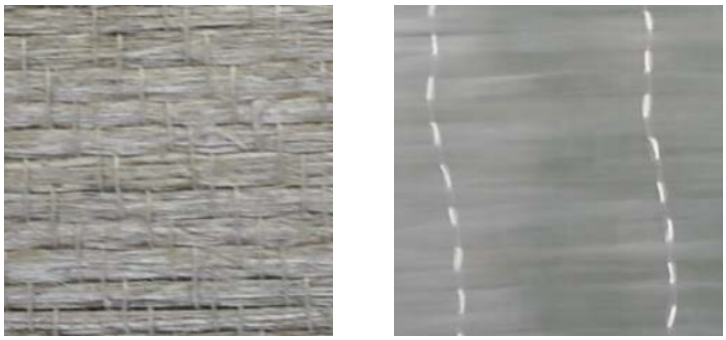
La matrice utilisée pour élaborer les composites de l'étude est la résine époxyde SR8100, associée au durcisseur SD8822. Ces deux produits sont commercialisés par la société SICOMIN.

Les cycles de polymérisation du mélange résine-durcisseur préconisés par le fabricant consistent d'abord à effectuer une réticulation de 24 heures à température ambiante suivie d'une deuxième de 24 heures à $40^{\circ}C$ ou de 16 heures à $60^{\circ}C$. Ces cycles de polymérisation permettent d'obtenir une réticulation complète du mélange résine-durcisseur. Les propriétés physiques et mécaniques du mélange résine-durcisseur, fournies par le fabricant, sont récapitulées dans le Tableau 3.1.

Les renforts fibreux utilisées dans ce travail sont fournis sous forme de rouleaux de tissus secs quasi UD. Les tissus de lin (Figure 3.1-a) et de verre (Figure 3.1-b) sont commercialisés, respectivement, par les entreprises DEPESTELLE et SICOMIN. Le Tableau 3.2 présente les propriétés mécaniques des deux fibres.

Tableau 3.1 – Propriétés de la résine époxyde (données du fabricant) [8].

Propriétés	Epoxyde SR 8100/SD 8822	Normes
Masse volumique (Kg/m^3)	1158	NF EN ISO 2811-1
Viscosité à $20^{\circ}C$ (M.Pa.s)	390±100	NF EN ISO 2811-1
Température de Transition vitreuse ($^{\circ}C$)	66	ISO 11357-2 :1900
Module d'Young (MPa)	3000	
Résistance maximale en traction (MPa)	70	NF T51-034
Résistance à la rupture en traction (MPa)	63	
Allongement à la rupture en traction (%)	3,8	



(a) (b)

FIGURE 3.1 – Renforts UD utilisés pour l'élaboration des composites de l'étude : (a) lin et (b) verre.

Tableau 3.2 – Propriétés mécaniques des fibres de lin et de verre.

Fibre	Module (GPa)	Contrainte à la rupture (MPa)	Allongement à la rupture (%)	Référence
Lin	12-85	600-2000	1-4	[92]
Verre	74	2500	4,8	Fabricant

3.2.2 Mise en œuvre des matériaux composites

Le procédé de l'infusion de résine est une technique qui utilise une pression sous vide pour infuser la résine [93]. L'emploi de la pression sous vide permet d'imprégner les fibres avec la résine et donne la forme finale aux composites au moyen d'un moule (figure 3.2). La dépression imposée au matériau permet aussi d'améliorer le rapport fibre/résine et expulse les bulles d'air du composite. Étant donné que les équipements spécifique à cette technique sont disponibles à l'IUT de Troyes, ceci nous a permis de fabriquer nous même les matériaux de l'étude par le procédé de l'infusion.

Pour ce faire, la préforme composée de différentes couches de tissus secs est drapée sur un moule rigide et étanche. Ensuite, un tissu d'arrachage est placé sur cette préforme pour faciliter son démoulage après durcissement. Enfin, un tissu drainant et micro-perforé est déposé sur l'ensemble afin d'améliorer la répartition de la résine et favoriser son écoulement. L'étanchéité de la préforme est assurée par une bâche qui recouvre l'ensemble (figure 3.2). L'aspiration de la résine est assurée également par divers accessoires : raccords, robinets d'arrêt en plastique, serpentin flexible, gaine pour recouvrement du serpentin et tuyaux de différents diamètres.

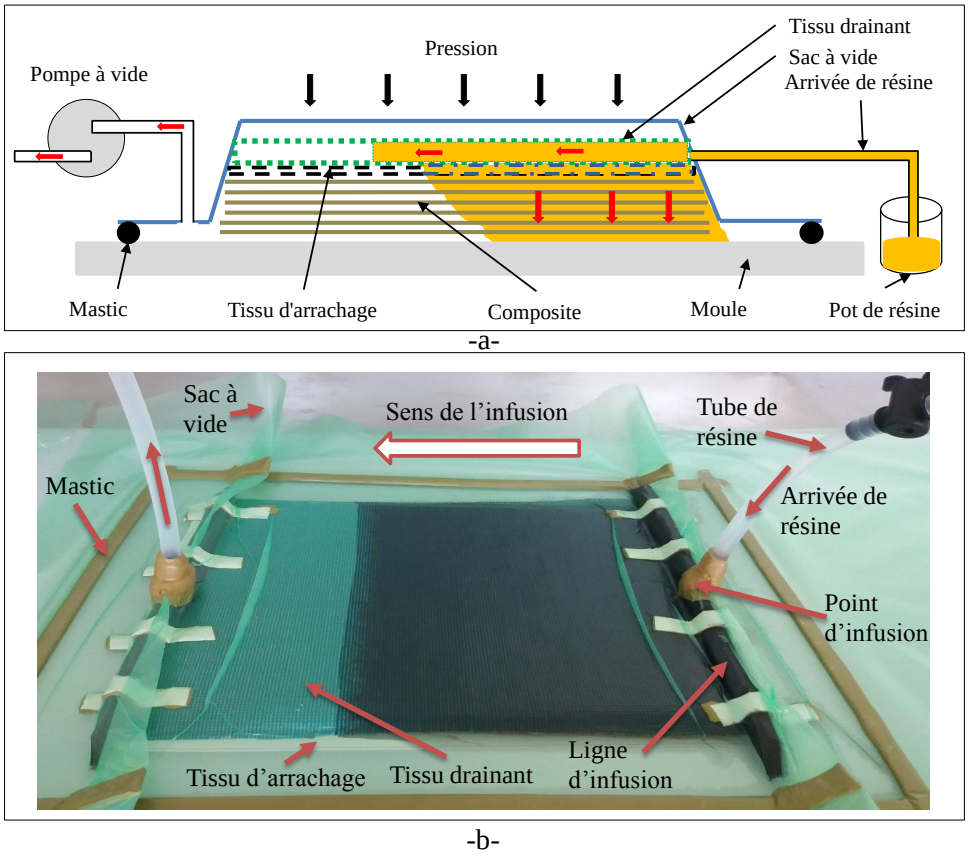


FIGURE 3.2 – Schéma d'explication (a) et montage expérimental (b) du procédé de l'infusion de résine.

Après 24h, les plaques composites ont été démoulées puis placées dans un four à 40°C durant 24h, afin d'obtenir une polymérisation complète. Ensuite, elles sont découpées sous forme d'éprouvettes rectangulaires de dimension 20x250x2mm à l'aide d'une tronçonneuse à diamant. Le tableau 3.3 présente les différents matériaux composites UD non hybrides élaborés ainsi que leurs caractéristiques principales et le tableau 3.4 renseigne les caractéristiques des matériaux hybrides.

3.2.3 Essai mécanique et procédure d'acquisition des signaux d'EA

a) Dispositif expérimental

Pour caractériser les composites non-hybrides et hybrides, des essais de traction ont été effectués sur des éprouvettes rectangulaires comme le préconise la norme ASTM D3039-76. Ces essais ont été menés avec une vitesse de 1 mm/min, en utilisant une machine universelle Instron 3382 de capacité 100 kN (Figure 3.3). Cette machine est dotée de deux extensomètres pour mesurer les déformations longitudinale et transversale. L'identification des principaux mécanismes d'endommagement provoquant la ruine du matériau est possible à partir d'éprouvettes modèles UD et résine. Ces éprouvettes sont testés lors d'essais de traction spécifiques, pour deux orientations différentes à 0° et à 90° (angle entre la direction de l'effort et celle des fibres).

Le suivi de l'endommagement est réalisé à l'aide de la technique de l'EA. Un système d'acquisition commercialisé par la société Mistras Groupe, comprenant deux voies avec une fréquence d'échantillonnage de 5 MHz, est utilisé. Il est composé d'une carte d'acquisition PCI reliée à un micro-ordinateur. Les signaux d'EA sont ensuite enregistrés par deux capteurs piézoélectriques de type Micro-80, de bande passante 100 kHz–1 MHz et de 10 mm de diamètre, positionnés sur les éprouvettes avec une graisse silicone. Ces signaux sont amplifiés par un préamplificateur, avec un gain de 40 dB. Ils sont ensuite convertis par la carte PCI, qui les numérise et procède à l'extraction des divers paramètres temporels et fréquentiels. Les données d'EA sont ensuite transmises au micro-ordinateur pour afficher en temps réel le stockage et le post-traitement des

Tableau 3.3 – Nombre de plis, épaisseur et fraction volumique des fibres dans les matériaux composites UD non hybrides.

Matériau	Nombre de plis	Épaisseur (mm)	Fraction volumique des fibres
Lin	8	2.90	0,35
Verre	8	1.28	0,54

Tableau 3.4 – Nombre de plis, séquence d'empilement, épaisseur et fraction volumique des fibres dans les matériaux composites UD hybrides.

Matériau	Nombre de plis (Lin/Verre)	Séquence d'empilement L : Lin, V : Verre	Épaisseur (Lin/Verre)(mm)	Fraction volumique (Lin/Verre) (%)
$[V_2/L_2]_s$	4/4	[V-V-L-L-L-L-V-V]	1.40/0.60	26/22
$[L_2/V_2]_s$	4/4	[L-L-V-V-V-V-L-L]	1.40/0.60	25/21

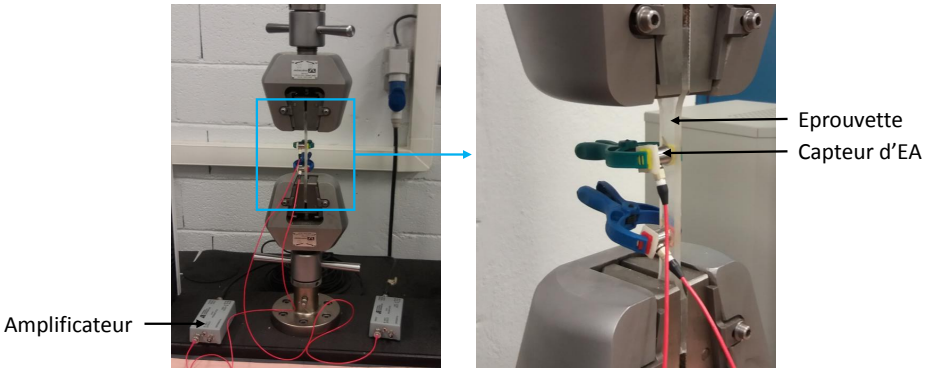


FIGURE 3.3 – Machine de traction et dispositifs de mesure de l'activité acoustique.

données. Le seuil de détection des signaux acoustiques a été fixé à 36 dB pour éliminer les signaux dus au bruit environnant. De plus, des tests de "cassé de mine" ont été effectués avant les essais pour s'assurer à la fois du bon fonctionnement de la chaîne d'acquisition et de la bonne définition des paramètres d'acquisition.

Le système d'acquisition, utilisé dans cette étude, demande un paramétrage afin de mieux détecter et enregistrer les signaux d'EA. Ce paramétrage consiste à fixer les valeurs adéquates de trois fenêtres temporelles, appelées PDT (*Peak Definition Time*), HDT (*Hit Definition Time*) et HLT (*Hit Lockout Time*).

b) Paramétrage du logiciel d'acquisition (les paramètres : PDT, HDT et HLT)

Le système d'acquisition nécessite un réglage des paramètres d'acquisition des salves d'EA. Les fenêtres temporelles PDT, HDT et HLT sont utilisées afin de définir lors de l'acquisition d'un signal, l'amplitude maximale, la durée et le nombre de coups. Le réglage de ces fenêtres temporelles peut avoir une influence sur le calcul des paramètres de la salve. Il dépend du type de matériaux composites à caractériser .

Le PDT permet de bien identifier le pic maximal d'une salve, comme le montre la figure 3.4. Un PDT mal réglé peut entrainer une mauvaise évaluation du maximum. Le PDT est déclenchable par toute alternance supérieure au seuil.

Le HDT est une constante de temps qui permet de déterminer la fin d'une salve, notamment la durée et le nombre de coups comme le montre la figure 3.5. Le réglage de ce paramètre peut donc avoir une influence sur le calcul de la durée et du nombre de coups.

Le HLT est une constante de temps qui correspond à une plage d'inhibition de l'acquisition permettant d'inhiber des échos secondaires. Le HLT débute à la fin du dernier HDT.

Les valeurs recommandées dans la littérature pour le test des matériaux composites sont présentées sur le Tableau 3.5 :

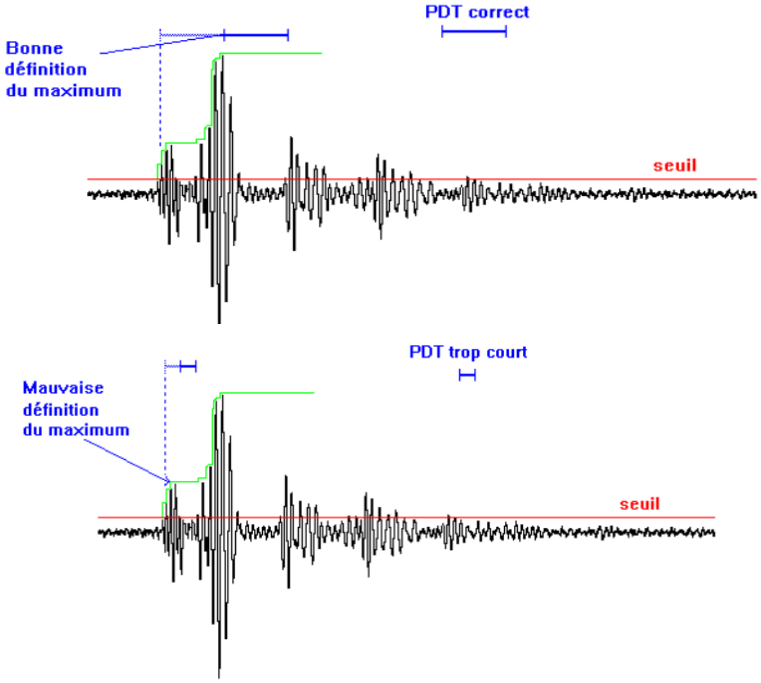


FIGURE 3.4 – Influence du PDT sur une salve typique d'EA pour définir le pic maximal de la salve.

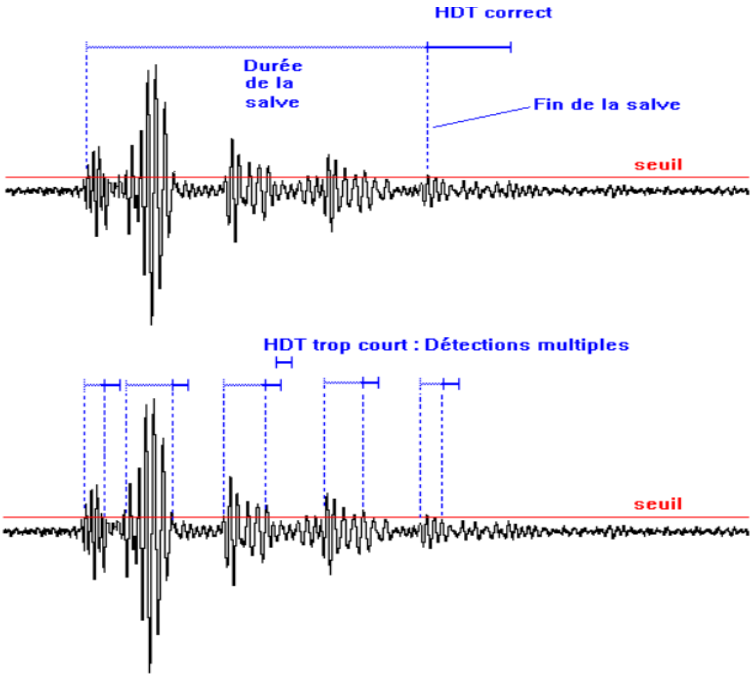


FIGURE 3.5 – Influence du HDT sur une salve typique d'EA pour définir la durée de la salve.

Tableau 3.5 – Valeur des paramètres du logiciel d'acquisition recommandées pour le test des matériaux composites.

Paramètre	Valeur (μs)
PDT	20-50
HDT	100-200
HLT	300

c) Caractérisation des signaux d'EA

Afin de pouvoir caractériser les différents types de signaux, des paramètres temporels et fréquentiels sont définis à partir des salves d'EA. La plupart de ces paramètres sont identifiés par rapport à un seuil d'acquisition. Ce dernier permet d'éliminer aussi le bruit de fond. Avant d'effectuer les essais mécaniques, les signaux d'EA dus au bruit de fond ont été filtré en augmentant progressivement ce seuil, jusqu'à atteindre une valeur optimale. Les essais effectués au laboratoire ont abouti à un seuil optimal d'environ 36dB. Les principaux paramètres temporels et fréquentiels caractérisant une salve d'EA sont présentés dans la figure 3.6 et définis et dans le Tableau 3.6.

3.3 Démarche de classification

L'analyse bibliographique présentée au chapitre 2 a montré que la plupart des approches de classification de signaux d'EA sont basées sur les méthodes non-supervisées, à savoir la méthode de k-moyennes (le principe de cette méthode est présenté dans la section 3.3.3). Ces méthodes de classification nécessitent une étape d'identification des groupes obtenus. Cette identification se fait la plupart du temps en comparant leurs caractéristiques avec celles des signaux identifiés dans la littérature. Les caractéristiques comparées sont

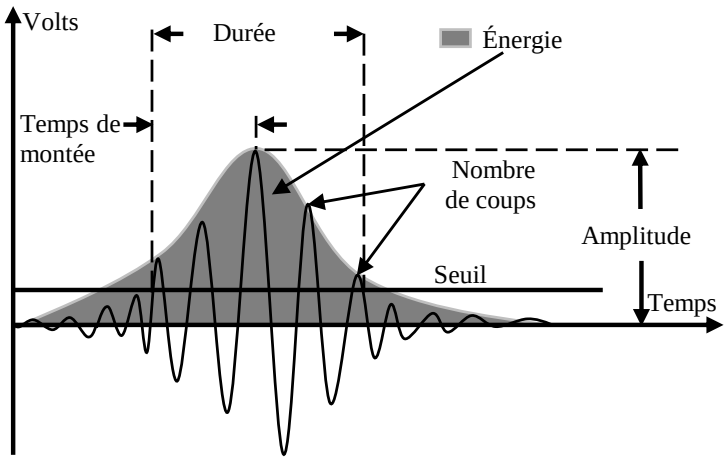


FIGURE 3.6 – Forme typique d'une salve d'EA caractérisée par les principaux paramètres temporels.

Tableau 3.6 – Les paramètres temporels et fréquentiels définis sur une salve d'EA.

Paramètre	définition
Amplitude	Amplitude maximale de la salve, son unité est le décibel (dB). $\text{Amplitude(dB)} = 20\log_{10}(\frac{V_{max}}{1\mu volt})$
Nombre de Coups (NC)	Correspond au nombre de franchissement de seuil sur la durée de la salve
Energie Absolue (ENA)	$\text{ENA(attoJoules)} = \int_T \frac{V^2}{R} dt$ avec $R = 10k\Omega$, est une résistance de référence
Durée	Le temps entre le premier dépassement de seuil et le dernier dépassement de seuil du signal. Son unité est la microseconde (µs)
Temps de Montée (TM)	Le temps entre le premier dépassement de seuil et le temps jusqu'à l'amplitude maximum du signal. Son unité est la microseconde (µs).
Nombre de Coups Au Pic (NCAP)	correspond au nombre de franchissements de seuil par le signal entre le premier dépassement de seuil et l'amplitude maximale
Fréquence Moyenne (FM)	$FM = \frac{NC}{Duree}$
Fréquence de Réverbération (FR)	$FR = \frac{NC-NCAP}{Duree}$
Fréquence d'Initiation (FI)	$FI = \frac{NCAP}{TM}$
Fréquence de Pic (FP)	Il correspond au point dans le spectre de fréquence du signal auquel l'amplitude maximale survient
Fréquence Centroïde (FC)	Le centre de masse du spectre de fréquence du signal, défini par : $FC = \frac{sum(amplitude*frequence)}{sum(amplitude)}$

souvent l'amplitude et la fréquence de pic. Cependant, les signaux d'EA sont de nature non-stationnaire, ce qui rend cette comparaison difficile. Ainsi, le but de ce travail est de remédier à ce problème en construisant une base de données d'apprentissage de signaux d'EA où chaque groupe est associé à un mode d'endommagement bien identifié. Pour ce faire, nous avons développé la méthode incrémentale. Le principe de cette méthode est présenté dans la section suivante.

L'objectif sera donc de combiner la base de données d'apprentissage avec un classifieur supervisé afin d'identifier les signaux d'EA des composites plus complexes. Pour ce faire, nous avons choisi d'utiliser le classifieur k-PPV. Le principe de ce classifieur est présenté dans la section 3.3.4.

D'autre part, l'évaluation des performances d'un classifieur supervisé est un enjeu de grande importance. En effet, ces performances peuvent être utilisées pour l'identification de nouveaux éléments ou pour l'optimisation des paramètres du classifieur. Les performances d'un classifieur peuvent être mesurées quantitativement par le biais de plusieurs critères. Parmi ces critères, on trouve la courbe ROC et l'aire sous la courbe ROC. Le principe de ces critères est présenté dans la section 3.3.5.

3.3.1 Méthode incrémentale

La méthode proposée dans ce travail permet de regrouper les signaux d'EA sous forme incrémentale. La mise en œuvre de cette méthode est possible par le fait que les signaux engendrés par des mécanismes d'endommagements de nature différente ne se chevauchent pas entre eux dans l'espace des descripteurs choisis. De ce fait, il est nécessaire au préalable de constituer des bases de données d'EA contenant un,

deux, trois ou plusieurs types d'endommagement. À titre d'exemple, le principe est décrit dans la figure 3.7. Deux bases de données A et B y sont superposées, la première étant uniquement caractérisée par un seul type d'endommagement (de type E_1) et la seconde contenant un type d'endommagement supplémentaire (de types E_1 et E_2). À l'aide de la superposition, du principe de voisinage et de l'algorithme 1, il est possible d'identifier les signaux d'EA de l'endommagement du type E_2. Les signaux de la seconde base proches de ceux de la première base, sont affectés à E_1 et les signaux éloignés de ceux de la première base, sont affectés à E_2. Cette opération est ensuite répétée d'une manière incrémentale afin d'identifier l'ensemble des modes d'endommagement.

Le voisinage d'un point x_i est défini par une N-sphère du centre x_i et de rayon R. Un point y_j est un voisinage de x_i si :

$$\sum_{k=1}^{n} (y_j^k - x_i^k)^2 \leq R^2. \tag{3.1}$$

Algorithme 3.1 *méthode de regroupement d'une base de données d'EA par la méthode incrémentale*

Entrées : *Base de donnée A contient N signaux d'EA représentant l'endommagement* E_1, *Base de donnée B contient M signaux d'EA représentant deux mécanismes d'endommagement* (E_1, E_2), $R > 0$.

Sorties : les signaux d'EA des endommagements E_1 et E_2.

1. *Pour* $x_i \in A$, $i \in \{1, 2, 3, ..., N\}$.
2. *Pour* $y_j \in B$, $j \in \{1, 2, 3, ..., M\}$.
3. *Si* y_j est un voisin de x_i *stocker* y_j *dans un tableau.*
4. *Fin des éléments de* B : passage vers 1.
5. *Fin des éléments de* A.
6. *Enregistrement du tableau.*

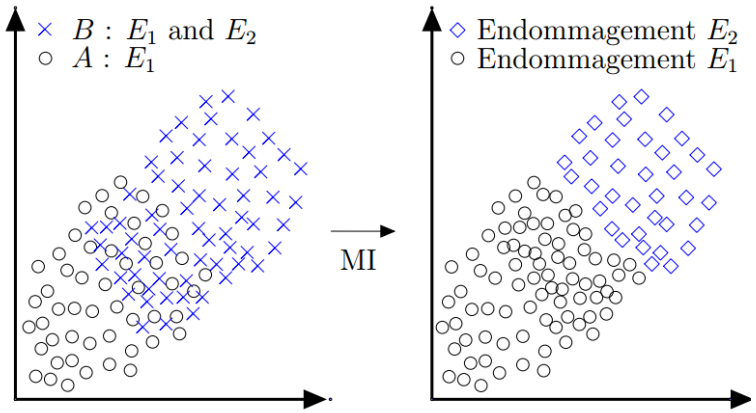


FIGURE 3.7 – Principe de la méthode incrémentale (MI)

À la fin de l'algorithme, L signaux de B vont être attribués aux mécanisme d'endommagement E_1 . Ainsi, $N + L$ signaux vont être attribués à l'endommagement E_1 et $M - L$ à l'endommagement E_2.

Avant d'appliquer cette méthode, il est nécessaire de construire une base de données de signaux d'EA. Celle-ci doit intégrer les différents mécanismes d'endommagement des matériaux composites, en particulier les 3 catégories décrites dans le chapitre 2. Pour ce faire, nous avons choisi d'effectué des essais sur la résine époxyde seule puis sur un composite UD sollicités à 0 ° (UD0) et 90° (UD90) par rapport à l'axe des fibres. L'avantage de ces derniers est la connaissance préalable des mécanismes d'endommagement qui peuvent être distingués lors d'essais mécaniques spécifiques [13, 94, 95]. Le tableau 3.7 présente les différents mécanismes dans chaque type de matériau lors d'une sollicitation de traction.

Pour regrouper les signaux d'EA représentant le même mode d'endommagement, l'algorithme de la méthode incrémentale est appliqué en exploitant les différentes bases de données. Le logigramme de la figure 3.8 présente les différentes étapes de la méthode incrémentale pour la construction d'une base de données d'apprentissage, permettant de distinguer les 3 catégories d'endommagement : E. matrice, E. interfaces et E. fibres.

Tableau 3.7 – Matériaux composites UD et ses mécanismes d'endommagements

Matériau composite	**Sollicitation**	**Mécanismes d'endommagements**
Matrice seule (résine époxyde)	Traction uniaxiale	Endommagement de la matrice
UD90	Traction uniaxiale perpendiculairement à la direction des fibres	Endommagement de la matrice et Endommagement des interfaces fibres/matrice
UD0	Traction uniaxiale dans la direction des fibres	Endommagement de la matrice, Endommagement des interfaces fibres/matrice et endommagement des fibres

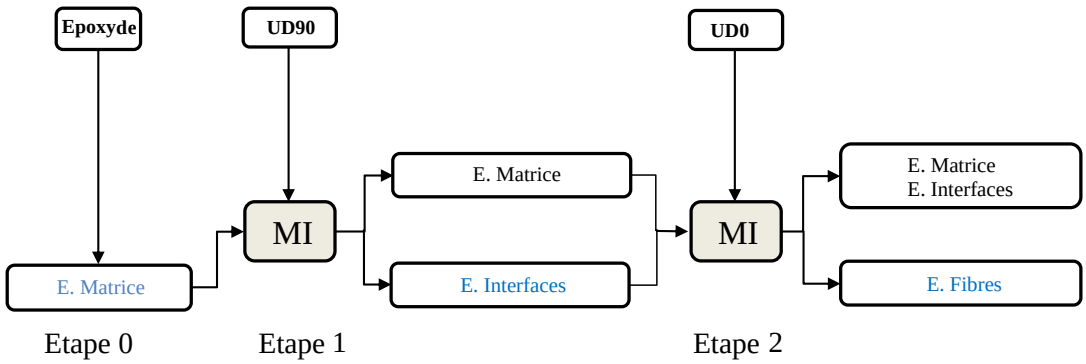


FIGURE 3.8 – Logigramme du regroupement par la Méthode Incrémentale (MI)

Dans l'algorithme de la méthode incrémentale, les données sont regroupées par rapport à une sphère caractérisée par un rayon R. Pour un choix optimal de ce rayon, la méthode 'Support Vector Data Description (SVDD)' est appropriée [96]. La SVDD fait partie des méthodes de classification à une seule classe 'One-Class Classification (OCC)' [97]. Ces méthodes visent à construire des modèles de classification lorsqu'un jeu de données contient une seule classe. Cette situation unique contraint l'apprentissage des classifieurs en définissant la limite de la classe juste avec la connaissance de la classe positive. Le problème de l'OCC a été étudié et appliqué avec succès sous de nombreux thèmes de recherche, tels que la détection des valeurs aberrantes [98]. La méthode SVDD et ses variantes permet de trouver un choix optimal du rayon R. Dans la suite, le principe de cette méthode sera détaillé afin de pouvoir la combiner avec la méthode incrémentale.

3.3.2 Support Vector Data Description

Le SVDD est une méthode de classification inspirée du principe de séparateur à vaste marge. Son objectif est de trouver une frontière de forme sphérique autour d'un jeu de données. La frontière peut être rendu flexible en utilisant les fonctions noyaux.

Pour introduire le principe de la méthode SVDD, considérons une base de données d'apprentissage $\{x_i \in \mathbb{R}^n\}, i = 1, 2, ..., m$, l'idée de base est de résoudre le problème d'optimisation suivant

$$\begin{cases} \min\limits_{R,a} \ R^2 \\ \|x_i - a\|^2 \leq R^2 \ , i = 1, ..., m \end{cases} \tag{3.2}$$

Avec :

a et R sont respectivement le centre et le rayon, les paramètres caractéristiques de la sphère. Les distances entre les données x_i et le centre de la sphère a ne devront pas être supérieurs à R^2. Par contre, les très grandes distances doivent être pénalisées en introduisant les variables ξ_i . Le problème d'optimisation devient donc :

$$\begin{cases} \min\limits_{R,a,\xi} \ (R^2 + C \sum\limits_{i=1}^{m} \xi_i) \\ \|x_i - a\|^2 \leq R^2 + \xi_i \ , i = 1, ..., m. \\ \xi \geq 0 \end{cases} \tag{3.3}$$

Le paramètre C contrôle le compromis entre le volume et les erreurs.

La fonction à optimiser et les contraintes peuvent être fusionnées dans une seule équation à l'aide des coefficients de Lagrange :

$$L(R,a,\alpha_i,\gamma_i,\xi_i) = R^2 + C\sum_{i=1}^{m}\xi_i - \sum_{i=1}^{m}\alpha_i\{R^2+\xi_i-(\|x_i\|^2 - 2a.x_i + \|a\|^2)\} - \sum_{i=1}^{m}\gamma_i\xi_i. \tag{3.4}$$

Avec $\alpha_i \geq 0$ et $\gamma_i \geq 0$ sont les coefficients de Lagrange.

Les dérivées partielles de L permettent d'avoir les contraintes suivantes :

$$\frac{\partial L}{\partial R} = 0 \rightarrow \sum_{i=1}^{m}\alpha_i = 1. \tag{3.5}$$

$$\frac{\partial L}{\partial a} = 0 \rightarrow a = \frac{\sum_{i=1}^{m}\alpha_i x_i}{\sum_{i=1}^{m}\alpha_i} = \sum_{i=1}^{m}\alpha_i x_i. \tag{3.6}$$

$$\frac{\partial L}{\partial \xi_i} = 0 \rightarrow C - \alpha_i - \gamma_i = 0. \tag{3.7}$$

De l'équation $C - \alpha_i - \gamma_i = 0$ et parce que $\alpha_i \geq 0$ et$\gamma_i \geq 0$, les coefficients de Lagrange γ_i peuvent être supprimés lorsqu'on exige que :

$$0 \leq \alpha_i \leq C. \tag{3.8}$$

En replaçant les résultats des expressions (3.5), (3.6) et (3.7) dans l'équation (3.4), on obtient le problème d'optimisation suivant :

$$\begin{cases} L = \sum_{i=1}^{m}\alpha_i(x_i.x_j) - \sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j(x_i.x_j) \\ 0 \leq \alpha_i \leq C \end{cases}. \tag{3.9}$$

L'optimisation (maximisation) du problème (3.9) permet d'avoir les coefficients α_i. Pour tout x_i, on a trois cas possibles :

— Si $\|x_i - a\|^2 < R^2 \Rightarrow \alpha_i = 0 \; et \; \gamma_i = 0$,

— Si $\|x_i - a\|^2 = R^2 \Rightarrow 0 < \alpha_i < C \; et \; \gamma_i = 0$,

— Si $\|x_i - a\|^2 > R^2 \Rightarrow \alpha_i = C \; et \; \gamma_i > 0$.

Dans l'équation (3.6), on remarque que le centre de la sphère est une combinaison linéaire des données d'apprentissage. L'évaluation des nouveaux éléments nécessite seulement l'utilisation des données x_i ayant une $\alpha_i > 0$. Ces données sont appelés Support Vectors (SV).

La fonction de décision pour tester un nouvel objet z est définie par la distance entre le centre de la sphère a et l'objet. Ce dernier est accepté dans le cas où cette distance est inférieure ou égale au rayon.

$$\|z-a\|^2 = (z.z) - 2\sum_{i=1}^{l}\alpha_i(z.x_i) + \sum_{i=1}^{l}\alpha_i\alpha_j(x_i.x_j) \leq R^2. \tag{3.10}$$

Par définition, R^2 est la distance entre le centre de la sphère a et sa frontière (les vecteurs de support). Les vecteurs de support avec un $\alpha = C$ sont exclus. Donc, le rayon de la sphère est défini pour tout $x_k \in SV_{\alpha_i^*}$ (les vecteurs de support dont $\alpha < C$), par :

$$R^2 = (x_k.x_k) - 2\sum_{i=1}^{l}\alpha_i(x_i.x_k) + \sum_{i,j=1}^{l}\alpha_i\alpha_j(x_i.x_j). \tag{3.11}$$

On remarque que dans toutes les équations (3.9, 3.10 et 3.11) les objets x_i n'apparaissent que sous forme de produits scalaire avec d'autres objets $(x_i.x_j)$. Analogue au principe de Séparateurs à Vastes Marges (SVM), les produits scalaires peuvent être remplacés par une fonction noyau pour obtenir des méthodes plus souples. Le remplacement du produit scalaire $x_i.x_j$ par une fonction noyau $k(x_i, x_j) = (\phi(x_i).\phi(x_j))$ permet de projeter implicitement les données vers un espace de redescription (éventuellement de grande dimension). L'utilisation des fonctions noyau a pour objectif la construction d'une hyper-sphère plus souple dans l'espace de resdescription.

La forme de la sphère dépend du type de la fonction noyau adopté. Parmi les noyaux les plus utilisées dans le domaine de la classification, on trouve des fonction de type :

— Polynomiale : $K(x_i, x_j) = (1 + x_i^t x_j)^p$,

— Gaussien $K(x_i, x_j) = \exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2})$,

— Sigmoïde : $K(x_i, x_j) = \tanh(\beta_0 x_i^t x_j + \beta_1)$.

3.3.3 Méthode de k-moyennes

La méthode des k-moyennes [39, 99] fait partie de la famille d'algorithmes de classification dit à "centres mobiles". C'est une méthode itérative de partition des données par minimisation de la variance intra-classe. La segmentation des données ou la division en classes est une tâche non-supervisée. Il est seulement nécessaire de préciser le nombre de classes k qu'on désire obtenir. Cependant, il n'est pas nécessaire de connaître la partition en classes des données. Une description plus précise de l'algorithme des k-moyennes est donné ci-dessous.

Algorithme 3.2. *Regroupement par la méthode de k-moyennes.*

Entrés : *un ensemble de données* $\xi = \{x_i \in \mathbf{R}^n, i \in \{1, 2, ..., N\}, (n, N) \in \mathbf{N}^*\}$, *le nombre de groupe souhaité (noté par k) et un mesure de similarité (noté par d).*

Sorties : *une partition en k groupes et leurs centres* C_k.

1. *Initialisation aléatoire de k centres* C_k.
2. *Calculer la distance entre les centres des k classes et chaque objet* x_i : $d_i^k = |x_i - C_k|$.
3. *Générer une nouvelle partition en assignant chaque objet au groupe dont le centre est le plus proche :* $x_i \in C_k$ *si* $\forall_j\ d_i^k = \min\ d_i^j$.
4. *Évaluation des nouveaux centres des classes :* $C_k = \frac{1}{M_k} \sum_{x \in C_k} x_i$.
5. *Si les centres des classes sont stables, alors l'algorithme a convergé et la procédure est terminée ; dans le cas contraire, on répète les itérations de 2 à 4.*

L'algorithme de k-moyennes est très populaire du fait qu'il est très facile à comprendre et à mettre en œuvre. Sa simplicité conceptuelle et sa rapidité permettent de l'appliquer à des données de grandes tailles, et aussi à tout type de données, en choisissant une bonne mesure de similarité. En revanche, le point faible de l'algorithme de k-moyennes est que le résultat dépend de tirage initial des centres des classes. Les classes sont construites par rapport à des objets inexistants (les milieux). Après l'application de l'algorithme, d'autres techniques ou des expertises doivent être utilisées pour dégager les significations physiques des classes.

3.3.4 Méthode de *k*-PPV

L'algorithme de k-PPV [100] est l'un des algorithmes d'apprentissage automatique les plus simples. En supposant qu'une base de données d'apprentissage est correctement étiquetée, cette méthode fournit de très bons résultats de classification. Le principe de la méthode de k-PPV est d'assigner de nouveaux objets non classifiés à la classe à laquelle appartient la majorité de ses k plus proches voisins. La similarité de deux objets est déterminée en utilisant une distance, comme décrit dans l'algorithme 1. La figure 3.9 illustre le principe de la méthode de k-PPV dans le cas d'un problème de classification à 2 classes. Sur cet exemple, parmi les *3*-PPV de l'objet étoile, 2 font partie de la classe C_2 (classe des objets cercle). Donc, l'objet étoile sera assigné à la classe C_2.

Algorithme 3.3 *Classification d'un objet par le k-PPV*

Entrés : *Base de données d'apprentissage* $\xi = ((x,y)|x \in \mathbb{R}^n, y \in C = (1,2,...,N), (n,N) \in \mathbf{N}^*)$, *un objet inconnu* z, *le nombre de plus proches voisins à considérer* k *et une mesure de similarité* d.

Sorties : *l'étiquète de la classe de l'objet* Z.

1. *Pour chaque* $\{(x,y) \in \xi\}$ *calcule de* $d(z,x)$.
2. *Pour chaque* $\{x \in kNN\,\{z\}\}$ *comptage de nombre d'occurrences de chaque classe.*
3. *Attribution de l'objet* Z *à la classe la plus fréquente (règle de majorité).*

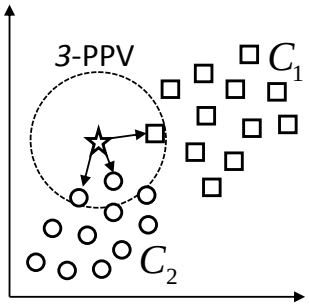


FIGURE 3.9 – Illustration du principe de la méthode de k-PPV avec un problème de classification à deux classes.

3.3.5 Mesures statistique pour évaluer les performances des classifieurs (Courbe ROC, Validation croisée)

Évaluer les performances d'une méthode de classification (classifieur) est un enjeu de grande importance car ses performances sont utilisées pour l'apprentissage et la classification des données. Dans cette partie le principe de deux méthodes pour évaluer les performances des classifieurs vont être présentées. Elles s'agissent des courbes ROC et la validation croisée.

a) Courbe ROC

Une Courbe ROC (*Receiver Operating Characteristics*) [101] est utilisée comme une technique pour visualiser, organiser et sélectionner les classifieurs en fonction de leurs performances. La courbe ROC a d'abord été utilisé dans la théorie de détection du signal pour représenter le compromis entre le taux de réussite et le taux de fausses alarmes. La courbe ROC est largement étudiée et appliquée dans le diagnostic médical depuis les années 1970.

Spackman [102] a été l'un des premiers chercheurs à avoir adopté les courbes ROC dans l'apprentissage automatique. Il a montré que les courbes ROC pouvaient être utilisées pour évaluer et comparer différents algorithmes. Au cours des dernières années, l'utilisation des courbes ROC dans la recherche sur l'apprentissage automatique a augmenté parce que cette méthode graphique est simple et donne un bon indicateur pour évaluer la performance en classification. L'aire sous la courbe ROC, en abrégé AUC (Area Under Curve), fournit une valeur scalaire représentant la performance des courbes ROC décrites en deux dimensions. Un bref aperçu du principe de la courbe ROC et de ses caractéristiques est donné ci-dessous.

Pour introduire le principe des courbes ROC, prenons le cas d'un problème de classification en utilisant seulement deux classes. Formellement, chaque instance I est mappée à un élément de l'ensemble $\{p, n\}$ des étiquettes de classes positives et négatives. Pour distinguer les instances prédites des instances réelles, les étiquettes $\{p', n'\}$ ont été utilisées pour les classes produites par le modèle. Étant donné un classifieur et une

Tableau 3.8 – Matrice de confusion pour un problème de classification à deux classes.

		Vérité	
		p	n
Décision	p'	Vrai Positif	Faux Positif
	n'	Faux Négatif	Vrai Négatif

instance, il y a quatre résultats possibles :

— Vrai Positive (VP) : si l'instance est positive et est classée comme positive.

— Faux Positive (FP) : si l'instance est négative et est classée comme positive.

— Vrai Négative (VN) : si l'instance est négative et est classée comme négative.

— Faux Négative (FN) : si l'instance est négative et est classée comme positive.

En considérant un classifieur et un ensemble d'instances (l'ensemble de test), les résultats de la classification sont représentés dans une matrice de confusion (Tableau 3.8). A partir de cette matrice, il est possible de déduire plusieurs mesures traditionnelles de la performance en classification :

— Taux de vrais positifs (TVP) = (nombre de VP)/ (nombre totale des instances positives).

— Taux de faux positifs (TVP) = (nombre de FP)/ (nombre totale des instances positives).

— Précision = (nombre de VP+VN)/ (nombre totale des instances positives et négatives).

— Sensibilité = (nombre de VP)/ (nombre de VP + nombre de FN).

— Spécificité = (nombre de VN)/ (nombre de VN + nombre de FP).

L'utilisation des méthodes scalaires, comme la précision, engendre une vue trop particulière de la performance du classifieur parce qu'elles sont sensibles à la disproportion des classes. Ces méthodes exploitent dans le calcul les éléments correctement identifiés par le classifieur (VP et VN) sans la prise en compte des éléments mal identifiés (FP et FN). En revanche, la courbe ROC utilise des mesures normalisées par les populations des classes : le TVP et le TFP. La courbe ROC est un graphe bidimensionnel où l'évolution du TVP (représenté sur l'axe des Y) est donnée en fonction du TFP (représenté sur l'axe des X). La figure 3.10 donne un exemple des courbes ROC de deux classifieurs A et B. Le classifieur A est meilleur que B puisque pour un taux de FP donné, son taux de VP est supérieur de celui de B.

Comparer la performance de deux classifieurs équivaut à comparer deux courbes, ce qui n'est pas évident. En effet, selon l'espace sous la courbe ROC, la performance des classifieurs peut varier car une zone d'espace peut être meilleure qu'une autre. Pour cette raison, il est préférable de réduire la courbe ROC à une valeur scalaire afin de faciliter la comparaison entre classifieurs [103–105]. La valeur scalaire qui est généralement extraite de la courbe ROC est l'AUC. Dans le cas le plus simple, pour un classifieur binaire, cette zone est définie par la région située entre la courbe ROC et deux segments. Le premier est défini par les 2 points (0,0), (1,0) et le second par les 2 points (1,0), (1,1).

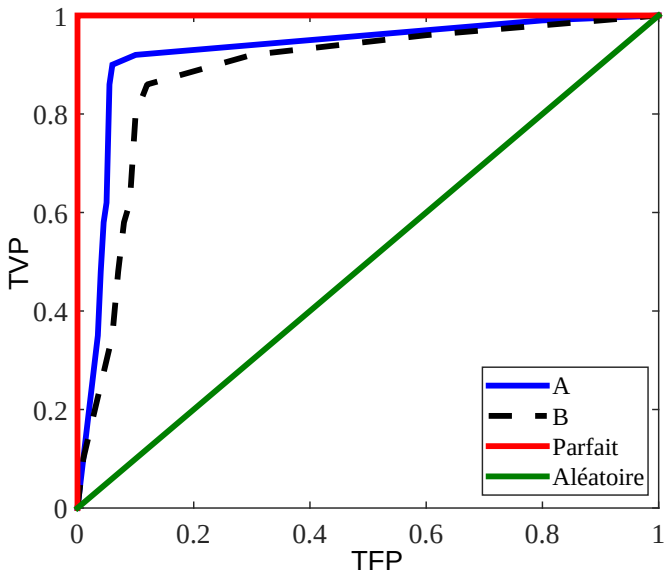


FIGURE 3.10 – Courbe ROC pour les classifieurs : A, B , parfait et aléatoire. le classifieur A est meilleur par rapport à B.

L'AUC est devenue une meilleure alternative à la précision pour évaluer un classifieurs. L'AUC est équivalente à la probabilité qu'un classifieur donne un meilleur rang à un élément positif par rapport à un élément négatif, tous deux choisis au hasard dans l'ensemble de données. Le classifieur parfait possède une AUC égale à 1 et le classifieur aléatoire possède une AUC égale à 0,5. Le but est d'approcher au classifieur parfait, donc à une AUC égale à 1. Dans la figure 1, l'AUC du classifieur A est plus grande que celle du classifieur B. Donc, le classifieur A est considéré comme ayant la meilleure performance.

b) Validation croisée

La validation croisée [106] est une méthode statistique d'évaluation et de comparaison des algorithmes d'apprentissage automatique en divisant les données en deux segments : l'un utilisé pour l'apprentissage du modèle et l'autre utilisé pour la validation du modèle. Dans la validation croisée, les ensembles d'apprentissage et de validation doivent se croiser dans des tours successifs de telle sorte que chaque point de données a une chance d'être validé. La forme de base de la validation croisée est la validation à k segments «$k-foldCross-validation$». Les autres formes de validation croisée sont des cas spéciaux de la forme de base, par exemple la validation « $Leave-one-outcrossvalidation$».

Dans la validation croisée $k-fold$, les données sont d'abord divisées en k segments de taille égale. Par la suite, k itérations d'apprentissage et de validation sont effectuées de telle sorte qu'à chaque itération un segment différent des données est retenu pour la validation alors que $k-1$ segments restants sont utilisés pour l'apprentissage. Dans l'apprentissage automatique, la validation croisée avec $(k=10)$ est la plus fréquente.

