



Université du Québec
à Rimouski

**ÉTUDE DES TECHNIQUES D'EXTRACTION DE
CARACTÉRISTIQUES DES SONS RESPIRATOIRES EN VUE
D'OPTIMISER LEUR CLASSIFICATION AUTOMATIQUE**

MÉMOIRE PRÉSENTÉ

dans le cadre du programme de maîtrise en ingénierie

en vue de l'obtention du grade de maître ès sciences appliquées (M.Sc.A.)

PAR

© ABDERAHMANE TAOUG

Septembre 2022

Composition du jury :

Raef Cherif (Ph.D.), président du jury, Université du Québec à Rimouski

Mohammed Bahoura (Ph.D.), directeur de recherche, Université du Québec à Rimouski

Abdellah Chehri (Ph.D.), examinateur externe, Université du Québec à Chicoutimi

Dépôt initial le 14 juin 2022

Dépôt final le 27 septembre 2022

UNIVERSITÉ DU QUÉBEC À RIMOUSKI
Service de la bibliothèque

Avertissement

La diffusion de ce mémoire ou de cette thèse se fait dans le respect des droits de son auteur, qui a signé le formulaire « *Autorisation de reproduire et de diffuser un rapport, un mémoire ou une thèse* ». En signant ce formulaire, l'auteur concède à l'Université du Québec à Rimouski une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de son travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, l'auteur autorise l'Université du Québec à Rimouski à reproduire, diffuser, prêter, distribuer ou vendre des copies de son travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de la part de l'auteur à ses droits moraux ni à ses droits de propriété intellectuelle. Sauf entente contraire, l'auteur conserve la liberté de diffuser et de commercialiser ou non ce travail dont il possède un exemplaire.

À mon père

À ma mère

À mes frères

À mes sœurs

REMERCIEMENTS

Je voudrais tout d'abord adresser toute ma gratitude à mon directeur de recherche monsieur Mohammed Bahoura, professeur au Département de Mathématiques d'Informatique et de Génie (DIMG) de l'Université du Québec à Rimouski (UQAR) pour sa patience, sa disponibilité et surtout ses judicieux conseils de qualités, qui ont contribué à alimenter ma réflexion, de m'avoir guidé dans mon travail et de m'avoir aidé à trouver des solutions pour avancer.

Je souhaiterais remercier également les membres du jury, le professeur Raef Cherif et le professeur Abdellah Chehri, qui m'ont fait l'honneur de prendre le temps d'étudier mon travail avec attention.

Je tiens aussi à remercier Abdelkrim Semmad et Pape Sémou Diouf de m'avoir guidé durant mes débuts dans ce projet et de me mettre sur la bonne voie.

Enfin, je remercie mes parents ainsi que tous les membres de ma famille, qui m'ont soutenu durant toute la période de l'accomplissement de mon projet.

RÉSUMÉ

L'asthme, la maladie pulmonaire obstructive chronique (MPOC) et de nombreux autres types de maladies pulmonaires constituent une cause majeure de décès dans le monde, selon l'Organisation mondiale de la santé. Dû à ces maladies, des sons respiratoires anormaux, tels que les crépitations et les sifflements, sont surajoutés aux sons respiratoires normaux. Compte tenu de la nature subjective de l'information obtenue à partir de l'auscultation par stéthoscope, nous sommes dans le besoin de réaliser un système de classification objective des sons respiratoires basé sur l'apprentissage machine.

Notre projet consiste à la réalisation d'un système d'aide au diagnostic permettant la classification automatique des sons respiratoires, en 4 classes (Normaux, sifflements, crépitations et le cas contenant les sifflements et les crépitations simultanément) utilisant la base de données ICBHI 2017. Globalement le système opère en deux étapes principales : l'extraction des caractéristiques des sons respiratoires et ensuite leurs classifications.