La validation croisée est utilisée pour évaluer ou comparer les algorithmes de classification comme suit : dans chaque itération, un ou plusieurs algorithmes de classification utilisent des $k-1$ segments de données pour l'apprentissage en suite les modèles sont invités à faire des prédictions sur les données dans le segment de validation. La performance de chaque algorithme de classification sur chaque segment peut être suivie en utilisant une métrique de performance prédéterminée comme la précision ou l'AUC.

On note que cette méthode est très utile pour faire tourner et évaluer les algorithmes d'apprentissage dans le cas où on se dispose d'un nombre faible d'échantillon d'apprentissage.

3.4 Résultats

Cette partie expose dans un premier temps le choix et les démarches pour sélectionner les descripteurs pertinents de signaux d'EA et les paramètres de la méthode de classification. Dans un second temps, une démarche de construction des vérités terrain de signaux d'EA, liés aux mécanismes d'endommagement des matériaux composites à fibres de lin et à fibres de verre, à l'aide de la méthode incrémentale est présentée. Ensuite, une comparaison entre les résultats obtenus par la méthode incrémentale et celle de k-moyennes est conduite. Cette comparaison permet de mettre en évidence les inconvénients de la méthode de k-moyennes dans la classification de signaux d'EA. Enfin, les deux bases de données d'apprentissage sont combinées pour former une base de données globale qui contient les signaux d'EA de cinq classes d'endommagement (matrice, interfaces de lin, interfaces de verre, fibres de verre et fibres de lin). Cette base est utilisée pour l'apprentissage d'un classifieur supervisé afin de discriminer les mécanismes d'endommagement des matériaux composites complexes à renforts hybrides lin-verre.

3.4.1 Sélection des descripteurs d'EA

Pour réaliser une classification efficace, les descripteurs décrivant les signaux d'EA doivent être sélectionnés d'une manière optimale. Chaque signal peut être décrit par plusieurs descripteurs, dont certains peuvent être corrélés et donc redondants. De plus, les descripteurs doivent être en nombre suffisant pour éviter de perdre des informations des signaux, mais pas en nombre élevé pour ne pas alourdir le calcul. Il est donc nécessaire de choisir judicieusement ces descripteurs.

Par ailleurs, une des limites des méthodes actuelles de classification de signaux d'EA est l'utilisation de la Fréquence de Pic (FP) pour décrire le contenu fréquentiel d'un signal d'EA. La FP est déterminée à partir d'un point unique dans le spectre de fréquence : le point de grandeur maximale. Il donnerait une description appropriée du contenu fréquentiel uniquement si le signal contenait une seule composante de fréquence, ce qui est rarement le cas pour les signaux d'EA. Lorsque la FP est utilisée pour caractériser les signaux d'EA

avec plusieurs composantes fréquentielles, une faible valeur de FP ne signifie pas nécessairement qu'il n'y avait pas de contenu à haute fréquence du signal, et vice versa.

La figure 3.11 présente les spectres fréquentiels de quatre signaux d'EA obtenus à partir de nos matériaux. D'après ces spectres, on constate la présence de plusieurs composantes fréquentielles. Donc, décrire un signal d'EA par sa valeur de FP ne donne pas une indication significative de son contenu fréquentiel et peut perturber la classification. Ainsi, ce descripteur n'a pas été choisi pour caractériser les signaux d'EA des matériaux composites dans cette étude.

Pour chaque signal d'EA, nous avons défini 10 descripteurs temporels et fréquentiels (Amplitude, Nombre de Coups(NC), Energie Absolue (ENA), Durée, Temps de Montée (TM), Nombre de Coups Au Pic (NCAP), Fréquence Moyenne (FM), Fréquence de Réverbération (FR), Fréquence d'Initiation (FI), Fréquence Centroïde (FC)). Ensuite, nous avons effectué une analyse de corrélation entre ces descripteurs. Le degré de corrélation entre les 10 descripteurs est calculé par le coefficient de corrélation de $Pearson$:

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i - \mu_X}{\sigma_X}\right)\left(\frac{Y_i - \mu_Y}{\sigma_Y}\right). \tag{3.12}$$

Où $X = \{X_1, ..., X_n\}$ et $Y = \{Y_1, ..., Y_n\}$ sont deux vecteurs de descripteurs contenant n valeurs. μ_X et σ_X sont la moyenne et l'écart-type de X, respectivement. De même, μ_Y et σ_Y sont la moyenne et l'écart-type de Y.

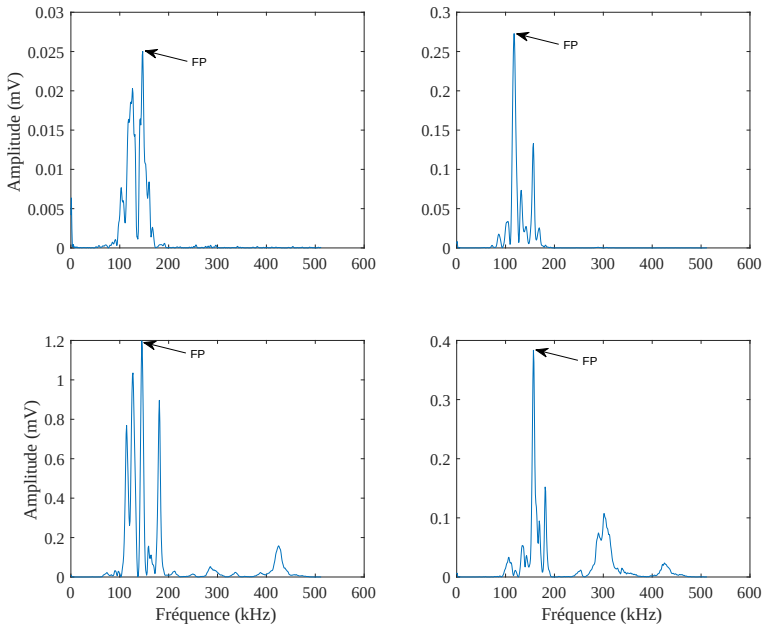


FIGURE 3.11 – Spectres fréquentiels de quatre signaux d'EA.

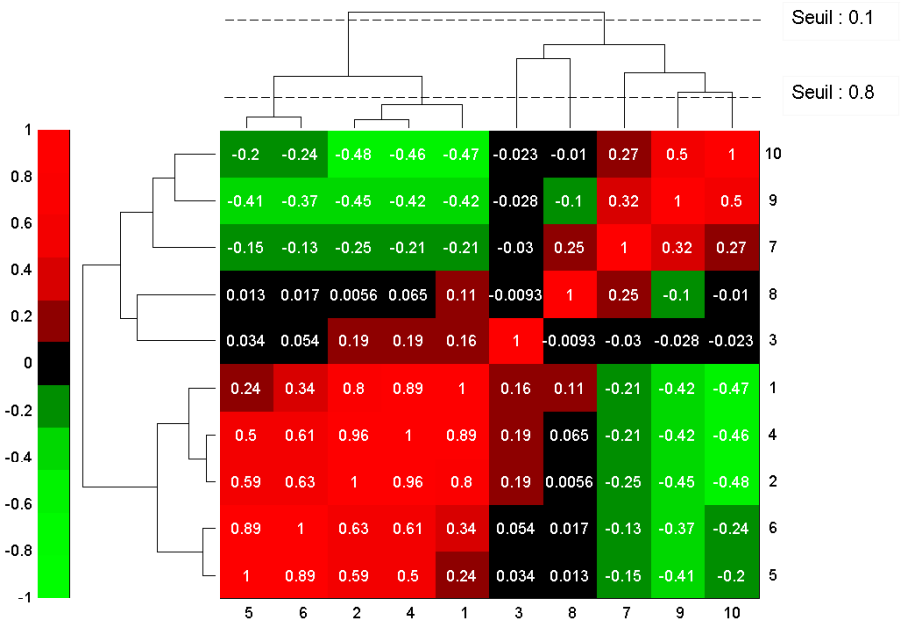


FIGURE 3.12 – Dendrogramme obtenu à partir de la matrice de corrélation.

Le résultat de l'analyse de corrélation entre les 10 descripteurs est présenté sous deux formes, matrice (tableau 3.9) et dendrogramme (figure 3.12).

Afin de sélectionner les descripteurs les moins corrélés, un seuil de 0.8 est choisi. En effet, les descripteurs sont considérés comme fortement corrélés si leurs indices de corrélation sont supérieurs au seuil, et vice versa. Ainsi, la matrice et le dendrogramme montrent une forte corrélation (un indice supérieur à 0.8) entre l'amplitude, la durée et le NC. En effet, l'un de ces trois descripteurs doit être utilisé et suffira pour l'analyse. Pour la suite, la durée a été choisie parmi ces trois descripteurs de manière arbitraire. Le TM et le NCAP sont également fortement corrélés entre eux (un indice supérieur à 0.8), donc seul le TM est sélectionné.

Tableau 3.9 – Matrice de corrélation entre les descripteurs de signaux d'EA.

	1-Amplitude	2-durée	3-ENA	4-NC	5-TM	6-NCAP	7-FM	8-FR	9-FI	10-FC
1-Amplitude	1,00	0,8017	0,1566	0,8863	0,2412	0,3438	-0,2068	0,1107	-0,4229	-0,4739
2-durée	0,8017	1,0000	0,1902	0,9569	0,5850	0,6254	-0,2536	0,0056	-0,4450	-0,4808
3-ENA	0,1566	0,1902	1,0000	0,1877	0,0343	0,0544	-0,0304	-0,0093	-0,0284	-0,0234
4-NC	0,8863	0,9569	0,1877	1,0000	0,4995	0,6091	-0,2060	0,0648	-0,4231	-0,4636
5-TM	0,2412	0,5850	0,0343	0,4995	1,0000	0,8950	-0,1535	0,0129	-0,4084	-0,1987
6-NCAP	0,3438	0,6254	0,0544	0,6091	0,8950	1,0000	-0,1287	0,0166	-0,3715	-0,2374
7-FM	-0,2068	-0,2536	-0,0304	-0,2060	-0,1535	-0,1287	1,0000	0,2515	0,3178	0,2712
8-FR	0,1107	0,0056	-0,0093	0,0648	0,0129	0,0166	0,2515	1,0000	-0,1041	-0,0103
9-FI	-0,4229	-0,4450	-0,0284	-0,4231	-0,4084	-0,3715	0,3178	-0,1041	1,0000	0,5022
10-FC	-0,4739	-0,4808	-0,0234	-0,4636	-0,1987	-0,2374	0,2712	-0,0103	0,5022	1,0000

Le reste de descripteurs représente une décorrélation (un indice inférieur à 0) ou bien une faible corrélation (un indice inférieur à 0.8) entre eux. Par conséquent, 7 descripteurs (Durée, ENA, TM, FM , FR, FI et FC) sont sélectionnés pour caractériser les signaux d'EA. Ainsi, la méthode de classification sera effectuée avec un ensemble limité de descripteurs de manière à obtenir une meilleure qualité de la base de données d'apprentissage.

Afin de rendre les descripteurs indépendants de l'unité, nous les avons normalisés (centré et réduit) par leurs écart types et moyennes à l'aide de la formule suivante :

$$X_i = \frac{x_i - \mu_i}{\sigma_i} \tag{3.13}$$

Où x_i , μ_i et σ_i sont, respectivement, le vecteur de valeurs, la moyenne et l'écart type de descripteur i. X_i est le vecteur de valeur normalisé de descripteur i.

3.4.2 Construction de la base de données d'apprentissage

La méthode incrémentale et le principe de la SVDD son combinés afin de construire deux bases de données d'apprentissage de signaux d'EA, représentant distinctivement les endommagements des composites à fibres de verre et ceux des composites à fibres de lin. Une fonction gaussienne est utilisée comme noyau pour la méthode SVDD. Ce noyau nécessite le réglage du paramètre sigma (σ). En effet, un choix optimal de ce dernier est d'une importance majeure pour construire une base d'apprentissage efficace.

a) Choix du paramètre σ

Dans le but de choisir une valeur optimale du paramètre σ, plusieurs bases de données d'apprentissage sont construites par la méthode incrémentale. Chaque signal dans une base d'apprentissage est identifié avec une étiquette qui correspond à un endommagement précis. Pour estimer la qualité de chaque base d'apprentissage, les performances d'un classifieur supervisé entrainé par ces bases sont évaluées. La qualité de chaque base est donc dépendante des bonnes ou mauvaises performances du classifieur. La méthode de *1*-PPV est utilisée comme classifieur supervisé, dont les performances sont évaluées par le calcul de l'aire sous la courbe ROC.

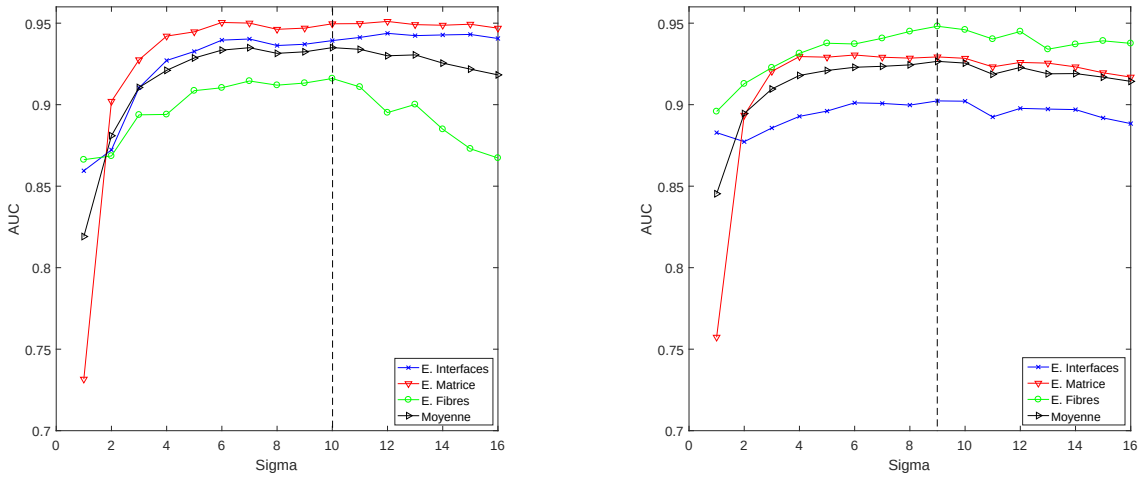
Pour chaque base de données, 50% de signaux sont sélectionnés par un tirage au hasard et sont utilisés pour l'apprentissage du classifieur. Les 50% de signaux restant sont utilisés pour le test. L'AUC de chaque classe et la moyenne de l'AUC entre les classes sont calculées sur la base de données en fonction de la décision du classifieur. Au total, 10 tirages sont réalisés pour chaque base de données. La figure 3.13 montre les courbes des performances (AUC dans chaque classe et AUC moyenne de classes) obtenus par le classifieur *1*-PPV sur les différentes bases de données en fonction de la valeur de σ. Le tableau 3.10 résume les valeurs d'AUC

Tableau 3.10 – AUC moyenne en fonction de la valeur de sigma. Performances obtenus par le classifieur *1*-PPV sur différentes base de données des composites à fibres de verre et à fibres de lin.

	AUC moyenne	
σ	composite à fibre de lin	composite à fibre de verre
1	0,8191	0,8453
2	0,8809	0,8945
3	0,9105	0,9096
4	0,9210	0,9179
5	0,9286	0,9209
6	0,9334	0,9229
7	0,9349	0,9235
8	0,9314	0,9244
9	0,9324	0,9265
10	0,9349	0,9255
11	0,9339	0,9186
12	0,9299	0,9228
13	0,9305	0,9189
14	0,9254	0,9190
15	0,9218	0,9168
16	0,9182	0,9143

moyennes obtenues par le classifieur sur chacune des deux bases de données (composites à fibres de verre et ceux à fibres de lin).

D'après la figure 3.13 et le tableau 3.10, on constate une augmentation de de l'AUC moyenne en fonction de σ jusqu'à une valeur maximale, puis diminue légèrement. En effet, les valeurs de sigma qui donnent les meilleures performances seront sélectionnées. Dans le cas des bases de données des composites à fibres de lin, les performances du classifeur sont maximales pour une valeur de σ égale à 9. Cette dernière a été choisie pour construire la base de données d'apprentissage des composites à fibres de lin par la méthode incrémentale. Concernant les bases de données des composites à fibres de verre, une valeur de σ de 10 permet d'obtenir



(a) (b)

FIGURE 3.13 – Performances en fonction du paramètre sigma de la base de données d'apprentissage de matériaux composites à : (a) fibres de verre et (b) fibres de lin.

les meilleures performances du classifieur. Donc, cette valeur optimale de σ va être utilisée pour construire la base de données d'apprentissage de matériaux composites à fibres de verre par la méthode incrémentale.

b) Construction de la base de données d'apprentissage

Dans le but de construire une base de données d'apprentissage pour le composite lin-époxyde et verre-époxyde, la méthode incrémentale combinée avec la méthode SVDD est utilisée pour classifier les signaux d'EA obtenus à partir des tests sur les éprouvettes modèles (résine, UD90, UD0) de ces composites. Le paramètre σ du noyau gaussien de la méthode SVDD est fixé à 10 pour le verre-époxyde et à 9 pour le lin-époxyde. Les figures 3.14 et 3.15 représentent la projection sur le plan engendré par le couple de descripteurs (durée, FC) de la base de données d'apprentissage de signaux d'EA des deux composites. Nous avons choisi spécifiquement la projection sur ce plan, parce qu'il offre plus de lisibilité aux données par rapport aux autres plans engendrés par les autres couples de descripteurs. Ainsi, ce choix nous a permis d'analyser plus aisément les résultats de classification.

A partir de la projection, on remarque qu'il y a du chevauchement entre les signaux des trois classes. Le taux de ce chevauchement n'est pas très élevé et est observé principalement entre les frontières des classes. Ce chevauchement est peut-être dû aux pertinences des descripteurs utilisés. Et il peut être expliqué par le fait que le regroupement par la méthode incrémentale a été réalisé dans l'espace engendré par les 7 descripteurs. Il est donc tout à fait normal d'observer du chevauchement quand les signaux sont projetés uniquement sur un plan formé de 2 de 7 descripteurs.

c) Comparaison entre les classes d'endommagements des matériaux composites à fibres de lin et à fibres de verre

La figure 3.16 illustre la projection pour chaque classe d'endommagement des signaux d'EA des deux bases de données d'apprentissage. On remarque que les deux classes d'endommagements liés aux fibres (figure 3.16-a) et aux interfaces fibre/matrice (figure 3.16-b) sont positionnées dans deux zones différentes. Par contre, le chevauchement entre les deux classes d'endommagement des interfaces est plus prononcé. Ce chevauchement peut être expliqué par la présence des signaux d'EA similaires entre les endommagements des interfaces des deux types de matériaux composites (fibres de lin et fibres de verre). La figure 3.16-c représente la projection des signaux d'EA de la classe d'endommagement de la matrice des deux bases d'apprentissages. Cette fois-ci, les deux classes sont superposées. Ce résultat est logique puisque la même résine a été utilisée lors de la mise en œuvre des deux composites.

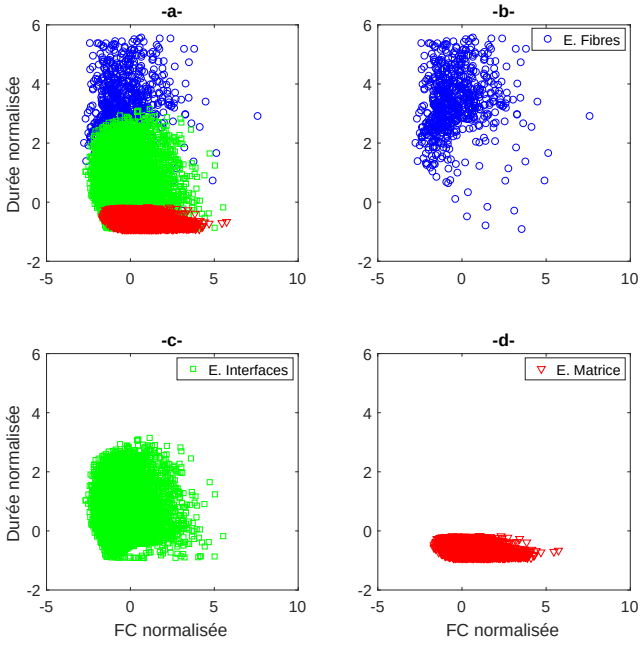


FIGURE 3.14 – Base de données d'apprentissage des composites à fibres de verre. Classification réalisée par la méthode incrémentale. Projection sur le plan (FC, Durée). -a- Les trois classes d'endommagements projeté au même endroit, -b- classe d'endommagement des fibres, -c- classe d'endommagement des interfaces, -d- classe d'endommagement de la matrice.

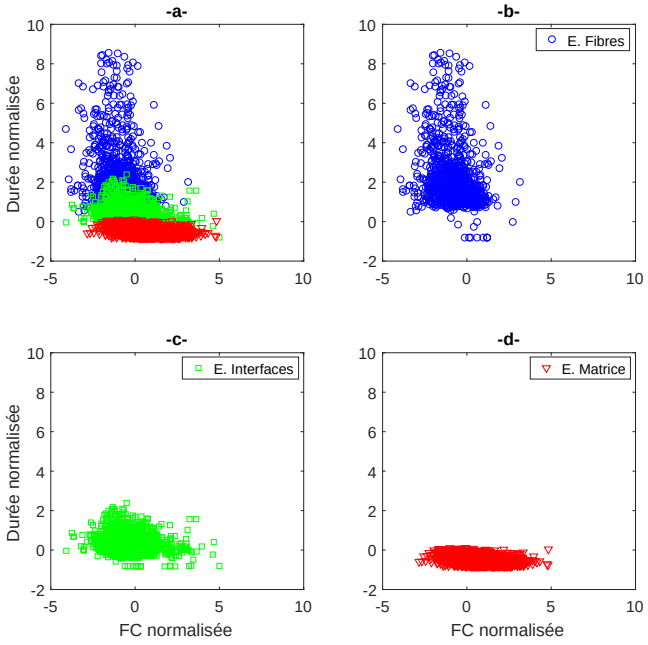


FIGURE 3.15 – Base de données d'apprentissage des composites à fibres de lin. Classification réalisée par la méthode incrémentale. Projection effectuée sur le plan (FC, Durée). -a- Les trois classes d'endommagements projeté au même endroit, -b- classe d'endommagement des fibres, -c- classe d'endommagement des interfaces, -d- classe d'endommagement de la matrice.

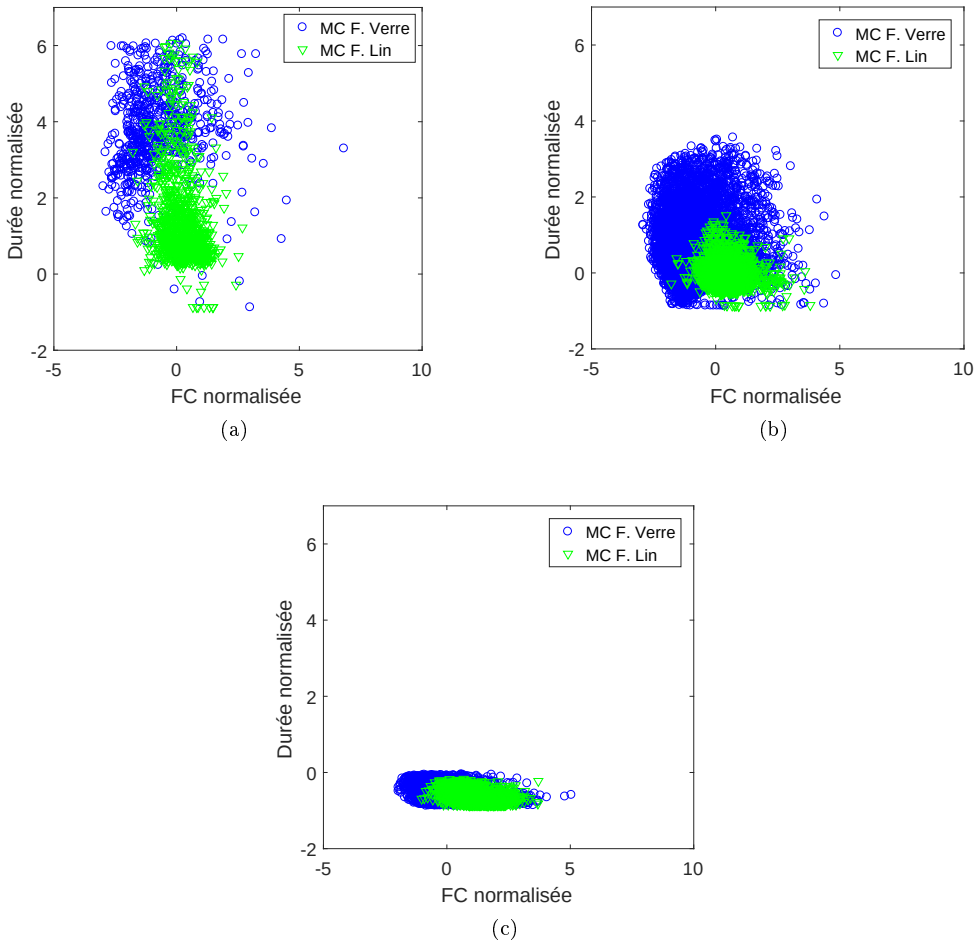


FIGURE 3.16 – Projection sur le plan (*FC, Durée*) pour chaque classe d'endommagement des signaux d'EA des deux bases de données d'apprentissage. (a) classe d'endommagement des fibres, (b) classe d'endommagement des interfaces et (c) classe d'endommagement de la matrice.

3.4.3 Performance de la base de données d'apprentissage.

L'objectif de cette partie est de mesurer les performances de la base de données d'apprentissage des composites lin-époxyde et verre-époxyde. D'abord, les deux bases de données ont été utilisées séparément pour l'apprentissage d'un classifieur supervisé, il s'agit du classifieur k-PPV. Les performances de ce classifieur ont été mesurées par l'AUC, la précision, la sensibilité et la spécificité. Ensuite, les deux bases de données d'apprentissage ont été regroupées pour former une nouvelle base de données d'apprentissage (base de données globale) avec cinq classes d'endommagement. Cette base de données sera utilisée pour l'apprentissage d'un classifieur supervisé afin d'identifier les signaux d'EA des composites hybrides. Enfin, les performances de la nouvelle base de données ont été aussi mesurées. Chaque base de données a été divisée en deux parties, la première pour l'apprentissage du classifieur et la deuxième pour le test.

Le tableau 3.11 résume les performances obtenues par le classifieur k-PPV sur les trois bases de données de tests. Concernant la base de données de matériaux composite à fibres de lin, le k-PPV classifié les signaux d'EA des mécanismes d'endommagement de la matrice, des fibres et des interfaces avec un couple (sensibilité, AUC) de (94,68%, 0.9363), (91,89%, 0.9559) et (87,96%, 0,9110), respectivement. Les signaux d'EA de tous les endommagements ont été classifiés avec une précision de 92,24% et AUC moyenne de 0.9344. Pour la base de données des matériaux composites à fibres de verres, le k-PPV permet de classifier les signaux d'EA des mécanismes d'endommagement de la matrice des fibres et des interfaces avec un couple (sensibilité, AUC) de (95,69%, 0.9501), (81,39%, 0.9051) et (93,17%, 0,9395), respectivement. Le classifieur atteigne une précision de 93,98% et un AUC moyenne de 0.9316 en distinguant les signaux d'EA de tous les endommagements. Ce qui concerne la base de données globale, le classifieur k-PPV classifié les signaux d'EA des mécanismes d'endommagement de la matrice, des fibres de lin, des fibres de verre, des interfaces de lin et des interfaces de verre avec un couple (sensibilité, AUC) de (92,09%, 0,9386), (80,58%, 0,8946), (80,58%, 0,9012), (83,72%, 0,8890) et (86,21%, 0,9049), respectivement. Les signaux d'EA de tous les endommagements ont été classifiés avec une précision acceptable de 87,17% et AUC moyenne de 0.9057.

A partir de ces résultats, on constate que les performances du classifieur k-PPV sont moins bonnes dans la base de données de test globale. Cette dégradation peut être due au fait qu'il y avait des signaux d'EA similaires entre les endommagements des matériaux composites à fibres de lin et ceux à fibres de verre. Ce raisonnement, peut être justifié par le fait que la dégradation a été beaucoup moins importante dans l'endommagement de la matrice. Le k-PPV classifié les signaux d'EA des endommagements de la matrice avec un AUC de l'ordre de 0.9400 dans toutes les bases de données de tests.

3.4.4 Comparaison entre la méthode incrémentale et la méthode de k-moyennes

Cette partie est consacrée à une comparaison entre la méthode de k-moyennes et celle incrémentale, en terme de résultat de classification. Dans un premier temps, les limites de la méthode de k-moyennes sont présentées à travers l'application de celle-ci sur une base de données de signaux d'EA. Dans un second temps, les résultats obtenus par la méthode de k-moyennes sont comparés avec ceux de la méthode incrémentale. A noter que les mêmes bases de données et descripteurs de signaux d'EA sont utilisés pour les deux méthodes.

a) Limites de la méthode de k-moyennes dans la classification de signaux d'EA

L'étude bibliographique présentée dans le chapitre 2 a révelé que la méthode la plus utilisée dans la classification de signaux d'EA est la méthode de k-moyennes. Cette méthode a été utilisée pour discriminer les modes d'endommagement de différents matériaux composites à fibres synthétiques et à fibres végétales. Cependant, la classification de signaux d'EA par cette méthode présente certains inconvénients. Entre autres, le nombre de clusters doit être déterminé avant la mise en œuvre de la méthode. De plus, les résultats peuvent être sensibles aux conditions initiales. En effet, le point de départ sélectionné pour l'initialisation des centres de clusters peut produire des résultats différents. Ces deux inconvénients entrainent une remise en question de l'homogénéité des clusters fournis par la méthode de k-moyennes. Le troisième inconvénient réside dans le fait qu'après avoir appliqué la méthode sur une base de données d'EA, des expertises supplémentaires doivent être utilisées pour corréler les modes d'endommagement aux groupes obtenus. Ce dernier est l'inconvénient majeur de l'utilisation de la méthode de k-moyennes pour discriminer les modes d'endommagement.

Pour illustrer ces inconvénients, la méthode de k-moyennes est appliquée sur une base de données d'EA collecté à partir de l'essai de traction sur un matériau composite UD0 de lin. Le nombre de cluster est fixé à $k = 3$. Ensuite, la méthode de k-moyennes est appliquée plusieurs fois (changement des conditions

Tableau 3.11 – Performances obtenues par le classifieur k-PPV sur les bases de données de test.

Base de données	Classe d'endommagement	Sensibilité	Spécificité	AUC	AUC moyenne	Précision
M. C. à fibre de lin	Matrice	94,68	94,42	0,9363	0,9344	92,24
	Fibres	91,89	94,24	0,9559		
	Interfaces	87,96	87,61	0,9110		
M. C. à fibre de verre	Matrice	95,69	94,20	0,9501	0,9316	93,98
	Fibres	81,39	89,15	0,9051		
	Interfaces	93,17	94,09	0,9395		
Globale	Matrice	92,09	92,34	0,9386	0,9057	87,17
	Fibres de lin	80,58	81,76	0,8946		
	Fibres de verre	80,58	84,82	0,9012		
	Interfaces de lin	83,72	81,32	0,8890		
	Interfaces de verre	86,21	87,30	0,9049		

initiales à chaque application) sur la même base de données. La figure 3.17 représente les différents résultats de classification. On constate que la méthode de k-moyennes fournit un regroupement très différent sur la même base de données d'EA. Cette différence confirme que le regroupement par cette méthode est fortement lié aux conditions initiales.

b) Comparaison entre la méthode de k-moyennes et la méthode incrémentale

La méthode de k-moyennes et celle incrémentale sont comparées à partir de leurs résultats de classification sur une base de données de signaux d'EA d'un matériau composite UD0 lin. Les 7 mêmes descripteurs d'EA (Durée, ENA, TM, FC, FM, FR, FI) sont utilisés pour les deux méthodes. Pour simplifier la présentation des résultats de classification, les signaux d'EA sont projetés seulement sur un plan, celui engendré par FC et la durée. Les résultats obtenus à partir de ces deux méthodes, en utilisant 3 groupes, sont illustrés sur la figure 3.18 (a). Les projections individuelles des 3 groupes sont illustrées sur les même figures (b, c et d). Pour la méthode de k-moyennes, les résultats présentés sur la figure 3.17-b sont choisis pour être comparés avec ceux de la méthode incrémentale.

La figure 3.18 montre une différence notable entre la méthode de k-moyennes et celle incrémentale. En effet, la partition en groupe est différente entre les deux méthodes. La classification obtenue par la méthode de k-moyennes montre la présence de trois groupes. Au contraire, le regroupement des mêmes signaux obtenu par la méthode incrémentale montre la distribution de trois mécanismes d'endommagement, lié à la matrice, aux interfaces et aux fibres.

Dans tous les cas, la difficulté d'utilisation de la méthode de k-moyennes réside dans l'étape d'interprétation et de validation des résultats obtenus.

La figure 3.19 représente l'évolution du nombre de salves des événements d'EA en fonction du temps. Les résultats de la méthode incrémentale montrent que l'endommagement engendré par la matrice commence au début de l'essai et totalise environ 63% des événements d'EA. Les endommagements des interfaces apparaissent également dès le début et leur nombre augmente pendant le test et atteint 30% d'événements d'EA en cumulé. Malgré leur augmentation significative à la fin de l'essai, les signaux d'EA dus à l'endommagement des fibres restent faibles car leur nombre cumulé n'excèdent pas 7%. L'endommagement des fibres conduit à la ruine finale du matériau. S'agissant des résultats obtenus par la méthode de k-moyennes, les trois clusters indiquent un taux d'événements d'EA d'environ 35%, 1% et 64% pour les clusters 1, 2 et 3, respectivement.

En analysant séparément les taux d'évènement d'EA et la projection des groupes sur le plan, on constate que les résultats obtenus par les deux méthodes sont cohérents. Par contre, si les deux analyses sont combinées, on constate cette fois-ci que les résultats des deux méthodes sont totalement contradictoires. Pour mieux comprendre cette contradiction, des liens sont établis entre les classes d'endommagement obtenus par la

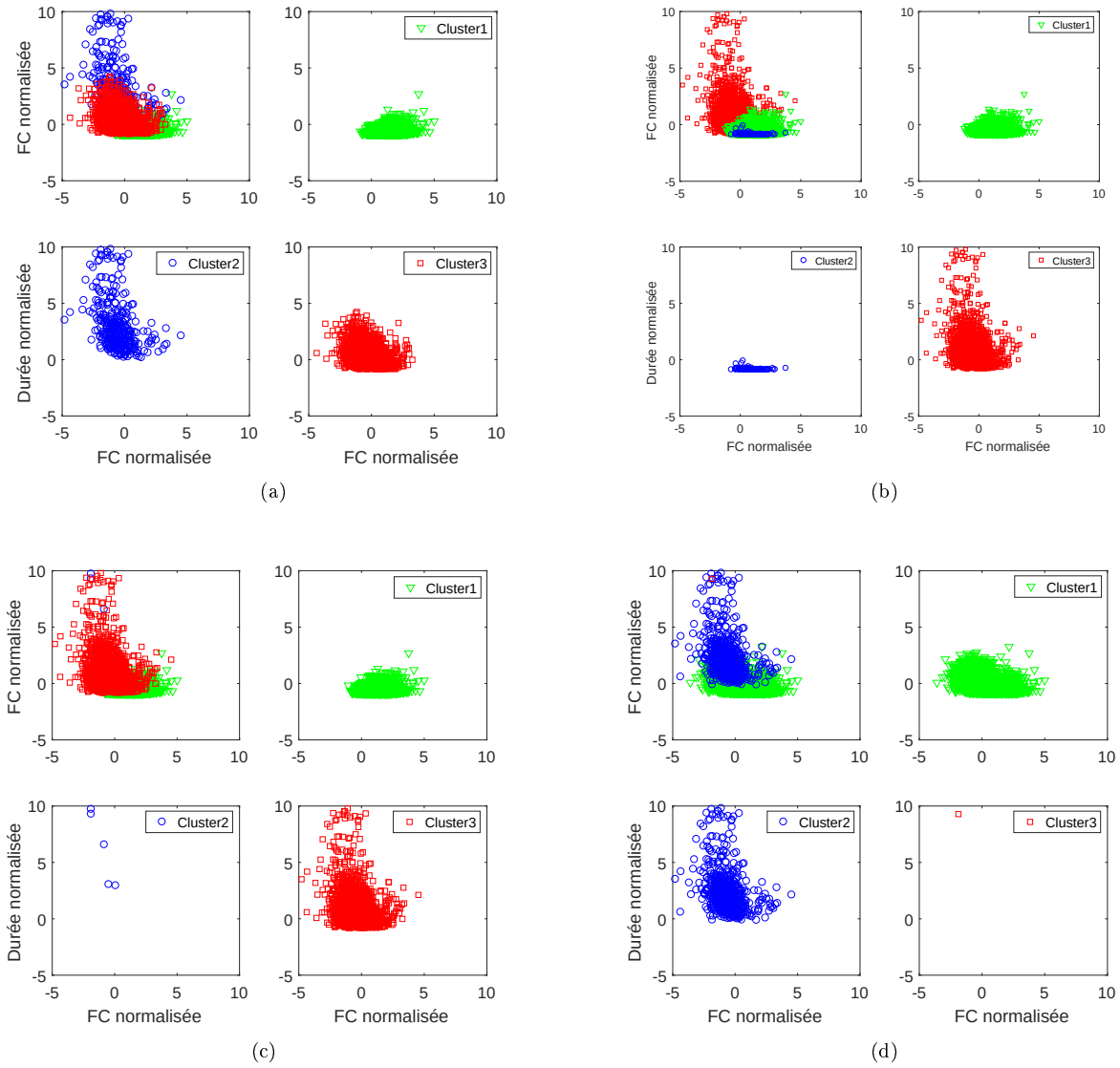


FIGURE 3.17 – Projection sur le plan (*FC, Durée*) des clusters obtenus par la méthode de k-moyennes sur une base de données d'un composite UD0 de lin. La méthode de 3-moyennes est appliquée plusieurs fois (changement des conditions initiales à chaque application) sur la même base de données : (a) première, (b) deuxième, (c) troisième et (d) quatrième application.

méthode incrémentale et les clusters obtenus par la méthode de k-moyenne. Ces liens sont effectués en fonction : Cas 1) du taux d'événement (Figure 3.19), Cas 2) de la projection sur le plan (Figure 3.18).

— Cas 1) En fonction du taux d'événement, on constate que les mécanismes d'endommagement de la matrice, des interfaces et des fibres, sont en correspondance avec le cluster 3, le cluster 1 et le cluster 2, respectivement. En se référant maintenant à la projection sur le plan, on constate que les couples d'endommagement (matrice, cluster 3) et (fibres, cluster 2) n'occupent pas la même zone dans le plan. A noter aussi que le cluster 2 de la méthode de k-moyennes a moins détecté de signaux associés aux endommagements des fibres. Le cluster 1 a cependant plus d'évènement lié aux endommagements de l'interface. Ainsi, certains évènements, liés aux fibres, détectés par la méthode incrémentale sont considérés comme des endommagements des interfaces avec la méthode de k-moyennes.

— Cas 2) En fonction de la projection, on constate cette fois-ci que les mécanismes d'endommagement de la matrice, des interfaces et des fibres sont en correspondance avec le cluster 2, le cluster 1 et le cluster 3, respectivement. De même, si en se référant au taux d'événement, on constate que l'endommagement de la matrice contient beaucoup plus d'événement d'EA (63%) par rapport au cluster 2 (1%). Cependant ces deux classes censées être en correspondance. De plus, Le cluster 3 a plus d'événement (64%) par rapport à l'endommagent des fibres (7%). Cependant, ces deux classes doivent aussi être en correspondance.

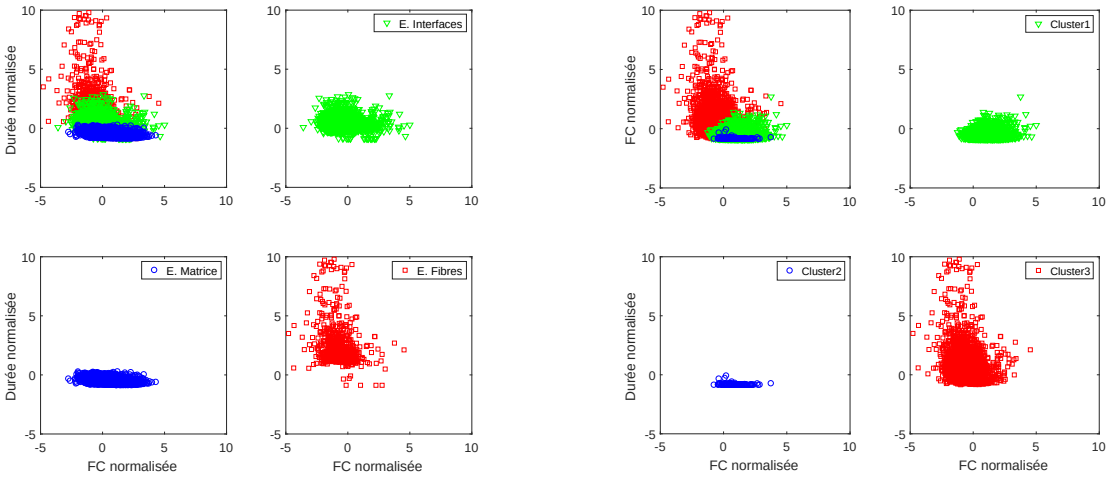
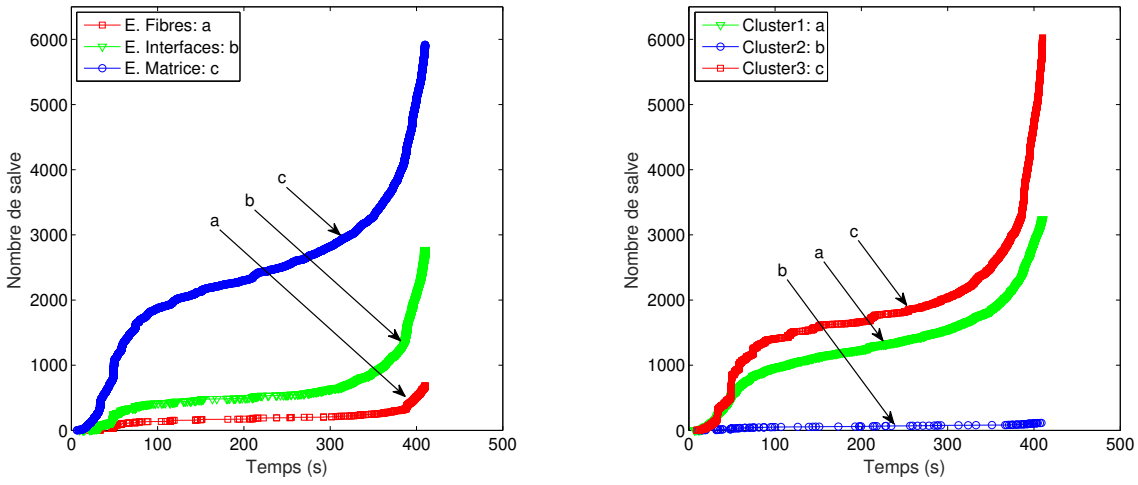


FIGURE 3.18 – Projection sur le plan (FC, Durée) des trois clusters obtenus par : (a) Méthode incrémentale et (b) Méthode de k-moyennes. Base de données d'EA obtenu par essai mécanique sur un matériau composite UD0 de lin.



(a) (b)

FIGURE 3.19 – Evolution du nombre de salves d'EA en fonction du temps dans les trois cluster obtenus par : (a) la méthode incrémentale, (b) la méthode de k-moyennes. Base de données d'EA obtenu par essai mécanique sur un matériau composite UD0 de lin.

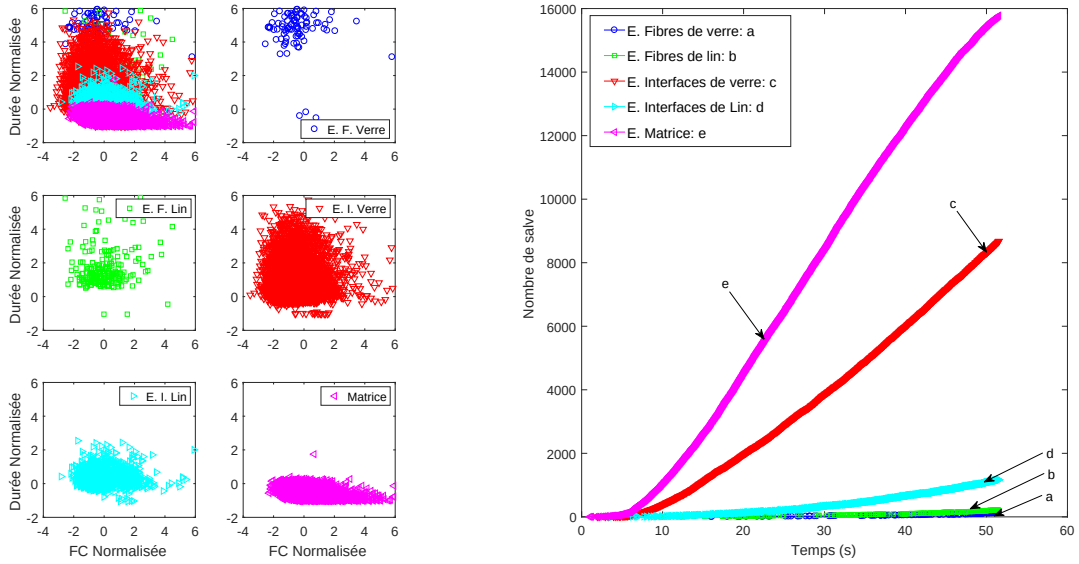
3.4.5 Application de la classification aux composite hybrides lin-verre

L'objectif de cette partie est de classifier les signaux d'EA des composites hybrides lin-verre. Pour ce faire, deux matériaux composites avec deux séquences d'empilement différentes ont été testés. Le premier est un composite stratifié symétrique composé d'un empilement de 2 plis de verre, 4 plis de lin et 2 plis de verre, référence $[V_2/L_2]_s$. Le second est caractérisé par un empilement inversé, de 2 plis de lin, 4 plis de verre et 2 plis de lin, référence $[L_2/V_2]_s$. Ensuite, des essais de traction, associés à la technique de l'EA, ont été effectués afin d'identifier les différents mécanismes d'endommagement de ces deux matériaux composites. Pour ce faire, les bases de données d'apprentissages construites auparavant par la méthode incrémentale sont utilisées pour l'apprentissage d'un classifieur supervisé afin de construire un rôle de décision. Ce dernier est utilisé pour la classification des signaux d'EA inconnus.

La base de données d'apprentissage qui contient 5 classes d'endommagement (fissuration de la matrice, interfaces de verre, interfaces de lin, fibres de verre et fibres de lin) est utilisée pour l'apprentissage du classifieur *1*-PPV. Ensuite, les signaux d'EA des bases de données des deux composites hybrides sont classifiés par le principe de la méthode de *1*-PPV dans l'une de 5 classes. Les figures 3.20 et 3.21 montrent les résultats de cette classification. Le tableau 3.12 donne le pourcentage de salves d'EA engendré par chaque mode d'endommagement.

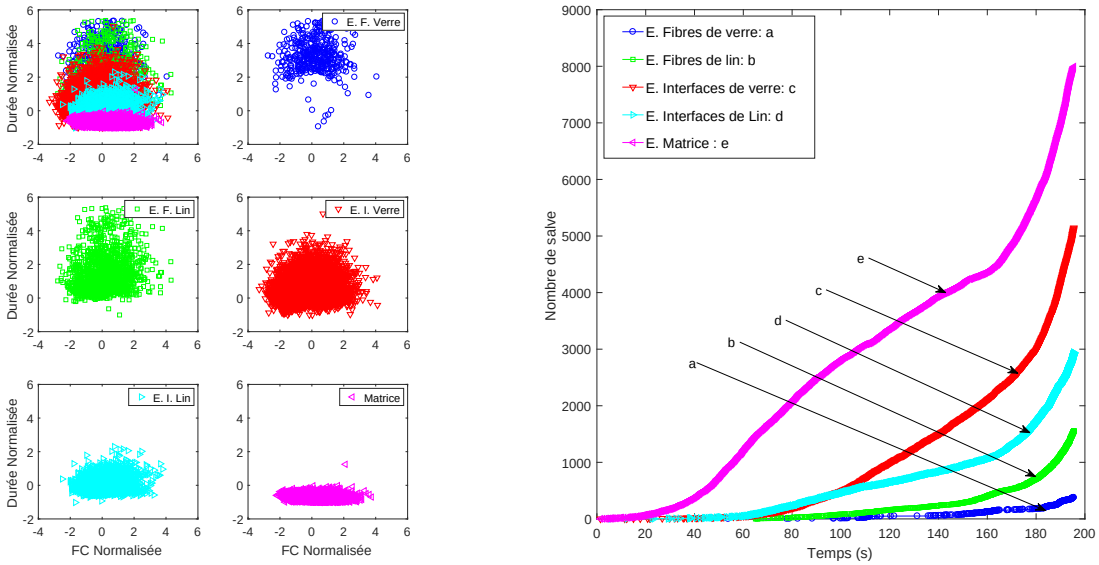
D'après le tableau 3.12 et les figures 3.20 et 3.21, on remarque que :

- Pour le composite $[L_2/V_2]_s$, les endommagements dominants sont la fissuration de la matrice et l'interface



(a) (b)

FIGURE 3.20 – (a) Projection sur le plan (FC, durée) et (b) évolution du nombre de salve des signaux d'EA des classes d'endommagement du matériaux composite hybride $[V_2/L_2]_s$. Les signaux d'EA sont classifiés par la méthode de 1-PPV en s'appuient sur la base d'apprentissage.



(a) (b)

FIGURE 3.21 – (a) Projection sur le plan (FC, durée) et (b) évolution du nombre de salve des signaux d'EA des classes d'endommagement du matériaux composite hybride $[L_2/V_2]_s$. Les signaux d'EA sont classifiés par la méthode de 1-PPV en s'appuient sur la base d'apprentissage.

Tableau 3.12 – Nombre d'évènement d'EA en pourcentage produite par chaque mode d'endommagement. Classification des bases de données des matériaux composites hybrides ($[L_2/V_2]_s$,$[V_2/L_2]_s$).

Matériau composite hybride	Pourcentage des évènement d'EA par mécanisme d'endommagement				
	Matrice	Interfaces de verre	Interfaces de lin	Fibres de verre	Fibres de lin
$[L_2/V_2]_s$	45	29	16	3	7
$[V_2/L_2]_s$	59	32	6	2	1

entre la fibre de verre et la matrice avec un taux de salves d'EA de $\simeq 45\%$ et $\simeq 29\%$, respectivement. Le nombre de salves des endommagements de la matrice augmente significativement à partir de 50s du début de l'essai. Cependant, le nombre de salves des endommagements des interfaces de verre n'augmente significativement qu'à partir de 100s du début de l'essai. Les endommagements des interfaces de lin génèrent un taux de salves d'EA de l'ordre de 16% et leur nombre augmente linéairement entre 60 et 160 puis significativement à partir de 160s de début du l'essai. Les endommagements des fibres de lin quant à eux, produisent un taux de salves de $\simeq 7\%$. Par contre, leur nombre ne commence à augmenter qu'à partir de 150s de début de l'essai. Les endommagements qui fournissent le taux le plus faible de salves (3%) sont les endommagements liés aux fibres de verre. Cette catégorie d'endommagement n'apparaît qu'à la fin de l'essai et conduit le matériau à la rupture finale. En analysant le taux de salves d'EA produit par chaque mécanisme d'endommagement, on constate que les endommagements des fibres de lin produisent plus de salves par rapport à ceux des fibres de verre. Cela peut être expliqué par le fait que la contrainte à la rupture des fibres de lin est inférieure à celle des fibres de verre. Donc, les fibres de lin sont moins résistant et leur endommagement commence avant celui des fibres de verre. d'autre part, les endommagements des interfaces de lin produisent beaucoup moins de salves par rapport à ceux des interfaces de verre. Deux hypothèses peuvent l'expliquer. La première est la présence de signaux d'EA similaires entre les deux types d'endommagement. La deuxième plus possible est liée au fait que les fibres de verre adhérent mieux avec la matrice et l'interface résultante verre/matrice devient plus rigide et résistant que l'interface lin/matrice. Donc, dans un composite hybride lin-verre, il y a plus d'interface verre/matrice que d'interface lin/matrice.

- Pour le composite $[V_2/L_2]_s$, les endommagements dominants sont toujours la fissuration de la matrice et les interfaces de verre avec un taux de salves d'EA plus élevé que le composite $[L_2/V_2]_s$, $\simeq 59\%$ et $\simeq 32\%$, respectivement. Cette fois-ci, le nombre de salves produites par ces deux modes d'endommagement augmente significativement dès le début de l'essai (10s). Les endommagements des interfaces et des fibres de lin génèrent un taux de salves d'EA plus faible par rapport au composite $[L_2/V_2]_s$ ($\simeq 6\%$ pour les interfaces et $\simeq 1\%$ pour les fibres). Cette diminution est due au fait que les fibres de lin ont été protégées par les couches de verre placées à l'extérieur. L'endommagement des fibres de verre garde approximativement le même taux de salves

(2%). Cela est toujours due au fait que les fibres de verre résistent aux contraintes et leur endommagement conduit le matériau à la rupture finale.

3.5 Conclusion

Le premier objectif de ce travail doctoral est de proposer une solution aux deux premières limites des approches actuelles de classification de signaux d'EA. Ces limites sont liées à l'utilisation des méthodes de classification non-supervisée et à l'étape d'identification des groupes obtenus par ces méthodes. Pour ce faire, nous avons développé une méthode de regroupement dite méthode incrémentale afin de construire une base de données d'apprentissage de signaux d'EA pour pouvoir utiliser les méthodes supervisées. L'avantage de cette méthode par rapport à la méthode de k-moyennes se situe dans le fait que l'étape d'identification des classes obtenus n'est pas nécessaire. Chaque groupe obtenu par la méthode incrémentale doit être en correspondance avec un et un seul endommagement précis. Le principe de la méthode incrémentale consiste à combiner deux bases de données de signaux d'EA dont l'un d'entre eux contient un type d'endommagement supplémentaire par rapport à l'autre. L'objectif est de séparer les signaux correspondant à l'endommagement supplémentaire et ceux liés aux autres endommagements. Pour construire la base de données d'apprentissage, les signaux d'EA sont d'abord collectés à partir d'essais mécaniques spécifiques sur les matériaux composites UD. Ils sont ensuite regroupés à l'aide de la méthode incrémentale à partir de leurs descripteurs. Enfin, les groupes identifiés sont utilisés comme base de données d'apprentissage pour séparer les mécanismes d'endommagement dans des matériaux plus complexes. L'application de la méthode s'est avérée efficace pour détecter tous les types d'endommagement dans deux composites à renfort hybride lin-verre. Elle a permis de dissocier les endommagements de chacun des constituants, à savoir la matrice, les interfaces fibres de lin/matrice et fibres de verre/matrice, les fibres de lin et de verre.

Deuxième partie

Approche de classification basée sur une mesure de dissimilarité : application aux signaux EEG d'épilepsie et aux images

Chapitre 4

Introduction

Sommaire

4.1 Généralité **77**

4.2 Représentation par dissimilarité **79**

4.3 Objectif de la deuxième partie **80**

4.4 Organisation de la deuxième partie **81**

4.1 Généralité

Le défi de la reconnaissance automatique des formes consiste à développer des méthodes informatiques qui apprennent à identifier un nouvel objet, pour lequel on ne dispose pas d'information a priori, en se basant sur la connaissance préalable de certain nombre de classes (une base de données d'objets d'apprentissage). Tout d'abord, une représentation appropriée des objets doit être trouvée. Ensuite, une règle de décision peut être construite. Cette règle de décision doit être capable de distinguer entre les différentes classes d'objets d'apprentissage et capable de se généraliser (atteindre une grande précision lors de l'identification de nouveaux objets).

D'une part, les objets sont classiquement représentés par des caractéristiques (descripteurs). Cependant, les descripteurs pertinents peuvent être difficiles à obtenir ou inefficients pour construire la règle de décision. Par exemple, lorsque les experts ne peuvent pas définir les descripteurs d'une manière directe ou lorsque les objets sont de grande dimension [107]. Un autre inconvénient de l'utilisation des descripteurs est que des objets différents peuvent avoir la même représentation s'ils diffèrent principalement par des propriétés qui ne sont pas exprimées dans l'ensemble des descripteurs choisies [108].

D'autre part, l'une des représentations possibles est basée sur une mesure de dissimilarité entre les objets.

Lorsqu'elle est correctement définie, elle peut être avantageuse pour construire les règles de décisions et résoudre les problèmes d'identification d'objets. La représentation des objets par dissimilarité [107–109] a été proposée à l'origine comme une représentation plus flexible que celle basée sur les descripteurs. Dans cette approche, principalement conçue à des fins de classification, les objets sont représentés par leurs dissimilarités avec un ensemble d'objets représentatifs de chaque classe. En effet, si l'on suppose que les objets dits « similaires » peuvent être regroupés pour former une classe, une « classe » n'est rien de plus qu'un ensemble de ces objets « similaires ». S'appuyant sur cette idée, Duin et al. [109] affirment que la notion de dissimilarité est en fait plus fondamentale que celle de descripteurs pour définir une classe. Ainsi, le fait que les dissimilarités devraient être plus petites pour les objets similaires (même classe) et plus larges pour des objets différents suggère qu'elles pourraient être utilisées comme des descripteurs plus discriminants.

La figure 4.1 représente la différence par rapport à l'espace de représentation entre l'approche classique basée sur les descripteurs et l'approche basée sur la dissimilarité [110].

La méthode du plus proche voisin est traditionnellement appliquée dans les représentations par dissimilarité. Bien que cette règle de décision soit basée sur des voisinages locaux, c'est-à-dire un ou quelques voisins, elle est toujours coûteuse en calculs, puisqu'il faut trouver les dissimilarités avec tous les objets d'apprentissage. Un autre inconvénient est que ses performances sont potentiellement diminuées lorsque le jeu d'apprentissage est petit. Pour surmonter ces limitations et améliorer la précision de la classification, Pekalska et Duin ont proposé de remplacer cette méthode par une règle de décision plus globale [107]. Un tel classifieur est construit à partir d'un ensemble d'apprentissage représenté par les dissimilarités avec un ensemble de représentation (un ensemble de prototypes ou d'objets représentatifs pour chaque classe). Si ce

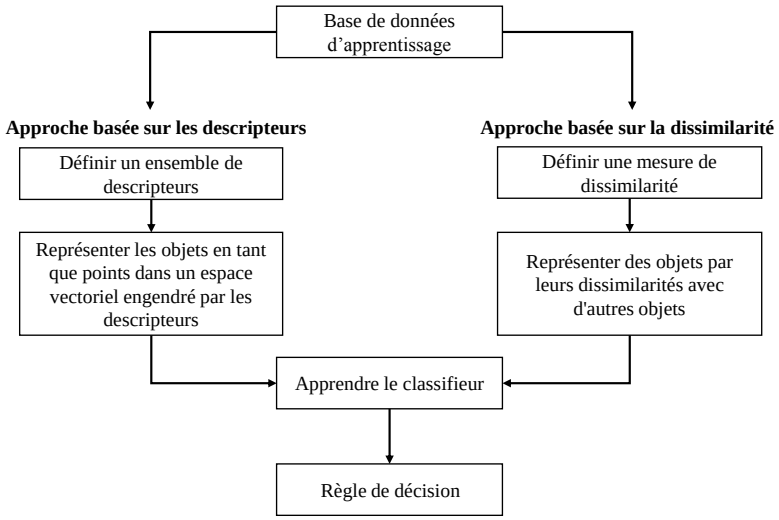


FIGURE 4.1 – Différence entre l'approche basée sur les descripteurs et l'approche basée sur la dissimilarité.

dernier est petit, le classifieur a l'avantage que seul un petit nombre de dissimilarité doit être calculé pour son apprentissage et son évaluation. Cependant, le classifieur peut encore bénéficier de la précision offerte par un grand ensemble d'apprentissage.

4.2 Représentation par dissimilarité

Afin d'introduire le principe de cette approche, on considère un ensemble d'objets d'apprentissage $T = \{o_1, ..., o_n\}$, un ensemble de représentation $R = \{r_1, ..., r_k\}$ et une mesure de dissimilarité, la distance entre chaque objet o_i à chaque objet r_k sera définie comme $d(x_i, r_k)$. L'ensemble de représentation R peut être un sous-ensemble de T ou T lui-même. Il est à noter que T et R peuvent être des ensembles complètement différents. Un objet de l'ensemble d'apprentissage est alors représenté par un vecteur de dissimilarité $d(o_i, R) = [d(o_i, r_1), ...(o_i, r_k)]^T$. Par conséquent, à la place de la matrice de descripteurs classique $T \in \mathbb{R}^{pxn}$ où n représente les objets et p représente les variables, l'ensemble d'apprentissage est maintenant représenté par la matrice de dissimilarité $D(T, R)$ de taille $n \times r$, qui associe tous les objets de l'ensemble d'apprentissage à tous les objets de l'ensemble de représentation :

$$D = \begin{pmatrix} d(o_1, r_1) & d(o_1, r_2) & ... & d(o_1, r_k) \\ d(o_1, r_1) & d(o_2, r_2) & ... & d(o_2, r_k) \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ d(o_n, r_1) & d(o_n, r_2) & ... & d(o_n, r_k) \end{pmatrix}$$

Nous construisons à partir de cette matrice un espace vectoriel de dissimilarité. Les objets sont représentés dans cet espace par les vecteurs lignes de la matrice de dissimilarité, de telle sorte que chaque dimension correspond aux dissimilarités avec l'un des objets de représentation (figure 4.2). En utilisant la représentation par dissimilarité, les classifieurs sont implémentés dans l'espace de dissimilarités entre les objets, au lieu de l'espace de descripteurs classique. Par conséquent, les liens de dissimilarités entre tous les objets de l'ensembles d'apprentissage et de représentation sont utilisés pour la classification. Si une mesure de dissimilarité adéquate est choisie, la discrimination vient du fait que les dissimilarités sont faibles pour les objets similaires (appartenance à la même classe) et grandes pour les objets distincts (appartenant à des classes différentes). Par conséquent, il devrait être plus facile pour les classifieurs de distinguer entre les classes, de sorte que les classifieurs linéaires dans l'espace de dissimilarité puissent correspondre à un classifieur non linéaire dans l'espace de descripteurs. En général, tout classifieur fonctionnant sur des descripteurs peut être utilisé.

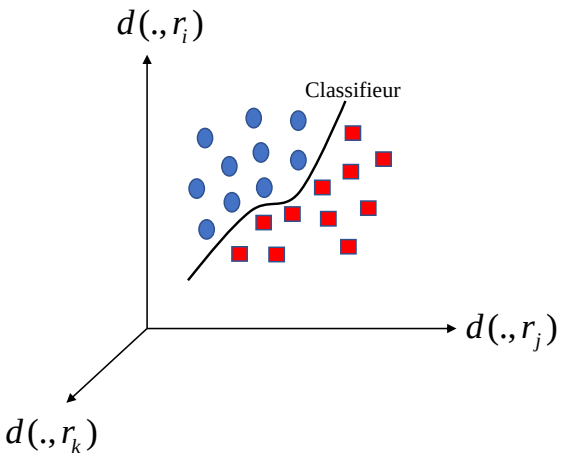


FIGURE 4.2 – Espace vectoriel de dissimilarités

Étant donné un ensemble de tests $S = \{s_1, ..., s_m\}$, ces objets sont classés dans l'espace de dissimilarité, en utilisant leurs distances par rapport aux prototypes R, ce qui constitue une matrice $D(S, R)$ de taille m x r.

4.3 Objectif de la deuxième partie

L'objectif du présent travail est d'exploiter et de tester cette approche avantageuse pour la classification de signaux et d'images.

D'une part, il est aujourd'hui bien admis que les représentations d'un signal conjointement en temps et en fréquence offrent un réel intérêt. En effet, elles permettent une description des signaux non-stationnaires, c'est à dire l'analyse des lois de comportement fréquentiel du signal au cours du temps. Ainsi, plusieurs méthodes de classification ont été développé pour la classification de signaux dans le plan T-F. La plupart de ces méthodes sont basées soit sur l'approche de l'extraction caractéristiques ou bien sur le principe de plus proches voisins. Par ailleurs, aucune attention n'a été accordée pour combiner un espace de dissimilarités avec l'ATF pour la classification de signaux non-stationnaire. Ainsi, cette idée est l'objet principal d'étude du chapitre 6.

D'autre part, et contrairement aux signaux, plusieurs travaux de recherche ont été conduits dans l'espace de dissimilarités pour la classification des images. Les résultats de ces travaux ont montré les capacités discriminatives d'un classifieur construit dans cet espace. Pourtant, les classifieurs ont nécessité beaucoup de données d'apprentissage afin de converger vers les meilleures précisions. Ceci peut être due aux mesures de dissimilarité utilisées. Ces mesures étaient dans la plupart du temps globale et ne prennent pas en considération les dissimilarités locales dans les images. Contrairement à ces mesures, la Carte de Dissimilarité Locale

(CDL) regroupe l'ensemble des mesures de dissimilarités locales faites en différents endroits des images. Ceci permet d'avoir accès à leur répartition spatiale. Ainsi, la CDL fera l'objet du chapitre 7. Nous monterons, à travers deux applications, les capacités discriminatives d'un classifieur construit dans l'espace de dissimilarités engendré par la CDL.

4.4 Organisation de la deuxième partie

Cette deuxième partie est organisée en trois chapitres. Le chapitre 5 introduit les concepts de base de l'approche basée sur la dissimilarité. Les chapitres 6 et 7 quant à eux sont dédiés à une étude expérimentale de l'application de cette approche pour la détection de l'épilepsie (classification de signaux EEG d'épilepsie) et la classification des images (chiffres manuscrits et lettrines).

Chapitre 5. Pour pouvoir situer les travaux de cette partie, ce chapitre présente les principales approches de classification basées sur une mesure de dissimilarité. Trois approches différentes peuvent être distinguées. La première est une extension de l'espace de dissimilarités, comme le prolongement vers un espace de Hilbert afin d'appliquer les classifieurs à noyau. La deuxième approche est basée sur le traitement des dissimilarités en tant que distances dans un espace euclidien et utiliser le classifieur k-PPV. Le concept du dernier est de traiter les dissimilarités entre objets comme des vecteurs de descripteurs pour utiliser n'importe quel classifieur.

Le chapitre 6 est consacré à une application de l'approche de dissimilarité pour la classification de signaux non-stationnaires. Dans un premier temps, un état de l'art sur les travaux de recherche qui ont été déployés pour le traitement et l'analyse des signaux non-stationnaires, s'avère nécessaire. Ensuite, nous présenterons une nouvelle approche de classification de ces signaux, nommée approche de classification basée sur une mesure de dissimilarité entre les RTF. Les classifieurs dans cette approche combinent l'ATF et l'espace de dissimilarités pour construire leurs règles de décisions. Ensuite, nous introduirons le principe de l'ATF comme étant l'approche la plus adaptée pour le traitement de signaux non-stationnaires. En même temps, nous présenterons une liste non exhaustive des mesures de dissimilarité utilisées pour la classification dans le plan T-F. Enfin, nous appliquerons notre approche, sans perte de généralité, pour la classification d'une base de données de signaux EEG d'épilepsie.

Le chapitre 7 présente la deuxième application de l'approche de dissimilarité pour pour classifier des images. Premièrement, une CDL pour les images en niveau de gris est développée et une mesure de dissimilarité est déduite. Ensuite, cette mesure est combinée avec l'approche de dissimilarité pour classifier deux bases de données, une des chiffres manuscrits et l'autre de lettrines. Enfin, Les résultats de l'approche proposée sont comparés avec les approches de la littérature.

Chapitre 5

Représentation et classification par dissimilarité

Sommaire

5.1 Introduction 83

5.2 Représentation des données 84

5.3 Mesure de dissimilarité 85

5.4 Représentation par dissimilarité 86

5.4.1 Prolongement des matrices de dissimilarités 87

5.4.2 Le classifieur k-PPV 87

5.4.3 Espace de dissimilarités 88

5.5 Conclusion 89

5.1 Introduction

Ce chapitre présente les principales approches de classification par dissimilarité qui vont être utilisées tout au long de cette partie. Premièrement, on présente les principaux types de présentation des objets et l'avantage de l'utilisation d'une représentation par dissimilarités. Deuxièmement, on introduit les techniques de classification basées sur une représentation par dissimilarités. Ainsi, on évoque la méthode de prolongement de dissimilarités et le classifieur k-PPV comme étant la méthode la plus simple et la plus directe dans ce contexte. Également on décrit le principe de construction d'un classifieur directement dans un espace de dissimilarités comme étant l'approche la plus globale. Finalement, nous finirons avec une conclusion pour

pouvoir situer les travaux de cette partie.

La plupart des concepts présentés dans ce chapitre peuvent être trouvés dans [107, 110–112].

5.2 Représentation des données

Les systèmes automatiques de reconnaissance des objets comme des images, des vidéos, des signaux, des spectres, etc., sont conçus par apprentissage à partir d'exemples des objets étiquetés suivant leur classe d'appartenance. Ce système de reconnaissance se réalise en suivant deux étapes principales [113] :

Représentation : les objets individuels doivent être décrits de telle sorte que leurs caractéristiques soient extraites et encodées d'une façon numérique ou symbolique. Cette représentation est en fait une simplification, mais elle doit « capter » les caractéristiques les plus importantes, les plus discriminantes des objets d'une classe à une autre. Les représentations les plus courantes utilisent des espaces vectoriels euclidiens, des chaînes de caractères, des graphes, etc. Une bonne représentation doit être compacte. C'est-à-dire que dans l'espace de représentation, les points associés à des objets appartenant à la même classe doivent former des partitions compactes. C'est l'hypothèse de compacité qui stipule que les objets qui se ressemblent dans le monde physique, doivent aussi se ressembler dans leur espace de représentation [113]. Les principaux types de représentation sont résumés dans le paragraphe ci-dessous.

Généralisation : à partir d'un choix de représentation, il s'agit d'apprendre la relation entre les objets à travers leur représentation et leur classe d'appartenance. Ensuite, l'apprentissage sur les objets déjà étiquetés doit permettre l'identification de nouveaux objets non étiquetés. C'est la généralisation de la règle de décision pour des nouvelles observations.

Le thème de la généralisation a été intensément étudié suivant différentes approches telles que la théorie statistique de l'apprentissage [114], la reconnaissance des formes [115], les réseaux de neurones [116] et les machines à vecteurs de support [117–119].

Les principaux types de représentation des objets sont les suivants :

— **Représentation par caractéristiques** : les caractéristiques décrivant les objets doivent permettre de les différencier en classes. Ces caractéristiques sont définies par des experts du domaine d'application afin de mieux exploiter leurs connaissances de l'application. Un inconvénient de l'utilisation des caractéristiques est que des objets différents peuvent avoir la même représentation s'ils diffèrent principalement par des propriétés qui ne sont pas exprimées dans l'ensemble des caractéristiques choisies. Cela se traduit par une erreur de classification intrinsèque qui pourrait être corrigée par l'ajout de nouvelles caractéristiques.

— **Représentation par données brutes** : en prenant l'exemple d'un signal ou d'une image, il s'agit

directement des valeurs d'acquisition, disposées en vecteur. L'utilisation directe des données brutes se fait souvent en l'absence de bonnes caractéristiques. Cette représentation est peu efficace car elle aboutit à une représentation vectorielle des objets, complètement éclatée, sans aucune invariance et dans un espace à très grandes dimensions.

— **Représentation par modèles de probabilité** : les caractéristiques peuvent être reliées à un modèle de probabilité donné. Ces modèles peuvent être fondés sur des connaissances d'experts ou formés à partir des exemples.

— **Représentation par modèles structurels** : au lieu d'utiliser des probabilités, des modèles d'objets peuvent être basés sur une description structurelle telle que des chaînes, des contours, des séquences de temps, des arbres et autres données dépendantes d'un ordre particulier. Des procédures automatiques pour la construction des descriptions structurelles sont encore à développer mais une première idée peut être trouvée dans [120].

— **Représentation par dissimilarité** : Une alternative à l'utilisation des caractéristiques et des données brutes est la représentation par dissimilarité, basée sur des comparaisons directes des objets par paires [107]. C'est ce mode de représentation qui sera utilisé pour cette partie de la thèse

5.3 Mesure de dissimilarité

Pour obtenir une description robuste des objets, une mesure de dissimilarité doit intégrer les invariances nécessaires pour un meilleur regroupement intra-classe et éloignement interclasse. Pour ce faire, il est nécessaire que la dissimilarité intègre des invariances propres au problème posé. Par exemple, pour la reconnaissance d'objets dans les images, la dissimilarité doit être invariante à la translation, à la rotation, à l'échelle spatiale, aux variations d'illumination et robuste au bruit et aux occlusions. La mesure de dissimilarité est spécifiée par l'utilisateur expert du domaine, de manière à intégrer ses connaissances de l'application.

Une mesure de dissimilarité entre deux objets o_i et o_j d'un ensemble d'apprentissage de n objets peut avoir une ou plusieurs des propriétés suivantes pour tout i, j, $k \leqslant n$.

— Non négativité $d(o_i, o_j) \geq 0$

— Réflectivité $d(o_i, o_j) = 0$

— Symétrie $d(o_i, o_j) = d(o_j, o_i)$

— L'inégalité triangulaire $d(o_i, o_j) + d(o_j, o_k) \geq d(o_i, o_k)$

Une mesure de dissimilarité vérifiant les quatre propriétés est une métrique et on peut parler de distance. La dissimilarité ayant les propriétés de réflectivité et de symétrie et obéissant à l'inégalité triangulaire est une distance pseudo-métrique et si elle a les propriétés de non négativité, symétrie et réflectivité mais elle

n'obéit pas à l'inégalité triangulaire alors elle est semi-métrique.

La plupart du temps des métriques sont employées dans les tâches de classifications, en particulier pour leur bonnes propriétés qui permettent des optimisations en coût de calcul. Cependant des mesures non métriques, asymétriques par exemple, sont plus fidèles au jugement humain de la similarité [121, 122]. L'emploi de mesures non métriques peut ainsi permettre des performances accrues par rapport aux mesures métriques [123]. La difficultés réside alors dans la restriction des méthodes de classifications applicables.

5.4 Représentation par dissimilarité

L'utilisation des dissimilarités pour représenter les objets ouvre de nouvelles possibilités en apprentissage statistique [110] car les dissimilarités peuvent capturer à la fois les informations statistiques et structurelles des objets [124]. Des techniques appropriées de classification sont alors développées [107, 125, 126]. Les dissimilarités peuvent être construites directement à partir des données brutes (ex. dissimilarités entre des images et distributions T-F) ou à partir de caractéristiques extraites.

Trois ensembles d'objets peuvent être distingués pour construire une représentation :

— **Un ensemble de représentation** $R = \{r_1, ..., r_k\}$**.** Ce sont les objets auxquels nous nous référons. Les dissimilarités par rapport à l'ensemble de représentations doivent être calculées pour les objets d'apprentissage ainsi que pour les objets de test utilisés pour l'évaluation, ou tout objet à classer ultérieurement. Parfois, les objets de l'ensemble R sont appelés prototypes. Ce mot peut suggérer que ces objets sont en quelque sorte des exemples typiques des classes. Cela peut être le cas mais ce n'est pas nécessaire. Ainsi, des prototypes peuvent être utilisés pour la représentation, mais l'ensemble de représentation peut également être constitué de d'autres objets.

— **Un ensemble d'apprentissage** $T = \{o_1, ..., o_n\}$. Ce sont les objets utilisés pour l'apprentissage du classifieur afin de construire un rôle de décision. Dans de nombreuses applications, nous utilisons $T = R$, mais R peut aussi être juste un (petit) sous-ensemble de T, ou être entièrement différent de T.

— **Un ensemble de test** S. Ce sont les objets qui sont utilisés pour évaluer la procédure de classification. Ils doivent être représentatifs des objets cibles pour lesquels la procédure de classification est construite.

Après avoir déterminé ces trois ensembles d'objets, les matrices de dissimilarité $D(T,T)$, $D(T,R)$ et $D(S,R)$ doivent être calculées. Le prochain problème est de savoir comment utiliser ces trois matrices pour l'apprentissage et les tests. Trois procédures sont généralement considérées :

— **Espace de prolongement**. Il s'agit de déterminer un espace vectoriel sous-jacent aux données de dissimilarité $D(T,T)$. A partir de là, les classifieurs standard sur données vectorielles peuvent s'appli-

quer.

— **Le classifieur k-PPV.** C'est la manière traditionnelle pour classifier des nouveaux objets : Affecter ces objets à la classe (majoritaire) de ses (k) plus proches voisins. Cette procédure peut directement s'appliquer à $D(S,T)$. Les matrices d'apprentissages $D(T,T)$ et $D(T,R)$ ne sont pas utilisées.

— **Espace de dissimilarités.** Cet espace est postulé comme un espace vectoriel euclidien défini par les vecteurs de dissimilarité $d(.,R) = [d(.,r_1),...,\ d(.,r_k)]^T$, calculés sur l'ensemble de représentations R. Réellement, les vecteurs de dissimilarité sont utilisés comme des caractéristiques numériques (descripteurs). Ainsi, ces vecteurs de dissimilarité peuvent être entrées pour n'importe quel classifieur.

Dans le paragraphe qui suit on présentera les principes de ces trois procédures.

5.4.1 Prolongement des matrices de dissimilarités

La procédure de prolongement de dissimilarités vise à représenter au mieux des objets dans un espace visualisable, de façon à ce que les distances entre les objets dans cet espace soient aussi proches que possible des dissimilarités initiales. Une telle configuration se trouve généralement dans un espace euclidien, bien que tout autre espace $\Re^p$, avec $p \geqslant 1$, puisse aussi être considéré. Par conséquent, le résultat d'un algorithme de prolongement est une représentation spatiale des données.

Cette procédure a été récemment étudié d'une manière approfondie. En particulier, une question intéressante était sur le sujet de prolongement des dissimilarités non-euclidien dans un espace vectoriel de sorte que les distances obtenues sont suffisamment précises par rapport aux dissimilarités originales. Les opérations de projection sont simples et linéaires, mais elles peuvent donner des solutions avec des distances négatives par rapport aux autres objets, même si les dissimilarités d'origine sont non négatives. Ainsi, plusieurs aspects ont été étudiés : pourquoi les dissimilarités non-euclidiennes apparaissent-elles, sont-elles informatives et comment les traiter. Ces questions sont difficiles, à la fois du point de vue théorique et mathématique. De plus, le prolongement s'appuie sur une matrice de dissimilarité carrée $D(T,T)$. En effet, les dissimilarités entre tous les objets d'apprentissage doivent être prises en compte. Ainsi, le calcul de cette matrice et la procédure de prolongement elle-même peuvent être ainsi des opérations exigeant du temps et de la mémoire.

5.4.2 Le classifieur k-PPV

Dans le cas des k-PPV, la règle est fondée sur le vote majoritaire. Un objet inconnu devient membre de la classe la plus fréquente parmi ses k plus proches voisins. Habituellement, k est supposé être impair pour éviter les conflits. Lorsque k est fixé aucune étape d'apprentissage n'est nécessaire.

Dans un espace euclidien de prolongement, le classifieur k-PPV est basé sur les distances calculées dans cet

espace. Par définition, il s'agit des dissimilarités originales (à condition que les objets de test soient prolongés avec l'ensemble des objets d'apprentissage). Donc, sans le processus de prolongement, ce classifieur peut directement être appliqué à une matrice de dissimilarité donnée. C'est le classifieur traditionnellement utilisé par de nombreux chercheurs dans le domaine de la reconnaissance des formes structurales. Ce classifieur n'utilise pas les dissimilarités dans la matrice d'apprentissage $D(T, R)$ (ou $D(T, T)$). La classification est entièrement basée sur les dissimilarités d'un objet de test avec les objets de l'ensemble de d'apprentissage $D(S, T)$ (ou de représentation $D(S, R)$) uniquement.

Le classifieur k-PPV est couramment utilisé parce qu'il est simple et a tendance à bien fonctionner sur de grands ensembles d'apprentissage. Cependant, il souffre de plusieurs limitations, à savoir, une complexité de calcul élevée, une perte de précision potentielle lorsqu'un petit ensemble de prototypes est utilisé et une sensibilité au bruit. De plus, la règle du plus proche voisin souffre de décisions locales et ne fonctionne pas pour une mesure de dissimilarité asymétrique.

Pour surmonter ces limitations et améliorer la précision de la classification, le classifieur k-PPV peut être remplacé par une règle de décision plus globale. Une telle décision peut être obtenu par un classifieur construit dans un espace de dissimilarités.

5.4.3 Espace de dissimilarités

L'espace de dissimilarités [107, 110, 127] est un espace vectoriel dans lequel les dimensions sont définies par les vecteurs de dissimilarité. Les éléments de ces vecteurs sont les dissimilarités entre un objet donné et les objets de l'ensemble de représentations R. Les vecteurs de dissimilarité sont considérés comme caractéristiques pour les objets dans la base d'apprentissage. Par conséquent, un tel espace surmonte tous les problèmes qui surviennent habituellement avec les mesures de dissimilarité non-euclidienne, simplement en négligeant la nature de la dissimilarité. En effet, cet espace de représentation peut être utilisé avec des dissimilarités métriques et non métriques. Il construit un espace de caractéristiques, qui sont toutes de même nature (dissimilarité à un prototype donné), contrairement à un espace de caractéristiques standards où sont présentées très souvent des caractéristiques de nature différente.

Ainsi, les classifieurs peuvent être implémentés dans cet espace. La discrimination vient du fait que les dissimilarités sont faibles pour les objets similaires (appartenance à la même classe) et grandes pour les objets distincts (appartenant à des classes différentes). Ainsi l'ensemble de représentation doit être choisi afin que pour deux objets similaires o_i et o_j, leurs vecteurs représentatifs et soient corrélés, c'est-à-dire proches dans l'espace de dissimilarités. Les prototypes r_k sont des observations typiques des classes et non pas des observations aberrantes.

En revanche, dans le cas où tous les objets d'apprentissage T sont utilisés pour la représentation, la dimension de l'espace de dissimilarité est égale à $\mid T \mid$. Bien que, en principe, tout classifieur basé sur l'espace de caractéristique puisse être appliqué à l'espace de dissimilarités, certains d'entre eux seront mal définis ou surentraînés pour un ensemble de représentation aussi grand. En plus, le calcul de la matrice de dissimilarité $D(T,T)$ peut ainsi être une opération exigeant du temps et de la mémoire. La réduction de dimension, par exemple la sélection de prototypes peut donc être une issue importante dans cette approche [128, 129].

5.5 Conclusion

Les dissimilarités sont un moyen naturel de représenter des objets. Certains les considèrent comme plus fondamentales que les caractéristiques [130]. Les dissimilarités ont été étudiées dans [107, 110] pour l'apprentissage supervisé et non supervisé comme une alternative à l'utilisation des caractéristiques pour la construction des représentations. Elles sont particulièrement utiles dans deux contextes : tout d'abord, en l'absence de propriétés claires qui peuvent devenir des caractéristiques et, d'autre part, lorsque les objets peuvent être comparés au niveau global comme par exemples des formes dans les images, les représentations temps-fréquence ou de spectres. Il existe trois principales approches pour la classification dans des données de dissimilarités.

La première est la méthode de prolongement (projection) linéaire ou non linéaire de la matrice de dissimilarités. Avec la projection d'une matrice de dissimilarités dans un espace avec une métrique donnée, la nature des dissimilarités est préservée. Il est naturel de chercher un prolongement dans un espace euclidien car la métrique euclidienne est supposée, soit implicitement ou explicitement dans de nombreux systèmes de classification et dans ce cas-là, on peut utiliser les classifieurs traditionnels. Les distances non-euclidiennes ne peuvent être prolongées que d'une manière approximative dans un espace euclidien. Parfois ce n'est pas suffisant et on arrive à construire un espace pseudo euclidien.

La deuxième est la règle de k-PPV. Cette règle est la technique la plus simple et facile à appliquer directement dans l'espace des dissimilarités. La règle k-PPV fonctionne bien, mais c'est une technique locale, qui a besoin d'un vaste ensemble d'apprentissage pour être efficace, elle est sensible au bruit et ne convient pas pour des dissimilarités qui ne respectent pas l'inégalité triangulaire.

La troisième approche de gérer la matrice de dissimilarités, le postuler comme un espace vectoriel, soulève moins de problèmes et présente un grand intérêt pour des applications pratiques. Il peut, sans problèmes, être utilisé pour presque n'importe quel type de mesure de dissimilarité. De plus, il offre la possibilité d'une complexité de calcul réglable.

Ainsi, la deuxième et la troisième approches sont les cibles de cette partie et elles vont être utilisées pour

la classification de signaux non-stationnaires (chapitre 6) et les images (chapitre 7).

Chapitre 6

Approche de classification basée sur une mesure de dissimilarité entre les RTF : application aux signaux EEG d'épilepsie

Sommaire

6.1 Introduction 92

6.2 Approche de classification basée sur une mesure de dissimilarité entre les RTF 94

6.2.1 Classification dans le plan T-F 94

6.2.2 Approche de classification basée sur une mesure de dissimilarité entre les RTF . . 95

6.3 Analyse temps-fréquence 97

6.3.1 Préambule 97

6.3.2 Définition de la non-stationnarité 102

6.3.3 Définition de la DSP 102

6.3.4 Distribution de Wigner-Ville 102

6.3.5 La classe de Cohen de RTF quadratique 104

6.3.6 Synthèse 107

6.4 Mesure de dissimilarité entre les RTF 107

6.5 Application de l'approche sur une base de données de signaux EEG d'épilepsie109

6.5.1 Préambule 109

6.5.2 Base de données des signaux EEG d'épilepsie 111

6.5.3 Résultats 112

6.6 Conclusion 125

6.1 Introduction

Au cours des dernières décennies, des efforts considérables ont été déployés pour développer des méthodes de traitement du signal numérique efficaces pour la représentation, l'analyse et le traitement des signaux non-stationnaires [131, 132]. Les méthodes basées sur l'ATF sont fréquement préférées pour analyser les signaux non-stationnaires car elles permettent de représenter les composants constituant le signal et suivre leurs variations dans le domaine T-F. Ces méthodes sont souvent utilisées pour l'extraction de caractéristiques (descripteurs) et la classification de signaux dans de nombreux domaines. Parmi ceux-ci, on trouve par exemple l'analyse de signaux biomédicaux [131, 133–147], l'analyse de signaux mécaniques [79, 82, 148], le traitement de signaux radar et sonore [144, 149–156], le traitement de signaux électrique [157–160], le traitement de la parole et les signaux musicaux [161–167, 167–169]. L'approche et les méthodes typiques d'analyse T-F utilisées pour l'extraction et la classification sont illustrées dans la figure 6.1. L'idée principale est de trouver et de reconnaître des motifs dans la représentation T-F du signal qui indiqueraient, par exemple, la présence d'un défaut ou une anomalie [170]. S'il y a une anomalie, alors il faut identifier et la classifier.

Parmi les procédures de traitement du signal basées sur l'analyse T-F est la concentration d'énergie [171]. L'idée est d'analyser le comportement de la distribution d'énergie, c'est-à-dire la concentration d'énergie à un instant donné ou à une certaine bande de fréquence ou plus généralement, dans une région T-F particulière. Une telle analyse est capable de révéler plus d'informations sur un phénomène particulier à des fins de diagnostic. Cependant, si la concentration d'énergie dans le domaine T-F est utilisée pour l'extraction de descripteurs, il est bien connu que le choix de ces descripteurs affecte la précision de la classification. Pour diminuer ces effets, différentes représentations T-F ont été introduites pour l'extraction de descripteurs. Par exemple, les représentations basées sur la décomposition du signal [172, 173] ou sur l'idée de la classe de Cohen [174, 175].

L'étape qui suit l'extraction de descripteurs est la classification. Cette étape consiste essentiellement en l'apprentissage d'un classifieur sur les vecteurs de descripteurs afin d'établir une règle de décision. Cette dernière est généralement basée sur des écarts statistiques (SVM, Analyse Discriminante Linéaire (ADL), Réseau de Neurones artificiels (RNA),...) ou sur une mesure de distance entre les vecteurs de descripteurs (k-PPV). La performance d'un classifieur est fortement liée à la pertinence des descripteurs. Cependant, Les descripteurs pertinents peuvent être difficiles à obtenir ou inefficient pour l'apprentissage. Par exemple, lorsque les experts ne peuvent pas définir des caractéristiques d'une manière directe ou lorsque les données sont de grande dimension et consistent en un grand ensemble d'attributs comme les représentations T-F.

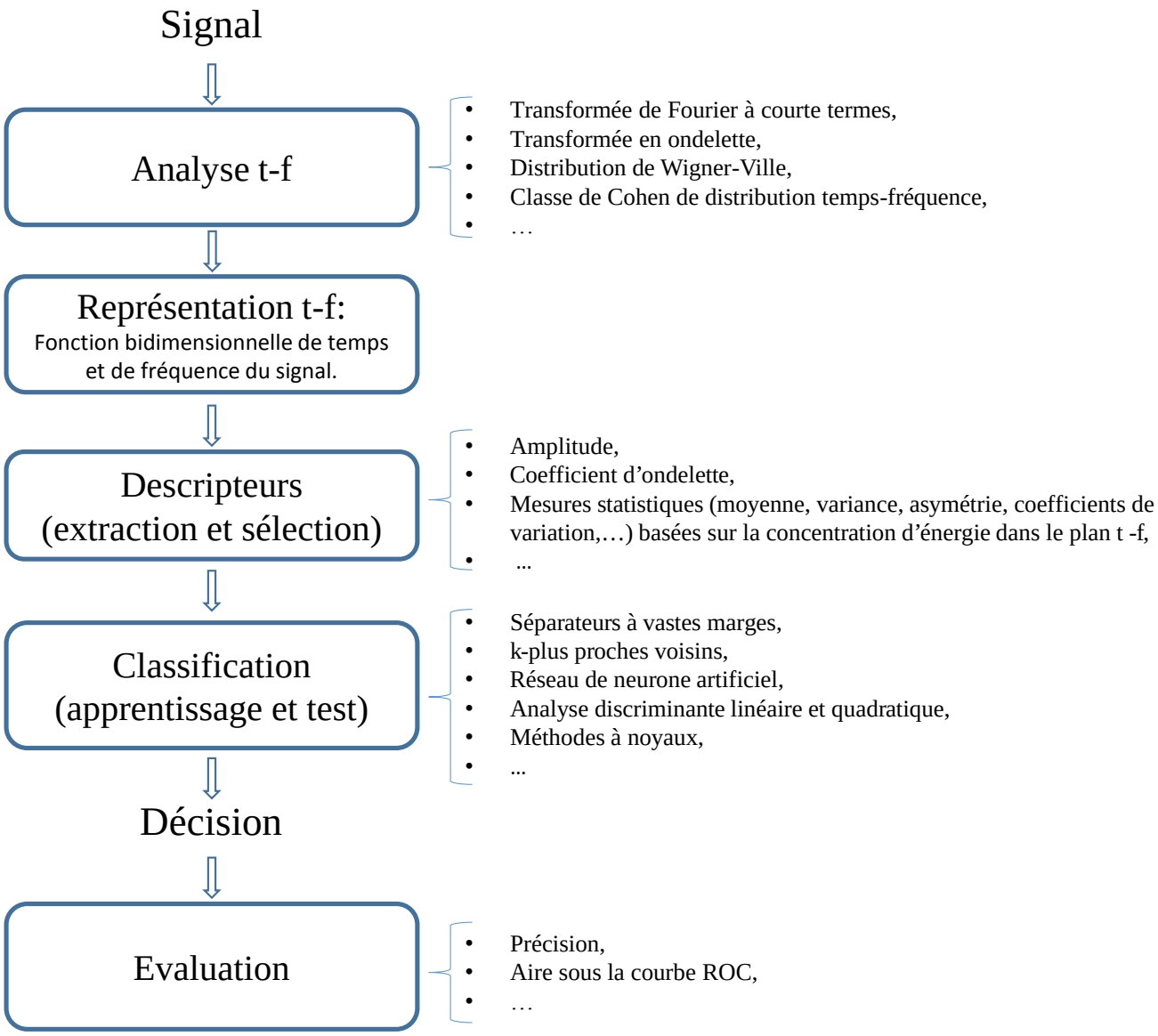


FIGURE 6.1 – Approche typique d'analyse et classification de signal.

En plus, l'extraction de descripteurs peut être une tâche laborieuse et longue. Les descripteurs peuvent être dépendant de la nature, de l'origine des signaux non-stationnaires et du domaine d'application ce qui peut limiter leurs généralisation dans d'autre application. Heureusement, l'approche basée sur l'extraction de descripteurs n'est pas la seule approche sur laquelle des problèmes de classification peuvent être résolus. D'autres moyens peuvent être trouvés en construisant des règles de décision dans l'espace de dissimilarités (chapitre 5).

Ainsi, le but du travail présenté dans ce chapitre est d'exploiter cette approche avantageuse pour la classification de signaux. Nous proposons une approche de classification qui combine les mesures de dissimilarités et les Représentations Temps-Fréquence (RTF) basées sur la distribution de l'énergie, nommée 'Approche de classification basée sur une mesure de dissimilarité entre les RTF'. Cette approche est destinée à être applicable à tous les types de signaux non-stationnaires, indépendamment de leur nature ou de leur origine, mais sans perte de généralité, nous considérerons les signaux EEG d'épilepsie comme application.

Pour ce faire, nous présenterons dans un premier temps le principe de l'approche de classification basée sur une mesure de dissimilarité entre les RTF. Ensuite, nous montrons que l'analyse d'un signal non-stationnaire passe nécessairement par une représentation conjointe en temps et en fréquence. Après avoir donné le cadre général, nous mettrons en évidence l'intérêt de l'élaboration d'une ATF. Cela nous conduit à étudier les RTF de la classe de Cohen. Il ressort de cette étude que ces descriptions tentent dans une certaine mesure de localiser au mieux les différentes composantes du signal dans le plan T-F. Puis, nous introduirons une liste non exhaustive de mesures de dissimilarités adéquatent pour la classification dans le plan T-F. Enfin, nous mettrons en évidence les avantages de l'approche proposée en l'appliquant à un problème de classification de signaux EEG d'épilepsie.