Pour atteindre cet objectif, nous proposons de tester quinze systèmes obtenus par l'utilisation de cinq techniques d'extraction de caractéristiques (les coefficients cepstraux à l'échelle de Mel (MFCC), les coefficients cepstraux en fréquences Gammatone (GFCC), la transformée par paquets d'ondelettes (WPT), une combinaison des coefficients MFCC et WPT, et une combinaison des coefficients GFCC et WPT), en combinaison avec les trois modèles de classification (le perceptron multicouche (MLP), les machines à vecteur de support (SVM) et le k-plus proches voisins (k-NN)). Une phase de prétraitement a été effectuée sur les sons respiratoires avant de les présenter à ces systèmes de caractérisation/classification.

Les meilleurs résultats ont été obtenus avec la combinaison de la technique d'extraction de caractéristiques basé sur les coefficients GFCC-WPT et le classifieur SVM, avec plus de 90 % d'exactitude et de score ICBHI. Les classifieurs à base de réseaux de neurones MLP et de k-NN ont également révélé des résultats prometteurs pour la classification, avec plus de 87 % d'exactitude et de score ICBHI.

Mots clés : Sons respiratoires, MPOC, crépitations, sifflements, extraction des caractéristiques, GFCC, WPT, Classification automatique, MLP, SVM, k-NN.

ABSTRACT

Asthma, chronic obstructive pulmonary disease (COPD) and many other types of lung diseases are a major cause of death across the world, according to the World Health Organization. Due to these diseases, abnormal breath sounds, such as crackles and wheezes, are added to the normal breath sound. Given the subjective nature of the information obtained from stethoscope auscultation, we need an objective classification system for respiratory sounds based on machine learning.

Our project consists in the realization of a diagnostic aid system allowing the automatic classification of respiratory sounds, in 4 classes (Normal, Wheezes, Crackles and the case containing wheezes and crackles simultaneously) using the ICBHI 2017 database. Overall, the system operates in two main steps: feature extraction of the respiratory sounds and then their classifications.

To achieve this goal, we propose to test fifteen systems obtained by using five feature extraction techniques (Mel-scale cepstral coefficients (MFCC), Gammatone frequency cepstral coefficients (GFCC), the wavelet packet transform (WPT), a combination of the MFCC and WPT coefficients, and a combination of the GFCC and WPT coefficients), in combination with the three classification models (the multilayer perceptron (MLP), the support vector machines (SVM), and the k-nearest neighbor (k-NN)). A preprocessing phase was performed on the breath sounds before presenting them to these characterization/classification systems.

The best results were obtained with the combination of the GFCC-WPT coefficient-based feature extraction technique and the SVM classifier, with over 90% accuracy and ICBHI score. The MLP and k-NN based neural network classifiers also showed promising results for classification, with over 87% accuracy and ICBHI score.

Keywords: Breath sounds, COPD, crackles, wheezes, feature extraction, GFCC, WPT, Automatic classification, MLP, SVM, k-NN.

TABLE DES MATIÈRES

REMERCIEMENTS.....	ix
RÉSUMÉ.....	xi
ABSTRACT.....	xiii
TABLE DES MATIÈRES.....	xv
LISTE DES TABLEAUX.....	xix
LISTE DES FIGURES.....	xxi
LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES.....	xxiii
INTRODUCTION GÉNÉRALE.....	1
CHAPITRE 1 DÉFINITION DES SONS RESPIRATOIRES ET LEURS CLASSIFICATION AUTOMATIQUE.....	3
1.1 NOMENCLATURE DES SONS RESPIRATOIRES.....	3
1.1.1 Sons respiratoires normaux.....	3
1.1.2 Sons respiratoires adventices.....	4
1.1.3 Base de données utilisée.....	7
1.2 PROBLÉMATIQUE DE RECHERCHE.....	9
1.3 OBJECTIF.....	10
1.4 HYPOTHÈSE.....	10
1.5 MÉTHODOLOGIE.....	11
1.5.1 Extraction de caractéristiques des sons respiratoires.....	11
1.5.2 Classification automatique des sons respiratoires.....	12
1.6 CONTRIBUTION.....	12
CHAPITRE 2 EXTRACTION DES CARACTÉRISTIQUES.....	15
2.1 DÉFINITION ET PRINCIPE DE L'EXTRACTION DES CARACTÉRISTIQUES.....	15