6.2 Approche de classification basée sur une mesure de dissimilarité entre les RTF

6.2.1 Classification dans le plan T-F

La RTF est un espace discriminant pour la classification de signaux non-stationnaires. Cette idée n'est pas nouvelle, et l'on peut distinguer trois grandes familles de méthodes : La première famille considère le plan T-F dans son ensemble et compare les RTF à l'aide d'une mesure de dissimilarité [176–179].

Pour introduire le rôle de décision utilisé dans cette famille des méthodes, on suppose disposer d'un signal $s(t)$ à classifier et d'une base de données d'apprentissage χ de N classes, tel que $\chi_i = \{s_1^i, ..., s_{n_i}^i\}$ représente les différentes classes $\omega_i, i = 1, ..., N$. Chaque classe est constitué de n_i signaux $s_k^i(t),\ k = 1, ..., n_i$.

Soient $\mathrm{RTF}_{s_k^i}(t,f),\ k=1,...,n_i$ et $\mathrm{RTF}_s(t,f)$ les RTF des signaux d'apprentissage et du signal à classifier, respectivement.

Ainsi, le signal s est assigné à la classe ω_i selon le rôle de décision suivant :

$$\omega_i = \arg \min_{i=1,...,N} d(\mathrm{RTF}_s, \mathrm{RTF}_{\omega_i}) \tag{6.1}$$

avec, $d(.,.)$ est une mesure de dissimilarité et RTF_{ω_i} est une RTF caractérisant la classe ω_i.

RTF_{ω_i} est définie dans la plupart de temps comme la moyenne des RTF des signaux d'apprentissage de la classe :

$$\mathrm{RTF}_{\omega_i}(t,f) = \frac{1}{n_i} \sum_{k=1}^{n_i} \mathrm{RTF}_{s_k^i}(t,f). \tag{6.2}$$

La deuxième famille tire parti des spécificités des RTF pour extraire des caractéristiques discriminantes, permettant de réduire ainsi la dimension de l'espace de décision [136–138, 140–146, 148, 153–157, 160, 168, 169, 180]. Les rôles de décisions de cette famille de méthodes sont construits dans l'espace de caractéristiques. En effet, tous les classifieurs de l'apprentissage automatique peuvent être utiliser.

La troisième famille fait appel aux méthodes à noyau pour la mise en œuvre d'un classifieur dans le plan T-F [181–183]. En effet, à chaque distribution T-F nous pouvons associer un noyau reproduisant et un espace de Hilbert à noyau reproduisant, ce qui permet de recourir aux deux clés de voûte que sont le coup du noyau et le théorème de représentation [184].

Dans tous les cas, ces études montrent l'importance du choix de la RTF, de la distance utilisée, ainsi que de la règle de décision [177–179, 185].

Dans ce travail, nous nous inscrivons plutôt dans la première famille de méthodes. L'idée est de classifier les signaux à l'aide de leurs RTF dans l'espace de dissimilarités. À notre connaissance, cette idée n'a jamais été utilisée dans la littérature.

6.2.2 Approche de classification basée sur une mesure de dissimilarité entre les RTF

Nous avons combinés l'espace de dissimilarités (Chapitre 5) et les RTF pour classifier les signaux non-stationnaires. Ainsi, Le processus de classification que nous avons proposé de mettre en œuvre est présenté sur la figure 6.2. Compte tenu de trois ensembles de signaux : l'ensemble de représentation R, l'ensemble d'apprentissage T et l'ensemble de test S, l'algorithme commence par les étapes d'apprentissage (première colonne de la figure) et le calcul des RTF des signaux d'apprentissage et de représentation.

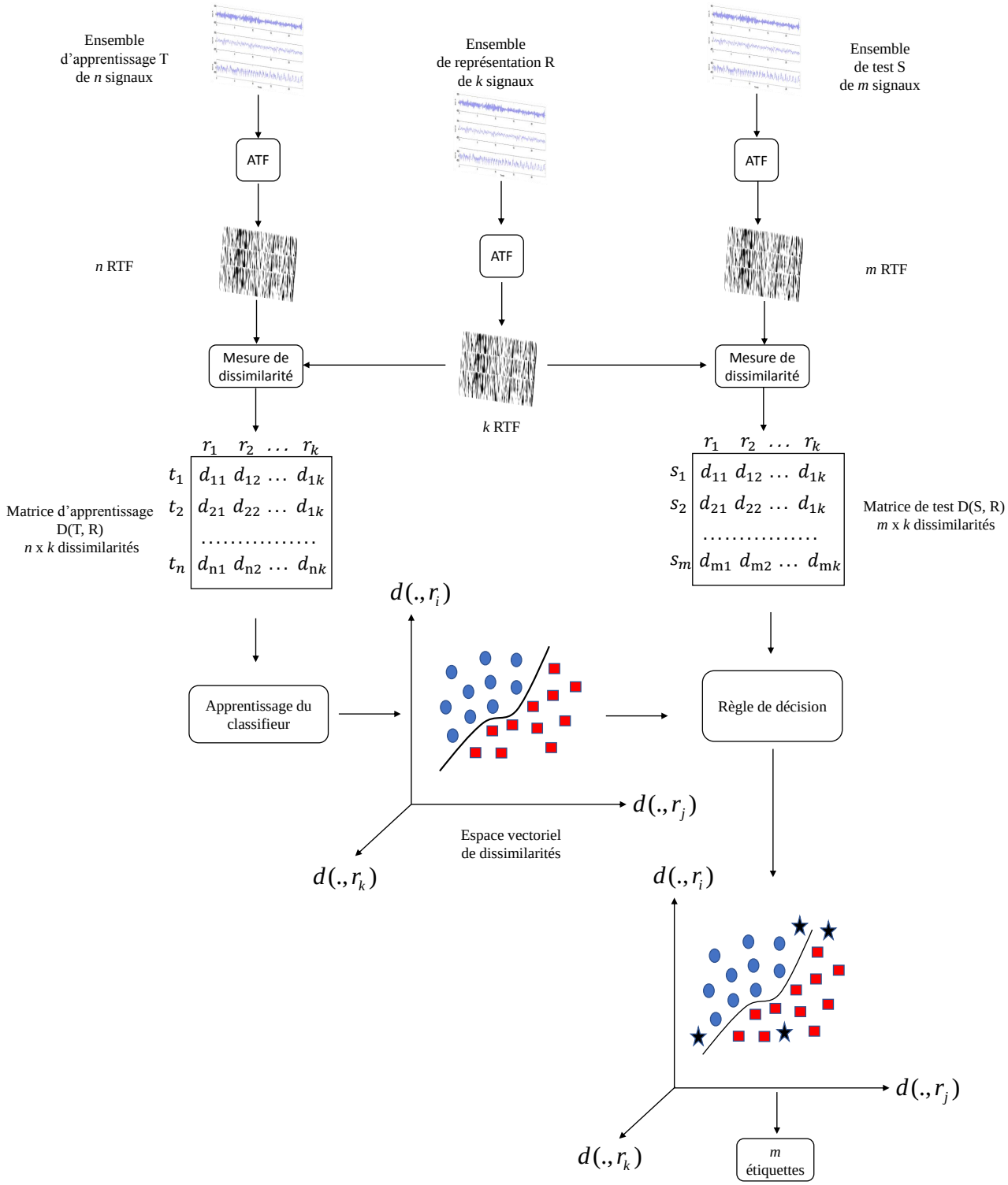


FIGURE 6.2 – Approche de classification basée sur une mesure de dissimilarité entre les RTF.

À partir de n and k signaux, une matrice d'apprentissage $D(T, R)$ de n x k dissimilarités est construite. Un signal donné est donc décrit par k dissimilarités, codant sa proximité aux autres signaux. Chacune des k colonnes de la matrice est ensuite injectée dans l'étape d'apprentissage du classifieur pour construire une fonction de décision. Après cette étape, le classifieur peut fonctionner pour tester un signal inconnu (seconde colonne de la figure). Les RTF des signaux de test sont calculés. Ces RTF sont comparés aux RTF de signaux de représentation en calculant la matrice de test $D(S, R)$ de m x k dissimilarités. Ces dissimilarités sont injectées dans le classifieur pour prendre la décision, produisant ainsi les étiquettes des signaux de test.

La réussite de cette approche dépend de deux concepts fondamentaux, à savoir le choix de la mesure de dissimilarité et de la RTF. Pour un « bon » choix de ces concepts, nous proposons une procédure dont le principe est exposé sur la figure 6.3. Premièrement, on retiendra la mesure de dissimilarité donnant les meilleurs résultats pour une RTF donnée. Deuxièmement, on déterminera la RTF donnant les meilleurs résultats pour la mesure de dissimilarité retenue.

En revanche, diverses méthode d'ATF ont été proposées. Ces méthodes comprennent les RTF linéaires telles que la TFCT, la TO et les RTF quadratique, par exemple, les distributions de la classe de Cohen et la classe affine basées sur la Distribution de Wigner-Ville (Wigner-Ville Distribution (WVD))... Ces méthodes ont pour objectif l'identification des structures d'un signal non-stationnaire.

Aussi, plusieurs mesures de dissimilarité utilisées en T-F peuvent être trouvées dans la littérature. Elles ont comme point commun de regrouper en une seule valeur des comparaisons effectuées point par point du domaine concerné. Parmi ces dissimilarités, on trouve les distances p-norme, la distance quadratique, les divergences (qui voit les RTF comme des densités de probabilité) et les déviations spectrales logarithmique...

Ainsi, nous nous restreindrons à une liste non exhaustive de mesures de dissimilarité et quelques DTF la classe de Cohen. Plus de détails sur l'ATF et les mesures de dissimilarité sont donnés dans les sections suivantes.

6.3 Analyse temps-fréquence

6.3.1 Préambule

La clé pour la description d'un signal non-stationnaire était de trouver une forme qui unit les variables temps et fréquence. Le domaine temporel indique comment l'amplitude du signal varie dans le temps et le domaine fréquentiel indique à quelle fréquence ces variation ont lieu. L'outil correspondant au temps et à la fréquence était la transformée de Fourier. Depuis son introduction, la transformée de Fourier est devenue l'un des outils d'analyse de signaux les plus utilisés dans de nombreuses disciplines de la science et de l'ingénierie.

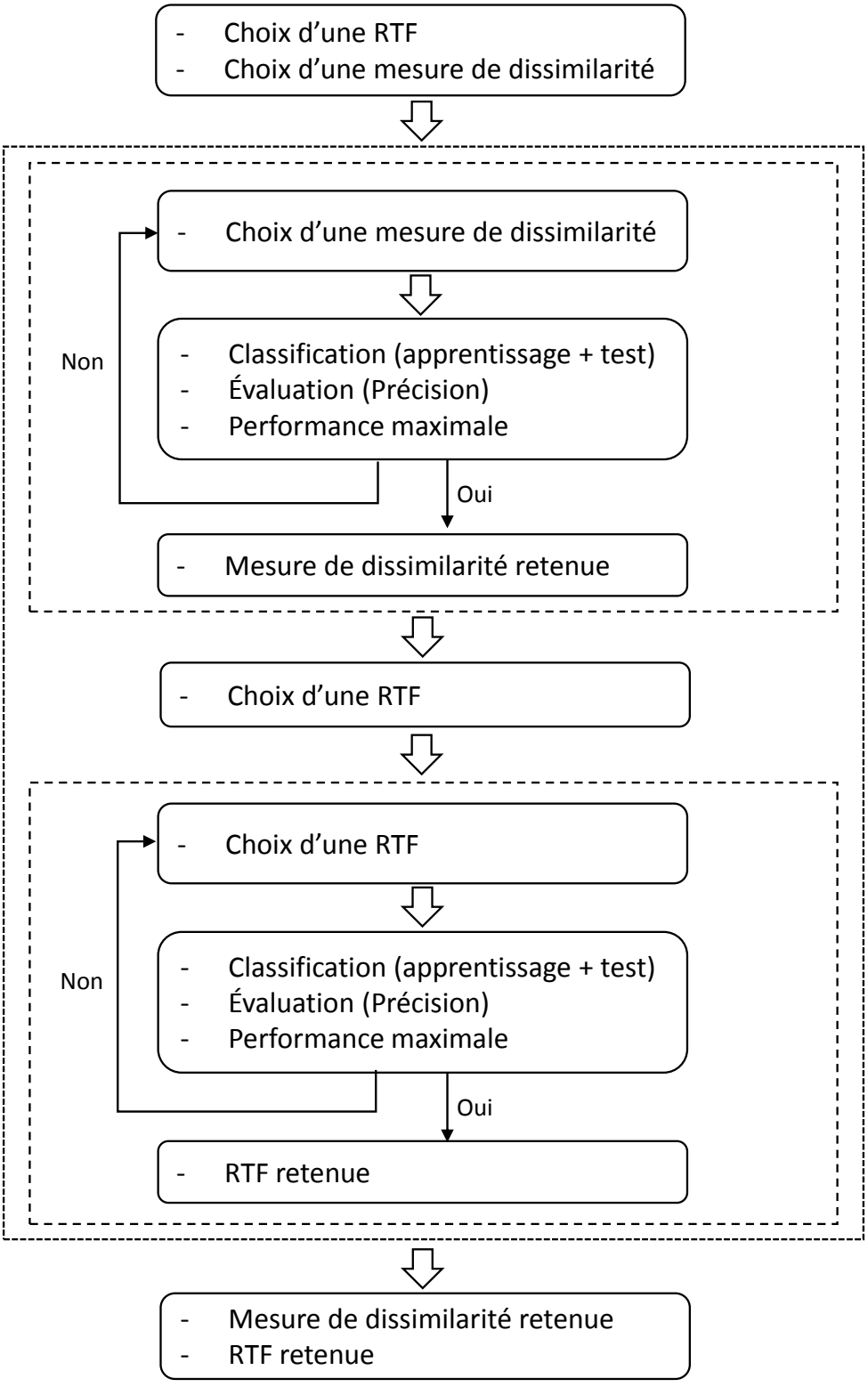


FIGURE 6.3 – Procédure pour choisir le couple (Mesure de dissimilarité, RTF) utiliser pour l'approche de classification basée sur une mesure de dissimilarité entre les RTF.

Quoique parfaite du point de vue mathématique, la transformée de Fourier fait perdre toute notion du temps. Quand on utilise la transformée de Fourier pour estimer le spectre fréquentiel d'un signal, on suppose que le contenu fréquentiel de ce signal est relativement stable durant l'intervalle temporel d'observation. Si le contenu fréquentiel d'un signal se change avec le temps (cas d'un signal non-stationnaire), il est impossible de surveiller comment cette variation avait lieu dans la fonction temporelle. En conséquence le spectre fréquentiel est utile quand on veut représenter le signal mais il n'est pas adéquat pour représenter les caractéristiques actuelles du signal.

Pour illustrer les limites de la transformée de Fourier, prenons l'exemple de deux signaux différents avec la même amplitude. Les deux signaux contiennent trois sinusoïdes de fréquences 10Hz, 20Hz et 30Hz. Ces signaux sont différents en ce sens que les intervalles d'apparitions des trois sinusoïdes sont différents (Les signaux en fonction du temps de la figure 6.4). Les fréquences des trois sinusoïdes sont retrouvées par la transformée de Fourier (Les deux figures en bas de la Figure 6.4). Cependant, cette information est incomplète dans la mesure où la transformée de Fourier ne nous renseigne pas sur la durée temporelle et les instants d'émission de chaque sinusoïde. En plus, nous constatons ici que des signaux différents peuvent avoir des transformées de Fourier identiques. Cet exemple montre que l'analyse fréquentiel masque l'information temporelle présente dans le signal. En conséquence, l'absence d'information temporelle rend la transformée de Fourier inapplicable dans le cadre d'un processus d'identification des signaux non-stationnaires.

Pour surmonter les limites de la transformées de Fourier, une représentation conjointe T-F combinant simultanément les deux aspects temporel et fréquentiel devrait permettre une meilleure compréhension des différents phénomènes présents dans un signal non-stationnaire. Ainsi, une RTF cartographie un signal uni-

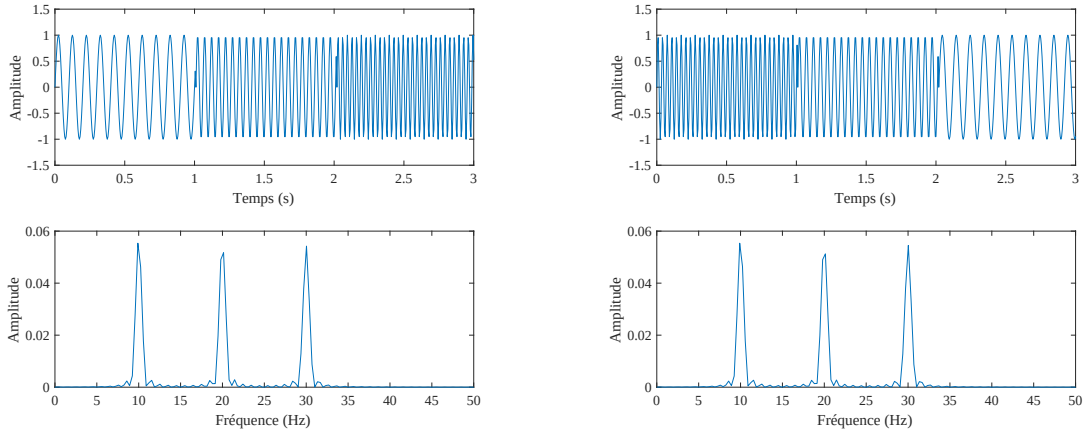


FIGURE 6.4 – deux signaux contiennent trois sinusoïdes et leurs transformées de Fourier.

dimensionnel du temps $s(t)$, en une fonction bidimensionnelle de temps et de fréquence $\mathrm{RTF}_{s(t)}(t, f)$ (figure 6.5). Contrairement à la TF, qui donne le contenu fréquentiel global d'un signal, les RTF décrivent l'évolution des propriétés fréquentielles du signal au cours du temps.

Les RTF peuvent être classées selon les approches en deux catégories. La première, appelée RTF linéaire, considère implicitement un signal non-stationnaire comme une succession de situations localement stationnaires. Historiquement la TFCT [186] et la TO [187, 188] sont des exemples typiques de cette catégorie. Les RTF de cette catégorie utilisent un atome d'analyse, appelée fenêtre d'analyse dans le cas de la TFCT et ondelette mère dans le cas de la TO. Cet atome sélectionne une portion du signal et se déplace par translation le long de l'axe temporel. À chaque déplacement on associe le spectre instantané de la portion de signal analysé par l'atome à la position temporelle de l'atome.

Cependant, la résolution temporelle et la résolution fréquentielle des RTF linéaires ne peuvent pas être bonnes simultanément (c'est le principe d'incertitude d'Heisenberg-Gabor). Ces résolutions sont déterminées par la localisation T-F des atomes utilisés. Dans le cas de la TFCT, la localisation parfaite d'un atome consisterait à pouvoir diminuer infiniment sa taille. Or, ce n'est pas possible à cause de la dualité T-F. Cette

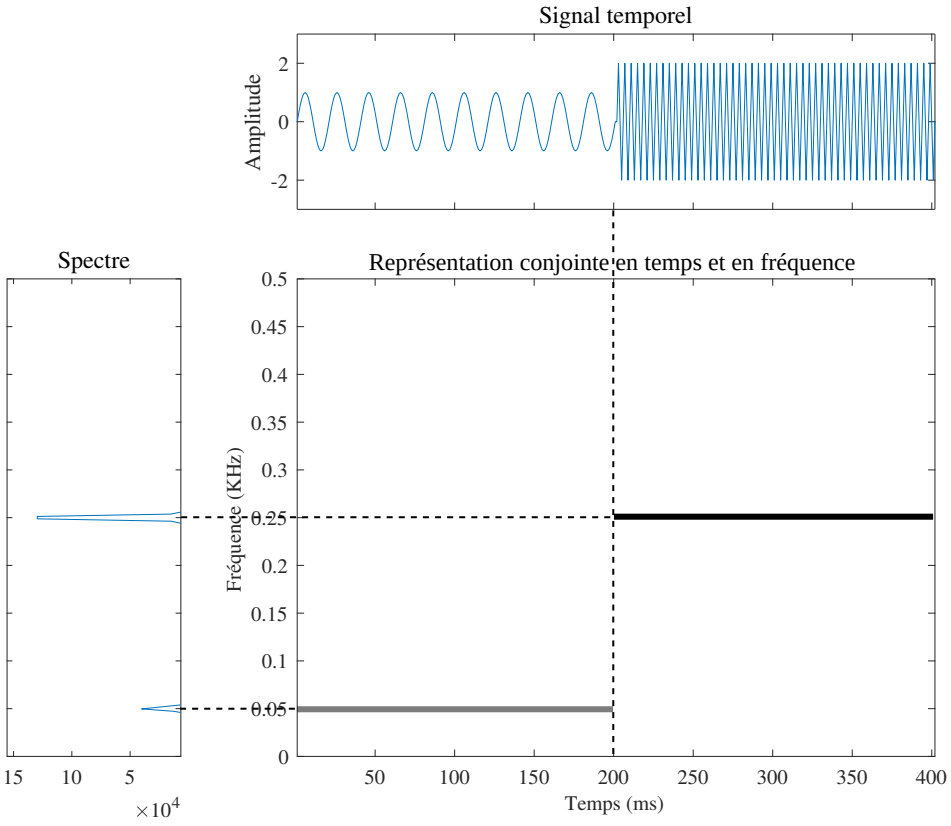


FIGURE 6.5 – RTF idéal d'un signal.

dernière s'exprime par le fait que toute diminution du support temporel provoque un élargissement du support fréquentiel du signal, et réciproquement. La dualité T-F repose sur le principe d'incertitude d'Heisenberg-Gabor [189] qui postule qu'il n'est pas possible de connaitre en précision la localisation d'un signal à la fois en temps et en fréquence.

Principe d'incertitude d'Heisenberg-Gabor :

Les supports du signal en temps et en fréquence n'en peuvent pas être changés indépendamment (figure 6.6), car ils sont reliés par l'inégalité suivante :

$$\Delta \mathrm{t}.\Delta\omega \geq \frac{1}{4\pi} \tag{6.3}$$

avec Δt et $\Delta\omega$ désignant les largeurs, mesurées par les variances, de l'atome d'analyse [188, 190].

Pour combler les lacunes des RTF linéaires, idéalement, on souhaiterait aboutir à une densité d'énergie dans le plan T-F sans perte de résolution. Les grandeurs physiques pertinentes sont de type énergétique, ce qui conduit à rechercher des décompositions de l'énergie du signal et non du signal lui-même. La densité spectrale de puissance (DSP) indique la distribution de l'énergie d'un signal dans le domaine fréquentiel. Cependant que la transformé de Fourier est une fonction linéaire, La DSP est une fonction quadratique. Alors, les RTF qui sont un dérivé de la DSP sont appelé RTF quadratique et représente la deuxième catégorie. L'objectif des RTF de cette catégorie est de déterminer les caractéristiques d'un signal en se basant sur sa distribution d'énergie dans le plan T-F [132, 191, 192].

On compte la distribution de Wigner parmi les RTF quadratique qui, avec sa résolution T-F optimale, sera étudiée par la suite. On s'intéressera ensuite à la classe de Cohen des distributions covariantes par translation dans le plan T-F. Mais, avant de poursuivre et d'étudier ces distribution T-F, on reviendra sur la notion de stationnarité et la définition de la DSP.

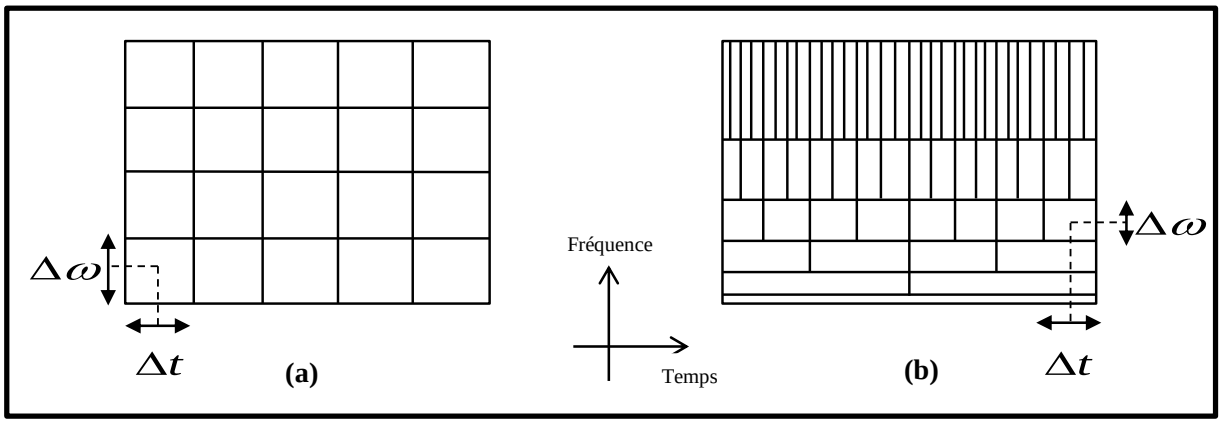


FIGURE 6.6 – Echantillonnage du plan T-F : (a) TFCT et (b) Ondelette.

6.3.2 Définition de la non-stationnarité

La nature nous montre dans nos expériences quotidiennes que les deux variable temps et fréquence, sont habituellement interconnectés dans les signaux, tels que le discours humain, la musique enregistrée ou l'émission acoustique. De tels signaux sont décrits comme non-stationnaire, parce que leur contenu fréquentiel varie avec le temps. Plus généralement, un signal dont les caractéristiques spectrales ne varient pas au cours du temps est dit stationnaire, a contrario, un signal dont les caractéristiques spectrales se modifient au cours du temps est dit non-stationnaire.

6.3.3 Définition de la DSP

La DSP d'un signal $s(t)$ est la grandeur $|S(f)|^2$, qui est le carrée de la transformé de Fourier de ce signal. La DSP peut être également exprimé par la transformée de Fourier de la fonction d'autocorrélation de s(t) :

$$|S(f)|^2 = \int R(\tau) e^{-2j\pi f \tau} d\tau \tag{6.4}$$

avec :

$$R(\tau) = \int s(t) s^*(t-\tau) dt. \tag{6.5}$$

6.3.4 Distribution de Wigner-Ville

La WVD est une RTF quadratique fondamentale. Cette distribution est un puissant outil pour l'analyse des signaux non-stationnaires, elle a été développé dans un première lieu dans la mécanique quantique par Wigner en 1932 [193], en suite elle a été introduite dans le traitement du signal par Ville [194]. La WVD associé à un signal $s(t)$ d'énergie finie est la fonction $\mathrm{WVD}(t, f)$ des deux variables, temps et fréquence :

$$\mathrm{WVD}_s(t, f) = \int s(t + \frac{\tau}{2}) s^*(t - \frac{\tau}{2}) e^{-2j\pi\tau f} d\tau. \tag{6.6}$$

La WVD vérifie un nombre important de propriétés mathématiques désirables [195]. Outre le fait qu'elle est à valeurs réelles, elle préserve l'énergie et le produit scalaire entre domaine temporel et T-F, respecte les marginales temporelles et fréquentielles, et est covariante par translation en temps et en fréquence [196, 197]. De plus, la WVD est un bon indicateur sur comment l'énergie du signal se distribue dans le plan T-F.

Cependant la WVD répond à un principe de superposition quadratique et non linéaire. Ainsi, la transformée de Wigner-Ville de deux signaux ne se réduit pas à la somme des distributions individuelles de ces signaux. Par exemple la WVD d'un signal $z(t) = x(t) + y(t)$ est définie par :

$$\mathrm{WVD}_z(t,f) = \mathrm{WVD}_x(t,f) + \mathrm{WVD}_y(t,f) + 2\mathrm{Re}(\mathrm{WVD}_{x,y}(t,f)). \tag{6.7}$$

Avec $\mathrm{WVD}_{x,y}(t,f)$ est la DWV croisée entre les signaux $x(t)$ et $y(t)$ définie par :

$$\mathrm{WVD}_{x,y}(t,f) = \int x(t+\frac{\tau}{2})y^*(t-\frac{\tau}{2})e^{-2j\pi f\tau}d\tau. \tag{6.8}$$

Bien que la WVD vérifie de nombreuses propriétés, sa lisibilité est généralement compromise par la présence de nombreux termes d'interférence. Ceux-ci introduisent de plus des composantes négatives qui ne permettent pas d'interpréter cette distribution comme une densité d'énergie. On illustre ce phénomène dans la figure 6.7 par la WVD d'un signal comportant deux sinusoïdes. On distingue les termes d'interférence à mi-distance entre les deux composantes du signal dans le plan T-F. Afin de remédier à ce problème de lisibilité, une solution consiste à appliquer un lissage dans le plan T-F afin de les atténuer. Ainsi la distribution de pseudo Wigner-Ville (Pseudo Wigner-Ville Distribution (PWVD)) effectue un lissage en fréquence tandis que la distribution de pseudo Wigner-Ville lissée (Smoothed Pseudo Wigner-Ville Distribution (SPWVD)) applique un lissage bidimensionnel séparable en temps et en fréquence.

Ainsi, diverses distributions faisant intervenir un lissage ont été proposées. Elles appartiennent à une vaste famille de distributions T-F : la classe de Cohen.

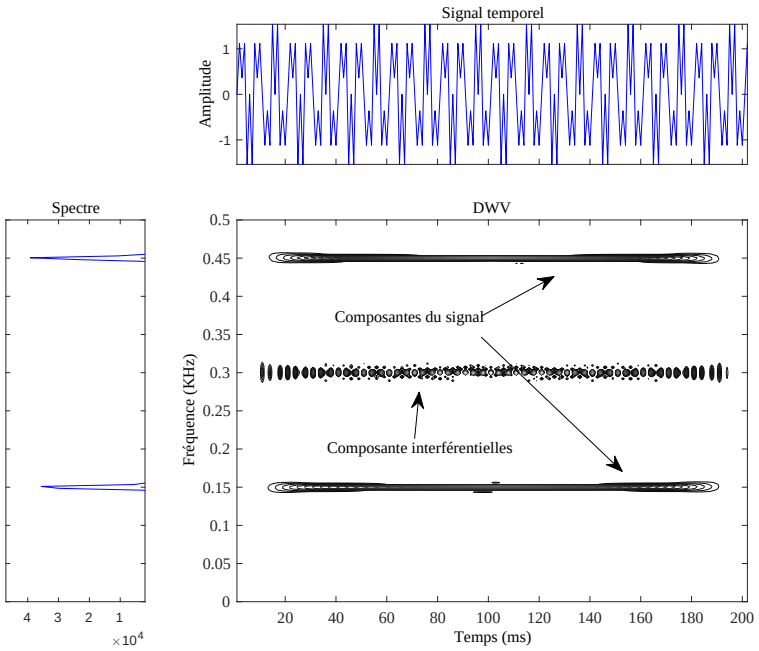


FIGURE 6.7 – WVD d'un signal avec deux composantes sinusoïdale. Illustration des termes d'interférences.

6.3.5 La classe de Cohen de RTF quadratique

En plus de la DWV, plusieurs autres RTF quadratique ont été proposées par plusieurs chercheurs pour l'analyse des signaux non-stationnaires. En 1966, Cohen [198] a trouvé une formule générale qui permet de rassembler toutes les RTF. Cette formule facilite l'écriture des nouvelles représentations. Ces représentations sont connues sous le nom de la classe de Cohen.

L'expression générique de la classe de Cohen peut s'écrire :

$$\mathrm{C}_s(t,f) = \int\int\int_{\mathbb{R}^3} s(u+\frac{\tau}{2})s^*(u-\frac{\tau}{2})\phi(v,\tau)e^{-2j\pi f\tau - j2\pi tv + j2\pi uv}\mathrm{d}u\mathrm{d}\tau\mathrm{d}v \tag{6.9}$$

où $s(t)$ est le signal à analyser, $s^*(t)$ est son conjugué complexe, et $\phi(v,\tau)$ est une fonction noyau (fonction de paramétrisation). Chaque DTF de la classe de Cohen a son propre noyau.

Ainsi, il existe de nombreuse représentation appartenant à la classe de Cohen comme par exemple la WVD (et ses extensions : PWVD et SPWVD), Rihaczeck, Margenau-Hill, ou encore Choi-Williams (Tableau 6.1) [132, 191, 192]

Toutefois, par leur bilinéarité, ces distributions engendrent des termes d'interférences pouvant nuire à la lisibilité des représentations obtenues. L'une des solutions consiste à appliquer un lissage dans le plan T-F afin de les atténuer. Une autre solution permettant d'améliorer la lisibilité des distributions de la classe de Cohen (et de la classe affine) est basée sur la méthode de la réallocation [191]. Le principe est de réarranger le plan T-F pour y améliorer la localisation des composantes d'un signal. Une dernière solution, consiste à

Tableau 6.1 – Quelques RTF de la classe de Cohen et leurs fonctions noyaux.

RTF	Noyau ($\phi(v,\tau)$)
WV	1
PWVD	$h(\tau)$
	$h(\tau)$: window function
SPWVD	$\widehat{g}(v)h(\tau)$
	$g(v), h(\tau)$: window functions
Rihaczek (RIH) [199]	$e^{j\pi v\tau}$
Margenau-Hill (MH) [200]	$cos(\pi v\tau)$
Pseudo Margenau-Hill (PMH) [201]	$h(\tau)e^{j\pi v\tau}$
	$h(\tau)$: window function
Choï-Williams (CW)	$e^{-\frac{1}{2}(\frac{\pi v\tau}{\sigma})^2}$
	σ : scaling factor
Born-Jordan (BJ)	$\frac{sin(\pi v\tau)}{\pi v\tau}$
Zhao-Atlas-Marks (ZAM)	$h(\tau)\frac{sin(\pi v\tau)}{\pi v\tau}$
	$h(\tau)$: window function
Generalized Rectangular (GRECT)	$\frac{sin(\frac{2\pi v\sigma}{\lvert\tau\rvert^\alpha})}{\pi v}$
	σ : scaling factor, α : dissemmetry ratio
Reduced Interference (RI)	$\int_{-\infty}^{+\infty} h(t)e^{-2j\pi v\tau t}\mathrm{d}t$
	$h(t)$: hamming window

utiliser les techniques de traitement d'image, telles que la reconstruction d'image morphologique et le filtrage non-linéaire... [202–204].

Pour illustrer certaines RTF, nous utilisons un signal non-stationnaire composé de la combinaison de quatre signaux chirp (un chirp est un signal avec une modulation fréquentielle linéaire). Ce signal, son spectre et les lois d'évolution théorique des fréquences au cours du temps sont illustrés dans la figure 6.8.

A partir de la représentation temporelle de ce signal, il est difficile de dire quelle sorte de modulation et quelles sont les fréquences contenues dans ce signal. Si nous considérons maintenant le spectre d'énergie de ce signal, nous ne pouvons toujours pas dire quoi que ce soit sur l'évolution dans le temps du contenu fréquentiel. En revanche, les quatre signaux sont identifiés sur la RTF idéalisée.

Quelques RTF de la même signal, obtenus par quelques distributions du tableau 6.1, sont présentées dans la figure 6.9.

La WVD permet de mettre en évidence les interférences entre les quatre composantes du signal (Figure 6.9(a)). L'une des solutions pour réduire ces interférences est d'utiliser la PSWVD qui lisse les interférences (en temps et en fréquence).

Les quatre dernières figures montrent le comportement de quatre autres distributions de la classe de Cohen. On peut observer que selon les caractéristiques du noyau de la RTF les interférences ne sont pas traitées de la même manière.

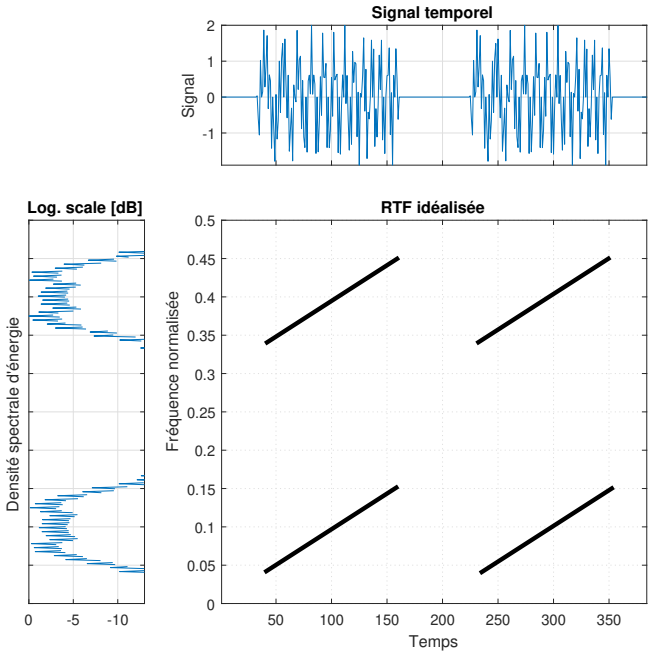


FIGURE 6.8 – Représentations (temps, fréquence, RTF idéalisée) d'un signal composé de quatre chrip

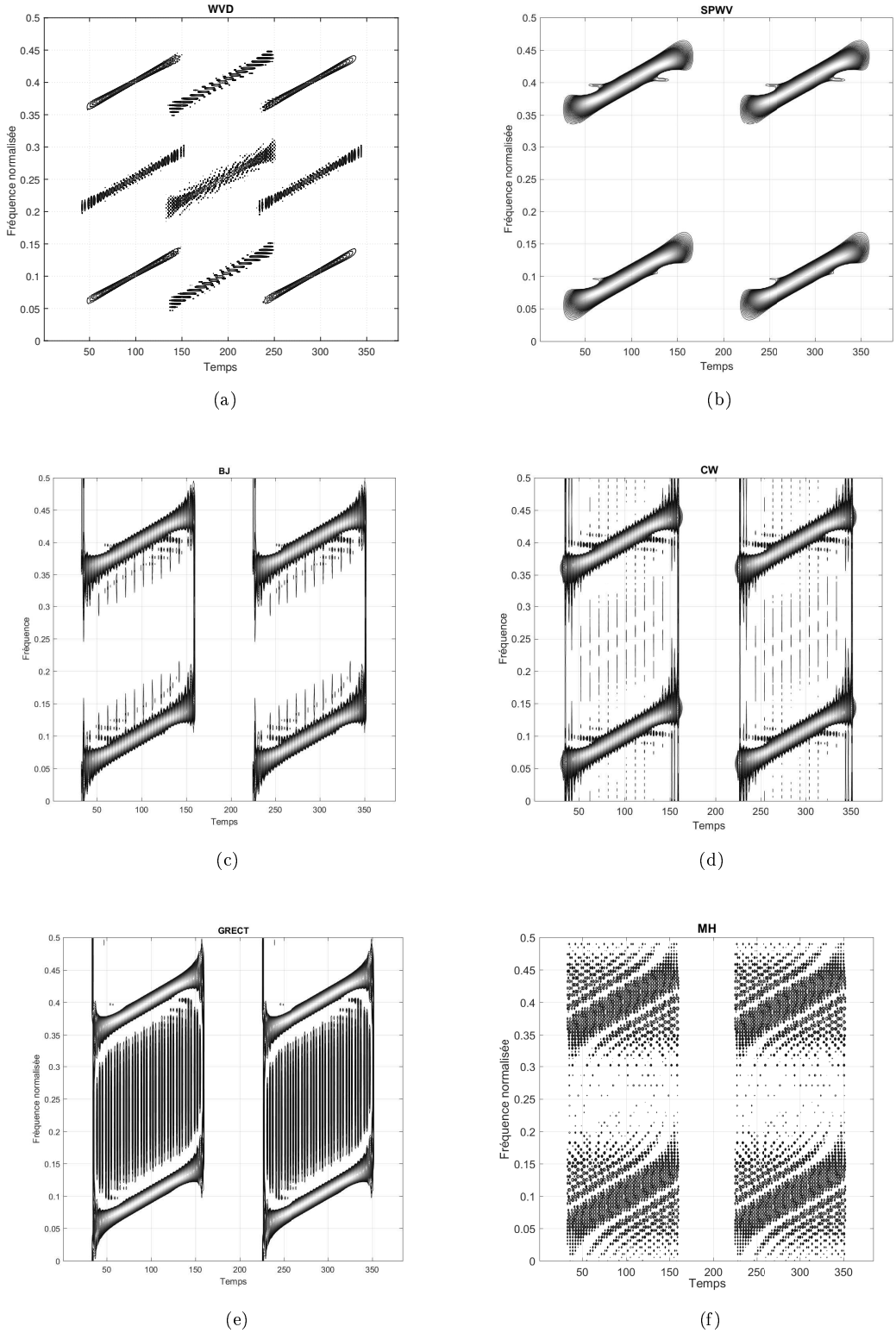


FIGURE 6.9 – Exemples de RTF quadratiques de la classe de Cohen du signal à quatre chirp

6.3.6 Synthèse

Les représentations d'un signal conjointement en temps et en fréquence offrent un réel intérêt : elles permettent de mettre en évidence les non-stationnarités qui caractérisent le signal, c'est à dire l'analyse des lois de comportement fréquentiel du signal au cours du temps. Ainsi, une RTF permet la mise en correspondance d'une fonction monovariable (en t) avec une fonction a deux variables : le temps t et la fréquence f. Contrairement à la transformé de Fourier, qui donne le contenu fréquentiel global d'un signal, les RTF situent l'énergie dans le plan T-F.

De manière générale, les RTF, quoique différentes reposent sur un concept identique à savoir la réécriture du noyau de lissage selon des critères prédéfinis à l'avance. Ces critères sont souvent choisis selon les caractéristiques connues des signaux à analyser ou de la problématique envisagée (détection, classification ...).

Ainsi, il n'existe pas de RTF universelle et unique pour analyser les signaux non-stationnaires. Aucune RTF n'est ni meilleure ni moins bonne. En fait, tout dépend des signaux à analyser et la problématique envisagée.

Dans cette partie de cette étude doctorale, nous intéresserons aux RTF de la classe de Cohen. Ces RTF vont être combinées avec l'approche de classification basée sur une mesure de dissimilarité.

6.4 Mesure de dissimilarité entre les RTF

La dissimilarité est une outil utile pour mesurer le taux de ressemblance de deux objets quelconques. Il est souvent défini comme une fonction distance d.

Ainsi l'approche de classification basée sur une mesure de dissimilarité entre les RTF fait appel à des distances. Nous présentons ici une liste non exhaustive de distances utilisées en T-F, mais valables aussi dans les domaines temps, fréquence, ... Elles ont comme point commun de regrouper en une seule valeur des comparaisons effectué point par point du domaine concerné.

Pour introduire la définition de chaque distance, prenons deux signaux (S_1, S_2) et leurs RTF (RTF_{S_1}, RTF_{S_2}).

a) p-norme

La distance p-norme (notée par d_p) est définie par :

$$d_p(\mathrm{RTF}_{s_1}, \mathrm{RTF}_{s_2}) = \left[\int \int |\mathrm{RTF}_{s_1}(t,f) - \mathrm{RTF}_{s_2}(t,f)|^p \, \mathrm{d}t\mathrm{d}f \right]^{\frac{1}{p}}. \tag{6.10}$$

Nous notons que, pour des valeurs données de p, il existe des cas particuliers de distances :

— $p = 1$: Distance de Manhattan.

— $p = 2$: Distance euclidienne.

— $p = \infty$: Distance de Chebyshev.

La distance p-norme est rarement utilisée pour les valeurs de p autres que 1, 2 et l'infini.

b) Distance quadratique

La distance quadratique est déduite de la distance p-norme (d_p) en faisant d_p^2.

c) La corrélation

La distance de corrélation (notée par d_{corr}) est définie par :

$$d_{corr}(\mathrm{RTF}_{s_1}, \mathrm{RTF}_{s_2}) = \frac{[d_2(\mathrm{RTF}_{s_1}, \mathrm{RTF}_{s_2})]^2}{\int\int |\mathrm{RTF}_{s_1}(t,f)|^2 \mathrm{d}t\mathrm{d}f + \int\int |\mathrm{RTF}_{s_2}(t,f)|^2 \mathrm{d}t\mathrm{d}f}. \tag{6.11}$$

L'une des faiblesses des fonctions distance p-norme est que si l'un des attributs (descripteur, pixel,...) d'entrée d'un objet a une portée relativement grande, alors il peut dominer les autres attributs. Par conséquent, les RTF sont souvent normalisées par :

$$\mathrm{RTFN}(t,f) = \frac{|\mathrm{RTF}(t,f)|}{\int\int |\mathrm{RTF}(t,f)| \mathrm{d}t\mathrm{d}f}. \tag{6.12}$$

d) p-norme normalisée

La distance p-norme entre deux RTF normalisés donne une nouvelle famille de distance (noté d_{pn}) :

$$d_{pn}(\mathrm{RTF}_{s_1}, \mathrm{RTF}_{s_2}) = d_p(\mathrm{RTFN}_{S_1}, \mathrm{RTFN}_{S_2}). \tag{6.13}$$

Nous notons également que, pour un cas particulier de $p = 1$ dans la distance d_{pn}, la distance obtenue est la distance de Kolmogorov.

e) f-divergence

La f-divergence [205] permet d'avoir une famille de distances paramétrée par deux fonctions. La fonction f est une fonction convexe continue sur $\mathbb{R}^+$ et g une fonction croissante sur $\mathbb{R}$. La f-divergence entre deux densités de probabilité $p_1(x)$ et $p_2(x)$ est définie par :

$$d_{f-div}(p_1, p_2) = g\left\{\mathrm{E}_1\left[f\left(\frac{p_2(x)}{p_1(x)}\right)\right]\right\}. \tag{6.14}$$

Où E_1 est l'opérateur d'attente par rapport à p_1.

L'ATF quadratique présente la distribution d'énergie d'un signal dans le domaine T-F, donc elle peut être considérée comme une densité de probabilité. La RTF normalisée est positive, donc la définition de la f-divergence peut être appliquée comme :

$$d_{f-div}(\mathrm{RTFN}_{S_1}, \mathrm{RTFN}_{S_2}) = g\Big[\int\int f\left(\frac{\mathrm{RTFN}_{S_2}(x)}{\mathrm{RTFN}_{S_1}(x)}\right)\mathrm{RTFN}_{S_1}\mathrm{d}t\mathrm{d}f\Big]. \tag{6.15}$$

Le tableau 6.2 présente une liste de quelques f-divergences couramment utilisées dans le traitement du signal pour la classification des RTF [206].

La divergence de Kullback-Leibler n'est pas symétrique. Cette mesure peut être adaptée aux RTF comme une distance comme suit :

$$d(\mathrm{RTF}_{S_1}, \mathrm{RTF}_{S_2}) = d_{f-div}(\mathrm{RTFN}_{S_1}, \mathrm{RTFN}_{S_2}) + d_{f-div}(\mathrm{RTFN}_{S_2}, \mathrm{RTFN}_{S_1}). \tag{6.16}$$

f) Déviation spectral logarithmique

La déviation spectral logarithmique [207] (noté d_{DSL_p}) est défini par la p-norme entre les logarithmes de deux RTF normalisées. Le logarithme est utilisé ici pour réduire l'influence des valeurs de crêtes élevées des RTFs.

$$d_{\mathrm{DSL}_p}(\mathrm{RTF}_{s_1}, \mathrm{RTF}_{s_2}) = d_p(\log(|\mathrm{RTFN}_{s_1}|), \log(|\mathrm{RTFN}_{s_2}|). \tag{6.17}$$

6.5 Application de l'approche sur une base de données de signaux EEG d'épilepsie

6.5.1 Préambule

L'Electroencéphalographie (EEG) est un outil clinique important utilisé pour l'étude et la mise en évidence de l'activité électrique du cerveau. Elle demeure incontournable pour étudier la fonction cérébrale humaine et les troubles neurologiques [208]. L'épilepsie est l'un des troubles les plus communs du système nerveux et touche environ 1% de la population humaine [209]. La détection des crises d'épilepsie dans les signaux EEG est un moyen efficace pour diagnostiquer l'épilepsie [210, 211]. En raison des interconnexions complexes entre des milliards de neurones du cerveau, les signaux EEG enregistrés sont complexes, non linéaires, non

Tableau 6.2 – f-divergences couramment utilisées dans le traitement du signal pour la classification des RTFs.

Divergence	$f(x)$	$g(x)$	$d_{f-div}(\mathrm{RTFN}_{S_1}, \mathrm{RTFN}_{S_2})$
Kolmogorov	x	$\lvert 1-x\rvert$	$\int\int \lvert\mathrm{RTFN}_{S_1}(t,f) - \mathrm{RTFN}_{S_2}(t,f)\rvert \mathrm{d}t\mathrm{d}f$
Matusita	$x^{1/2}$	$\lvert 1-x^{1/2}\rvert^2$	$[\int\int \lvert(\mathrm{RTFN}_{S_1}(t,f))^{1/2} - (\mathrm{RTFN}_{S_2}(t,f))^{1/2}\rvert^2 \mathrm{d}t\mathrm{d}f]^{1/2}$
Kullback	x	$(x-1)log(x)$	$\int\int \lvert\mathrm{RTFN}_{S_1}(t,f) - \mathrm{RTFN}_{S_2}(t,f)\rvert \log\frac{\mathrm{RTFN}_{S_1}(t,f)}{\mathrm{RTFN}_{S_2}(t,f)} \mathrm{d}t\mathrm{d}f$
Kullback-Leibler	x	$-log(x)$	$\int\int \mathrm{RTFN}_{S_1}(t,f) \log\frac{\mathrm{RTFN}_{S_1}(t,f)}{\mathrm{RTFN}_{S_2}(t,f)} \mathrm{d}t\mathrm{d}f$

stationnaires et de nature aléatoire.

De nos jours, de nombreuses techniques d'analyse et de traitement des signaux ont été étudiées pour l'identification de l'épilepsie. D'une manière générale, le processus de ces méthodes peut être classé en trois étapes, à savoir le prétraitement, l'extraction de descripteurs et la classification. Dans l'étape de prétraitement, la détection du bruit qui perturbent la lecture, l'interprétation et l'analyse des enregistrements d'EEG), suivi de sons élimination avec une perte minimale d'information est largement conduit. L'extraction de caractéristiques vise à capturer les caractéristiques distinctives significatives cachées dans les signaux EEG, ce qui domine directement l'exactitude de la classification finale. À notre connaissance, la plupart des études dans la littérature [135, 147, 210–228] sur la détection des crises d'épilepsie sont basées sur l'approche de l'extraction de caractéristiques et aucune étude n'utilise l'approche basée sur la dissimilarité.

Plusieurs méthodes ont été appliquées aux signaux EEG épileptiques pour l'extraction de descripteurs [135, 147, 210–228], telles que les représentations basées sur la transformé de Fourier, la Transformée en Ondelettes (TO) et les RTF. Les méthodes basées sur la TF [214] reposent sur des observations antérieures selon lesquelles les crises d'épilepsie entraînent des changements dans certaines bandes de fréquences telles que $\delta(0.4-4Hz)$, $\theta(4-8Hz)$, $\alpha(8-12Hz)$. Étant donné que la nature des signaux EEG épileptiques est non stationnaire et multi-composée, cette méthode ne convient pas à la décomposition en fréquence de ces signaux. Les méthodes basées sur les RTFs linéaire et quadratique ont montré des performances supérieures par rapport aux méthodes classiques d'analyse fréquentielle. De la diversité des approches disponibles, la TO est le plus largement utilisé pour l'extraction de caractéristiques des signaux EEG épileptique [210, 216, 221]. Ceci peut être expliqué par ses excellentes capacités en matière de représentation multi-résolution. Malgré les diverses résolutions qu'elle autorise, cet outil ne permet pas d'adapter à tous les signaux en particulier des signaux non stationnaires et non linéaires tels que des signaux EEG. Parce que la résolution temporelle et fréquentielle de la TO est déterminée par la localisation T-F de l'atome d'analyse. Les RTFs quadratiques ont été exploitées précédemment dans certains travaux [135, 213, 222] pour analyser les crises d'épilepsie. En fait, ces représentations ont l'avantage de caractériser les signaux par leurs distributions d'énergie dans le plan T-F. Comme le montre [171], les RTFs quadratiques fournissent un cadre puissant pour l'extraction de descripteurs et la classification de phénomènes non-stationnaires dans les signaux.

La classification est une autre procédure essentielle pour l'analyse de l'épilepsie. Les classifieurs tels que les SVM [215, 216, 219, 220, 223, 228, 229], le k-PPV [211], l'Analyse Discriminante Quadratique (ADQ) [218], le RNA [135, 147, 212, 224–227], sont largement utilisés dans la classification de l'épilepsie. Cependant, ces méthodes sont toutes appliquées avec l'approche de l'extraction de descripteurs.

L'objectif de cette partie est de mettre en évidence les avantages de l'approche de classification basée sur une mesure de dissimilarité entre les RTF à travers une application sur une base de données de signaux EEG

d'épilepsie. Pour ce faire, une description de la base de données des signaux EEG épileptiques accessible au public est réalisée. Cette base de données consiste en cinq ensembles différents de signaux EEG. De ces cinq ensembles, 13 problèmes de classification différents, y compris 2, 3 et 5 classes, sont définis. Afin d'évaluer les performances de l'approche basée sur la dissimilarité et les RTFs, différentes mesures statistiques ont été utilisées. Ces mesures incluent la précision, la sensibilité, la spécificité et l'aire sous la courbe ROC. La technique de validation croisée *10-flod* est également utilisée. Différentes mesures de dissimilarité et RTFs ont été comparées pour sélectionner les plus pertinentes pour les signaux EEG épileptiques. Quatre classifieurs, tels que RNA, SVM, ADL et k-PPV, ont été combinés avec l'approche proposée. La complexité en terme de temps de calcul et la robustesse de l'approche proposée sont aussi explorées. Les résultats obtenus avec l'approche proposée sous divers problèmes de classification sont comparés aux résultats rapportés dans la littérature avec la même base de données de signaux EEG d'épilepsie. Leur comparaison nous a permet de valider l'efficacité de cette approche pour la détection de l'épilepsie.

6.5.2 Base de données des signaux EEG d'épilepsie

Les signaux EEG utilisés dans cette étude sont la base de données accessible au public fournie par Andrzejak et al. [5]. Cette base de données consiste en cinq ensembles de signaux EEG différents. Les signaux du premier ensemble (notés A) ont été récupérés à partir d'enregistrements EEG de surface effectués sur cinq volontaires sains avec les yeux ouverts. Les signaux du second ensemble (noté B) ont été collectés à partir des enregistrements EEG de surface des mêmes cinq volontaires sains avec les yeux fermés. Les signaux du troisième ensemble (noté C) et du quatrième ensemble (noté D) ont été obtenus pendant des intervalles sans crises (périodes interictales) de cinq patients épileptiques. Les signaux de l'ensemble C ont été enregistrés à partir de la formation de l'hippocampe des hémisphères opposés du cerveau. Les signaux dans l'ensemble D ont été enregistrés dans la zone épileptogène. Les signaux du dernier ensemble (noté E) ont été enregistrés pendant l'activité de crise (périodes ictales). Chaque jeu consiste en 100 segments EEG monocanal d'une durée de 23,6 s, qui ont été numérisés à une fréquence d'échantillonnage de 173,61 Hz et une résolution A/N de 12 bits. La Figure 6.10 montre un exemple d'un signal EEG provenant de chaque ensemble.

L'étape de pré-traitement (réduction du bruit, alignement du signal, etc.) des signaux EEG de cette base de données n'est pas étudiée dans cet étude. Dans cette étude, nous suivons la même procédure de validation dans [135, 147, 210–228], où seuls les signaux bruts sont traités, sans pré-traitement.

Cette base de données est utilisée pour évaluer les performances de notre approche. 13 problèmes de classification, incluant 2, 3 et 5 classes sont considérés. La difficulté de chaque problème dépend du nombre de classe qui contient. Ainsi, le problème à 5 classes est considéré comme le problème le plus difficile. Les

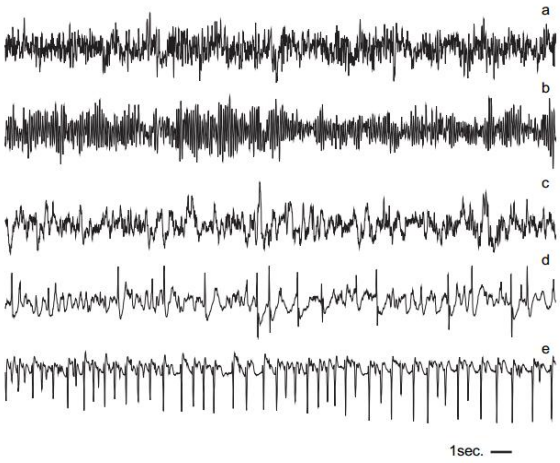


FIGURE 6.10 – Exemples de signaux EEG de la base de données épileptique [5].

instances de classification de chaque problème sont présentées dans le tableau 6.3.

6.5.3 Résultats

Dans la partie 6.2.2, nous avons vu que l'approche de classification basée sur une mesure de dissimilarité entre les RTF nécessite le choix du couple (mesure de dissimilarité, RTF). Ainsi, nous proposons une procédure pour un choix optimal. Cette procédure consiste en deux étapes :

— 1ère étape : Choisir une mesure de dissimilarité donnant les meilleurs résultats pour une RTF donnée.

— 2ème étape : Choisir une RTF donnant les meilleurs résultats pour la mesure de dissimilarité retenue dans la 1ère étape.

Tableau 6.3 – Les instances des trois problèmes de classification des signaux EEG.

Nombre de class	Instance de classification	classe 1	classe 2	classe 3	classe 4	classe 5
2	A, E	Normale	ictale	-	-	-
	B, E			-	-	-
	AB, E			-	-	-
2	C, E	interictale	ictale	-	-	-
	D, E			-	-	-
	CD, E			-	-	-
2	ACD, E	Normale + interictale	ictale	-	-	-
	BCD, E			-	-	-
	ABCD, E			-	-	-
3	A, C, E	Normale	interictale	ictale	-	-
	A, D, E				-	-
	AB, CD, E				-	-
5	A, B, C, D, E	Normale (yeux ouverts)	Normale (yeux fermés)	interictal (épileptogène)	interictale (hippocampe)	ictale

Ainsi, la première section de cette partie est consacrée au choix du couple (mesure de dissimilarité, RTF).

Nous avons introduit dans le chapitre 5 les différentes approches pour la classification par une mesure de dissimilarité. Ceux qui nous intéressent le plus sont la règle de k-PPV et l'espace vectoriel de dissimilarités. Parce que la règle de k-PPV est la technique la plus simple et facile à utiliser. Et le fait de postuler l'espace de dissimilarités comme un espace vectoriel offre la possibilité d'utiliser n'importe quelles mesures de dissimilarité et d'appliquer n'importe quels classifieurs. De ce fait, le but de la deuxième section est d'évaluer l'approche proposée en utilisant différents classifieurs.

En revanche, dans le cas d'un problème de classification avec une grande base de données, le calcul de la matrice d'apprentissage peut être une opération exigeant du temps et de la mémoire. Aussi, la grande dimension de l'espace vectoriel de dissimilarités peut augmenter la complexité en termes du temps de calcul des classifieurs. Ce paramètre peut limiter l'applicabilité de l'approche proposée à des problèmes de classification de petite base de données. Donc, la complexité en termes du temps de calcule de l'approche proposée fera l'objet de la troisième section.

Le bruit est un deuxième paramètre qui pourrait limiter l'applicabilité de l'approche proposée. Ainsi, l'étude de la robustesse aux bruits de cette approche est l'objectif principal de la quatrième section de cette partie.

Nous avons vu également que l'approche basée sur la dissimilarité nécessite l'utilisation d'une base de données d'apprentissage T et celle de représentation R. Cette dernière peut être un sous ensemble de T, ou T lui-même ou être entièrement diffèrent de T. Ainsi, tout au longue de nos expérimentation, nous considérons que T et R sont les mêmes ($T = R$).

a) Choix d'une mesure de dissimilarité et d'une RTF

Pour choisir une mesure de dissimilarité et une RTF appropriées à l'approche proposée, le classifieur k-PPV, le problème de classification à 5 classes, la précision et la validation croisée $10 - fold$ sont utilisés.

a.1) Choix d'une mesure de dissimilarité

Pour choisir une mesure de dissimilarité, la distribution SPWVD avec une fenêtre de lissage gaussienne est utilisée. Dix mesures de dissimilarité sont comparées : Manhattan, Euclidienne, Chebyshev, corrélation, Kolmogorov, Kulback, Matusita, Kulback-Leiber, d_{DSL_1} et d_{DSL_2}. Les résultats obtenus sont présentés dans le tableau 6.4.

Les précisions les plus élevées du classifieur k-PPV sont obtenues par la corrélation et les dissimilarités basées sur les RTF normalisées. Cependant, la meilleure précision du classifieur est obtenue par la déviation spectrale logarithmique d_{DSL_1}. Pour cela, la distance d_{DSL_1} est retenue pour choisir une RTF.

a.2) Choix d'une RTF

Tableau 6.4 – Précisions obtenues par le classifieur k-PPV en utilisant différentes mesures de dissimilarité. Évaluation de la performance en utilisant le problème de classification à 5 classes et la validation croisée *10-fold*.

Distance :	Manhattan	Euclidean	Chebyshev	Correlation	Kolmogorov
Précision (%) :	20	22.40	36.40	90.20	88.00
Distance :	Kulback	Matusita	Kullback-Leiber	d_{DSL_1}	d_{DSL_2}
Précision (%) :	90.20	89.60	90.20	94.40	93.60

Pour choisir une RTF approprié à l'approche proposée, la distance d_{DSL_1} est utilisée.

Les 11 RTFs présentées dans la section 6.3.5 sont divisés en trois catégories. La première comprend les RTF qui n'utilisent pas de fenêtres de lissage (MH, RIH, WV). La seconde inclut les ceux qui utilisent des fenêtres de lissage de fréquence (PMH, PWVD). La dernière englobe les RTF qui exploitent les fenêtres de lissage de fréquence et de temps (SPWVD, CW, GRECT, RI, ZAM, BJ). Pour les catégories deux et trois, une fenêtre de lissage gaussien est utilisée. Le classifieur k-PPV a été retourné pour chaque RTF et les précisions obtenues ont été notées. Les résultats obtenus sont présentés dans le tableau 6.5.

Pour la première catégorie de RTF, la plus grande précision (86,60%) est obtenue par la WVD. Pour la seconde catégorie, la plus grande précision (89,60%) est obtenue par PWVD. Enfin, pour la troisième catégorie de RTF, la grande précision (94,40%) est obtenue par SPWV. On remarque que les RTF qui utilisent à la fois une fenêtre de lissage fréquentielle et temporelle présentent d'excellents résultats de classification en termes de précision. Cela signifie que l'élimination des termes d'interférences améliore significativement la qualité de la classification. Dans toutes les catégories de RTF, les meilleures précisions ont été obtenues par la WVD et ses extensions (PWVD, SPWVD). Cependant, la meilleure précision dans toutes les catégories est obtenue par la distribution de SPWVD (94,40%).

Pour éliminer les termes d'interférences qui perturbent la lecture et l'interprétation des RTF, de nombreuses fenêtres de lissage de temps et de fréquence peuvent être utilisées. Pour choisir la meilleure, nous avons comparé cinq d'entre eux (Gaussian, Hamming, Hanning, Rectangulaire, Parzen et Kaiser) avec une longueur de la fenêtre de 128 points. La distribution SPWVD, le classifieur k-PPV, la précision, le problème de classification à 5 classes et la validation croisée *10-fold* sont utilisés. Les résultats obtenus sont présentés dans le tableau 6.6. Selon ces résultats, la meilleure précision (95,40%) est obtenue par les fenêtres de lissage Hanning et Parzen. Pour cette raison, la SPWVD avec une fenêtre de lissage Hanning est retenue pour évaluer

Tableau 6.5 – Précisions obtenues par le classifieur k-PPV en utilisant les trois catégories de RTF. La distance d_{DSL_1} est utilisée comme mesure de dissimilarité. Évaluation de la performance en utilisant le problème de classification à 5 classes et la validation croisée *10-fold*.

Categorie	1			2		3					
Distribution	WV	RIH	MH	PWVD	PMH	SPWVD	CW	GRECT	RI	ZAM	BJ
Précision (%)	86.60	66.80	65.80	89.60	74.20	94.40	86.60	86.40	85.00	83.80	88.00

Tableau 6.6 – Précisions obtenues par le k-PPV basée sur les mesures de dissimilarités entre les RTFs comme distances. La distribution SPWVD avec différentes fenêtres de lissage est utilisée. La distance d_{DSL_1} est exploitée comme mesure de dissimilarité. Évaluation des performances en utilisant le problème de classification à 5 classes (instances : A, B, C, D, E) et la validation croisée *10-flod*.

Fenêtre de lissage :	Gaussian	Hamming	Hanning	Rectangular	Parzen	Kaiser
Précision (%) :	94.20	88.80	95.40	79.40	95.40	95.00

l'approche proposée dans les sections suivantes.

a.3) Synthèse. Le couple (Distance d_{DSL_1}, SPWVD avec noyau Hanning) permet au classifieur k-PPV d'obtenir une précision maximale de 95.40%. Ainsi, ce couple sera utilisé pour l'approche proposée dans les sections suivantes de ce chapitre.

b) Évaluation des classifieurs

Le but de cette section est d'évaluer l'approche proposée en utilisant différents classifieur. Premièrement, les performances de ces classifieurs sont comparés en utilisant le problème de classification à 5 classes afin de choisir les meilleurs. Deuxièmement, les performances des classifieurs choisis sont évaluées sous les problèmes de classification à 2 et 3 classes. La SPWVD avec une fenêtre de lissage Hanning et le mesure de dissimilarité d_{DSL_1} sont utilisées. Les performances en termes de classification sont vérifiées en utilisant la validation croisée *10-fold*, l'AUC, la spécificité, la sensibilité et la précision.

Nous avons comparé quatre classifieurs. Les classifieurs : RNA, le SVM et l'ADL implémentés dans l'espace de dissimilarité et le classifieur k-PPV. Le nombre de voisins considéré pour le k-PPV, l'architecture du RNA, les paramètres du SVM et de l'ADL sont dérivés par essais et erreurs. La complexité du réseau de neurone est caractérisée par le nombre de couches cachées. Et il n'y a pas de règle générale pour la sélection du nombre approprié de couches cachées. Un noyau linéaire est utilisé pour le classifieur SVM.

Nous avons implémenté tous les classifieurs avec l'approche proposée en utilisant le logiciel MATLAB avec deux boîtes à outils : réseaux de neurones et apprentissages statistiques. Nous avons utilisé la plupart des paramètres par défaut des classifieurs pré-implémentés. Pour chaque classifieurs, les paramètres que nous avions modifiés et qui donnent les plus grandes précisions sont présentés dans le tableau 6.7. Et les résultats obtenus par chaque classifieurs sont résumés dans le tableau 6.8.

Tableau 6.7 – Les Valeurs numériques des paramètres de chaque classifieurs.

Classifieur	Paramètre	Valeur
RNA	Nombre de couches cachées	100
SVM	Le paramètre C qui contrôle le compromis entre la complexité de la règle de décision et le taux d'erreur d'apprentissage [230]	100
k-PPV	Le nombre de voisins	1
ADL	Pas de modification	

Tableau 6.8 – Classification d'une base de données de signuax EEG d'épélepsie. Les performances obtenues par le classifieurs RNA, SVM et l'ADN implémentés dans l'espace de dissimilarité et le classifieur k-PPV. La RTF de SPWVD avec une fenêtre de lissage Hanning et la distance d_{DSL_1} ont été utilisées. Évaluation des performances à l'aide du problème de classification à 5 classes et la validation croisée *10-flod*.

Classifieur	Classe	Spécificité (%)	Sensibilité (%)	AUC	AUC (moyenne)	Précision (%)
RNA	A	100	100	1	0.9850	97.60
	B	99.00	100	0.9988		
	C	96.80	92.00	0.9562		
	D	92.40	97.00	0.9750		
	E	100	99.00	0.9950		
SVM	A	84.70	100	1	0.9698	94.20
	B	99.00	99.00	0.9938		
	C	94.50	86.00	0.9238		
	D	96.70	87.00	0.9388		
	E	98.00	99.00	0.9925		
ADL	A	100	100	1	0.9875	98.00
	B	100	100	1		
	C	95.00	96.00	0.9738		
	D	95.00	95.00	0.9688		
	E	100	99.00	0.9950		
k-PPV	A	91.50	97.00	0.9738	0.9738	95.80
	B	96.80	90.00	0.9463		
	C	93.30	98.00	0.9813		
	D	97.90	95.00	0.9725		
	E	100	99.00	0.9950		

A partir du tableau 6.8, on peut observer que les meilleures performances de tous les classifieurs sont obtenues dans les classes A, B et E. La spécificité et la sensibilité obtenues dans ces classes sont d'environ 100% et l'AUC est d'environ 1. Cependant, les performances les plus bases de ces classifieurs sont obtenus dans les classes C et D. Selon la précision et l'AUC moyenne, il est évident que le classifieur ADL est classée en première place (98%, 09875). Tandis que le classifieur RNA est classée en second place (97.60%, 0.9850). Les classifieurs SVM et k-PPV sont placées en troisième place, ils avaient approximativement la même précision et AUC moyenne ($\sim$ 95%, $\sim$ 0,9700).

Les tableaux 6.9 et 6.10 récapitulatif les performances obtenues par les mêmes classifieurs sur les problèmes de classification à 2 et 3 classes. Nous pouvons clairement voir que tous les classifieurs ont des performances élevées dans toutes les instances étudiées. Pour tous les problèmes, tous les classifieurs avaient une AUC supérieure ou égale à 0,9900 et une précision supérieure ou égale à 99,30%.

b.1) Synthèse

Les résultats de classification confirment que les classifieurs implémentés dans l'espace de dissimilarités sont plus performant par rapport au classifieur k-PPV. Ceci est clairement prouvée par l'application de ces classifieurs sur le problème de classification à 5 classes. Une précision de 98.00% est achevé par le classifeiur ADL dans ce problème de classification par rapport à 95.80% achevé par le classifieur k-PPV.

Tableau 6.9 – Classification d'une base de données de signuax EEG d'épélepsie. Les performances obtenues par le classifieurs RNA, SVM et l'ADN implémentés dans l'espace de dissimilarité et le classifieur k-PPV. La RTF de SPWVD avec une fenêtre de lissage Hanning et la distance d_{DSL_1} ont été utilisées. Évaluation en utilisant les instances du problème de classification à 2 classes. Évaluation de la performance en utilisant la précision et la validation croisée *10-flod*.

classifieur	RNA		ADL		kNN		SVM	
Instance	AUC	Précision (%)	AUC	Accuracy (%)	AUC	Précision (%)	AUC	Précision (%)
A, E	**1**	**100**	**1**	**100**	**1**	**100**	**1**	**100**
B, E	**1**	**100**	**1**	**100**	0.9950	99.50	**1**	**100**
C, E	0.9500	99.50	**1**	**100**	**1**	**100**	0.9950	99.50
D, E	0.9950	99.50	0.9950	99.50	**1**	**100**	0.9950	99.50
AB, E	0.9950	99.70	**1**	**100**	0.9950	99.70	**1**	**100**
CD, E	0.9950	99.70	0.9950	99.70	**1**	**100**	0.9950	99.70
ACD, E	0.9933	99.50	0.9950	99.80	**1**	**100**	0.9950	99.80
BCD, E	0.9933	99.50	**0.9950**	**99.80**	**0.9950**	**99.80**	**0.9950**	**99.80**
ABCD, E	0.9938	99.60	**0.9950**	**99.80**	**0.9950**	**99.80**	0.9900	99.60

Tableau 6.10 – Classification d'une base de données de signuax EEG d'épélepsie. Les performances obtenues par le classifieurs RNA, SVM et l'ADN implémentés dans l'espace de dissimilarité et le classifieur k-PPV. La RTF de SPWVD avec une fenêtre de lissage Hanning et la distance d_{DSL_1} ont été utilisées. Evaluation en utilisant les instances du problème de classification à 3 classes. Évaluation de la performance en utilisant la précision et la validation croisée *10-flod*.

classifieur	RNA		ADL		k-PPV		SVM	
Instance	AUC	Précision (%)	AUC	Accuracy (%)	AUC	Précision (%)	AUC	Précision (%)
A, C, E	**0.9975**	**99.70**	**0.9975**	**99.70**	**0.9975**	**99.70**	**0.9975**	**99.70**
A, D, E	**0.9975**	**99.70**	**0.9975**	**99.70**	**0.9975**	**99.70**	**0.9975**	**99.70**
AB, CD, E	0.9942	99.40	**0.9964**	**99.60**	0.9950	99.40	0.9938	99.20

En revanche, pour tous les classifieurs, la plupart des erreurs obtenus dans le problème de classification à 5 classes sont les taux d'erreurs entre les classes C et D. Etant donné que les signaux EEG de ces classes sont enregistrés au cours des crises épileptique des mêmes patients, il convient de mentionner que cela n'a pas d'impact majeur sur l'applicabilité de l'approche. La seule différence entre les signaux EEG de la classe C et D est la zone du cerveau à partir de laquelle ils ont été enregistrés. Ce qui signifie que les signaux EEG de ces deux ensembles sont plus similaires par rapport aux signaux des autres ensembles (A, B et E). Et ceci est parfaitement prouvé par l'application de l'approche proposée sur les problèmes de classification à 2 et 3 classes. Les résultats présentés dans les tableaux 6.9 et 6.10 montre que lorsque les ensembles C et D sont considérés comme une même classe, tous les classifieurs ont atteint une précision très satisfaisante dans la plupart des problèmes de classification.

c) Complexité de l'approche proposée

La complexité d'un algorithme consiste en l'étude de deux quantités de ressource, à savoir le temps et l'espace mémoires nécessaires à l'exécution de cet algorithme. Généralement cette complexité consiste à évaluer le nombre d'opération élémentaires de l'algorithme en fonction de la taille de données.

Dans cette étude, nous avons choisi d'évaluer d'une manière expérimentale la complexité de l'algorithme de l'approche proposée. D'abord, nous avons étudié le temps d'exécution et l'espace mémoire nécessaires à l'exécution de cet algorithme en fonction de la taille de la base de données d'apprentissage. Ensuite, Nous avons tracé et ajuster les courbes mesurées. Enfin, nous avons utilisé la notion Big-O pour approximer la complexité de l'algorithme.

Tout au long de ces expérimentations, nous avons utilisé un logiciel Matlab pour programmer l'algorithme de l'approche proposée. Ce logiciel est installé sur un ordinateur avec les caractéristiques suivantes : Intel(R) Xeon(R) CPU E5-1620 v2-3.70GHz, RAM-8Go, système d'exploitation Windows 8 (64bits).

C.1) Complexité en temps

L'algorithme que nous avons développé pour l'approche proposée consiste en trois étapes. La première vise à extraire les RTF des signaux EEG. La matrice d'apprentissage est calculée en utilisant les RTF dans la deuxième étape. La troisième étape comprend les algorithmes d'apprentissage et de prédiction du classifieur.

D'abord, nous avons évalué séparément la complexité de chaque étape afin d'avoir une idée sur l'étape qui consomme le plus de temps. Ensuite, nous avons évalué la complexité de l'ensemble de l'algorithme.

Pour estimer la complexité de chaque étape, le temps d'exécution est mesuré en fonction du nombre de signaux dans la base de données d'apprentissage. Les figures 6.11 et 6.12 présentent les courbes mesurés et ajustée. Le tableau 6.11 résume la notion Big-O et le nom de la fonction de la complexité en temps de chaque étape.

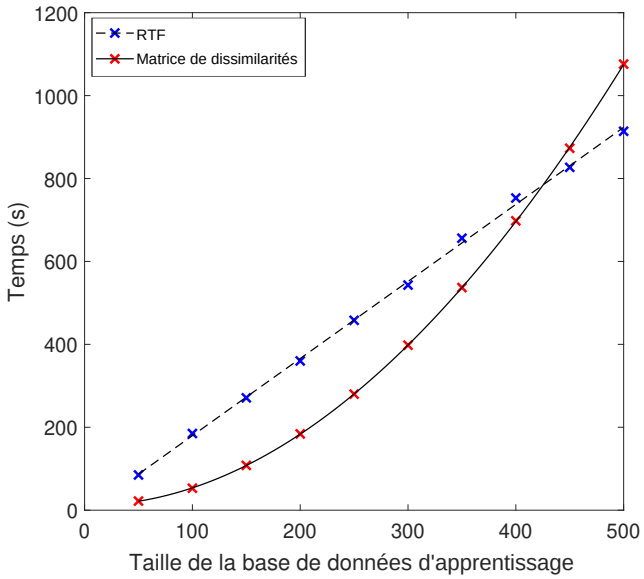


FIGURE 6.11 – Le temps d'exécution de la première (calcule des RTF) et la deuxième (calcule de la matrice d'apprentissage) étapes en fonction du nombre de signaux dans la base de données d'apprentissage

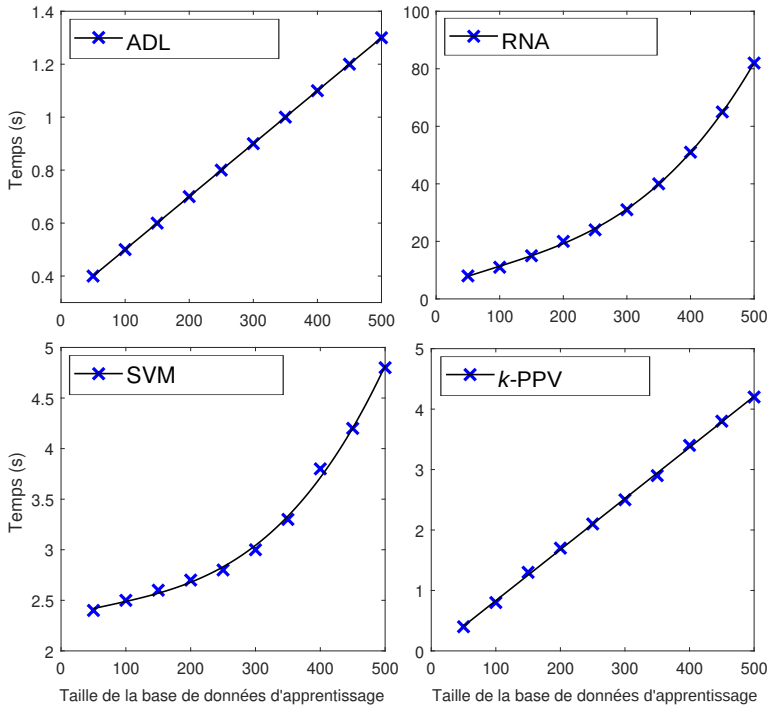


FIGURE 6.12 – Le temps d'exécution de la troisième étape en fonction du nombre de signaux dans la base de données d'apprentissage. L'approche est combinée avec les classifieurs : RNA, SVM, ADL et k-PPV.

Tableau 6.11 – La notation Big-O de la complexité en temps des trois étapes de l'algorithme de l'approche de classification basée sur une mesure de dissimilarité.

Etape de l'algorithme	RTF	Matrice de dissimilarités	ADL	RNA	SVM	k-PPV
La notation Big-O	$O(n)$	$O(n^3)$	$O(n)$	$O(n^3)$	$O(n^3)$	$O(n)$
Nom de la fonction	linéaire	Quadratique	linéaire	Quadratique	Quadratique	linéaire

Les résultats obtenus montrent que la complexité de la première et la seconde étape sont $O(n)$ et $O(n^3)$, respectivement. La complexité de la troisième étape dépend du type du classifieur utilisé. On remarque que la première et les secondes étapes sont les plus consommateurs de temps. Par exemple, avec une base de données d'apprentissage de 400 signaux, 914 secondes sont nécessaires pour extraire les RTF et 1076 secondes pour calculer la matrice d'apprentissage (matrice de dissimilarité).

Afin d'estimer la complexité en temps de l'approche proposée, le temps d'exécution de l'algorithme entier (les trois étapes) est mesuré en fonction du nombre de signaux dans la base de données d'apprentissage. La figure 6.13 expose les courbes mesurés et ajustés du temps d'exécution en fonction du nombre de signaux de la base de données d'apprentissage. On peut clairement noter que la complexité en temps de l'approche proposée est $O(n^3)$ quel que soit le classifieur. C'est parce que l'étape de calcul de la matrice de dissimilarité d'apprentissage est celle qui détermine cette complexité.

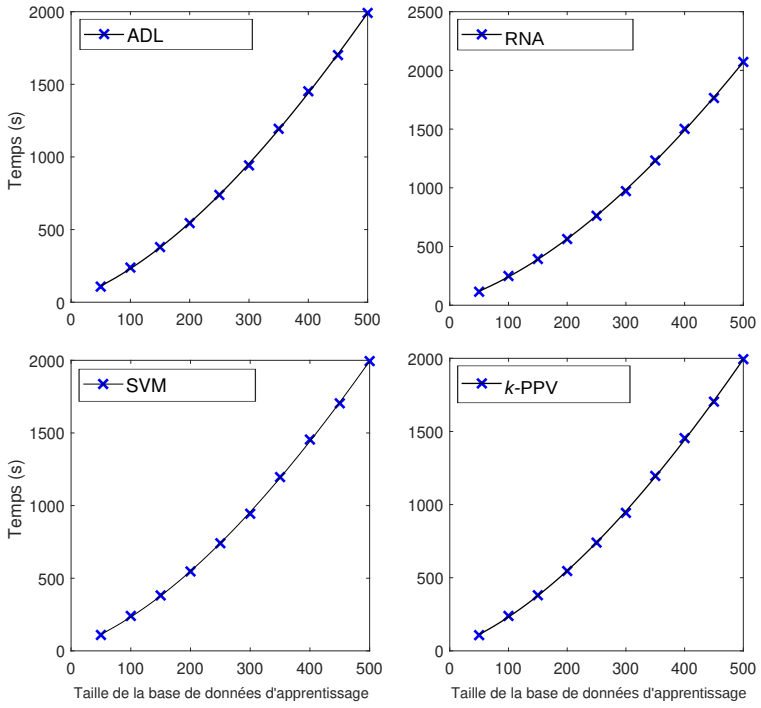


FIGURE 6.13 – Le temps d'exécution de l'algorithme complet de l'approche proposée en fonction du nombre de signaux dans la base de données d'apprentissage. L'approche est combinée avec les classifieurs : RNA, SVM, ADL et k-PPV.

C.2) Complexité en espace mémoire

L'algorithme de l'approche proposée utilisent des données numériques comme les RTF et la matrice d'apprentissage. Ces données sont d'abord stockées dans la mémoire et sont ensuite utilisées par l'algorithme. Pour estimer la complexité en espace mémoire de l'approche proposée, nous avons étudié la mémoire requise pour stocker ces données en fonction de la taille de la base de données d'apprentissage (Figure 6.14). Ces données sont stockées sous forme de valeurs numériques avec un virgule flottante à double précision. A l'aide de l'ajustement de la courbe mesuré, on constate que la complexité en espace mémoire de l'approche proposée est O(n).

d) Robustesse aux bruits de l'approche

Pour avoir une idée de l'influence du bruit sur l'applicabilité de l'approche proposée, les performances des classifieurs sont évaluées en fonction du rapport signal sur bruit (SNR) des signaux EEG propres. Le SNR est définie par $SNR(dB) = 10log(P_s/P_n)$ où P_s et P_n sont respectivement la puissance du signal et du bruit. Nous avons étudié la base de données de signaux EEG corrompus avec un bruit. Ainsi, Nous avons ajouté un bruit additif blanc gaussien à tous les signaux EEG de la base de données. Nous avons considéré un SNR allant de 0dB à 50 dB. Ensuite, nous avons utilisé ces signaux bruités pour l'apprentissage et le test des classifieurs.

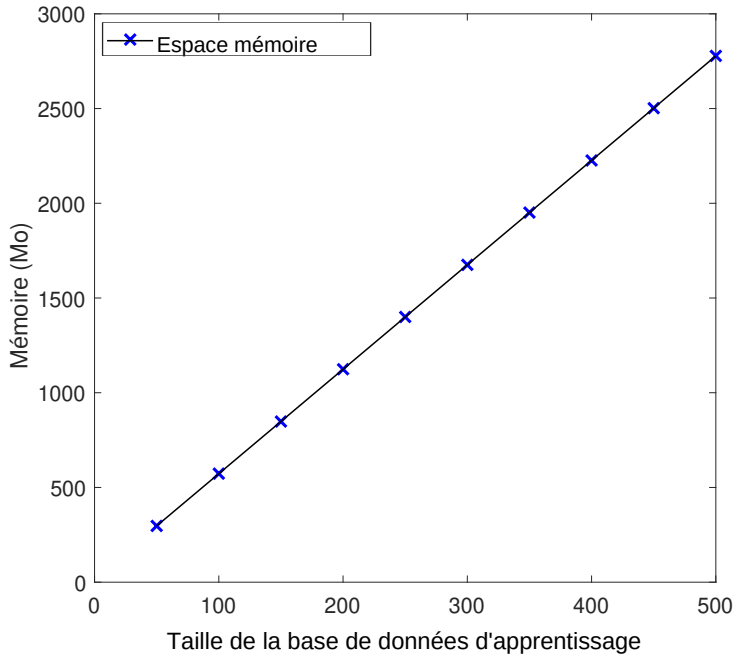


FIGURE 6.14 – Espace mémoire nécessaire pour l'exécution de l'algorithme de l'approche proposée en fonction de la taille de la base de données d'apprentissage (nombre de signaux).

Les classifieurs sont tournés en utilisant trois problèmes de classification, un problème à 2 classes (instance : ABCD-E), un problème à 3 classes (instance : AB-CD-E) et le problème à 5 classes. Les performances de ces classifieurs sont vérifiées par la validation croisée $10 - flod$ et la précision. Les graphes de la figure 6.15 représentent les précisions obtenues en fonction du SNR.

Comme le montrent les figures 6.15-a et 6.15-b, l'approche proposée donne de meilleurs résultats dans les problèmes de classification à 2 et 3 classes. La précision de la classification se détériorant généralement de manière non significative lorsque le SNR diminue. Par exemple, à un SNR de 0dB, la détérioration est inférieure à 3% et 10% pour les problèmes de classification à 2 et 3 classes, respectivement. De plus, pour chaque classifieur, il n'y a pas de changement significatif dans les précisions obtenues entre la base de données bruitée avec un SNR de 1dB et la base de données propre.

Comme le montre la figure 6.15-c, la performance de l'approche proposée dans le problème de classification à 5 classes est loin d'être satisfaisante lorsque le SNR est inférieur à 10 dB. Les performances des classifieurs sont significativement détériorées lorsque le SNR diminue. Cela en raison de la forte dépendance entre les classes C et D. Cependant, pour chaque classifieur, il n'y a pas un changement significatif dans les précisions obtenues entre la base de données bruitée avec un SNR de 20dB et la base de données propre.

Ces résultats indiquent clairement que l'approche proposée est robuste au bruit. Bien qu'une précision insatisfaisante soit obtenue dans le problème de la classification à 5 classes, un SNR faible ne peut pas influer sur l'applicabilité de l'approche proposée pour distinguer les signaux EEG des personnes en bonne santé et des patients épileptiques.

e) Comparaison avec les résultats de la littérature

Nous avons comparé quatre classifieurs. Les classifieurs : RNA, le SVM et l'ADL implémentés dans l'espace de dissimilarité et le classifieur k-PPV. Les performances de ces classifieurs sont évaluées par rapport à 13 problèmes de classification. Les résultats obtenus ont montré que ces classifieurs atteignent des précisions maximales dans la plupart des problèmes de classification. Cependant, le classifieur ADL a obtenu la meilleure précision dans le problème de classification à 5 classes. Ainsi, les résultats de notre approche qui ont été comparés aux résultats de la littérature sont ceux obtenus par le classifieur ADL.

La comparaison des résultats en termes de précisions rapportés dans la littérature pour les approches de classification de signaux EEG épileptiques publiées de 2008 à 2018 [135, 147, 210–228] avec celles de l'approche proposée est présentée dans les tableaux 6.12, 6.13 et 6.14. Toutes ces approches sont évaluées sur la même base de données [5] que pour l'approche proposée. Bien que Dhiman et al. [220] et Das et al. [223] ont rapporté un taux de classification de 100% pour les instances de classification A-E, B-E, C-E, D-E, AB-E, ABCD-E et A-D-E, respectivement. L'approche proposée surpasse approximativement les mêmes

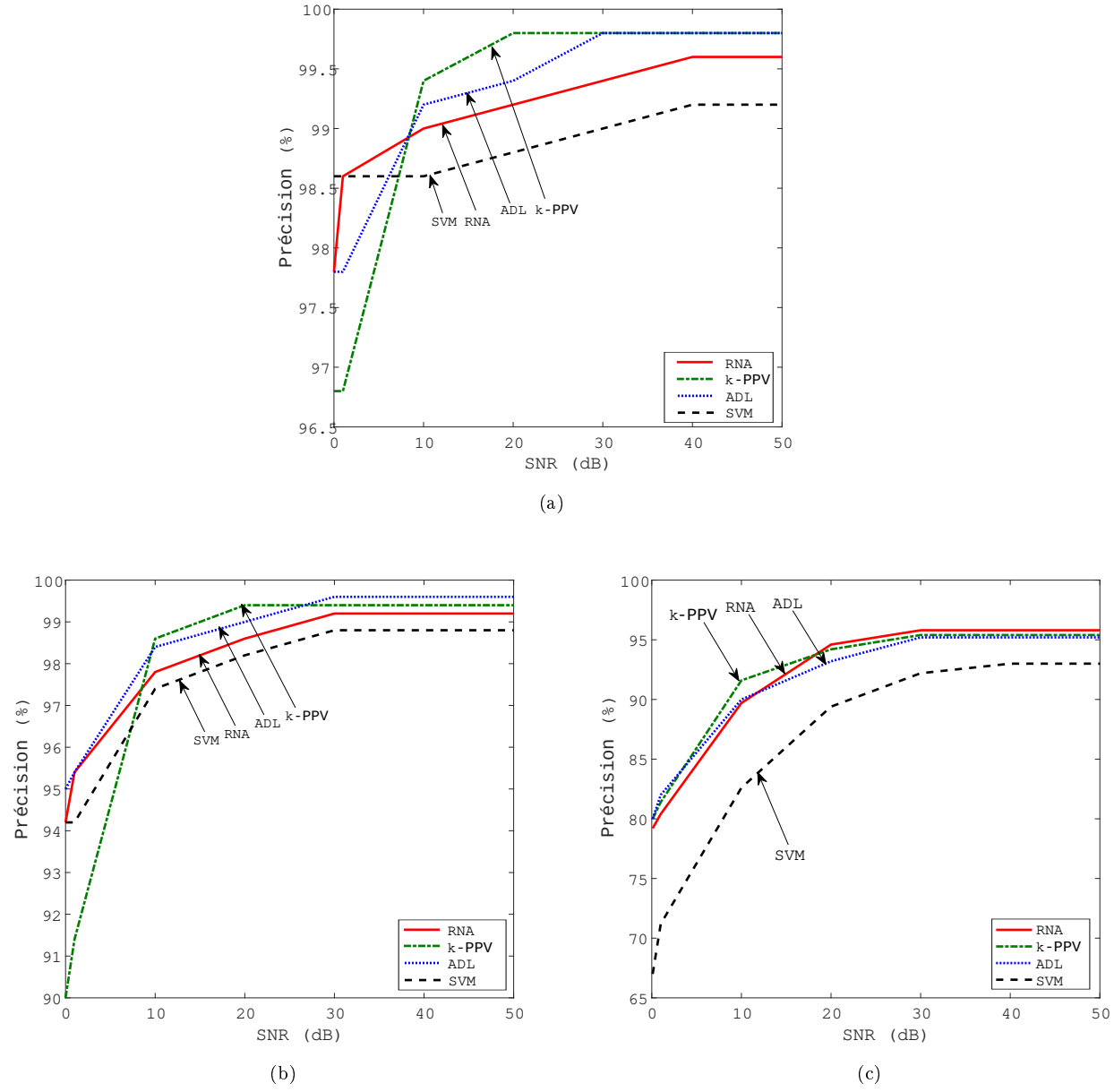


FIGURE 6.15 – Précision de la classification en fonction du SNR. Les classifieurs sont évalués en utilisant : (a) le problème de classification à 2 classes, (b) le problème de classification à 3 classes et (c) le problème de classification à 5 classes.

Tableau 6.12 – Comparaison de la précision (en %) obtenue par différentes méthodes de détection des crises d'épilepsie appliquées aux données de [39]. Comparaison des résultats précédemment rapportés pour un problème de classification à 2 classes. Instances de classification définies à partir de (A, B, C, D, E).

Étude	A, E	B, E	C, E	D, E	AB, E	CD, E	ACD, E	BCD, E	ABCD, E
[216]	**100**	**100**	99.60	95.80	-	-	98.15	98.22	97.38
[218]	**100**	**100**	99.86	97.97	-	-	-	-	-
[220]	**100**	**100**	**100**	**100**	-	-	-	-	**100**
[221]	**100**	-	**100**	**100**	-	-	-	-	99.60
[223]	**100**	-	**100**	**100**	-	-	-	-	**100**
[135]	-	-	-	-	-	99.33	-	-	-
[224]	-	-	-	**100**	-	-	-	-	-
[147]	**100**	**100**	99	98.50	**100**	98.67	-	-	99.20
[229]	-	-	-	-	**100**	99.45	-	-	99.31
[227]	99.82	99.25	99.10	99.07	-	98.88	-	-	98.72
Cette étude	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**99.80**	99.80

Tableau 6.13 – Comparaison de la précision (en %) obtenue par différentes méthodes de détection des crises d'épilepsie appliquées aux données de [39]. Comparaison des résultats précédemment rapportés pour un problème de classification à 3 classes. Instances de classification définies à partir de (A, B, C, D, E).