2.2	ANALYSE DE FOURIER.....	16
2.3	ANALYSE PAR ONDELETTES.....	17
2.3.1	Transformée en ondelette continue (CWT)	18
2.3.2	Transformée en ondelette discrète (DWT)	19
2.3.3	Transformée par paquets d'ondelettes (WPT)	20
2.4	ANALYSE CEPSTRAL	22
2.4.1	Coefficients cepstraux à l'échelle de Mel (MFCC)	24
2.4.2	Coefficients cepstraux des fréquences Gammatone (GFCC)	27
2.5	CARACTÉRISTIQUES STATESTIQUES.....	29
2.6	RÉDUCTION DE DIMENSIONNALITÉ DES CARACTÉRISTIQUES.....	31
2.7	APPROCHE PROPOSÉE	32
CHAPITRE 3 MÉTHODES DE CLASSIFICATION AUTOMATIQUE DES SONS RESPIRATOIRES		35
3.1	DÉFINITION ET PRINCIPE DE LA CLASSIFICATION AUTOMATIQUE	35
3.2	PERCEPTRON MULTICOUCHE MLP (MULTI-LAYER PERCEPTRON)	36
3.2.1	Perceptrons.....	37
3.2.2	Principe de rétro-propagation (back propagation)	39
3.3	MACHINE A VECTEUR DE SUPPORT SVM (<i>SUPPORT VECTOR MACHINE</i>).....	41
3.3.1	Introduction aux SVM	41
3.3.2	Espace caractéristique linéairement séparable.....	42
3.3.3	Espace caractéristique non linéairement séparable.....	44
3.4	K-PLUS PROCHES VOISINS K-NN (<i>K-NEAREST NEIGHBORS</i>)	45
3.5	MODÈLES DE MÉLANGE GAUSSIEN GMM (GAUSSIAN MIXTURE MODELS)	47
3.6	TECHNIQUES DE VALIDATION DES DONNÉES	49
3.6.1	La méthode de retenue (<i>Holdout</i>)	50
3.6.2	La validation croisée à k-blocs (<i>K-Fold Cross Validation</i>)	50
3.6.3	La validation croisée avec exclusion	51
3.6.4	Approche proposée pour la classification	52
CHAPITRE 4 RÉSULTATS EXPÉRIMENTAUX.....		53
4.1	OUTILS ET MATÉRIELS UTILISÉS	53

4.2	PRÉTRAITEMENT DE LA BASE DE DONNÉES.....	53
4.3	CRITÈRES D'ÉVALUATION.....	56
4.4	ETUDE PARAMETRIQUE DES TECHNIQUES D'EXTRACTION DE CARACTÉRISTIQUES.....	57
4.5	RÉSULTATS DE CLASSIFICATION PAR MLP	59
	4.5.1 Étude paramétrique.....	60
	4.5.2 Utilisation de la totalité la base de données.....	66
	4.5.3 Reproductibilité des résultats	70
4.6	RÉSULTATS DE LA CLASSIFICATION PAR SVM.....	71
4.7	RÉSULTATS DE LA CLASSIFICATION PAR K-NN.....	73
4.8	COMPARAISON DES RÉSULTATS DES TROIS MODÈLES DE CLASSIFICATION RÉALISER	75
4.9	COMPARAISON AVEC DES APPROCHES SIMILAIRES	76
	CONCLUSION GÉNÉRALE.....	79
	RÉFÉRENCES BIBLIOGRAPHIQUES.....	81