Étude	A,C,E	A,D,E	AB,CD,E
[212]	91-96	-	-
[223]	-	**100**	96.28
[221]	-	98.67	98.40
[225]	-	98.78	-
[229]	-	-	98.80
[227]	-	98.22	-
[228]	-	99.60	-
cette étude	**99.70**	99.70	**99.60**

résultats pour les mêmes instances mais sans extraction de caractéristiques. Pourtant, l'approche proposée a le meilleur taux de classification pour les instances de classification A-B-C-D-E (98%), ACD-E (99,80%), BCD-E (99,80%), CD-E (99,80%), A-C-E (99,70%), et AB-CD-E (99,60%) comparé aux résultats rapportés par Ahnaf et al. [221], Mousavi et al [212], et Kumar et al. [211].

Tableau 6.14 – Comparaison de la précision obtenue par différentes méthodes de détection des crises d'épilepsie appliquées aux données de [39]. Comparaison des résultats précédemment rapportés pour un problème de classification à 5 classes (instance A-B-C-D-E).

Étude	Instance	Précision (%)
[213]	A, B, C, D, E	89
[215]	A, B, C, D, E	85.90
[222]	A, B, C, D, E	96.40
[226]	A, B, C, D, E	87.20
Cette étude	A, B, C, D, E	**98.00**

f) Discussion

Le tableau 6.15 présente un bref résumé des différentes approches qui ont été exploitées dans la détection des crises d'épilepsie. Il est très important de noter que différents classifieurs ont été utilisés pour la détection des crises d'épilepsie dans les signaux EEG. Tous ces classifieurs ont été réalisés en utilisant l'approche de l'extraction de caractéristiques. Cependant, lorsque les caractéristiques ne sont pas pertinentes, les performances obtenues par le système de classification sont insatisfaisantes. En raison des caractéristiques inappropriés, l'algorithme de classification ne peut pas générer de bonnes performances. Par conséquent, il est obligatoire de trouver et d'extraire les caractéristiques pertinentes à partir de signaux bruts afin de pouvoir obtenir de bons résultats de classification. D'autre part, notre étude a montré que la détection des crises d'épilepsie dans les signaux EEG peut désormais se faire en utilisant l'approche basée sur une mesure de dissimilarité.

D'autre par les résultats de la classification en termes de précision obtenus par notre approche sont comparable voir supérieur à ceux obtenus par les approches de la littérature. L'approche proposée a permis d'obtenir des précisions élevées dans les 13 problèmes de classification étudiés (problèmes de classification à 2, 3 et 5 classes). Pour les problèmes de classification à 2 classes, la précision obtenue par notre approche est la meilleure présentée pour les instances de classification (A-E, B-E, C-E, CD-E, ACD-E et BCD-E) et est la deuxième présentée pour les instances D-E et ABCD-E. Concernant les problèmes de classification à 3 classes, la précision obtenue avec notre approche est la meilleure présentée pour les instances de classification (AB-CD-E et A-C-E) et est la deuxième présentée pour l'instance de classification A-D-E. Pour le problème de classification à 5 classes, la précision obtenue à l'aide de notre approche est la meilleure présentée.

Par ailleurs, les classifieurs que nous avons combinés avec notre approche ont atteint de très bonnes performances. Et les principales erreurs de ces classifieur étaient le taux d'erreur entre les classes C et D. Cela peut s'expliquer par la forte dépendance entre ces deux classes. Cependant, cela n'affecte pas l'applicabilité de notre approche pour discriminer :

— Les signaux EEG des personnes en bonne santé et des patients en état de crise.

— Les signaux EEG des personnes en bonne santé et des patients en absence d'état de crise.

— Les signaux EEG des patients en état de crise et ceux en absence d'état de crise

6.6 Conclusion

La classification des signaux non-stationnaires est un problème difficile et très étudié. D'une part, la non-stationnarité exclut la classification dans le domaine temporel ou fréquentiel. D'autre part, les RTF, bien qu'adaptées aux signaux non-stationnaires, ont une grande dimension. Les RTF ont été utilisées auparavant pour la classification des signaux non stationnaires. Cependant, ces représentations sont souvent combinées

Tableau 6.15 – Diverses approches utilisées pour la détection de crise épileptique appliquée à la base de données [5]. Travaux publiés entre 2004 et 2018.

Auteurs	Année	Descripteurs	Classifieur
Mousavi et al.[212]	2008	AR model , wavelet decomposition	Multi-Layer Perceptron(MLP)
Tzallas et al. [213]	2009	Features based TFD	SVM
Liang et al.[215]	2010	Fast Fourier Transform, approximate entropy analysis	SVM
Kumar et al. [216]	2014	Digital Wavelet Transform (DWT) based fuzzy approximate entropy	SVM
Dhiman et al. [220]	2016	Wavelet Packet Transform	Twin SVM
Ahnaf et al. [221]	2016	Tunable-Q factor wavelet transform	Bootstrap aggregating
Das et al. [223]	2016	Dual-tree complex wavelets, inverse Gaussian	SVM
Alcin et al. [222]	2016	STFT, Gray Level Co-occurrence Matrix, Fisher Vector	Extreme Learning Machine
Bhati et al. [135]	2017	Time–frequency localized three-band synthesis filter bank	MLPNN
Biju et al. [224]	2017	Energy/entropy of IA and SD of IF from Hilbert-Huang Transform	Artificial Neural Network (ANN)
Li et al. [225]	2017	DWT-based EA	Neural Network Ensemble (NNE)
Sharma et al. [147]	2017	analytic time-frequency flexible wavelet transform (ATFFWT) and fractal dimension (FD)	LS-SVM
Tiwari et al. [229]	2017	Key-Point Descriptor of EEG Signals	SVM
Zahra et al. [226]	2017	Multivariate EMD	ANN
Jaiswal et al. [227]	2017	Local Neighbor Descriptive Pattern and One-dimensional Local Gradient Pattern feature extraction techniques	ANN
Li and al. [228]	2018	wavelet packet transform (WPT)	Least squares support vector machine
Cette étude	**2018**	**Mesure de dissimilarité entre les RTF**	**RNA, ADL, k-PPV, SVM**

avec l'approche basée sur l'extraction de caractéristiques ce qui peut limiter les performances de ces méthodes de classification. Pourtant, Les RTF peuvent être combinés avec l'approche de classification basée sur une mesure de dissimilarité. Ainsi, cette idée est l'objectif principal de la deuxième partie de cette étude doctorale.

Dans ce chapitre, nous avons proposé une nouvelle approche de classification de signaux non-stationnaires, nommée approche de classification basée sur une mesure de dissimilarité entre les RTF. Cette approche diffère des approches classiques de classification où les signaux à classer sont représentés par des vecteurs de descripteurs déduites à partir des RTF. Dans cette approche, les signaux sont représentés par un vecteur de leurs dissimilarités par rapport à une collection de signaux représentatif de chaque classe. Ainsi, Le fait de postuler l'espace engendré par les dissimilarités en tant qu'espace vectoriel, permet l'utilisation de tous classifieurs. Si une mesure de dissimilarité appropriée est choisie, la propriété de compacité (les RTF des signaux de la même classe doivent être similaires et les RTF des signaux de différentes classes doivent être différents) des classes doivent être plus prononcées. Par conséquent, il devrait être plus facile pour les classifieurs de les distinguer.

Ensuite, nous avons montré que la représentation la plus complète pour analyser les signaux non-stationnaires est la RTF, parce qu'elle fournit des informations simultanées sur les variables temps et fréquence d'un signal. La RTF permet d'analyser le comportement de la concentration d'énergie des composantes fréquentielles d'un signal à différents instants. Il est donc intéressant d'étudier un signal non stationnaire dans le domaine T-F afin d'avoir des informations pertinentes et complètes. Ces informations sont largement utilisées pour la classification et l'identification des signaux non-stationnaires dans de nombreux domaines, tels que l'ingénierie biomédicale, l'ingénierie mécanique, le traitement de signaux radar et sonore, etc. En même temps, nous avons présenté la WVD comme étant la RTF qui offre, théoriquement, une localisation T-F parfaite. Cependant, en raison de sa nature quadratique, l'utilisation et l'interprétation de cette représentation est rendue difficile par la présence des termes d'interférence. Par conséquent, plusieurs méthodes ont été présentées pour remédier à cet inconvénient, à savoir la méthode qui consiste à appliquer un lissage dans le plan T-F afin d'atténuer les termes d'interférences. Ainsi, cette méthode a permis d'obtenir de nombreux nouvelles distributions T-F appartenant à la classe de Cohen. Parmi celles-ci, on trouve la SPWVD qui fait intervenir un lissage double en temps et en fréquence. Il devient alors assez simple d'ajuster ces deux degrés de libertés, pour obtenir, en fonction de la structure du signal, un bon compromis entre la localisation des termes propres et l'atténuation des termes d'interférences. Grâce à cette simplicité, cette représentation est certainement l'un des éléments de la classe de Cohen qui fournit des représentations des signaux non-stationnaires à la fois précises et faciles à interpréter.

Enfin, nous avons choisi d'évaluer les performances de notre approche sur un problème de classification de signaux EEG d'épilepsie. La base de données de ces signaux a été obtenus à partir d'enregistrements cliniques

d'EEG, effectués dans le département Epileptology de l'université de Bonn, sur à la fois des patients normaux et épileptiques pendant les périodes d'activité et d'inactivité des crises. L'état de l'art que nous avons effectué sur un nombre important de travaux de recherche utilisent cette base de données à des fins de classification, nous a révélé que la seule approche utilisée est celle basée sur l'extraction de caractéristiques. De plus, 13 problèmes de classification sont généralement utilisés, comprendre des problèmes à 2 classes ((normale, active) (active, inactive) et (normale + active, inactive)), des problèmes à 3 classes (normale, active, inactive) et un problème à 5 classes (active (yeux ouverts), normale (yeux fermés), inactive (zone épileptogène), inactive (zone hippocampe), active). Ainsi, pour mettre en évidence le potentiel de notre approche, les résultats obtenus dans chaque problème de classification de ces travaux de recherche sont comparés avec les résultats obtenus par notre approche.

Ainsi, les résultats expérimentaux de cette étude montrent que l'approche proposée atteint des performances satisfaisantes dans la plupart des problèmes de classification. Une précision de 98% est obtenue pour le problème de classification à 5 classes. En outre, dans la plupart des problèmes de classification avec 2 et 3 classes, il donne également une précision satisfaisante d'environ 100%. La comparaison entre l'approche proposée et les études de la littérature a confirmé que l'approche proposée avait un potentiel dans la classification des signaux EEG et la détection des crises épileptiques.

D'une part, nous avons étudié également la robustesse de l'approche proposée avec l'ajout de bruit aux signaux EEG à différents SNR. Les résultats expérimentaux montrent que, lorsque le SNR$>=$ 1 dB, la précision de la classification est supérieure à 98%, 95% et 82% pour les problèmes de classification à deux classes, à 3 classes et à 5 classes, respectivement. Ces résultats indiquent clairement que l'approche proposée est robuste au bruit. Et ce dernier ne peut pas affecter l'applicabilité de cette approche pour le diagnostic de l'épilepsie en utilisant les signaux EEG.

D'autre part, le principal problème de l'approche proposée est la grande dimension de l'espace engendré par les dissimilarités. Ce dernier peut augmenter considérablement la complexité en termes de temps de calcule du classifieur utilisé. Comme le montrent les résultats expérimentaux, les complexités de l'approche proposée est : $O(n^3)$ pour la complexité en temps et $O(n)$ pour la complexité en espace mémoire, quel que soit le classifieurs.

Afin de résoudre ce problème, deux solutions peuvent être proposées. La première consiste à construire l'espace de dissimilarité en utilisant tous les signaux disponibles, puis à appliquer des techniques de réduction de dimensionnalité (analyse en composantes principales, t-distributed stochastic neighbour embedding, ...). La seconde consiste à utiliser une approche de sélection de prototypes pour réduire la dimension de l'espace de dissimilarité [129]. Cependant, le fait de travailler avec un nombre réduit de prototype peut probablement dégrader les performances du classifieur. En effet, il faut trouver une solution pour que le classifieur converge

vers la même précision malgré le nombre réduite de signaux d'apprentissage. Ainsi, la propriété de compacité d'une mesure de dissimilarité peut être une solution concevable.

Par ailleurs, l'approche que nous avons proposée nécessite l'utilisation d'une RTF et d'une mesure de dissimilarité pour calculer une matrice de dissimilarité. Ainsi, un classifieur combiné avec cette approche utilise la matrice de dissimilarité afin de construire une règle de décision. En effet, la réussite de ce classifieur est directement lié à un choix judicieux d'une RTF et à une mesure de dissimilarité. Un mauvais choix de l'un de ces deux derniers peut considérablement dégrader la précision du classifieur. Par ailleurs, la matrice de dissimilarité représente les dissimilarités par paires entre les RTF des signaux d'apprentissage. Ainsi, le choix de la RTF et la mesure de dissimilarité peuvent dépendre des signaux d'apprentissage.

En conséquence, deux pistes sont envisageables pour améliorer les performances de notre approche. Le premier est de ne pas restreindre la recherche d'une RTF plus appropriée au problème étudié à la classe de Cohen et d'élargir la recherche à d'autre classe de RTF, tels que la classe affine de RTF. Le deuxième, nous l'avons évoqué ci-dessus, est en relation avec la propriété de compacité de la mesure de dissimilarité choisi. Ainsi, la recherche d'une mesure de dissimilarité avec une propriété de compacité plus importante (réduire la distance entre les signaux de la même classe et l'augmenter entre les signaux de deux classes différentes), est la deuxième solution. Pour ce faire, nous avons élargé la recherche d'une mesure de dissimilarité aux mesures utilisés dans le domaine de traitement d'image. Ainsi, cette idée est l'objet principale de l'étude du chapitre suivant de cette partie.

Chapitre 7

Approche de classification basée sur la carte de dissimilarité locale : application aux images

Sommaire

7.1 Introduction 132

7.2 Approche de classification basée sur une mesure de dissimilarité 133

7.3 Carte de dissimilarité locale 134

7.3.1 CDL pour image binaire 134

7.3.2 CDL pour image en niveaux de gris 135

7.3.3 Seuils non régulièrement espacés pour la CDL 136

7.4 Base de données 137

7.5 Résultats et discussion 137

7.5.1 Performance de la CDL en fonction de nombre de coupes et les seuils 138

7.5.2 CDL contre la distance euclidienne et la SSIM 140

7.5.3 Classification dans l'espace de dissimilarités 145

7.5.4 Autre application : classification d'une base de données de lettrines 147

7.6 Conclusion 148

7.1 Introduction

Cette partie de cette étude doctorale a pour objectif de mettre au goût du jour l'approche de classification basée sur une mesure de dissimilarité. Nous avons montré, dans le chapitre précédent, l'avantage d'implémenter un classifieur dans un espace de dissimilarité au lieu d'un espace de caractéristiques classique. Dans cette approche, l'espace de dissimilarité est construit à partir d'un petit ensemble de prototypes. Dans la phase d'apprentissage, une base de données d'apprentissage T avec n objets et un ensemble de représentation R avec $r(n >> r)$ prototypes sont utilisés pour calculer une matrice de dissimilarité $D(T, R)$ établissant une relation entre les objets d'apprentissage et les prototypes. Dans ce nouvel espace, appelé espace de dissimilarité, chaque entrée de la matrice correspond à la distance d'un objet d'apprentissage à un prototype. Ainsi, la réussite de cette approche dépend d'un choix approprié de la mesure de dissimilarité.

D'une part, la dissimilarité est une mesure relative calculée entre deux objets qui ont la même nature ou les mêmes caractéristiques, exprimant essentiellement une notion de ressemblance entre les objets comparés. Pour obtenir une description robuste des objets, une mesure de dissimilarité doit vérifier la propriété de compacité, c'est-à-dire, elle doit intégrer les invariances nécessaires pour un meilleur regroupement intra-classe et éloignement interclasse. Cette propriété doit permettre au classifieur, implémenté dans un espace de dissimilarité, de converger vers sa précision maximale le plus rapidement possible en fonction du nombre d'objets d'apprentissage. Ainsi, le premier bénéfice d'une telle mesure de dissimilarité est la résolution des problèmes de classification toute en utilisant un nombre réduit de prototypes et en conséquence un espace de dissimilarité de petite dimension. Ce dernier va réduire considérablement le temps de résolution d'un problème de classification. Le second bénéfice d'une telle mesure de dissimilarité est la résolution des problèmes de classification disposant d'une petite base d'apprentissage.

D'autre part, plusieurs mesures de dissimilarité ont été proposées dans la littérature [231]. Ces mesures peuvent être divisées en deux grandes familles. Les mesures de la première famille utilisent la distance euclidienne, bien que cette mesure ne soit pas nécessairement la plus appropriée. Les mesures du second tentent d'intégrer des informations statistiques sur les objets, de sorte que les valeurs mesurées reflètent les caractéristiques intrinsèques de chaque classe. Cependant, les mesures de dissimilarité de cette famille sont souvent globales et ne permettent pas d'avoir accès aux répartitions spatiales entre les objets.

Par ailleurs, de nombreuses mesures de dissimilarités ont été utilisées dans le domaine du traitement des images (comparaison, débruitage, compression, qualité, etc). Parmi elles, on trouve la CDL [232]. Cette méthode permet une évaluation bien définie des dissimilarités locales entre deux images, ce qui n'est pas le

cas des autres méthodes. Cette mesure a été développée dans un premier temps pour comparer les images binaires, ensuite une extension de la méthode a été adaptée pour les images en niveaux de gris. La CDL a montré de bonnes performances par rapport aux autres méthodes classiques de comparaison d'images.

Ainsi, le but principal de ce chapitre est de combiner la CDL avec l'approche de classification basée sur une mesure de dissimilarité. Ensuite évaluer les performances en termes de classification de cette approche sur deux bases de données d'images (une base de données de chiffres manuscrits et une base de données de lettrines). Pour ce faire, le principe de l'approche de classification basée sur une mesure de dissimilarité est rappelé. Ensuite, le principe de la CDL pour la comparaison des images binaires et en niveaux de gris est présenté. Enfin, la partie résultats est consacrée à une comparaison de notre proposition et l'approche de l'état de l'art.

7.2 Approche de classification basée sur une mesure de dissimilarité

Dans cette partie, nous rappelons brièvement le principe de l'approche de classification basée sur les mesures de dissimilarité comme descripteurs. Pour plus de détails, nous référons au chapitre 5.

Lorsqu'une image est présentée pour être classée, elle est comparée aux autres images et le résultat de cette comparaison ne peut être que «similaire» ou «dissimilaire». Pour mesurer le degré de similarité entre deux images, on considère la fonction d, définie par : $d: \mathrm{E} \times \mathrm{E} \rightarrow [0, \infty]$.

En effet, plus la mesure d entre deux images tend vers zéro plus les deux images sont similaires. Ces mesures peuvent donc être traitées comme une description d'un espace de caractéristique spécifique, où chaque dimension correspond à une image. Cela ne modifie pas essentiellement l'approche basée sur les descripteurs. En général, tout classifieur supervisé fonctionnant sur des descripteurs peut être ainsi utilisé. Dans le processus d'apprentissage (Figure. 7.1), les classifieurs sont construits sur la matrice de distance D d'une taille $m \times m$ (où chaque valeur de donnée exprime une mesure de dissimilarité entre deux images d'apprentissage) : $D(k,l) = d(I_k, I_l)$, d est la distance entre les deux images I_k et I_l. Les p images de test sont classés en utilisant leurs distances aux m images d'apprentissage (la taille des données d'essai est $(p \times m)$). Chaque image est ainsi représentée par l'ensemble de ses distances aux autres images. Le travail du classifieur est donc de déterminer les seuils permettant de distinguer les images similaires des images dissimilaires.

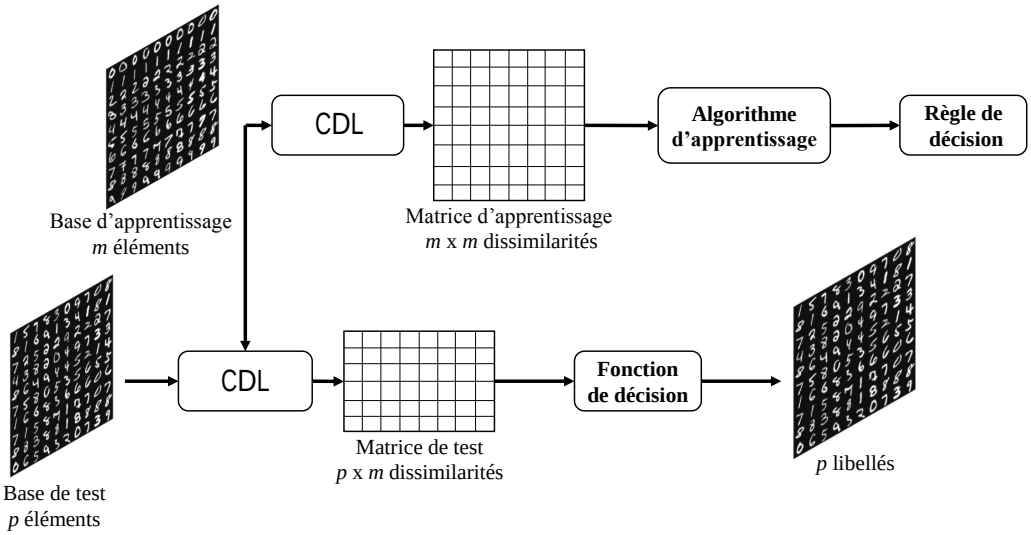


FIGURE 7.1 – Processus des méthodes de classification basées sur une mesure de dissimilarité.

7.3 Carte de dissimilarité locale

7.3.1 CDL pour image binaire

Nous rappelons ici brièvement le principe de la CDL entre deux images binaires [232]. L'objectif est de comparer localement deux images binaires avec une mesure locale bien définie. La carte de dissimilarité locale regroupe l'ensemble des mesures de dissimilarités locales faites en différents endroits des images. Ceci permet d'avoir accès à leur répartition spatiale. Ces informations sont rarement accessibles via les méthodes de comparaison d'images existantes qui sont souvent globales. Les images binaires étant représentées comme des ensembles de points, une distance locale efficace est une version modifiée de la distance de Hausdorff. La définition suivante de la CDL permet d'obtenir ses valeurs avec un coût de calcul faible :

Définition 1. CDL binaire

Soient A et B deux images binaires, la CDL est définie de $\mathbb{R}^2 \times \mathbb{R}^2$ dans $\mathbb{R}^2$ par

$$\mathrm{CDL}(A,B)(p) = |A(p) - B(p)| \max(d(p,A), d(p,B)), \tag{7.1}$$

$p = (x, y)$ étant une coordonnée pixel et $d(p, A)$ la transformée en distance de A c'est-à-dire la distance entre le pixel p et le pixel de A non-nul (donc définissant la forme) le plus proche.

7.3.2 CDL pour image en niveaux de gris

Plusieurs approches sont possibles pour généraliser la CDL aux images en niveaux de gris. Une solution consiste à étendre la définition de la transformée en distance aux images en niveaux de gris. La plupart des propositions présentent l'inconvénient de mêler distances pures et écarts de luminance. Les distances obtenues ne sont plus des écarts de position entre deux pixels (x, y). La proposition de Molchanov se distingue par la définition d'une distance résultant de l'accumulation de distances entre les sous-images binaires successives des deux images en niveaux de gris (Real Valued Distance Transform (RVDT)) [233] :

$$\mathrm{RVDT}(p, A) = \frac{1}{b-a} \sum_{i=a}^{b} d(p, A_i)\omega_i \tag{7.2}$$

avec $A_i = \{p : A(p) \geq i\}$, $0 <= a < b <= 255$ et $\omega_i = 1$ dans le cas le plus simple. Une possibilité serait donc d'exploiter cette RVDT en l'injectant dans l'eq. (7.1). Néanmoins, des distances entre niveaux de gris et entre positions (x, y) sont alors mélangées, rendant délicate une éventuelle interprétation et posant le problème des dynamiques des dimensions.

Nous proposons ici de conserver l'esprit de l'approche de Molchanov en découpant les images en niveaux de gris en images binaires (avec des seuils donnés). Les images binaires peuvent être comparées avec la CDL de l'équation (7.1). Les CDL sont alors accumulées en une moyenne selon les seuils.

Définition 2. Proposition de CDL pour image en niveaux de gris

Soient A et B deux images en niveaux de gris, N est un entier représentant le nombre de coupes, la $\mathrm{CDL_N}$ est définie de $\mathbb{R}^2 \times \mathbb{R}^2$ dans $\mathbb{R}^2$ par :

$$\mathrm{CDL_N(A, B)}(p) = \frac{1}{N} \sum_{i=1}^{N} \mathrm{CDL(A_i, B_i)(p)}, \tag{7.3}$$

où A_i (resp. B_i) est l'image binaire de A (resp. B) pour la i-ème coupe donnée par la sous-image de A :

$$A_i = \{p : A(p) \geq s_i\}\,. \tag{7.4}$$

Parmi les multiples stratégies pour déterminer les seuils donnant les sous-images, nous choisissons ici une solution simple qui consiste à prendre des seuils régulièrement espacés entre 0 et le maximum des deux images. Ce qui est équivalent à une normalisation en luminosité des deux images :

$$s_i = \frac{i}{N} m_{AB}, i \in [1..N] \tag{7.5}$$

avec $m_{AB} = \max(\max_p A(p), \max_p B(p))$.

La CDL finale entre deux images en niveaux de gris est obtenue en augmentant le nombre de coupes :

$$\mathrm{CDL}(A,B)(p) = \lim_{N\to\infty} \mathrm{CDL}_N(A,B)(p). \tag{7.6}$$

Cette définition généralise complètement la définition de la CDL binaire. Si les images sont binaires, toutes les coupes sont identiques et l'équation (7.6) revient à l'équation (7.1).

En pratique, si les niveaux de gris sont codés sur 8 bits, on peut poser $N = 255$. La figure 7.2 montre trois exemples de CDL obtenus à partir de trois paires d'images en niveaux de gris. Nous avons considéré $N = 10$ pour calculer chaque CDL.

La CDL compare entre les structures des images. Lorsque deux images ont la même structure mais différentes luminosités, leur CDL sera nulle. Donc la luminosité moyenne n'a aucune influence sur la comparaison des images par la CDL.

Pour pouvoir utiliser la CDL comme mesure de dissimilarité, nous avons défini une distance à partir de celle-ci. Ainsi, la distance entre deux images A et B ayant la CDL (A, B) est définie par :

$$d_{\mathrm{CDL}}(A,B) = ||\mathrm{CDL}(A,B)||_2 \tag{7.7}$$

7.3.3 Seuils non régulièrement espacés pour la CDL

Lors de la comparaison de deux images par la CDL, les seuils régulièrement espacés ne prennent pas en considération la dispersion de la luminosité des pixels. Ce qui peut influer sur la qualité de la CDL. Ainsi, nous proposons une stratégie qui consiste à choisir les seuils en fonction de l'histogramme cumulé des deux images. Cette stratégie est présentée sur l'algorithme suivant :

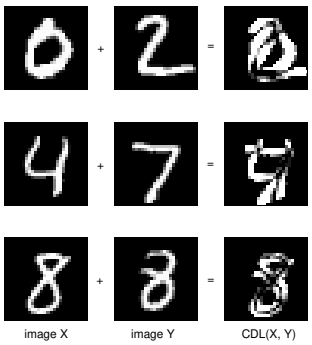


FIGURE 7.2 – CDL obtenues à partir de trois paires d'images en niveaux de gris.

Algorithme 7.1 : *génération des seuils*

Entrés : *deux images A et B, le nombre de coupes N*

Sorties : *les seuils S*

1. $C = (A + B)/2$
2. $[nbrPixel, lumstPixel] = histogramme(C)$
   // nbrPixel et lumstPixel sont deux vecteurs, tel que nbrPixel(i) correspond au nombre de pixels ayant la luminosité lumnstPixel(i)
3. *Q=cumule (nbrPixel)*
4. *m=min(Q)*
5. *M=mzx(Q)*
6. *pour j allant de 1 à N faire*
7. *t=m+j*(M-m)/(N+1)*
8. *posval=trouver la position de t dans Q*
9. *S(j)= lumstPixel(posval)*
10. *retourner S*

7.4 Base de données

La base de données MNIST (Figure 7.3) contient un grand nombre de chiffres manuscrits, et se compose de 10 classes (chiffres de 0 à 9). Chaque chiffre est une image de 28x28 pixels. La base MNIST est disponible à partir de la page web `http://yann.lecun.com/exdb/mnist`, où de nombreux résultats de classification de différentes méthodes sont également répertoriés. Les images sont des images en niveaux de gris. La base dispose d'un ensemble d'apprentissage de 60 000 exemples et d'un ensemble de tests de 10 000 exemples.

7.5 Résultats et discussion

La CDL, pour les images en niveau de gris, que nous avons proposé de mettre en œuvre nécessite le choix de deux paramètres, à savoir le nombre de coupes et le seuil. Ainsi, le premier objectif de cette partie est de faire un choix optimal de ces deux paramètres. Ensuite, la comparaison entre les performances de la CDL avec les performances de deux autres mesures de dissimilarité, telles que la distance euclidienne et la similarité structurale moyenne (Structural SIMilarity (SSIM)), est le deuxième objectif de cette partie.



FIGURE 7.3 – Exemple de chiffres manuscrits de la base de données MNIST.

Enfin, la comparaison entre un classifieur implémenté dans l'espace de dissimilarités engendré par la CDL et le classifieur k-PPV basée sur la distance de la CDL est le dernier objectif de cette partie.

7.5.1 Performance de la CDL en fonction de nombre de coupes et les seuils

a) Nombre de coupes

La qualité de la CDL pour les images en niveaux de gris dépend du choix de nombre de coupes. Pour un choix optimal de ce dernier, les performances de la CDL comme distance dans un algorithme d'apprentissage (le classifieur k-PPV) ont été évaluées en fonction du nombre de coupes. Les performances du classifieur k-PPV ont été mesurées par la précision. Une base de données d'apprentissage des chiffres manuscrit (contient 10 classes, chaque classe contient 100 individus) a été utilisée pour apprendre le classifieur k-PPV. Une deuxième base de données de ces mêmes chiffres (10 classes et chaque classe contient 10 individus) a été utilisée pour le test. Le nombre de coupes varie entre 1 et 10 avec un pas de 1. La figure 7.4 représente la précision obtenue par le classifieur et la durée de simulation en fonction de nombre de coupes.

On remarque que la précision du classifieur variée entre 96% et 98% en fonction de nombre de coupes. La meilleure précision (98%) est obtenue pour une seule coupe. En principe, la précision devrait être augmentée en fonction de nombre de coupes. Ce qui n'est pas le cas pour la base de données des chiffres manuscrits. Au contraire, on remarque une petite dégradation de la performance quand le nombre de coupes de la CDL est supérieur à 1. Cette dégradation est due au fait que les images des chiffres manuscrits de cette base de données sont quasi binaires. Sur les histogrammes de 4 images de ces chiffres (figure 7.5), on constate que la luminosité de la majorité des pixels de ces images est concentrée autour de 0 et 1. Donc, une seule coupe a

été suffisante pour que la CDL puise différencier deux images. En effet, un nombre de coupes supérieur à 1 a ajouté de l'information inutile dans la CDL, ce qui permet de perturber le classifieur. Sur la courbe de la durée de simulation en fonction de nombre de coupes, on constate que la durée de la simulation augmente linéairement en fonction de nombre de coupes.

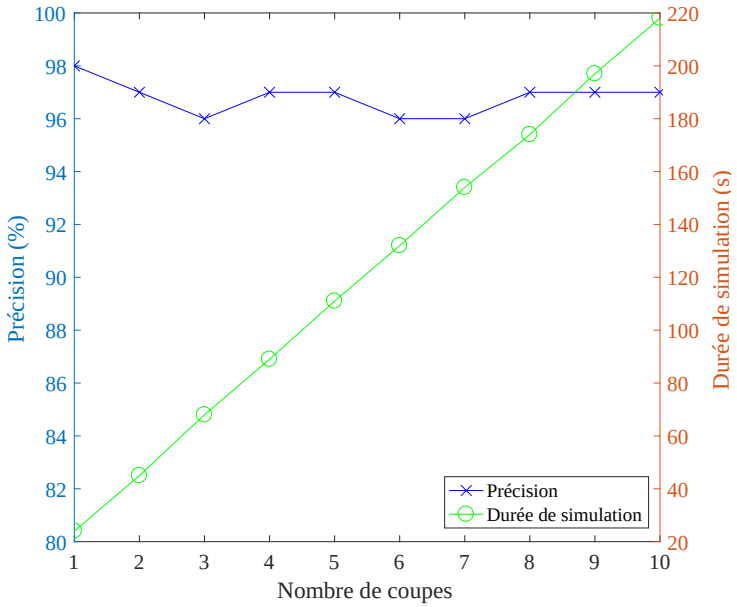


FIGURE 7.4 – Précision et durée de la simulation en fonction de nombre de coupes de la CDL. Seuils régulièrement espacée. Performances évaluées par le classifieur k-PPV.

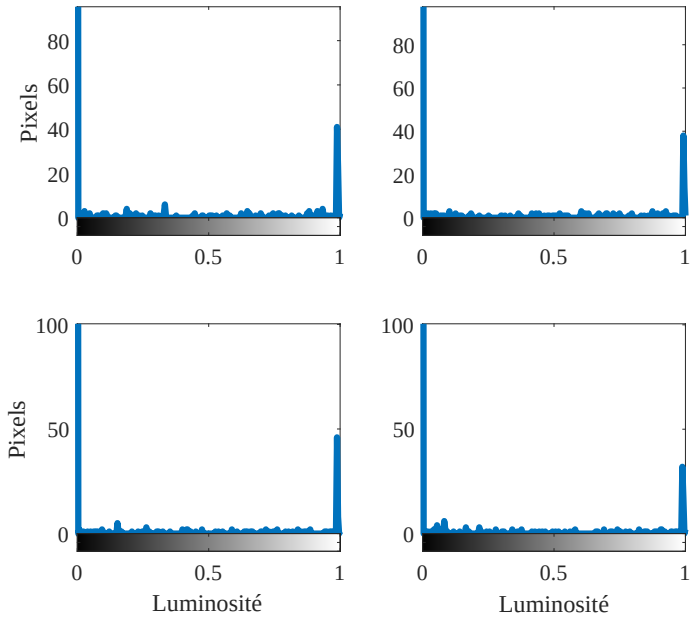


FIGURE 7.5 – Histogrammes de 4 images de chiffres manuscrits.

Synthèse : Le nombre de coupes influence sur la qualité de la CDL et en conséquence sur la performance du classfieur. L'augmentation de nombre de coupes permet d'augmenter la durée de simulation d'une manière linéaire. Le nombre de coupes dépend des histogrammes des images à classifier. Donc, le choix du nombre de coupes est d'une importance majeure pour aboutir à une précision maximale toute en optimisant la durée de simulation.

b) Les seuils

Un deuxième paramètre qui pourra influencer sur la qualité de la CDL pour les images en niveaux de gris est le seuil de chaque coupe. Dans la définition de la CDL, pour déterminer les seuils, nous avons choisi de prendre des seuils régulièrement espacés entre 0 et le maximum des deux images. Par contre cette stratégie ne prend pas en compte la luminosité des deux images. Une deuxième stratégie consiste à prendre des seuils non régulièrement espacés, mais en fonction des histogrammes des deux images. Pour ce faire, nous avons développé l'algorithme 7.1 qui prend comme arguments d'entrées deux images et le nombre de coupes souhaité, et qui permet de fournir les seuils adéquats en fonction de ces arguments. L'algorithme est ensuite introduit dans le programme de la CDL. Pour évaluer les performances de la CDL avec cette nouvelle stratégie dans le choix des seuils, le classifieur k-PPV est utiliser avec la base de données des chiffres manuscrits. Les performances de CDL ont été aussi évaluées en fonction de nombre de coupes. La figure 7.6 représente la précision obtenue par le classifieur et la durée de simulation en fonction de nombre de coupes.

On remarque que les courbes ont suivis la même tendance que celles obtenus avec des seuils régulièrement espacés (Figure 7.4). Par contre, dans ce cas la précision du classifieur est variée entre 94% et 99%. La meilleure précision de 99% est obtenue aussi pour une seule coupe.

Synthèse : l'utilisation des seuils non régulièrement espacés, nous a permet d'améliorer la précision du classifieur avec 1%. Par contre, la durée de simulation a légèrement augmenté. Le choix de nombre de coupes est toujours d'une importance majeure, quelle que soit la stratégie adaptée pour les seuils.

7.5.2 CDL contre la distance euclidienne et la SSIM

L'objectif de cette partie est de comparer les performances de la CDL avec celles de la distance euclidienne et la SSIM. Les trois mesures sont utilisées comme distances pour le classifieur k-PPV. Une seule coupe avec des seuils non régulièrement espacés est utilisée pour la CDL. Les distances sont définies par :

$$d_{\mathrm{CDL}}(A,B) = ||\mathrm{CDL}(A,B)||_2 \tag{7.8}$$

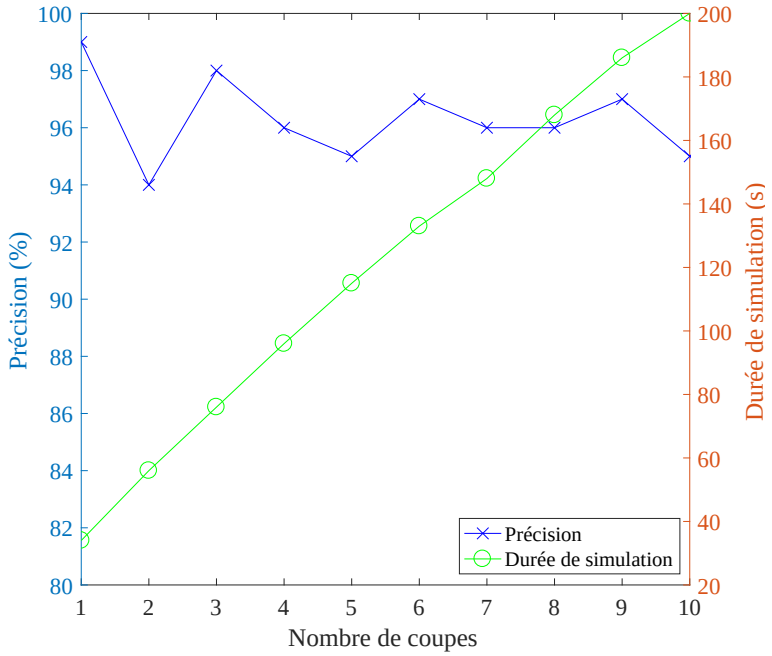


FIGURE 7.6 – Précision et durée de la simulation en fonction de nombre de coupes de la CDL. Seuils non régulièrement espacée. Performances évaluées par le classifieur k-PPV.

$$d_{\text{Euclidienne}}(A, B) = ||A - B||_2 \tag{7.9}$$

$$d_{\text{SSIM}}(A, B) = 1 - \text{SSIM}(A, B) \tag{7.10}$$

Deux types de comparions sont étudiées. La première consiste à comparer les performances des trois distances en fonction de nombre de classes de la base de données d'apprentissage. La deuxième comporte sur la comparaison des mêmes distances en fonction de nombre d'individus dans la base de données d'apprentissage. La base de données MNIST est utilisée comme application. Les performances du classifieur k-PPV sont mesurées par la précision.

a) Premier type de comparaison

Pour réaliser cette comparaison, deux bases de données constituées de 10 classes sont utilisées pour l'apprentissage et le test. Les classes de la base d'apprentissage contiennent 100 individus et ceux de la base de test 10 individus. La méthode de k-PPV est évaluée pour chaque distance en fait varié le nombre de classes dans la base d'apprentissage et de test de 2 à 10 avec un pas de 1. En effet, le problème à deux classes contient les chiffres (1 et 2), celui à trois classes les chiffres (1, 2 et 3) et ainsi de suite. Les figures 7.7 et 7.8

représentent les résultats obtenus par les trois distances. Les courbes de ces figures représentent les précisions et la durée de la simulation en fonction de nombre de classes, respectivement.

À partir de la figure 7.7, on remarque que, pour les trois courbes, la précision est une fonction décroissante du nombre de classes. La pente de décroissance de la courbe obtenue par la CDL est beaucoup moins importante à celles obtenues par la distance euclidienne et la SSIM. Les performances du classifieur k-PPV basée sur la CDL commencent à se dégrader pour un nombre de classes supérieur à 6. Par contre la dégradation des performances du même classifieur basé sur la distance euclidienne et la SSIM débute pour un nombre de classes supérieur à 2 et supérieur à 1, respectivement. Les précisions obtenues par le classifieur k-PPV basé sur la CDL, la distance euclidienne et la SSIM variées en fonction de nombre de classes dans les intervalles [98 ; 100], [85 ; 100] et [85 ; 100], respectivement. Pour un problème de classification à 2 classes, une précision de 100% est obtenue par le classifieur pour les trois distances. À partir de 3 classes, les meilleures précisions du classfieur sont obtenues avec la CDL comme distance.

À partir de la figure 7.8, on constate que pour les trois courbes, la durée de la simulation est une fonction croissante du nombre de classes. La courbe obtenue par la SSIM est une fonction exponentielle. Les courbes obtenues par la CDL et la distance euclidienne sont des fonctions linéaires. Par contre, la pente de la courbe obtenue par la CDL est très importante par rapport à celle de la distance euclidienne. Cette différence entre les trois méthodes est due à leurs complexités.

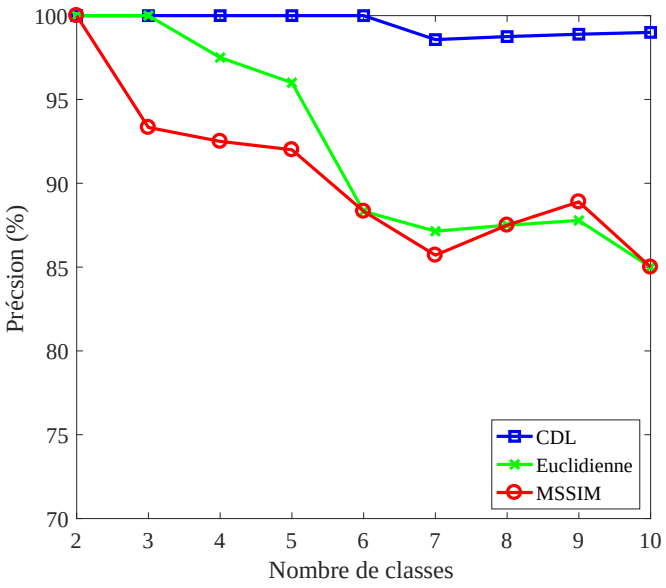


FIGURE 7.7 – Précision en fonction de nombre de classes obtenus par les trois distances (CDL, Euclidienne et SSIM). Performances évaluées par le classifieur k-PPV.

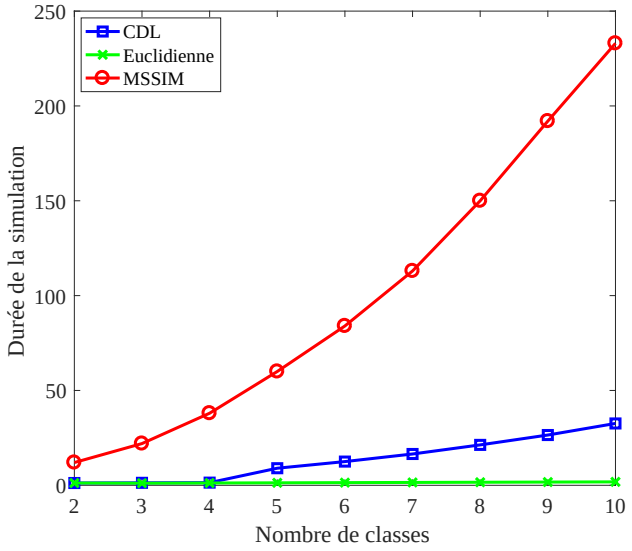


FIGURE 7.8 – Durée de la simulation en fonction de nombre de classes obtenus par les trois distances (CDL, Euclidienne et SSIM). Performances évaluées par le classifieur k-PPV.

b) Second type de comparaison

Pour effectuer cette comparaison, deux bases de données constituées de 10 classes sont utilisées pour l'apprentissage et le test. Pour mesurer les performances des trois distances, le classifieur k-PPV est évalué en fonction de nombre d'individus dans la base de données d'apprentissage (10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400 et 500 individus dans chaque classe) pour une base de données de test fixe (10 individus dans chaque classe). Les figures 7.9 et 7.10 représentent les résultats obtenus par les trois distances. Les courbes de ces figures représentent les précisions et la durée de la simulation en fonction de nombre d'individus.

D'après la figure 7.9, on constate que les précisions obtenues par le classifieur sont des fonctions croissantes du nombre d'individus. Donc, l'enrichissement de la base d'apprentissage permet d'améliorer les capacités discriminatives du classifeur. On remarque que la courbe de la précision obtenue par le classifieur en utilisant la CDL comme distance, convergent rapidement par rapport à la distance euclidienne et la SSIM. En utilisant la CDL, le classifeur k-PPV atteint sa précision maximale de 99% pour seulement 90 individus dans chaque classe de la base d'apprentissage. Par contre, en utilisant la distance euclidienne et la SSIM, le même classifieur atteint ses précisions maximales (90% et 94%) pour 300 individus dans chaque classe de la base d'apprentissage.

À partir de la figure 7.10, on constate que pour les trois courbes, la durée de la simulation est une fonction linéaire et croissante du nombre d'individus. La pente de la courbe obtenue par la SSIM est très importante par rapport à celles de la distance euclidienne et la CDL. Cette différence entre les trois méthodes est due à leurs complexités.

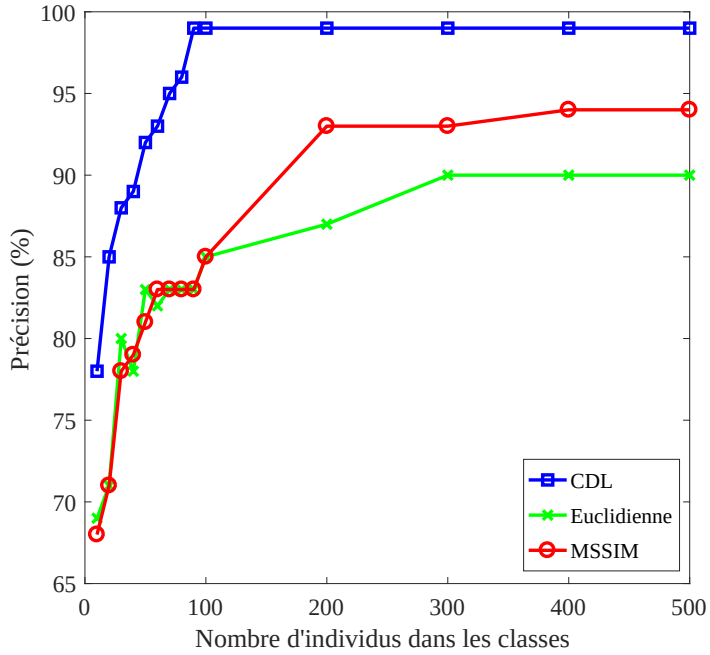


FIGURE 7.9 – Précision en fonction de nombre d'individus obtenus par les trois distances (CDL, Euclidienne et SSIM). Performances évaluées par le classifieur k-PPV.