LISTE DES TABLEAUX

Tableau 1.1 Exemple de nom de fichier de la base de données ICBHI 2017.....	8
Tableau 4.1 Effet de la technique de caractérisation sur les performances de classification.....	58
Tableau 4.2 Paramètres de la première expérience.....	59
Tableau 4.3 Résultats de classification avec 15 neurones cachés.....	62
Tableau 4.4 Résultats de classification avec 20 neurones	63
Tableau 4.5 Résultats de classification avec 30 neurones	63
Tableau 4.6 Résultats de classification avec 40 neurones	63
Tableau 4.7 Résultats de classification avec deux couches cachées de 10 et 20 neurones	64
Tableau 4.8 Résultats de classification avec deux couches cachées de 20 et 30 neurones	64
Tableau 4.9 Résultats de classification avec deux couches cachées de 45 et 55 neurones	64
Tableau 4.10 Combinaisons prises en compte pour l'utilisation de toutes les ressources de la base de données.....	67
Tableau 4.11 Résultats de simulation pour les 18 combinaisons de données avec la méthode GFCC-PCA et un réseau MLP 45-55.....	68
Tableau 4.12 Moyennes des performances obtenus pour les 18 combinaisons de données avec la méthode GFCC-PCA	68
Tableau 4.13 Résultats de simulation pour les 18 combinaisons de données avec la méthode WPT-GFCC et un réseau MLP 45-55	69
Tableau 4.14 Moyennes des performances obtenus pour les 18 combinaisons de données avec la méthode WPT-GFCC.....	69
Tableau 4.15 Les valeurs du Max, min et la moyenne des 1800 résultats.....	71

Tableau 4.16 Résultats de classification par SVM pour 18 combinaisons	72
Tableau 4.17 Résultats de classification par k-NN pour 18 combinaisons	75
Tableau 4.18 Comparaisons des résultats de classification.....	75
Tableau 4.19 Performances de classification présentées par des études utilisant la base de données ICBHI 2017	76

LISTE DES FIGURES

Figure 1.1 Représentation temporelle d'un son respiratoire normal et son spectrogramme	4
Figure 1.2 Représentation temporelle d'un sifflement et son spectrogramme	5
Figure 1.3 Représentation temporelle d'un son respiratoire contenant des crépitants et son spectrogramme.....	6
Figure 1.4 Emplacement où les sons respiratoires ont été recueillis	7
Figure 1.5 Étape de la classification des sons respiratoires.....	11
Figure 2.1 Résolutions temporelle et fréquentielle de la STFT et la WT.....	18
Figure 2.2 Décomposition en sous-bandes de la transformée en ondelettes discrète; $g[n]$ est un filtre passe-haut, $h[n]$ un filtre passe-bas	19
Figure 2.3 Décomposition en sous-bandes de la transformée par paquets d'ondelettes discrète; $h[n]$ est le filtre passe-haut, $g[n]$ le filtre passe-bas	20
Figure 2.4 Schéma de principe de l'extraction des coefficients cepstraux à l'échelle de Mel	24
Figure 2.5 Banc de filtres à l'échelle de Mel.....	25
Figure 2.6 Banc de filtres Gammatone	28
Figure 3.1 Représentation d'un réseau de neurone simple.....	36
Figure 3.2 Modèle d'un neurone simple	37
Figure 3.3 Fonction tangente hyperbolique et fonction sigmoïde	38
Figure 3.4 Description de la Marge	42
Figure 3.5 Règle de décision k-NN pour les valeurs $k = 3$ et $k = 9$	46
Figure 3.6 Exemple de 10-fold cross validation	50
Figure 4.1 Un enregistrement audio contenant six cycles respiratoires.....	54
Figure 4.2 Regroupement des sons de mêmes classes dans les mêmes enregistrements	54
Figure 4.3 Différentes combinaisons de segments de la base de données utilisées.....	55

Figure 4.4 Résultats de simulation initial (la sensibilité, la spécificité, l'exactitude, le score ICBHI et le score harmonique)	60
Figure 4.5 Résultats de simulation pour une fenêtre de taille 512 échantillons.....	61
Figure 4.6 Résultats de simulation pour une fenêtre de taille 2048 échantillons.....	61
Figure 4.7 Matrice de confusion pour la classification des quatre classes obtenues avec la méthode d'extraction des caractéristiques WPT-GFCC en combinaison avec le classifieur basés sur MLP	65
Figure 4.8 La Courbe ROC pour les 4 classes	66
Figure 4.9 Représentation du Max, du Min et de la moyenne des 1800 résultats.....	70
Figure 4.10 Résultats de simulation du classifieur SVM pour 18 combinaisons.....	72
Figure 4.11 Résultats de simulation du classifieur k-NN pour k = 1	73
Figure 4.12 Résultats de simulation du classifieur k-NN pour k = 3	74
Figure 4.13 Résultats de simulation du classifieur k-NN pour k = 5	74

LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES

COPD	: Chronic Obstructive Pulmonary Disease
STFT	: Short-time Fourier transform
WT	: Wavelet transform
CWT	: Continuous wavelet transform
DWT	: Discrete wavelet transform
WPT	: Wavelet packet transform
MFCC	: Mel Frequency Cepstral Coefficients
FFT	: Fast Fourier transform
DCT	: Discrete Cosine transforms
GFCC	: Gammatone Frequency Cepstral Coefficients
ERB	: Equivalent rectangular bandwidth
MLP	: Multilayer perceptron
ANN	: Artificial neural network
SVM	: Support-vector machine
MMH	: maximal margin hyperplane
K-NN	: k-nearest neighbours
GMM	: Gaussian mixture model
EM	: Expectation-Maximisation
VP	: Vrai Positif

VN	: Vrai Négatif
FP	: Faux positif
FN	: Faux négatif
SE	: Sensitivity
SP	: Specificity
ACC	: Accuracy
SM	: Score moyen
SH	: Score harmonique
ROC	: Receiver operating characteristic
ICBHI	: International Conference in Biomedical and Health Informatics
PCA	: Principal component analysis

INTRODUCTION GÉNÉRALE

L'asthme est actuellement la maladie respiratoire la plus courante au Canada, avec environ 2.4 millions de cas, tandis que la maladie pulmonaire obstructive chronique (MPOC) toucherait plus de 830 000 Canadiens. Ces maladies ont causé plus de 23 000 décès en 2020 selon Statistique Canada (2022). Cependant, le nombre de décès au niveau mondial est de 2.6 millions en 2019 selon l'Organisation mondiale de la santé OMS (2020). De nos jours, pas seulement l'asthme, mais diverses maladies respiratoires menacent la santé des gens à travers le monde tel que la pandémie du COVID-19 (maladie de coronavirus 2019). Selon les données recueillies à ce jour, les personnes asthmatiques seraient sur-représentées parmi les adultes hospitalisés pour la COVID-19 (Elissa *et al.*, 2020).

Actuellement, le stéthoscope est largement utilisé pour l'auscultation pulmonaire. Il permet aux médecins spécialistes d'écouter les sons respiratoires du patient. Malgré ses avantages, le diagnostic des sons respiratoires fourni par le stéthoscope repose sur la perception auditive et les connaissances du médecin (Boujelben, 2017). Afin de faire face aux limites de cette technique de diagnostic (analyse subjective), plusieurs approches basées sur le traitement du signal et l'apprentissage machine (approches objectives) ont été proposées dans la littérature pour la classification automatique des sons respiratoires (Lu, 2008).

Pour répondre à cette problématique, un projet de recherche a été lancé, il y a plusieurs années au Département de Mathématiques, d'Informatique et de Génie (DMIG) de l'Université du Québec à Rimouski (UQAR), dont l'objectif principal est de réaliser un système de classification automatique des sons respiratoires pour l'aide au diagnostic.

Notre travail de recherche a pour but de réaliser un système de classification automatique des sons respiratoires basé sur l'apprentissage machine. Le premier chapitre

présente une introduction aux sons respiratoires, la problématique et les objectifs du projet, ainsi que les hypothèses et la méthodologie de travail. Dans le second chapitre, nous présenterons les techniques d'extraction de caractéristiques les plus pertinentes rencontrées dans la littérature. Le troisième chapitre va englober également quelques techniques de classification rencontrées dans les recherches récentes. Le quatrième chapitre présente les différentes étapes effectuées pour répondre à la problématique, ainsi que les résultats obtenus avec les différentes combinaisons et les performances atteintes. Enfin, nous allons finir par une conclusion recapitulant les travaux accomplis et les perspectives d'amélioration.

CHAPITRE 1

DÉFINITION DES SONS RESPIRATOIRES ET LEURS CLASSIFICATION AUTOMATIQUE

Dans ce chapitre, nous allons décrire les sons respiratoires normaux et adventices ainsi que leurs caractéristiques. Nous allons présenter la problématique de classification des sons respiratoires et les objectifs de ce projet, nous présenterons également les hypothèses sur lesquelles nous nous sommes basées pour formuler notre méthodologie.

1.1 NOMENCLATURE DES SONS RESPIRATOIRES

L'auscultation est une procédure clinique qui consiste à écouter, avec un stéthoscope ou directement à l'oreille, les différents sons produits par le corps humain, tels que les sons pulmonaires et les battements de cœur. Cette méthode de diagnostic médical a été développée par René Laennec en 1816 (Lu, 2008). Toutefois, l'un des désavantages de cette méthode est la nature subjective des informations obtenues (Lu, 2008). Les bruits pulmonaires peuvent être classés en deux grandes catégories : normaux et adventices (anormaux). Le son pulmonaire normal correspond au son respiratoire d'un sujet sain, à peine audible sans stéthoscope tandis que le son adventice désigne la superposition d'un son supplémentaire au son normal (Sengupta *et al.*, 2016).

1.1.1 Sons respiratoires normaux

Le passage normal de l'air ambiant inspiré à travers les bronches peut être décrit par deux bruits respiratoires. Les bruits trachéo-bronchiques, entendus à la trachée du larynx ou sur les grandes voies aériennes, et les bruits vésiculaires, obtenus à la surface du thorax (Pelletier, 2006). Le premier son normal de respiration est un son continu d'intensité élevée, aperçu à la fois pendant l'inspiration et l'expiration. Le deuxième son normal est un souffle

continu, doux, et à intensité faible, entendue pendant la totalité de l'inspiration et de l'expiration (Boujelben, 2017). La figure 1.1 est une représentation d'un son de respiration normal.

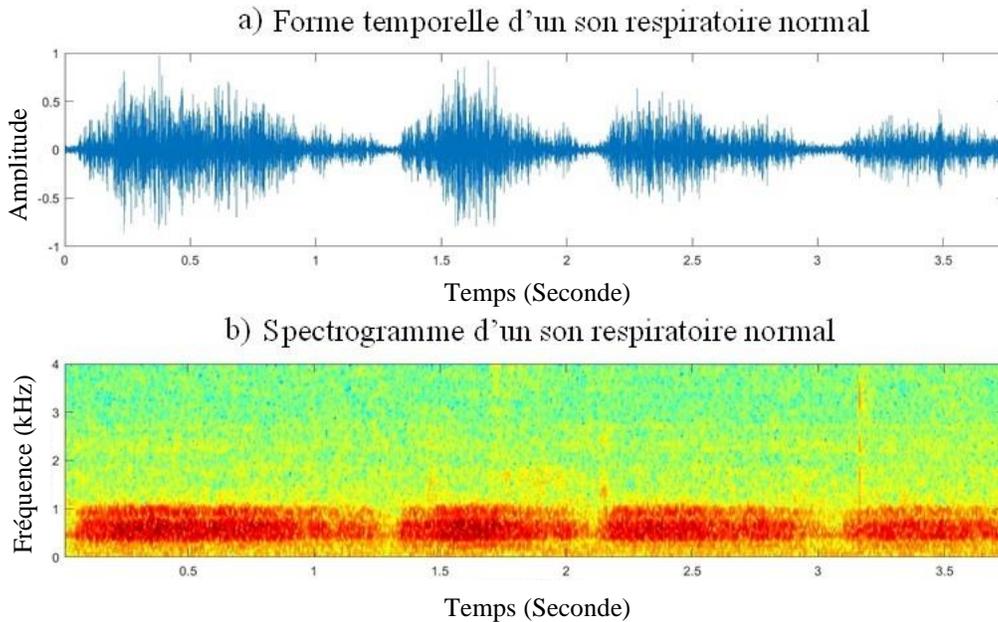


Figure 1.1 Représentation temporelle d'un son respiratoire normal et son spectrogramme

1.1.2 Sons respiratoires adventices

Les sons adventices sont divisés en sons continus et discontinus, dépendamment de leur durée. Les sons continus se subdivisent en sifflements et en rhonchus. Les sons discontinus sont divisés en crépitants fins et en gros crépitants (Bahoura, 2009; Sengupta *et al.*, 2016). Parmi les nombreux sons respiratoires adventices que nous rencontrons dans le domaine médical, on trouve le sifflement (sibilant), le crépitant, la toux, les rhonchus, le stridor, etc. Souvent, on constate que la plupart des autres bruits pulmonaires adventices mentionnés plus haut sont caractérisés par un sifflement ou un crépitant (Sengupta *et al.*, 2016).