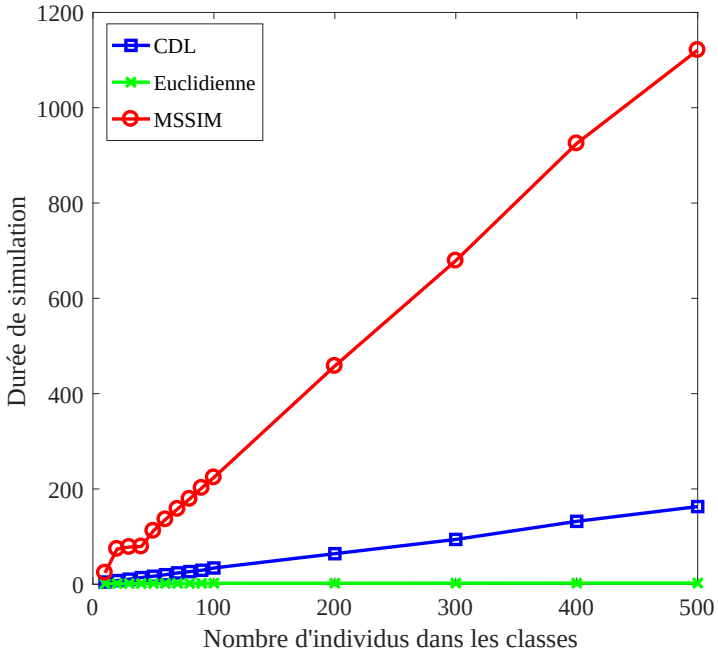


FIGURE 7.10 – Durée de la simulation en fonction de nombre d'individus obtenus par les trois distances (CDL, Euclidienne et SSIM). Performances évaluées par le classifieur kPP.

c) Synthèse :

L'augmentation du nombre de classes de la base d'apprentissage n'a pas beaucoup d'influence sur les capacités discriminatives de la CDL. Les meilleures précisions du classifieur k-PPV sont obtenues avec la CDL comme distance. L'enrichissement de la base d'apprentissage en termes de nombre d'individus permet d'améliorer les capacités discriminative du classifieur k-PPV. En utilisant la CDL comme distance, le classifieur k-PPV convergent rapidement. La complexité de la CDL augmente la durée de la simulation d'une manière linéaire en fonction de nombre de classe et d'individus dans la base d'apprentissage.

7.5.3 Classification dans l'espace de dissimilarités

Le premier objectif de cette partie est d'étudier les performances du classifieur RNA implémenté dans l'espace de dissimilarités. Le second objectif est de comparer entre les performances de ce classifieur avec celles du classifieur k-PPV basée sur la dissimilarité comme distance. Pour réaliser ces deux objectifs, la base de données des chiffres manuscrits est utilisée comme application. Les classifieurs sont construites sur différentes base de données d'apprentissages (de 10 à 100 avec un pas de 10 et de 100 à 500 avec un pas de 100 individus dans chaque classe) et leurs performances sont mesurées par la précision sur une base de test de 10 individus dans chaque classe. La CDL, la distance euclidienne et la SSIM sont utilisés comme mesures de dissimilarités. Les résultats obtenus sont présentés sur les figures de 7.11 et 7.12.

En analysant les résultats présentés sur les figures 7.11 et 7.12, on constate que le classifeur RNA atteint une précision maximale de 98% pour toutes les mesures de dissimilarités. Par contre, le classifieur converge rapidement en utilisant la CDL comme mesure de dissimilarité. Le temps de simulation augmente en fonction de nombre d'individus dans la base d'apprentissage. Cette augmentation est sous la forme d'une fonction exponentielle.

Le tableau 7.1 résume les performances en termes de précisions du classifieur RNA implémenté dans l'espace de dissimilarités et du classifieur k-PPV basé sur la dissimilarité comme distance. Une base de données de 1000 individus dans chaque classe est utilisée pour l'apprentissage et une base de données de 100 individus dans chaque classe est utilisée pour l'évaluation des performances des deux classifieurs. Les résultats indiquent que les meilleures précisions sont obtenus par le classifieur RNA quels que soit le mesure de dissimilarité. Donc, l'espace de dissimilarité permet aux classifieurs d'affranchir les meilleures performances. En revanche la meilleure précision est obtenue par le classifieur RNA implémenté dans l'espace de dissimilarités engendré par la CDL. Par contre, le fait d'implémenté un classifieur dans un espace de dissimilarité augmente le temps de simulation.

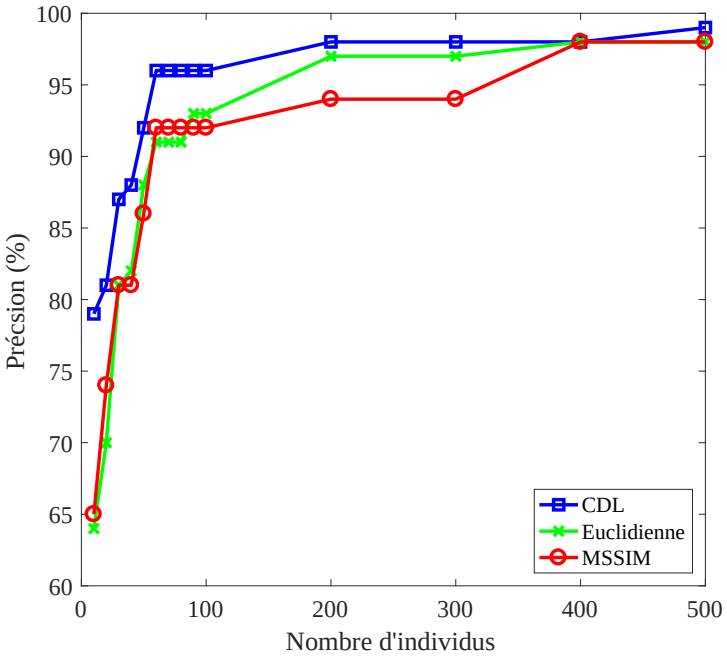


FIGURE 7.11 – Précision en fonction de nombre d'individus obtenus par les trois distances (CDL, Euclidienne et SSIM). Performances évaluées par le classifieur RNA basée sur la dissimilarité comme descripteur.

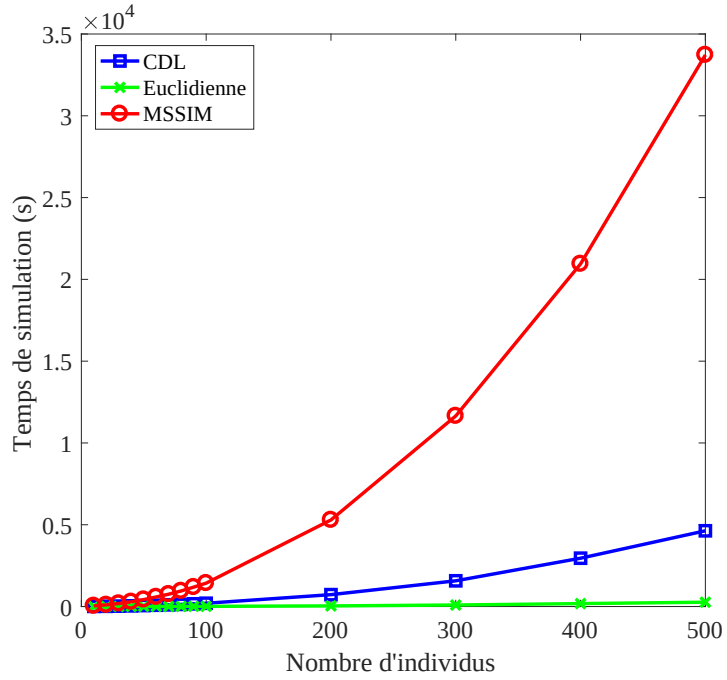


FIGURE 7.12 – Durée de la simulation en fonction de nombre d'individus obtenus par les trois distances (CDL, Euclidienne et SSIM). Performances évaluées par le classifieur RNA basée sur la dissimilarité comme descripteur.

Tableau 7.1 – Précisons et durées de simulation obtenues par le RNA et le k-PPV. Application sur la base de données de chiffre manuscrit.

	Précision (%)			Temps de simulation (s)		
	CDL	Euclidienne	SSIM	CDL	Euclidienne	SSIM
RNA	97,4	95,9	95,5	19721	1371	207676
k-PPV	96,8	95,1	93,2	3470	23	26190

7.5.4 Autre application : classification d'une base de données de lettrines

Les capacités discriminatives de la CDL ont été évaluées sur une base de données de chiffre manuscrit. La particularité de cette base est la simplicité de la forme des images. Sur cette base de données, les précisions obtenues par le classifieur RNA sont satisfaisantes quels que soit la mesure de dissimilarité. Pour mettre en évidence les avantages de la CDL, évaluer ces performances sur une deuxième base de données plus complexe est nécessaire. Pour ce faire, une base de données de lettrines comportant 13 lettres est utilisée. Cette base comportant donc 13 classes et chacune contient 50 lettrines. Un exemple de lettrine de chaque classe est présenté sur la figure 7.13.

Pour l'apprentissage et le test des classifieurs, la technique de validation croisée *10-fold* est utilisée. La performance de chaque classifieur est mesurée par la précision. Deux mesures de dissimilarités sont comparées, la CDL et la distance euclidienne. Les résultats obtenus sont présentés sur le tableau 7.2. Ainsi, on remarque que les meilleures performances des deux classifierus sont obtenues en utilisant la CDL comme mesure de dissimilarité. Ces résultats confirment les avantages de la CDL, malgré la complexité et pauvreté de la base

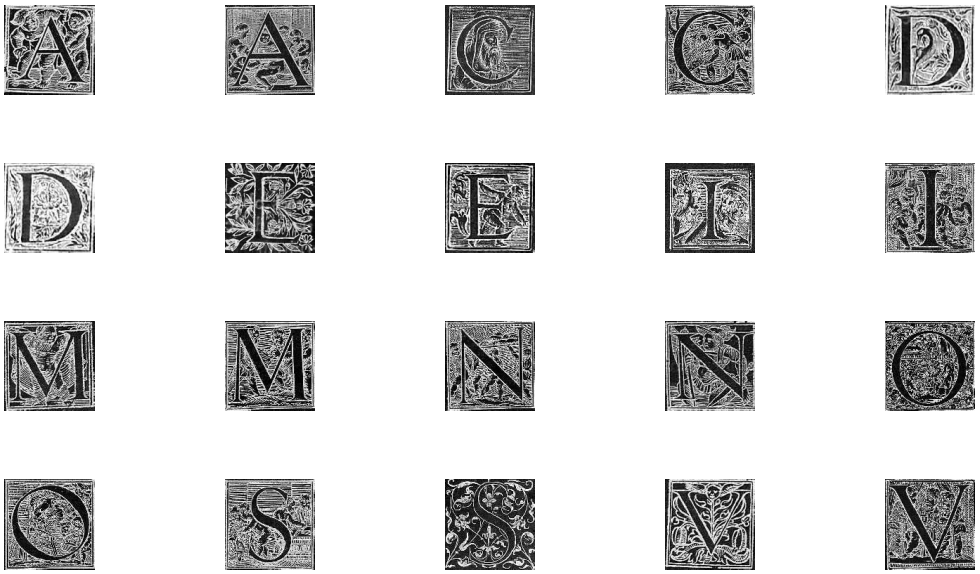


FIGURE 7.13 – Quelques exemples de lettrines de la base de données

Tableau 7.2 – Précisons et durées de simulation obtenues par le RNA et le k-PPV. Application sur la base de données de lettrines.

	Précision (%)	
	CDL	Euclidienne
RNA	86,60	75,00
k-PPV	85,40	40,60

de données en termes de nombre d'individus.

7.6 Conclusion

L'objectif de ce chapitre a été la mise en place d'une méthode de classification des images basée sur la CDL et sans exploitation de descripteurs. Deux méthodes de classification ont été exploitées pour mesurer les capacités discriminatives de la CDL (RNA et k-PPV). Les performances en terme de précision de la CDL ont été comparées avec celles de la distance euclidienne et la SSIM. Les résultats obtenus montrent que les méthodes de classification basées sur la CDL convergent plus rapidement avec une précision de 97,40% pour la méthode RNA-CDL.

Les résultats obtenus permettent de vérifier les capacités discriminatives de la CDL. L'avantage de la CDL par rapport aux autres distances est sa capacité à converger rapidement vers sa précision maximale. Cette rapidité permet à la CDL d'obtenir de bons résultats de classification même pour de faibles bases d'apprentissage (lettrines). Cet avantage permettra à la CDL de résoudre des problèmes de classification dans des domaines où il y a très peu d'exemples d'apprentissage, comme le domaine de la médecine.

Les méthodes de classifications basées sur une mesure de dissimilarité permettent d'exploiter la totalité de l'information présente dans les images. Ces méthodes permettent aussi d'éliminer la tâche d'extraction de caractéristiques sur laquelle les méthodes classiques sont basées.

On note que différentes méthodes ont été exploitées pour la classification de la base de données MNIST. La meilleure précision a été de l'ordre de 99,8% pour la base d'apprentissage complète. On note aussi que le travail présenté dans ce chapitre a pour principal objectif la proposition d'une mesure de dissimilarité qui permet à un classifieur implémenté dans l'espace de dissimilarité de converger le plus rapidement possible vers la précision maximale. Cet avantage, va nous permettre d'appliquer la solution de sélection de prototype afin de résoudre le problème de complexité de l'approche de classification basée sur une mesure de dissimilarité. De plus, cet avantage va rendre aussi la résolution d'un problème de classification disposant d'une petite base de données d'apprentissage possible et avec une précision acceptable.

Conclusion générale

Ce travail doctoral constitue une contribution à la compréhension des mécanismes d'endommagement de matériaux composites à fibres naturelles. En effet, le contrôle non destructif par émission acoustique (EA) est un outil efficace pour détecter les mécanismes d'endommagement en fonction des contraintes mécaniques. Généralement, les signaux d'EA enregistrés lors de la défaillance d'un matériau composite proviennent de plusieurs sources d'endommagement (mécanismes liés à la rupture des fibres, à la fissuration de la matrice et ceux couplés aux fibres et à la matrice). Par conséquent, l'identification des sources d'endommagement basée sur un descripteur unique (par exemple l'amplitude ou la fréquence du pic) du signal n'est pas possible et les techniques d'analyse statistique multivariable sont plutôt souhaitables. Ainsi, une méthode appropriée pour distinguer les différents types de signaux d'EA est l'application des techniques de reconnaissance de formes basées sur les descripteurs. Ces techniques peuvent être divisées en deux catégories de méthodes de classification, à savoir les méthodes non-supervisées et supervisées. Les méthodes non-supervisées consistent à regrouper les signaux d'EA, en se basant sur leurs descripteurs, en plusieurs groupes homogènes. Cette dernière est liée aux conditions initiales (le centre de chaque groupe) et le choix du nombre de groupe. Au-delà de ce problème d'homogénéité, ces méthodes non-supervisées nécessitent une étape supplémentaire afin de corréler chaque groupe avec un mécanisme d'endommagement. Cette corrélation se fait la plupart du temps en se basant sur les résultats de la littérature, en particulier sur les gammes d'amplitude. Les méthodes supervisées quant à elles consistent à identifier un signal inconnu en s'appuyant sur une base de données d'apprentissage pour construire la règle de décision. Cependant, la réussite de ces méthodes supervisées est liée à la qualité de la base de données d'apprentissage. Toutefois, un point commun entre les méthodes de classification non-supervisées et supervisées est que leurs fiabilités sont fortement liées à la pertinence des descripteurs.

De ce fait, l'objectif de ce projet de thèse était de proposer des solutions pour améliorer les démarches actuelles de classification de signaux d'EA. Ceci permet de mieux identifier les mécanismes d'endommagement afin de suivre la dégradation des matériaux en service et d'évaluer efficacement leur état de santé.

Pour ce faire, nous avons proposé dans une première étape de construire une base de données d'apprentis-

sage de signaux d'EA relatif aux matériaux composites. Ce qui permet d'exploiter les méthodes de classification supervisée. Cette étape était l'objet de la première partie de ce manuscrit et avait pour but de remédier aux problèmes liés aux méthodes de classification non-supervisée. Dans une seconde étape, pour illustrer le problème lié à la pertinence des descripteurs, une approche de classification de signaux non-stationnaires sans extraction de descripteurs a été développée. Étant donné que nous avons construit nous-mêmes la base de données d'apprentissage de signaux d'EA, il était judicieux de valider notre approche de classification sur une base de données connue au grand public. Ainsi, cette seconde étape était l'objet de la deuxième partie de ce manuscrit. Au regard des différents résultats obtenus, plusieurs conclusions et/ou enseignements ont été mis en évidence. Ceux-ci sont décrits ci-dessous en distinguant les deux parties du manuscrit.

Première partie : Discrimination des mécanismes d'endommagement de matériaux composites à fibres végétales par une classification de signaux d'émissions acoustiques.

Nous avons montré dans un premier temps le potentiel des matériaux composites à fibres végétales. En effet, ces matériaux constituent une alternative prometteuse aux composites conventionnels grâce à leurs propriétés mécaniques spécifiques intéressantes. Toutefois, leur développement dans des applications industrielles nécessite une meilleure compréhension de leur processus d'endommagement afin de mieux prédire leur durée de vie. À ce sujet, le contrôle non destructif par la technique d'EA couplé avec une méthode de classification permet d'apporter des informations pertinentes à ce sujet. Ainsi, l'étude bibliographique menée a porté sur les méthodes qui ont été utilisées jusqu'à présent pour classifier les signaux d'EA. Celle-ci était principalement dédiée aux matériaux composites à fibres usuels et/ou végétales. À l'issu de cette étude, trois grands problèmes liés aux méthodes actuelles de classification de signaux d'EA ont été soulignés. Le premier est lié à l'étape d'identification des groupes exigés par les classifications non-supervisées. Le deuxième est relatif à la pertinence de quelques descripteurs utilisés pour caractériser les signaux d'EA. Le dernier problème réside dans le fait qu'il n'existe pas une base de données d'apprentissage de signaux d'EA qui rend possible l'utilisation des méthodes supervisées.

En l'occurrence, ces problématiques, et surtout la dernière, nous ont conduits à mener une étude expérimentale complète pour recueillir les données indispensables. Les matériaux composites de l'étude et leurs caractéristiques ont d'abord été présentés. Il s'agit de deux matériaux composites unidirectionnels (UD) à fibres de lin et de verre et un composite UD hybride lin-verre. Ces matériaux sont testés mécaniquement afin de suivre la chronologie de leurs mécanismes d'endommagement à l'aide des capteurs d'émission acoustique. Des tests spécifiques nous ont permis de construire des bases de données de signaux d'EA relatifs aux matériaux composites non-hybrides (le composite à fibres de lin et celui à fibre de verre). Ensuite, le principe d'une nouvelle méthode de classification de signaux d'EA, nommée méthode incrémentale, a été proposé. Cette dernière a pour but de construire une base de données d'apprentissage de signaux d'EA. Les signaux de chaque

sous-ensemble de cette base de données doivent être en correspondance avec un seul mécanisme d'endommagement. Le principe de la méthode incrémentale est basé sur la superposition des signaux à classifier et le principe de voisinage. Sa mise en œuvre nécessite l'utilisation des bases de données de signaux d'EA relatif à des matériaux composites spécifiques. Il s'agit de la résine époxyde seule et un composite UD sollicités à 0° (UD0) et 90° (UD90) par rapport à l'axe des fibres. La spécificité de ces derniers est la connaissance préalable des mécanismes d'endommagement qui peuvent être distingués lors d'essais mécaniques spécifiques. Ainsi, la construction de la base d'apprentissage nous a permis d'utiliser une méthode de classification supervisée et par conséquent remédier aux problèmes liés à l'utilisation des méthodes non-supervisées.

Enfin, nous avons procédé à une démarche pour sélectionner les descripteurs les plus pertinents parmi les plus couramment utilisés pour caractériser un signal d'EA. Cette démarche consiste à effectuer une analyse de corrélation entre 11 descripteurs fréquentiels et temporels afin de choisir les moins corrélés. Cette démarche nous a permis de sélectionner sept descripteurs (durée, énergie absolue, temps de montée, fréquence moyenne, fréquence de réverbération, fréquence d'initiation, fréquence centroïde) comme étant les plus pertinents. Les résultats ont montré aussi que la fréquence de pic (FP) ne permet pas de décrire significativement le contenu fréquentiel d'un signal d'EA et son utilisation peut influencer les résultats de classification. En raison de la nature non-stationnaire et multi-composants d'un signal d'EA, une faible valeur de FP ne signifie pas forcément qu'il n'y a pas de contenu à haute fréquence, et vice versa. Puis, la méthode incrémentale a été appliquée, sur les bases de données de signaux d'EA construites à partir des éprouvettes modèles UD. Cette application a ensuite permis de construire deux bases de données d'apprentissage, représentant distinctivement les endommagements des composites à fibres de verre et ceux des composites à fibres de lin. L'évaluation des performances du classifieur k-PPV par les courbes ROC a montré la fiabilité de chaque base de données d'apprentissage. Les comparaisons entre les signaux d'EA par paires d'endommagements entre les matériaux composites à fibres de lin et ceux à fibres de verre nous ont révélé que les caractéristiques de ces signaux sont différentes. Toutefois, la possibilité de la présence de signaux similaires n'est pas exclue. Ceci peut être justifié par le fait que les performances du classifieur k-PPV ont diminué quand les deux bases de données d'apprentissage sont combinées en une seule base de données globale. Afin de mettre en évidence les avantages de la méthode incrémentale, nous l'avons comparé avec la méthode de k-moyennes, couramment utilisée dans la littérature. Cette comparaison a montré les avantages et la fiabilité de la méthode incrémentale et les limites de celle des k-moyennes. Enfin, les bases de données d'apprentissage sont combinées avec le classifieur k-PPV et utilisées pour séparer les mécanismes d'endommagement dans des matériaux plus complexes comme les composites hybrides lin-verre. L'application de la méthode s'est avérée efficace pour détecter tous les types d'endommagement dans ces composites. Elle a permis de dissocier les endommagements de chacun des constituants, à savoir la matrice, les interfaces fibres de lin/matrice et fibres de verre/matrice,

les fibres de lin et de verre.

Pour conclure, nous avons développé un outil afin de construire des bases de données d'apprentissage de signaux d'EA à partir d'éprouvettes modèles et relatif aux matériaux composites. Ces bases rendent ensuite possible l'utilisation des méthodes de classification supervisée pour identifier les mécanismes d'endommagement intervenant au sein de matériaux composites plus complexes. L'implémentation de la méthode a été effectuée dans l'espace basé sur des caractéristiques temporelles et fréquentielles des signaux d'EA. Cependant, ces caractéristiques peuvent être s'avérer inefficace dans le cas des signaux non-stationnaires comme les signaux d'EA. Ainsi, les techniques d'analyse temps fréquence (ATF) sont plus judicieuses dans ce cas. De plus, dans le cas où les descripteurs ne sont pas pertinents, l'utilisateur ne peut pas définir des descripteurs d'une manière directe ou lorsque les descripteurs sont définis d'une manière arbitraire, dans ces cas, l'espace basé sur ces descripteurs ne peut être approprié pour résoudre des problèmes de classification. Ainsi, dans ce cas l'approche de classification basée sur l'espace de dissimilarité est plutôt préférable. Pour toutes ces raisons, l'idée de développer une nouvelle approche de classification de signaux non-stationnaires s'est précisée en se basant sur l'ATF et l'approche de classification basée sur une mesure de dissimilarité. Cette approche était l'objet de la deuxième partie du manuscrit et a été validée sur une base de données de signaux EEG d'épilepsie. Son application sur les signaux d'EA, qui n'a pas pu être réalisée, constitue une des perspectives de ces travaux de thèse.

Deuxième partie : Approche de classification basée sur une mesure de dissimilarité : application aux signaux EEG épileptique et aux images.

Dans le cadre des études et développements menés dans différents domaines pour la détection et de la reconnaissance des signaux non-stationnaires, cette partie avait pour but de définir et concevoir une nouvelle approche de classification de ces signaux. L'objectif est de faire en sorte que cette approche soit généralisable et facile à mettre en œuvre pour des utilisateurs qui ne sont pas forcément spécialistes du domaine. De plus, celle-ci doit améliorer la précision des classifieurs, ce qui nous a conduits vers une approche de classification basée sur une mesure de dissimilarité entre les représentations temps-fréquence (RTF) de signaux non-stationnaires.

Nous avons introduit dans un premier temps la représentation par une mesure de dissimilarité comme une alternative à l'utilisation de la représentation par caractéristiques dans la reconnaissance d'objets. Au lieu d'une caractérisation absolue des objets par un ensemble de caractéristiques, l'utilisateur est invité à définir une mesure qui estime la dissimilarité entre des paires d'objets. Les mesures de dissimilarité permettent de capturer les informations statistiques et structurelles des formes et par conséquent elles forment un pont naturel entre ces deux approches.

En l'occurrence, l'ATF a été présenté comme étant un outil puissant pour le traitement des signaux non-stationnaires pour lesquels l'analyse temporelle et fréquentielle séparément n'est pas adéquate. En effet, les

caractéristiques spectrales des signaux non stationnaires changent avec le temps. Ainsi, pour des problèmes de classification impliquant ces signaux, les représentations temps-fréquence (RTF) se sont révélées un outil précieux en raison de leur capacité à mettre en évidence et à décrire les caractéristiques spectrales du signal au cours du temps. En particulier, les RTF quadratiques ont la capacité de décrire comment l'énergie du signal est répartie dans le plan temps-fréquence, ce qui permet d'identifier des descripteurs pertinents pour la classification. De ce fait, nos études se sont orientées vers les RTF de classe de Cohen et en particulier la distribution de Wigner-Ville et ses variances.

Ensuite, le potentiel de notre approche a été mis en évidence à partir d'un problème de classification de signaux réels utilisés dans la détection de l'épilepsie. La base de données de ces signaux a été obtenue à partir d'enregistrement cliniques d'EEG sur à la fois des patients souffrants d'épilepsie (pendant les périodes d'activité et d'inactivité des crises) et des personnes saines. Cette base de données consiste en cinq ensembles différents de signaux EEG. De ces cinq ensembles, 13 problèmes de classification différents ont été définis, y compris des problèmes à 2, 3 et 5 classes. Ainsi, plusieurs expériences ont été menées à partir de ces problèmes de classification. Dans un premier temps, nous avons proposé une procédure pour faire un choix optimal d'une mesure de dissimilarité et d'une RTF adéquates aux signaux EEG. En effet, nous avons trouvé que la distribution SPWVD et la déviation spectral logarithmique d'ordre 1 (d_{DSL_1}) sont les plus adaptées. Ensuite, notre approche a été combinée avec plusieurs classifieurs comme le k-plus proches voisins, le réseau de neurones artificiels, l'analyse discriminante linéaire et les séparateurs à vaste marge. Puis, l'utilisation des courbes ROC et la précision ont permis d'évaluer les performances de ces classifieurs. Les résultats obtenus ont montré que cette approche permet à tous les classifieurs d'atteindre de bonnes précisions. La robustesse de l'approche proposée a été également étudiée avec l'ajout de bruit aux signaux EEG à différents rapports signal/bruits. Les résultats obtenus indiquent que l'approche proposée est robuste au bruit et ce dernier ne peut pas affecter son applicabilité à des cas réels. Enfin, les résultats obtenus par notre approche sur les 13 problèmes de classification ont été comparés avec ceux obtenus par les approches de la littérature sur les mêmes problèmes. Cette comparaison a montré que notre approche est un outil performant pour résoudre des problèmes de classification de signaux EEG. En effet, une précision de 98% est obtenue dans le problème de classification à 5 classes. Également, dans les problèmes de classification avec 2 et 3 classes, une précision très satisfaisante d'environ 100% est révélée.

Enfin, nous avons étudié, d'une manière expérimentale, la complexité en temps et en espace mémoire de l'approche de classification basée sur une mesure de dissimilarité entre les RTF. Les résultats obtenus montrent que cette approche peut augmenter la complexité en temps du classifieur utilisé. Ainsi, ce problème peut limiter l'applicabilité de cette approche aux problèmes de classification disposant d'une base de données d'apprentissage réduite. Une des solutions envisageables pour remédier à ce problème consiste à utiliser une

approche de sélection de prototypes afin de réduire le nombre d'objets dans l'ensemble de représentation. Cependant, le fait de travailler avec un nombre réduit de prototypes peut probablement dégrader les performances du classifieur. En effet, il est nécessaire d'envisager l'utilisation d'une mesure de dissimilarité avec une bonne propriété de compacité pour aider le classifieur à converger rapidement vers sa précision maximale. C'est ainsi cette idée que nous avons développée dans le dernier chapitre de cette partie. Nous avons d'abord élargi la recherche d'une mesure de dissimilarité au domaine de traitement d'image puis étudié la carte de dissimilarité locale (CDL) en raison de ses propriétés remarquables. Ainsi, nous avons proposé une approche pour généraliser la CDL aux images en niveaux de gris. Ensuite, nous avons déduit une mesure de dissimilarité à partir de la CDL. Enfin, en étudiant les capacités de cette mesure dans deux applications, les résultats expérimentaux ont montré les capacités discriminatives de la CDL.

Les travaux de recherche réalisés au cours de cette thèse ouvrent un certain nombre de perspectives, dont certaines pourraient faire l'objet de développement à courts et moyens termes. Nous en citons quelques-unes :

— Élargir la recherche d'une représentation temps-fréquence au-delà de classe de Cohen. Pour améliorer les performances (augmenter la résolution) des représentations temps-fréquence, de nombreuses méthodes peuvent être trouvées dans la littérature, à savoir la méthode de la réallocation. Cette piste est envisageable afin d'améliorer la classification de signaux EEG d'épilepsie.

— Élargir la recherche d'une mesure de dissimilarité au-delà des dissimilarités métriques. Des mesures non métriques, asymétriques par exemple, sont plus fidèles au jugement humain de la similarité. Ainsi, l'emploi de mesures non métriques peut ainsi permettre des performances accrues par rapport aux mesures métriques.

— Étudier la possibilité d'adapter la carte de dissimilarité locale pour comparer les représentations temps-fréquence. Cette idée consiste à considérer les représentations temps-fréquence comme des images et étudier la possibilité de les comparer par la carte de dissimilarité locale.

— L'idéal pour résoudre un problème de classification dans un environnement non-stationnaire peut être donc une fertilisation croisée de l'approche basée sur une mesure de dissimilarité, l'analyse temps-fréquence et la carte de dissimilarité locale avec la méthode de sélectionne de prototype.

— Étendre l'application de l'approche basée sur une mesure de dissimilarité entre les représentations temps-fréquence à d'autres domaines et d'autres signaux non-stationnaires. En effet, les signaux non-stationnaires sont recueillis dans de nombreuses domaines, tels que le diagnostic des maladies (signaux EEG et ECG), l'analyse des signaux sismiques, l'analyse des signaux sonores, etc. Ces domaines peuvent être un terrain favorable pour mettre en évidence davantage l'efficacité de cette approche de classification et la possibilité de sa généralisation.

— Combiner l'approche de classification basée sur une mesure de dissimilarité avec la méthode incrémen-

tale pour construire les bases de données d'apprentissage de signaux d'émission acoustique. Cette idée consiste à implémenter la méthode incrémentale dans un espace de dissimilarités.

— Par analogie de l'idée précédente, la méthode de k-moyennes pourrait aussi combiner avec l'approche de classification basée sur une mesure de dissimilarités pour le regroupement de signaux d'émission acoustique. Ce qui permettra probablement de résoudre le problème du choix des descripteurs pertinents.

Pour conclure, ce mémoire de thèse pose les bases d'une nouvelle approche de classification de signaux non-stationnaires. Cette approche combine l'analyse temps-fréquence et l'approche de classification basée sur une mesure de dissimilarité. Les résultats obtenus dans ce travail indiquent que cette nouvelle approche est un outil performant et peuvent être généralisés à de nombreuses applications.

Production scientifique

Articles de Revue

— **Y. Ech-Choudany**, M. Assarar, D. Scida, F. Morain-Nicolier, B. Bellach. Unsupervised clustering for building a learning database of acoustic emission signals to identify damage mechanisms in unidirectional laminates. Applied Acoustics, Volume 123, August 2017, Pages 123–132. **État : paru**

— **Y. Ech-Choudany**, D. Scida, J. Landré, M. Assarar, B. Bellach, F. Morain-Nicolier. Similarity-based time-frequency analysis as features for epileptic EEG signals classification, Biomedical Signal Processing and Control 2018. **État : soumis**

Conférences Internationales et Nationales

Orale

— **Y. Ech-Choudany**, B. Bellach, F. Morain-Nicolier. Discrimination des mécanismes d'endommagement de matériaux composites à fibres végétales par une classification des signaux d'émissions acoustiques. 21èmes Rencontres de la Société Francophone de Classification. 2014, Rabat, Maroc.

— **Y. Ech-Choudany**, M. Assarar, D. Scida, F. Morain-Nicolier, B. Bellach. Identification des mécanismes d'endommagement des composites à renfort végétal à l'aide d'une classification supervisée. 1ère Conférence EuroMaghrébine des BioComposites. 2016, Marrakech, Maroc.

— **Y. Ech-Choudany**, M. Assarar, D. Scida, B. Bellach, F. Morain-Nicolier. Identification des mécanismes d'endommagement des composites à fibre végétale à l'aide d'une nouvelle méthode de classification non supervisée. 1ère édition de la conférence internationale sur : LES MATÉRIAUX INNOVANTS & LEURS APPLICATIONS. 2016, Oujda, Maroc.

Poster

— **Y. Ech-Choudany**, F. Morain-Nicolier , J. Landré, B. Bellach , M. Assarar, D. Scida. Méthodes de reconnaissance de formes basées sur la carte de dissimilarité local pour la classification des images. GRETSI 2017. Juan-les-Pins, France.

Communications dans des colloques internes

Orale

— **Y. Ech-Choudany**, F. Morain-Nicolier. Identification des mécanismes d'endommagement des composites à renfort végétal à l'aide d'une méthode de classification à noyau temps-fréquence. Séminaire des doctorants du CReSTIC 2017. Reims, France.

— **Y. Ech-Choudany**, F. Morain-Nicolier. Approche de classification basée sur une mesure de dissimilaritée entre les distributions temps-fréquence : Application aux signaux EEG épileptiques. Séminaire des doctorants du CReSTIC 2018. Reims, France.

Poster

— **Y. Ech-Choudany**, B. Bellach. Discrimination des mécanismes d'endommagement de matériaux composites à fibres végétales. Doctoriales 2015. Oujda, Maroc.

— **Y. Ech-Choudany**, F. Morain-Nicolier. Identification des mécanismes d'endommagement des composites à renfort végétal à l'aide d'une classification supervisée. Séminaire des doctorants du CReSTIC 2016. Reims, France.

Bibliographie

[1] Ashby, MF. The CES edupack database of natural and man-made materials. *Cambridge, UK : Cambridge University and Granta Design*, 2008.

[2] Busnel, Frédéric. *Contribution à l'étude de matériaux composites à matrice organique renforcés par des fibres de lin : influence des traitements chimiques sur la liaison interfaciale fibre-matrice.* Thèse de doctorat, Lorient, 2006.

[3] Berthelot, Jean-Marie. Matériaux composites. *Edition TEC et DOC, Paris,*, 1999.

[4] Huguet, Stéphane. *Application de classificateurs aux données d'émission acoustique : identification de la signature acoustique des mécanismes d'endommagement dans les composites à matrice polymère.* Thèse de doctorat, Villeurbanne, INSA, 2002.

[5] Andrzejak, Ralph G., Lehnertz, Klaus, Mormann, Florian, Rieke, Christoph, David, Peter, et Elger, Christian E. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity : Dependence on recording region and brain state. *Phys. Rev. E*, 2001. vol. 64, p. 061907.

[6] Faruk, Omar, Bledzki, Andrzej K, Fink, Hans-Peter, et Sain, Mohini. Biocomposites reinforced with natural fibers : 2000–2010. *Progress in polymer science*, 2012. vol. 37, n° 11, p. 1552–1596.

[7] Yan, Libo, Chouw, Nawawi, et Jayaraman, Krishnan. Flax fibre and its composites–a review. *Composites Part B : Engineering*, 2014. vol. 56, p. 296–317.

[8] Chilali, Abderrazak. *Étude expérimentale et modélisation de la durabilité des biocomposites à fibres de lin.* Thèse de doctorat, Reims, 2017.

[9] Kotsikos, George, Evans, JT, Gibson, AG, et Hale, J. Use of acoustic emission to characterize corrosion fatigue damage accumulation in glass fiber reinforced polyester laminates. *Polymer composites*, 1999. vol. 20, n° 5, p. 689–696.

[10] Kim, Sang-Tae et Lee, You-Tae. Characteristics of damage and fracture process of carbon fiber reinforced plastic under loading-unloading test by using ae method. *Materials Science and Engineering : A*, 1997. vol. 234, p. 322–326.

[11] Chen, O, Karandikar, P, Takeda, N, et Kishi Rcast, T. Acoustic emission characterization of a glass-matrix composite. *Nondestructive testing and evaluation*, 1992. vol. 8, n° 1-6, p. 869–878.

[12] Ceysson, O, Salvia, M, et Vincent, L. Damage mechanisms characterisation of carbon fibre/epoxy composite laminates by both electrical resistance measurements and acoustic emission analysis. *Scripta Materialia*, 1996. vol. 34, n° 8, p. 1273–1280.

[13] Barré, Sébastien et Benzeggagh, ML. On the use of acoustic emission to investigate damage mechanisms in glass-fibre-reinforced polypropylene. *Composites Science and Technology*, 1994. vol. 52, n° 3, p. 369–376.

[14] Zhuang, Xingmin et Yan, Xiong. Investigation of damage mechanisms in self-reinforced polyethylene composites by acoustic emission. *Composites Science and Technology*, 2006. vol. 66, n° 3, p. 444–449.

[15] Berthelot, JM et Rhazi, J. Acoustic emission in carbon fibre composites. *Composites Science and Technology*, 1990. vol. 37, n° 4, p. 411–428.

[16] Pappas, Yiannis Z et Kostopoulos, Vassilis. Toughness characterization and acoustic emission monitoring of a 2-d carbon/carbon composite. *Engineering fracture mechanics*, 2001. vol. 68, n° 14, p. 1557–1573.

[17] Heidary, Hossein, Karimi, Navid Zarif, Ahmadi, Mehdi, Rahimi, Abdolreza, et Zucchelli, Andrea. Clustering of acoustic emission signals collected during drilling process of composite materials using unsupervised classifiers. *Journal of Composite Materials*, 2015. vol. 49, n° 5, p. 559–571.

[18] Li, Li, Lomov, Stepan V, Yan, Xiong, et Carvelli, Valter. Cluster analysis of acoustic emission signals for 2d and 3d woven glass/epoxy composites. *Composite Structures*, 2014. vol. 116, p. 286–299.

[19] Wang, Xu, Zhang, Hui-Ping, et Yan, Xiong. Classification and identification of damage mechanisms in polyethylene self-reinforced laminates by acoustic emission technique. *Polymer Composites*, 2011. vol. 32, n° 6, p. 945–959.

[20] Sause, MGR, Gribov, A, Unwin, AR, et Horn, S. Pattern recognition approach to identify natural clusters of acoustic emission signals. *Pattern Recognition Letters*, 2012. vol. 33, n° 1, p. 17–23.

[21] Masmoudi, Sahir, El Mahi, Abderrahim, et Turki, Saïd. Fatigue behaviour and structural health monitoring by acoustic emission of e-glass/epoxy laminates with piezoelectric implant. *Applied Acoustics*, 2016. vol. 108, p. 50–58.

[22] Mohammadi, Reza, Najafabadi, Mehdi Ahmadi, Saeedifar, Milad, Yousefi, Jalal, et Minak, Giangiacomo. Correlation of acoustic emission with finite element predicted damages in open-hole tensile laminated composites. *complexity*, 2016. vol. 4, p. e10.

[23] Fotouhi, Mohamad, Sadeghi, Seyedali, Jalalvand, Meisam, et Ahmadi, Mehdi. Analysis of the damage mechanisms in mixed-mode delamination of laminated composites using acoustic emission

data clustering. *Journal of Thermoplastic Composite Materials*, 2017. vol. 30, n° 3, p. 318–340.

[24] Oskouei, Amir Refahi, Heidary, Hossein, Ahmadi, Mehdi, et Farajpur, Mehdi. Unsupervised acoustic emission data clustering for the analysis of damage mechanisms in glass/polyester composites. *Materials & Design*, 2012. vol. 37, p. 416–422.

[25] De Oliveira, R et Marques, AT. Health monitoring of frp using acoustic emission and artificial neural networks. *Computers & structures*, 2008. vol. 86, n° 3, p. 367–373.

[26] Al-Jumaili, Safaa Kh, Holford, Karen M, Eaton, Mark J, McCrory, John P, Pearson, Matthew R, et Pullin, Rhys. Classification of acoustic emission data from buckling test of carbon fibre panel using unsupervised clustering techniques. *Structural Health Monitoring*, 2015. vol. 14, n° 3, p. 241–251.

[27] Huguet, Stéphane, Godin, Nathalie, Gaertner, Roger, Salmon, L, et Villard, D. Use of acoustic emission to identify damage modes in glass fibre reinforced polyester. *Composites Science and Technology*, 2002. vol. 62, n° 10, p. 1433–1444.

[28] Kempf, M, Skrabala, O, et Altstädt, V. Acoustic emission analysis for characterisation of damage mechanisms in fibre reinforced thermosetting polyurethane and epoxy. *Composites Part B : Engineering*, 2014. vol. 56, p. 477–483.

[29] Pashmforoush, Farzad, Fotouhi, Mohamad, et Ahmadi, Mehdi. Acoustic emission-based damage classification of glass/polyester composites using harmony search k-means algorithm. *Journal of Reinforced Plastics and Composites*, 2012. vol. 31, n° 10, p. 671–680.

[30] El Mahi, Abderrahim, Salem, Imen Ben, Assarar, Mustapha, Berbaoui, Rachid, Poilane, Chritophe, et El Guerjouma, Rachid. Analyse par émission acoustique de l'endommagement des matériaux éco-composites. Dans *10ème Congrès Français d'Acoustique*. 2010 p. –.

[31] Chilali, Abderrazak, Zouari, Wajdi, Assarar, Mustapha, Kebir, Hocine, et Ayad, Rezak. Effect of water ageing on the load-unload cyclic behaviour of flax fibre-reinforced thermoplastic and thermosetting composites. *Composite Structures*, 2018. vol. 183, p. 309–319.

[32] Habibi, Mohamed, Laperriere, Luc, Lebrun, Gilbert, et Toubal, Lotfi. Combining short flax fiber mats and unidirectional flax yarns for composite applications : Effect of short flax fibers on biaxial mechanical properties and damage behaviour. *Composites Part B : Engineering*, 2017. vol. 123, p. 165–178.

[33] Assarar, M, Scida, D, El Mahi, A, Poilâne, C, et Ayad, R. Influence of water ageing on mechanical properties and damage events of two reinforced composite materials : Flax–fibres and glass–fibres. *Materials & Design*, 2011. vol. 32, n° 2, p. 788–795.

[34] Saidane, El Hadi, Scida, Daniel, Assarar, Mustapha, et Ayad, Rezak. Damage mechanisms assess-

ment of hybrid flax-glass fibre composites using acoustic emission. *Composite Structures*, 2017. vol. 174, p. 1–11.

[35] Monti, Arthur, El Mahi, Abderrahim, Jendli, Zouhaier, et Guillaumat, Laurent. Mechanical behaviour and damage mechanisms analysis of a flax-fibre reinforced composite by acoustic emission. *Composites Part A : Applied Science and Manufacturing*, 2016. vol. 90, p. 100–110.

[36] Assarar, Mustapha, Scida, Daniel, Zouari, Wajdi, Saidane, El Hadi, et Ayad, Rezak. Acoustic emission characterization of damage in short hemp-fiber reinforced polypropylene composites. *Polymer Composites*, 2014.

[37] Saidane, El Hadi. *Etude de l'effet du vieillissement hygrothermique sur le comportement mécanique d'éco-composites à renfort hybride*. Thèse de doctorat, Reims, 2015.

[38] John, George H, Kohavi, Ron, Pfleger, Karl, *et al.* Irrelevant features and the subset selection problem. Dans *Machine learning : proceedings of the eleventh international conference*. 1994 p. 121–129.

[39] Ng, Michael K. A note on constrained k-means algorithms. *Pattern Recognition*, 2000. vol. 33, n° 3, p. 515–519.

[40] Gurunathan, T, Mohanty, Smita, et Nayak, Sanjay K. A review of the recent developments in biocomposites based on natural fibres and their application perspectives. *Composites Part A : Applied Science and Manufacturing*, 2015. vol. 77, p. 1–25.

[41] Bourmaud, Alain, Morvan, Claudine, Bouali, Anis, Placet, Vincent, Perre, Patrick, et Baley, Christophe. Relationships between micro-fibrillar angle, mechanical properties and biochemical composition of flax fibers. *Industrial Crops and Products*, 2013. vol. 44, p. 343–351.

[42] Baley, Christophe et Bourmaud, Alain. Average tensile properties of french elementary flax fibers. *Materials Letters*, 2014. vol. 122, p. 159–161.

[43] Gourier, Clément, Le Duigou, Antoine, Bourmaud, Alain, et Baley, Christophe. Mechanical analysis of elementary flax fibre tensile properties after different thermal cycles. *Composites Part A : Applied Science and Manufacturing*, 2014. vol. 64, p. 159–166.

[44] Gning, PB, Liang, S, Guillaumat, L, et Pui, WJ. Influence of process and test parameters on the mechanical properties of flax/epoxy composites using response surface methodology. *Journal of materials science*, 2011. vol. 46, n° 21, p. 6801–6811.

[45] Le Duigou, A, Davies, Peter, et Baley, C. Seawater ageing of flax/poly (lactic acid) biocomposites. *Polymer Degradation and Stability*, 2009. vol. 94, n° 7, p. 1151–1162.

[46] Cañigueral, N, Vilaseca, F, Méndez, JA, López, JP, Barberà, L, Puig, J, Pèlach, MA, et Mutjé, P. Behavior of biocomposite materials from flax strands and starch-based biopolymer. *Chemical*

Engineering Science, 2009. vol. 64, n° 11, p. 2651–2658.

[47] Alix, S, Marais, S, Morvan, C, et Lebrun, L. Biocomposite materials from flax plants : preparation and properties. *Composites Part A : Applied Science and Manufacturing*, 2008. vol. 39, n° 12, p. 1793–1801.

[48] Oksman, Kristiina, Mathew, Aji P, Långström, Runar, Nyström, Birgitha, et Joseph, Kuruvilla. The influence of fibre microstructure on fibre breakage and mechanical properties of natural fibre reinforced polypropylene. *Composites Science and Technology*, 2009. vol. 69, n° 11, p. 1847–1853.

[49] Sathasivam, Kathiresan, Mas Haris, Mas Rosemal Hakim, et Noorsal, Kartini. The preparation and characterization of esterified banana trunk fibers/poly (vinyl alcohol) blend film. *Polymer-Plastics Technology and Engineering*, 2010. vol. 49, n° 13, p. 1378–1384.

[50] John, Maya Jacob et Anandjiwala, Rajesh D. Chemical modification of flax reinforced polypropylene composites. *Composites Part A : Applied Science and Manufacturing*, 2009. vol. 40, n° 4, p. 442–448.

[51] Poilâne, Christophe, Vivet, Alexandre, Momayez, Ladan, Doudou, B Ben, Ayachi, M'hamed, et Chen, Jun. Propriétés mécaniques de préimprégnés lin/époxyde. *Comptes Rendus des JNC*, 2009. vol. 16.

[52] Le Duigou, Antoine, Davies, Peter, et Baley, Christophe. Interfacial bonding of flax fibre/poly (l-lactide) bio-composites. *Composites Science and technology*, 2010. vol. 70, n° 2, p. 231–239.

[53] Le Duigou, Antoine, Davies, Peter, et Baley, Christophe. Macroscopic analysis of interfacial properties of flax/plla biocomposites. *Composites Science and Technology*, 2010. vol. 70, n° 11, p. 1612–1620.

[54] Arbelaiz, Aitor, Fernández, Borja, Cantero, Guillermo, Llano-Ponte, Rodrigo, Valea, Angel, et Mondragon, Iñaki. Mechanical properties of flax fibre/polypropylene composites. influence of fibre/matrix modification and glass fibre hybridization. *Composites Part A : Applied Science and Manufacturing*, 2005. vol. 36, n° 12, p. 1637–1644.

[55] Jawaid, MHPS et Khalil, HPS Abdul. Cellulosic/synthetic fibre reinforced polymer hybrid composites : A review. *Carbohydrate polymers*, 2011. vol. 86, n° 1, p. 1–18.

[56] Ramesh, M, Palanikumar, K, et Reddy, K Hemachandra. Mechanical property evaluation of sisal–jute–glass fiber reinforced polyester composites. *Composites Part B : Engineering*, 2013. vol. 48, p. 1–9.

[57] Jarukumjorn, Kasama et Suppakarn, Nitinat. Effect of glass fiber hybridization on properties of sisal fiber–polypropylene composites. *Composites Part B : Engineering*, 2009. vol. 40, n° 7, p. 623–627.

[58] Valente, Marco, Sarasini, Fabrizio, Marra, Francesco, Tirillò, Jacopo, et Pulci, Giovanni. Hybrid recycled glass fiber/wood flour thermoplastic composites : Manufacturing and mechanical characterization. *Composites Part A : Applied Science and Manufacturing*, 2011. vol. 42, n° 6, p. 649–657.

[59] Davoodi, MM, Sapuan, SM, Ahmad, D, Ali, Aidy, Khalina, A, et Jonoobi, Mehdi. Mechanical properties of hybrid kenaf/glass reinforced epoxy composite for passenger car bumper beam. *Materials & Design*, 2010. vol. 31, n° 10, p. 4927–4932.

[60] Amico, SC, Angrizani, CC, et Drummond, ML. Influence of the stacking sequence on the mechanical properties of glass/sisal hybrid composites. *Journal of Reinforced Plastics and Composites*, 2010. vol. 29, n° 2, p. 179–189.

[61] Bono, Pierre, Le Duc, Anne, Lozachmeur, Marie, et Day, Arnaud. Matériaux : les nouveaux champs de recherche et développement pour la valorisation des fibres végétales techniques (lin fibres et chanvre). *OCL*, 2015. vol. 22, n° 6, p. D613.

[62] Auld, Bertram Alexander. *Acoustic fields and waves in solids*. Ripol Classic Publishing House, 1973.

[63] Beattie, AG. Acoustic emission, principles and instrumentation. *Journal of acoustic emission*, 1983. vol. 2, n° 12, p. 95–128.

[64] Ono, Kanji. Acoustic emission. *Encyclopedia of Acoustics, Volume Two*, 1997. p. 797–809.

[65] Prosser, William H. Advanced AE techniques in composite materials research. *Journal of Acoustic Emission*, 1996. vol. 14, n° 3-4, p. S1–S–11.

[66] Gautschi, Dip-Ing ETH Gustav. Piezoelectric sensors. Dans *Piezoelectric Sensorics*, p. 73–91. Springer, 2002.

[67] Ativitavas, Nat, Fowler, T, et Pothisiri, T. Acoustic emission characteristics of pultruded fiber reinforced plastics under uniaxial tensile stress. *Proceedings of European WG on AE*, 2004. p. 447–454.

[68] Godin, Nathalie, Huguet, Stéphane, et Gaertner, Roger. Integration of the kohonen's self-organising map and k-means algorithm for the segmentation of the ae data collected during tensile tests on cross-ply composites. *NDT & E International*, 2005. vol. 38, n° 4, p. 299–309.

[69] De Groot, Peter J, Wijnen, Peter AM, et Janssen, Roger BF. Real-time frequency determination of acoustic emission for different fracture mechanisms in carbon/epoxy composites. *Composites Science and Technology*, 1995. vol. 55, n° 4, p. 405–412.

[70] Giordano, M, Calabro, A, Esposito, C, D'amore, A, et Nicolais, L. An acoustic-emission characterization of the failure modes in polymer-composite materials. *Composites science and technology*, 1998. vol. 58, n° 12, p. 1923–1928.

[71] Arumugam, V, Sajith, S, et Stanley, A Joseph. Acoustic emission characterization of failure modes in gfrp laminates under mode i delamination. *Journal of Nondestructive Evaluation*, 2011. vol. 30, n° 3, p. 213–219.

[72] Ramirez-Jimenez, CR, Papadakis, N, Reynolds, N, Gan, TH, Purnell, P, et Pharaoh, M. Identification of failure modes in glass/polypropylene composites by means of the primary frequency

content of the acoustic emission event. *Composites Science and Technology*, 2004. vol. 64, n° 12, p. 1819–1827.

[73] Gutkin, R, Green, CJ, Vangrattanachai, S, Pinho, ST, Robinson, P, et Curtis, PT. On acoustic emission for failure investigation in cfrp : Pattern recognition and peak frequency analyses. *Mechanical Systems and Signal Processing*, 2011. vol. 25, n° 4, p. 1393–1407.

[74] Eaton, M, May, M, Featherston, C, Holford, K, Hallet, S, et Pullin, R. Characterisation of damage in composite structures using acoustic emission. Dans *Journal of Physics : conference series*, vol. 305. IOP Publishing, 2011 p. 012086.

[75] Sause, MGR et Horn, S. Simulation of acoustic emission in planar carbon fiber reinforced plastic specimens. *Journal of Nondestructive Evaluation*, 2010. vol. 29, n° 2, p. 123–142.

[76] Maillet, Emmanuel, Baker, Christopher, Morscher, Gregory N, Pujar, Vijay V, et Lemanski, Joseph R. Feasibility and limitations of damage identification in composite materials using acoustic emission. *Composites Part A : Applied Science and Manufacturing*, 2015. vol. 75, p. 77–83.

[77] Qi, Gang, Barhorst, Alan, Hashemi, Javad, et Kamala, Girish. Discrete wavelet decomposition of acoustic emission signals from carbon-fiber-reinforced composites. *Composites Science and Technology*, 1997. vol. 57, n° 4, p. 389–403.

[78] Oskouei, AR, Ahmadi, M, et Hajikhani, M. Wavelet-based acoustic emission characterization of damage mechanism in composite materials under mode i delamination at different interfaces. *Express Polym Lett*, 2009. vol. 3, n° 12, p. 804–813.

[79] Bak, K Mohamed, Kalaichelvan, K, et Arumugam, V. A novel approach for classification of failure modes in single lap joints using acoustic emission data. *Journal of Composite Materials*, 2014. vol. 48, n° 24, p. 3003–3017.

[80] Han, WenQin et Zhou, JinYu. Acoustic emission characterization methods of damage modes identification on carbon fiber twill weave laminate. *Science China Technological Sciences*, 2013. vol. 56, n° 9, p. 2228–2237.

[81] Ni, Qing-Qing et Iwamoto, Masaharu. Wavelet transform of acoustic emission signals in failure of model composites. *Engineering Fracture Mechanics*, 2002. vol. 69, n° 6, p. 717–728.

[82] Hamdi, Seif E, Le Duff, Alain, Simon, Laurent, Plantier, Guy, Sourice, Anthony, et Feuilloy, Mathieu. Acoustic emission pattern recognition approach based on hilbert–huang transform for structural health monitoring in polymer-composite materials. *Applied Acoustics*, 2013. vol. 74, n° 5, p. 746–757.

[83] Romhany, G, Karger-Kocsis, J, et Czigany, T. Tensile fracture and failure behavior of technical flax fibers. *Journal of Applied Polymer Science*, 2003. vol. 90, n° 13, p. 3638–3645.

[84] Boussetta, Hajer, Beyaoui, Moez, Laksimi, Abdelouahed, Walha, Lassaad, et Haddar, Mohamed. Study of the filament wound glass/polyester composite damage behavior by acoustic emission data unsupervised learning. *Applied Acoustics*, 2017. vol. 127, p. 175–183.

[85] Bravo, Alencar, Toubal, Lotfi, Koffi, Demagna, et Erchiqui, Fouad. Characterization of tensile damage for a short birch fiber-reinforced polyethylene composite with acoustic emission. *International Journal of Material Science*, 2013.

[86] Czigány, Tibor. Special manufacturing and characteristics of basalt fiber reinforced hybrid polypropylene composites : mechanical properties and acoustic emission study. *Composites science and technology*, 2006. vol. 66, n° 16, p. 3210–3220.

[87] Elouaer, Abdelmonem, Aboura, Zoheir, Ayad, Rezak, Sabhi, Hamid, et Benzeggagh, Malik. Suivi de l'endommagement en fatigue des composites à base de fibres végétales= monitoring of fatigue damage in composites based fiber plant. Dans *JNC16*. AMAC, 2009 p. 10–p.

[88] Aslan, Mustafa. Investigation of damage mechanism of flax fibre lpet commingled composites by acoustic emission. *Composites Part B : Engineering*, 2013. vol. 54, p. 289–297.

[89] Dogossy, G et Czigány, T. Failure mode characterization in maize hull filled polyethylene composites by acoustic emission. *Polymer testing*, 2006. vol. 25, n° 3, p. 353–357.

[90] De Rosa, Igor Maria, Santulli, Carlo, Sarasini, Fabrizio, et Valente, Marco. Effect of loading-unloading cycles on impact-damaged jute/glass hybrid laminates. *Polymer Composites*, 2009. vol. 30, n° 12, p. 1879–1887.

[91] Akil, Hazizan Md, Santulli, Carlo, Sarasini, Fabrizio, Tirillò, Jacopo, et Valente, Teodoro. Environmental effects on the mechanical behaviour of pultruded jute/glass fibre-reinforced polyester hybrid composites. *Composites Science and Technology*, 2014. vol. 94, p. 62–70.

[92] BALEY, Christophe. *Fibres naturelles de renfort pour matériaux composites.* Ed. Techniques Ingénieur, 2005.

[93] Williams, Christopher, Summerscales, John, et Grove, Stephen. Resin infusion under flexible tooling (rift) : a review. *Composites Part A : Applied Science and Manufacturing*, 1996. vol. 27, n° 7, p. 517–524.

[94] Marec, Anne, Thomas, J-H, et El Guerjouma, R. Damage characterization of polymer-based composite materials : Multivariable analysis and wavelet transform for clustering acoustic emission data. *Mechanical systems and signal processing*, 2008. vol. 22, n° 6, p. 1441–1464.

[95] Godin, Nathalie, Huguet, Stéphane, Gaertner, Roger, et Salmon, L. Clustering of acoustic emission signals collected during tensile tests on unidirectional glass/polyester composite using supervised and unsupervised classifiers. *Ndt & E International*, 2004. vol. 37, n° 4, p. 253–264.

[96] Tax, David M.J. et Duin, Robert P.W. Support vector data description. *Machine Learning*, 2004. vol. 54, n° 1, p. 45–66.

[97] Khan, Shehroz S et Madden, Michael G. One-class classification : taxonomy of study and review of techniques. *The Knowledge Engineering Review*, 2014. vol. 29, n° 3, p. 345–374.

[98] Chandola, Varun, Banerjee, Arindam, et Kumar, Vipin. Anomaly detection : A survey. *ACM computing surveys (CSUR)*, 2009. vol. 41, n° 3, p. 15.

[99] Likas, Aristidis, Vlassis, Nikos, et Verbeek, Jakob J. The global k-means clustering algorithm. *Pattern recognition*, 2003. vol. 36, n° 2, p. 451–461.

[100] Hattori, Kazuo et Takahashi, Masahito. A new nearest-neighbor rule in the pattern classification problem. *Pattern recognition*, 1999. vol. 32, n° 3, p. 425–432.

[101] Fawcett, Tom. An introduction to roc analysis. *Pattern recognition letters*, 2006. vol. 27, n° 8, p. 861–874.

[102] Spackman, Kent A. Signal detection theory : Valuable tools for evaluating inductive learning. Dans *Proceedings of the sixth international workshop on Machine learning*. Elsevier, 1989 p. 160–163.

[103] Hand, David J. Measuring classifier performance : a coherent alternative to the area under the roc curve. *Machine learning*, 2009. vol. 77, n° 1, p. 103–123.

[104] Adams, Niall M et Hand, David J. Comparing classifiers when the misallocation costs are uncertain. *Pattern Recognition*, 1999. vol. 32, n° 7, p. 1139–1147.

[105] Bradley, Andrew P. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern recognition*, 1997. vol. 30, n° 7, p. 1145–1159.

[106] Kohavi, Ron. A study of cross-validation and bootstrap for accuracy estimation and model selection. Dans *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'95. Morgan Kaufmann Publishers Inc. ISBN 1-55860-363-8, 1995 p. 1137–1143.

[107] Pękalska, Elżbieta et Duin, Robert P.W. Dissimilarity representations allow for building good classifiers. *Pattern Recognition Letters*, 2002. vol. 23, n° 8, p. 943 – 956. ISSN 0167-8655.

[108] PW, Duin Robert et Elzbieta, Pekalska. *Dissimilarity Representation For Pattern Recognition, The : Foundations And Applications*, vol. 64. World scientific, 2005.

[109] Duin, Robert PW, Pekalska, Elzbieta, Paclik, P, et Tax, D. The dissimilarity representation, a basis for domain based pattern recognition. Dans *Pattern representation and the future of pattern recognition, ICPR 2004 Workshop Proceedings, Cambridge, UK*. 2004 p. 43–56.

[110] Pękalska, Elżbieta et Duin, Robert PW. *The dissimilarity representation for pattern recognition : foundations and applications*. World Scientific, 2005.

[111] Manolova, Agata. *Catégorisation par mesures de dissimilitude et caractérisation d'images en multi*

échelle. Thèse de doctorat, Université de Grenoble, 2011.

[112] Duin, Robert PW et Pękalska, Elżbieta. The dissimilarity representation for structural pattern recognition. Dans *Iberoamerican Congress on Pattern Recognition*. Springer, 2011 p. 1–24.

[113] Duin, Robert PW et Pekalska, E. The dissimilarity representation for non-euclidean pattern recognition, a tutorial. *Delft University of Technology, The Netherlands. Available at http ://rduin. nl/presentations/DisRep_ Tutorial. pdf*, 2011.

[114] Vapnik, Vladimir. *The nature of statistical learning theory*. Springer science & business media, 2013.

[115] Fukunaga, Keinosuke. *Introduction to statistical pattern recognition*. Elsevier, 2013.

[116] Schalkoff, Robert J. *Artificial neural networks*, vol. 1. McGraw-Hill New York, 1997.

[117] Alpaydin, Ethem. *Introduction to machine learning*. MIT press, 2009.

[118] Nasrabadi, Nasser M. Pattern recognition and machine learning. *Journal of electronic imaging*, 2007. vol. 16, n° 4, p. 049901.

[119] Cristianini, Nello, Shawe-Taylor, John, *et al. An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

[120] Goldfarb, Lev, Gay, David, Golubitsky, Oleg, et Korkin, Dmitry. What is a structural representation. *Techn. Report TR00*, 2004. vol. 137.

[121] Santini, Simone. *Exploratory image databases : content-based retrieval*. Elsevier, 2001.

[122] Tversky, Amos. Features of similarity. *Psychological review*, 1977. vol. 84, n° 4, p. 327.

[123] Chen, Xin et Brown, Frank K. Asymmetry of chemical similarity. *ChemMedChem : Chemistry Enabling Drug Discovery*, 2007. vol. 2, n° 2, p. 180–182.

[124] Bunke, Horst, Günter, Simon, et Jiang, Xiaoyi. Towards bridging the gap between statistical and structural pattern recognition : Two new concepts in graph matching. Dans *International Conference on Advances in Pattern Recognition*. Springer, 2001 p. 1–11.

[125] Chen, Yihua, Garcia, Eric K, Gupta, Maya R, Rahimi, Ali, et Cazzanti, Luca. Similarity-based classification : Concepts and algorithms. *Journal of Machine Learning Research*, 2009. vol. 10, n° Mar, p. 747–776.

[126] Pekalska, Elzbieta, Paclik, Pavel, et Duin, Robert PW. A generalized kernel approach to dissimilarity-based classification. *Journal of machine learning research*, 2001. vol. 2, n° Dec, p. 175–211.

[127] Duin, Robert PW et Pękalska, Elżbieta. The dissimilarity space : Bridging structural and statistical pattern recognition. *Pattern Recognition Letters*, 2012. vol. 33, n° 7, p. 826–832.

[128] Lozano, M, Sotoca, José Martínez, Sánchez, José Salvador, Pla, Filiberto, Pękalska, E, et Duin, Robert PW. Experimental study on prototype optimisation algorithms for prototype-based classification in vector spaces. *Pattern Recognition*, 2006. vol. 39, n° 10, p. 1827–1838.

[129] Pękalska, Elżbieta, Duin, Robert P.W., et Paclík, Pavel. Prototype selection for dissimilarity-based classifiers. *Pattern Recognition*, 2006. vol. 39, n° 2, p. 189 – 208. Part Special Issue : Complexity Reduction.

[130] Edelman, Shimon. Representation is representation of similarities. *Behavioral and Brain Sciences*, 1998. vol. 21, n° 4, p. 449–467.

[131] Boashash, Boualem, Azemi, Ghasem, et O'Toole, John M. Time-frequency processing of nonstationary signals : Advanced tfd design to aid diagnosis with highlights from medical applications. *IEEE Signal Processing Magazine*, 2013. vol. 30, n° 6, p. 108–119.

[132] Boashash, Boualem. *Time-frequency signal analysis and processing : a comprehensive reference*. Academic Press, 2015.

[133] Li, Yang, Liu, Qing, Tan, Si-Rui, et Chan, Rosa HM. High-resolution time-frequency analysis of eeg signals using multiscale radial basis functions. *Neurocomputing*, 2016. vol. 195, p. 96–103.

[134] Boashash, Boualem, Azemi, Ghasem, et Khan, Nabeel Ali. Principles of time-frequency feature extraction for change detection in non-stationary signals : Applications to newborn eeg abnormality detection. *Pattern Recognition*, 2015. vol. 48, n° 3, p. 616–627.

[135] Bhati, Dinesh, Sharma, Manish, Pachori, Ram Bilas, et Gadre, Vikram M. Time–frequency localized three-band biorthogonal wavelet filter bank using semidefinite relaxation and nonlinear least squares with epileptic seizure EEG signal classification. *Digital Signal Processing*, 2017. vol. 62, p. 259 – 273. ISSN 1051-2004.

[136] Avendano-Valencia, LD, Godino-Llorente, JI, Blanco-Velasco, M, et Castellanos-Dominguez, G. Feature extraction from parametric time-frequency representations for heart murmur detection. *Annals of Biomedical Engineering*, 2010. vol. 38, n° 8, p. 2716–2732.

[137] Bozkurt, Baris, Germanakis, Ioannis, et Stylianou, Yannis. A study of time-frequency features for cnn-based automatic heart sound classification for pathology detection. *Computers in Biology and Medicine*, 2018.

[138] Cömert, Zafer, Kocamaz, Adnan Fatih, et Subha, Velappan. Prognostic model based on image-based time-frequency features and genetic algorithm for fetal hypoxia assessment. *Computers in biology and medicine*, 2018.

[139] Fiscon, Giulia, Weitschek, Emanuel, Cialini, Alessio, Felici, Giovanni, Bertolazzi, Paola, De Salvo, Simona, Bramanti, Alessia, Bramanti, Placido, et De Cola, Maria Cristina. Combining eeg signal processing with supervised methods for alzheimer's patients classification. *BMC medical informatics and decision making*, 2018. vol. 18, n° 1, p. 35.

[140] Fraiwan, Luay, Lweesy, Khaldon, Khasawneh, Natheer, Wenz, Heinrich, et Dickhaus, Hartmut.

Automated sleep stage identification system based on time-frequency analysis of a single eeg channel and random forest classifier. *Computer methods and programs in biomedicine*, 2012. vol. 108, n° 1, p. 10–19.

[141] Bajaj, Varun et Pachori, Ram Bilas. Automatic classification of sleep stages based on the time-frequency image of eeg signals. *Computer methods and programs in biomedicine*, 2013. vol. 112, n° 3, p. 320–328.

[142] Martis, Roshan Joy, Acharya, U Rajendra, Tan, Jen Hong, Petznick, Andrea, Tong, Louis, Chua, Chua Kuang, et Ng, Eddie Yin Kwee. Application of intrinsic time-scale decomposition (itd) to eeg signals for automated seizure prediction. *International journal of neural systems*, 2013. vol. 23, n° 05, p. 1350023.

[143] Kumari, R Shantha Selva et Jose, J Prabin. Seizure detection in eeg using time frequency analysis and svm. Dans *Emerging Trends in Electrical and Computer Technology (ICETECT), 2011 International Conference on*. IEEE, 2011 p. 626–630.

[144] Zeng, Deguo, Zeng, Xiaodong, Cheng, Hao, et Tang, B. Automatic modulation classification of radar signals using the rihaczek distribution and hough transform. *IET Radar, Sonar & Navigation*, 2012. vol. 6, n° 5, p. 322–331.

[145] Tsinalis, Orestis, Matthews, Paul M, et Guo, Yike. Automatic sleep stage scoring using time-frequency analysis and stacked sparse autoencoders. *Annals of biomedical engineering*, 2016. vol. 44, n° 5, p. 1587–1597.

[146] Musselman, Marcus et Djurdjanovic, Dragan. Time–frequency distributions in the classification of epilepsy from eeg signals. *Expert Systems with Applications*, 2012. vol. 39, n° 13, p. 11413–11422.

[147] Sharma, Manish, Pachori, Ram Bilas, et Acharya, U Rajendra. A new approach to characterize epileptic seizures using analytic time-frequency flexible wavelet transform and fractal dimension. *Pattern Recognition Letters*, 2017. vol. 94, n° Supplement C, p. 172 – 179. ISSN 0167-8655.

[148] Gao, Huizhong, Liang, Lin, Chen, Xiaoguang, et Xu, Guanghua. Feature extraction and recognition for rolling element bearing fault utilizing short-time fourier transform and non-negative matrix factorization. *Chinese Journal of Mechanical Engineering*, 2015. vol. 28, n° 1, p. 96–105.

[149] Gaunaurd, Guillermo C et Strifors, Hans C. Signal analysis by means of time-frequency (wigner-type) distributions-applications to sonar and radar echoes. *Proceedings of the IEEE*, 1996. vol. 84, n° 9, p. 1231–1248.

[150] Turhan-Sayan, Gönül. Natural resonance-based feature extraction with reduced aspect sensitivity for electromagnetic target classification. *Pattern Recognition*, 2003. vol. 36, n° 7, p. 1449–1466.

[151] Xue, W et Sun, X-W. Target detection of vehicle volume detecting radar based on wigner-hough

transform. *Journal of Electromagnetic Waves and Applications*, 2007. vol. 21, n° 11, p. 1513–1523.

[152] Shi, Yu et Zhang, Xian-Da. A gabor atom network for signal classification with application in radar target recognition. *IEEE Transactions on Signal Processing*, 2001. vol. 49, n° 12, p. 2994–3004.

[153] Dennis, Jonathan, Tran, Huy Dat, et Li, Haizhou. Spectrogram image feature for sound event classification in mismatched conditions. *IEEE signal processing letters*, 2011. vol. 18, n° 2, p. 130–133.

[154] Ghoraani, Behnaz et Krishnan, Sridhar. Time-frequency matrix feature extraction and classification of environmental audio signals. *IEEE transactions on audio, speech, and language processing*, 2011. vol. 19, n° 7, p. 2197–2209.

[155] Su, Bo Yu, Ho, KC, Rantz, Marilyn J, et Skubic, Marjorie. Doppler radar fall activity detection using the wavelet transform. *IEEE Transactions on Biomedical Engineering*, 2015. vol. 62, n° 3, p. 865–875.

[156] Rakotomamonjy, Alain et Gasso, Gilles. Histogram of gradients of time-frequency representations for audio scene classification. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 2015. vol. 23, n° 1, p. 142–153.

[157] Mitiche, Imene, Morison, Gordon, Nesbitt, Alan, Hughes-Narborough, Michael, Stewart, Brian G, et Boreham, Philip. Classification of emi discharge sources using time–frequency features and multi-class support vector machine. *Electric Power Systems Research*, 2018. vol. 163, p. 261–269.

[158] Biswal, Milan et Dash, Pradipta K. Detection and characterization of multiple power quality disturbances with a fast s-transform and decision tree based classifier. *Digital Signal Processing*, 2013. vol. 23, n° 4, p. 1071–1083.

[159] Yong, Zhan, Haozhong, Cheng, Yifeng, Ding, Gan-Yun, Lv, et Yi-Bin, Sun. S-transform-based classification of power quality disturbance signals by support vector machines. *Proceedings of the CSEE*, 2005. vol. 25, n° 4, p. 51–56.

[160] Yuan, Yingjun, Huang, Zhitao, Wu, Hao, et Wang, Xiang. Specific emitter identification based on hilbert–huang transform-based time–frequency–energy distribution features. *IET Communications*, 2014. vol. 8, n° 13, p. 2404–2412.

[161] Degottex, Gilles, Lanchantin, Pierre, Roebel, Axel, et Rodet, Xavier. Mixed source model and its adapted vocal tract filter estimate for voice transformation and synthesis. *Speech Communication*, 2013. vol. 55, n° 2, p. 278–294.

[162] Stowell, Dan, Wood, Mike, Stylianou, Yannis, et Glotin, Hervé. Bird detection in audio : a survey and a challenge. Dans *Machine Learning for Signal Processing (MLSP), 2016 IEEE 26th International Workshop on*. IEEE, 2016 p. 1–6.

[163] Chunghsin, Y. *Multiple fundamental frequency estimation of polyphonic recordings*. Thèse de doctorat,

Ph. D. dissertation, University Paris 6, 2008.

[164] Jančovič, Peter et Köküer, Münevver. Automatic detection and recognition of tonal bird sounds in noisy environments. *EURASIP Journal on Advances in Signal Processing*, 2011. vol. 2011, n° 1, p. 982936.

[165] Potamitis, Ilyas, Ntalampiras, Stavros, Jahn, Olaf, et Riede, Klaus. Automatic bird sound detection in long real-field recordings : Applications and tools. *Applied Acoustics*, 2014. vol. 80, p. 1–9.

[166] Kim, Jonghwa et André, Elisabeth. Emotion recognition based on physiological changes in music listening. *IEEE transactions on pattern analysis and machine intelligence*, 2008. vol. 30, n° 12, p. 2067–2083.

[167] Belouchrani, A. et Amin, M.G. Time-frequency music. *IEEE Signal Processing Letters*, 1999. vol. 6, n° 5, p. 109–110.

[168] Hadjidimitriou, Stelios K et Hadjileontiadis, Leontios J. Toward an eeg-based recognition of music liking using time-frequency analysis. *IEEE Transactions on Biomedical Engineering*, 2012. vol. 59, n° 12, p. 3498–3510.

[169] Costa, Yandre MG, Oliveira, Luiz S, Koericb, Alessandro L, et Gouyon, Fabien. Music genre recognition using spectrograms. Dans *Intl. Conf. on Systems, Signal and Image Processing*. 2011 p. 1–4.

[170] Boashash, Boualem, Khan, Nabeel Ali, et Ben-Jabeur, Taoufik. Time–frequency features for pattern recognition using high-resolution tfds : A tutorial review. *Digital Signal Processing*, 2015. vol. 40, p. 1–30.

[171] Sejdić, Ervin, Djurović, Igor, et Jiang, Jin. Time-frequency feature representation using energy concentration : An overview of recent advances. *Digital Signal Processing*, 2009. vol. 19, n° 1, p. 153–183.

[172] Addison, Paul S. Wavelet transforms and the ecg : a review. *Physiological measurement*, 2005. vol. 26, n° 5, p. R155.

[173] Unser, Michael et Aldroubi, Akram. A review of wavelets in biomedical applications. *Proceedings of the IEEE*, 1996. vol. 84, n° 4, p. 626–638.

[174] Boashash, Boualem, Boubchir, Larbi, et Azemi, Ghasem. A methodology for time-frequency image processing applied to the classification of non-stationary multichannel signals using instantaneous frequency descriptors with application to newborn eeg signals. *EURASIP Journal on Advances in Signal Processing*, 2012. vol. 2012, n° 1, p. 117.

[175] Abed, Mansour, Belouchrani, Adel, Cheriet, Mohamed, et Boashash, Boualem. Time-frequency distributions based on compact support kernels : properties and performance evaluation. *IEEE Tran-*

sactions on Signal Processing, 2012. vol. 60, n° 6, p. 2814–2827.

[176] Zarjam, Pega, Azemi, Ghasem, Mesbah, Mostefa, et Boashash, Boualem. Detection of newborns' eeg seizure using time-frequency divergence measures. Dans *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, vol. 5. IEEE, 2004 p. V–429.

[177] Breakenridge, Calvin. Nonstationary signal classification using time-frequency optimization. Dans *Electronics, Circuits and Systems, 2003. ICECS 2003. Proceedings of the 2003 10th IEEE International Conference on*, vol. 1. IEEE, 2003 p. 132–135.

[178] Davy, Manuel, Doncarli, Christian, et Boudreaux-Bartels, G Faye. Improved optimization of time-frequency-based signal classifiers. *IEEE Signal processing letters*, 2001. vol. 8, n° 2, p. 52–57.

[179] Vincent, Isabelle, Doncarli, Christian, et Le Carpentier, E. Non stationary signals classification using time-frequency distributions. Dans *Time-Frequency and Time-Scale Analysis, 1994., Proceedings of the IEEE-SP International Symposium on*. IEEE, 1994 p. 233–236.

[180] Li, Bing, Zhang, Pei-lin, Tian, Hao, Mi, Shuang-Shan, Liu, Dong-Sheng, et Ren, Guo-Quan. A new feature extraction and selection scheme for hybrid fault diagnosis of gearbox. *Expert Systems with Applications*, 2011. vol. 38, n° 8, p. 10000–10009.

[181] Honeine, Paul, Richard, Cdric, et Flandrin, Patrick. Time-frequency learning machines. *IEEE Transactions on Signal Processing*, 2007. vol. 55, n° 7, p. 3930–3936.

[182] Honeine, Paul, Richard, Cédric, et Flandrin, Patrick. Reconnaissance des formes par méthodes à noyau dans le domaine temps-fréquence. Dans *20ème Colloque sur le traitement du signal et des images, FRA, 2005*. GRETSI, Groupe d'Etudes du Traitement du Signal et des Images, 2005 p. –.

[183] Honeine, Paul et Richard, Cédric. Signal-dependent time-frequency representations for classification using a radially gaussian kernel and the alignment criterion. Dans *Statistical Signal Processing, 2007. SSP'07. IEEE/SP 14th Workshop on*. IEEE, 2007 p. 735–739.

[184] Schölkopf, Bernhard, Herbrich, Ralf, et Smola, Alex J. A generalized representer theorem. Dans *International conference on computational learning theory*. Springer, 2001 p. 416–426.

[185] Honeiné, Paul, Richard, Cédric, Flandrin, Patrick, et Pothin, J-B. Optimal selection of time-frequency representations for signal classification : A kernel-target alignment approach. Dans *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, vol. 3. IEEE, 2006 p. III–III.

[186] Stankovic, Ljubiša. An analysis of some time-frequency and time-scale distributions. *Annales Des Télécommunications*, 1994. vol. 49, n° 9, p. 505.

[187] Daubechies, Ingrid. *Ten lectures on wavelets*. SIAM, 1992.

[188] Mallat, Stéphane. *A wavelet tour of signal processing*. Academic press, 1999.

[189] Gabor, Dennis. Theory of communication. part 1 : The analysis of information. *Journal of the Institution of Electrical Engineers-Part III : Radio and Communication Engineering*, 1946. vol. 93, n° 26, p. 429–441.

[190] Folland, Gerald B et Sitaram, Alladi. The uncertainty principle : a mathematical survey. *Journal of Fourier analysis and applications*, 1997. vol. 3, n° 3, p. 207–238.

[191] Hlawatsch, Franz, Auger, François, et OVARLEZ, Jean-Philippe. Temps-fréquence : concepts et outils. *Hermes, Lavoisier, Paris*, 2005.

[192] Boashash, Boualem et Barkat, Braham. Introduction to time-frequency signal analysis. Dans *Wavelet Transforms and Time-Frequency Signal Analysis*, p. 321–380. Springer, 2001.

[193] Wigner, Eugene. On the quantum correction for thermodynamic equilibrium. *Physical review*, 1932. vol. 40, n° 5, p. 749.

[194] Ville, Jean *et al.* Théorie et applications de la notion de signal analytique. *Cables et transmission*, 1948. vol. 2, n° 1, p. 61–74.

[195] Moyal, José E. Quantum mechanics as a statistical theory. Dans *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 45. Cambridge University Press, 1949 p. 99–124.

[196] Claasen, TACM et Mecklenbräuker, WFG. Time-frequency signal analysis. *Philips Journal of Research*, 1980. vol. 35, n° 4/5, p. 276–300.

[197] Flandrin, Patrick. *Time-frequency/time-scale analysis*, vol. 10. Academic press, 1998.

[198] Cohen, Leon. Generalized phase-space distribution functions. *Journal of Mathematical Physics*, 1966. vol. 7, n° 5, p. 781–786.

[199] Rihaczek, A. Signal energy distribution in time and frequency. *IEEE Transactions on Information Theory*, 1968. vol. 14, n° 3, p. 369–374.

[200] Margenau, Henry et Hill, Robert Nyden. Correlation between measurements in quantum theory. *Progress of Theoretical Physics*, 1961. vol. 26, n° 5, p. 722–738.

[201] Hippenstiel, Ralph D et De Oliveira, Paulo M. Time-varying spectral estimation using the instantaneous power spectrum (ips). *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1990. vol. 38, n° 10, p. 1752–1759.

[202] Gómez, Soledad, Naranjo, Valery, et Miralles, Ramón. Removing interference components in time–frequency representations using morphological operators. *Journal of Visual Communication and Image Representation*, 2011. vol. 22, n° 5, p. 401–410.

[203] Hasan, Syed Rashid. *Elimination of interference terms of the discrete Wigner distribution using nonlinear filtering*. Thèse de doctorat, University of Delaware, 1999.

[204] Khan, Nabeel Ali, Taj, Imtiaz Ahmad, Jaffri, M Noman, et Ijaz, Salman. Cross-term elimination

in wigner distribution based on 2d signal processing techniques. *Signal Processing*, 2011. vol. 91, n° 3, p. 590–599.

[205] Basseville, Michele. Distance measures for signal processing and pattern recognition. *Signal processing*, 1989. vol. 18, n° 4, p. 349–369.

[206] Aviyente, S. Divergence measures for time-frequency distributions. Dans *Seventh International Symposium on Signal Processing and Its Applications, 2003. Proceedings.*, vol. 1. 2003 p. 121–124.

[207] Vincent, I., Doncarli, C., et Carpentier, E. Le. Non stationary signals classification using time-frequency distributions. Dans *Proceedings of IEEE-SP International Symposium on Time- Frequency and Time-Scale Analysis*. 1994 p. 233–236.

[208] McSharry, P. E., He, T., Smith, L. A., et Tarassenko, L. Linear and non-linear methods for automatic seizure detection in scalp electro-encephalogram recordings. *Medical and Biological Engineering and Computing*, 2002. vol. 40, n° 4, p. 447–461.

[209] Mormann, Florian, Andrzejak, Ralph G., Elger, Christian E., et Lehnertz, Klaus. Seizure prediction : the long and winding road. *Brain*, 2006. vol. 130, n° 2, p. 314.

[210] Song, Jiang-Ling, Hu, Wenfeng, et Zhang, Rui. Automated detection of epileptic EEGs using a novel fusion feature and extreme learning machine. *Neurocomputing*, 2016. vol. 175, Part A, p. 383 – 391. ISSN 0925-2312.

[211] Kumar, T. Sunil, Kanhangad, Vivek, et Pachori, Ram Bilas. Classification of seizure and seizure-free EEG signals using local binary patterns. *Biomedical Signal Processing and Control*, 2015. vol. 15, p. 33 – 40. ISSN 1746-8094.

[212] Mousavi, SR, Niknazar, M, et Vahdat, B Vosoughi. Epileptic seizure detection using ar model on EEG signals. Dans *Biomedical engineering conference, 2008. CIBEC 2008. Cairo International*. IEEE, 2008 p. 1–4.

[213] Tzallas, A. T., Tsipouras, M. G., et Fotiadis, D. I. Epileptic seizure detection in EEGs using time-frequency analysis. *IEEE Transactions on Information Technology in Biomedicine*, 2009. vol. 13, n° 5, p. 703–710.

[214] Polat, Kemal et Güneş, Salih. Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast fourier transform. *Applied Mathematics and Computation*, 2007. vol. 187, n° 2, p. 1017 – 1026. ISSN 0096-3003.

[215] Liang, Sheng-Fu, Wang, Hsu-Chuan, et Chang, Wan-Lin. Combination of EEG complexity and spectral analysis for epilepsy diagnosis and seizure detection. *EURASIP Journal on Advances in Signal Processing*, 2010. vol. 2010, n° 1, p. 853434.

[216] Kumar, Yatindra, Dewal, M.L., et Anand, R.S. Epileptic seizure detection using DWT based fuzzy

approximate entropy and support vector machine. *Neurocomputing*, 2014. vol. 133, p. 271 – 279. ISSN 0925-2312.

[217] Fu, Kai, Qu, Jianfeng, Chai, Yi, et Dong, Yong. Classification of seizure based on the time-frequency image of EEG signals using HHT and SVM. *Biomedical Signal Processing and Control*, 2014. vol. 13, p. 15 – 22. ISSN 1746-8094.

[218] Kang, Jae-Hwan, Chung, Yoon Gi, et Kim, Sung-Phil. An efficient detection of epileptic seizure by differentiation and spectral analysis of electroencephalograms. *Computers in biology and medicine*, 2015. vol. 66, p. 352–356.

[219] Sharma, Rajeev et Pachori, Ram Bilas. Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions. *Expert Systems with Applications*, 2015. vol. 42, n° 3, p. 1106 – 1117. ISSN 0957-4174.

[220] Dhiman, Rohtash, Saini, J.S., et Priyanka. Biogeography based hybrid scheme for automatic detection of epileptic seizures from EEG signatures. *Applied Soft Computing*, 2017. vol. 51, p. 116 – 129. ISSN 1568-4946.

[221] Hassan, Ahnaf Rashik, Siuly, Siuly, et Zhang, Yanchun. Epileptic seizure detection in EEG signals using tunable-q factor wavelet transform and bootstrap aggregating. *Computer Methods and Programs in Biomedicine*, 2016. vol. 137, p. 247 – 259. ISSN 0169-2607.

[222] A., Omer F., Siuly, Siuly, Bajaj, Varun, Guo, Yanhui, Sengur, Abdulkadir, et Zhang, Yanchun. Multi-category EEG signal classification developing time-frequency texture features based fisher vector encoding method. *Neurocomputing*, 2016. vol. 218, p. 251 – 258. ISSN 0925-2312.

[223] Das, Anindya Bijoy, Bhuiyan, Mohammed Imamul Hassan, et Alam, SM Shafiul. Classification of EEG signals using normal inverse gaussian parameters in the dual-tree complex wavelet transform domain for seizure detection. *Signal, Image and Video Processing*, 2016. vol. 10, n° 2, p. 259–266.

[224] K.S., Biju, Hakkim, Hara Abdul, et M.G., Jibukumar. Ictal EEG classification based on amplitude and frequency contours of IMFs. *Biocybernetics and Biomedical Engineering*, 2017. vol. 37, n° 1, p. 172 – 183. ISSN 0208-5216.

[225] Li, Mingyang, Chen, Wanzhong, et Zhang, Tao. Classification of epilepsy EEG signals using dwt-based envelope analysis and neural network ensemble. *Biomedical Signal Processing and Control*, 2017. vol. 31, p. 357 – 365. ISSN 1746-8094.

[226] Zahra, Asmat, Kanwal, Nadia, ur Rehman, Naveed, Ehsan, Shoaib, et McDonald-Maier, Klaus D. Seizure detection from eeg signals using multivariate empirical mode decomposition. *Computers in Biology and Medicine*, 2017. vol. 88, n° Supplement C, p. 132 – 141. ISSN 0010-4825.

[227] Jaiswal, Abeg Kumar et Banka, Haider. Local pattern transformation based feature extraction

techniques for classification of epileptic eeg signals. *Biomedical Signal Processing and Control*, 2017. vol. 34, n° Supplement C, p. 81 – 92. ISSN 1746-8094.

[228] Li, Mingyang, Chen, Wanzhong, et Zhang, Tao. A novel seizure diagnostic model based on kernel density estimation and least squares support vector machine. *Biomedical Signal Processing and Control*, 2018. vol. 41, p. 233 – 241.

[229] Tiwari, A. K., Pachori, R. B., Kanhangad, V., et Panigrahi, B. K. Automated diagnosis of epilepsy using key-point-based local binary pattern of eeg signals. *IEEE Journal of Biomedical and Health Informatics*, 2017. vol. 21, n° 4, p. 888–896. ISSN 2168-2194.

[230] Cortes, Corinna et Vapnik, Vladimir. Support-vector networks. *Machine learning*, 1995. vol. 20, n° 3, p. 273–297.

[231] Deza, Michel Marie et Deza, Elena. Encyclopedia of distances. Dans *Encyclopedia of Distances*, p. 1–583. Springer, 2009.

[232] Baudrier, Étienne, Nicolier, Frédéric, Millon, Gilles, et Ruan, Su. Binary-image comparison with local-dissimilarity quantification. *Pattern Recognition*, 2008. vol. 41, n° 5, p. 1461–1478.

[233] Molchanov, Ilya S et Terán, Pedro. Distance transforms for real-valued functions. *Journal of mathematical analysis and applications*, 2003. vol. 278, n° 2, p. 472–484.

Analyse des signaux non-stationnaires à l'aide d'une nouvelle démarche de classification : application à l'identification de l'endommagement de matériaux composites par émission acoustique et à la détection de la crise d'épilepsie par EEG

La première partie de cette étude doctorale vise à développer une nouvelle méthode de classification de signaux d'émission acoustique (EA), appelée méthode incrémentale (MI), provoqués par les mécanismes d'endommagement des matériaux composites à fibres végétales et à fibres de verre. L'avantage de cette méthode est sa capacité d'identifier les signaux porteurs d'informations et de fournir les types de mécanismes d'endommagement sans avoir recours à une expertise supplémentaire. Les résultats expérimentaux ont montré que la MI basée sur les caractéristiques des signaux d'EA présente plus d'avantages pour les classifier et construire une base de données d'apprentissage. De plus, l'utilisation de la base de données d'apprentissage et un classifieur supervisé s'est avérée efficace pour détecter tous les types d'endommagement de matériaux composites hybrides. Étant donné que la représentation basée sur les caractéristiques peut s'avérer inefficace dans certains cas pour résoudre un problème de classification, c'est vers l'approche de classification basée sur une mesure de dissimilarité que notre étude s'est orientée pour mieux classifier les signaux. De plus, l'analyse temps-fréquence (ATF) est plus appropriée aux signaux non-stationnaires. Ainsi, le principe de combiner l'approche de classification basée sur une mesure de dissimilarité et l'ATF a été retenue, ce qui a constitué l'objectif principal de la deuxième partie de cette thèse. Afin de valider l'approche, des applications à des signaux EEG d'épilepsie de la littérature sont finalement considérées. Les résultats expérimentaux montrent que cette approche conduit à une amélioration notable des performances d'un classifieur comparée à l'approche basée sur les caractéristiques.

Mots clés : Emission acoustique, endommagement, fibres de lin, classification, k-moyennes, méthode incrémentale, k-PPV, SVM, RNA, ADL, Temps-fréquence, dissimilarité, EEG, signal, non-stationnaire, CDL, PWVL.

Analysis of non-stationary signals using a new classification approach: application to the identification of damage in composite materials by acoustic emission and to the detection of epilepsy by EEG

The first part of this thesis aims at developing a new clustering method of acoustic emission (AE) signals, called incremental method (IC), induced by damage mechanisms of glass and flax fibre-reinforced composite materials. The advantage of this method is the capability to identify the signals carrying information and to provide the types of damage mechanisms without using additional expertise. Experimental results showed that the IM based feature presented more advantages to successfully classify the AE signals and build a labelled learning database. Moreover, the use of the learning database and a supervised classifier has proven effective for detecting damage mechanisms of hybrid composite materials. However, traditional feature-based representation approach may be ineffective in some cases to solve a classification problem. We propose to use the dissimilarity-based approach for signals classification. Traditional signal processing approaches assume the stationarity of signals, which in practice is not often satisfied. Hence, time or frequency descriptions alone are insufficient to provide comprehensive information about such signals. On the contrary, time–frequency analysis (TFA) is more suitable for nonstationary signals. Thus, we have the idea to combine the classification-based dissimilarity approach and TFA. This idea is the main purpose of the second part of this thesis. We conducted experimental tests to validate the proposed classification approach using real signals. We are especially interested in the classification of EEG epileptic signals. The proposed approach leads to a significant improvement of the classifier performances compared to the feature-based approach.

Keywords: Acoustic emission, damage events, flax fibre, classification, k-means, incremental clustering, k-NN, SVM, ANN, LDA, time-frequency, dissimilarity, EEG, signal, non-stationary, LDM, SPWVD.

Discipline : AUTOMATIQUE, SIGNAL, PRODUCTIQUE, ROBOTIQUE

Spécialité : Traitement du signal et de l'image

Université Mohammed Premier Oujda-Maroc

LSE2I-ENSAO

Ecole Nationale des Sciences Appliquées
Complexe universitaire Hay El Qods
B. P. 669
60000 Oujda Maroc

Université de Reims Champagne-Ardenne -France

CRESTIC - EA 3804

UFR Sciences Exactes et Naturelles
Moulin de la Housse
51867 Reims France