1.1.2.1 Sons respiratoires adventices continus

Dans cette sous classe de sons respiratoires adventices, nous trouvons les sifflements et les rhonchus qui ont une forme d'onde continue. La principale différence entre les deux est que les sifflements ont une durée de 250 ms ou plus, tandis que les rhonchus ont une durée de 200 ms ou moins (Lu, 2008). Présentant de manière perceptible un caractère musical, les circonstances conduisant à une respiration sifflante sont nombreuses. Il s'agit de tous les mécanismes qui rétrécissent le diamètre des conduits respiratoires, tels que le bronchospasme, l'œdème de la muqueuse, la compression externe par une tumeur ou l'obstruction dynamique des conduits respiratoires (Bahoura, 2009). Habituellement, les sifflements et les rhonchus sont présents en phase d'inspiration et d'expiration. Les sifflements qui comportent une seule fréquence sont considérés comme « monophoniques » et ceux qui contiennent plusieurs fréquences sont considérés comme « polyphoniques ». Ils peuvent être perçus au niveau de la bouche, de la trachée et du thorax, mais à cause de la cage thoracique qui joue le rôle de filtre passe-bas, les sifflements de hautes fréquences peuvent être éliminés dans les sons perçus sur le thorax (Lu, 2008). La figure 1.2 représente un son respiratoire sifflant dans le domaine temporel ainsi que son spectrogramme où les sifflements sont délimités par des rectangles.

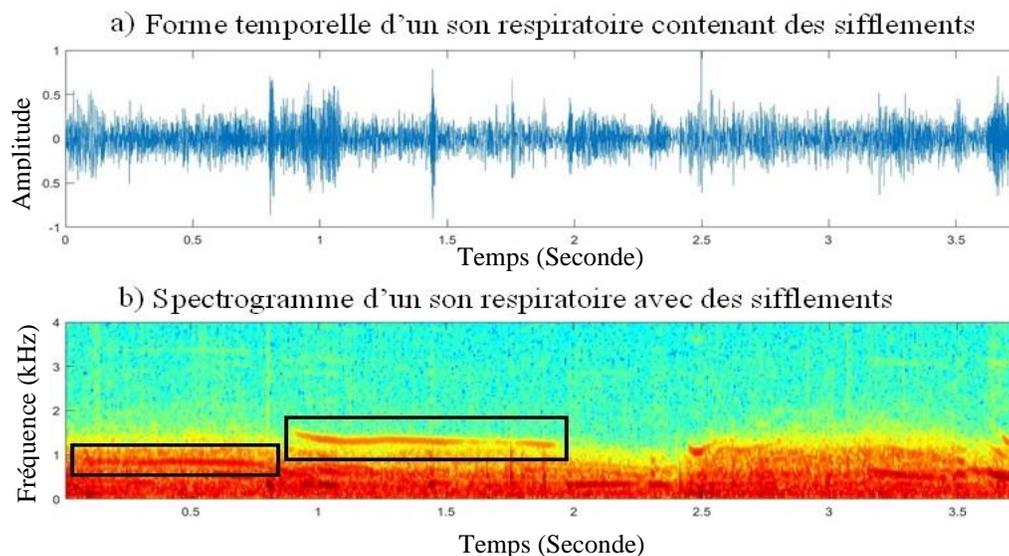


Figure 1.2 Représentation temporelle d'un sifflement et son spectrogramme

1.1.2.2 Sons respiratoires adventices discontinus

Dans cette sous classe, nous trouvons les crépitants qui durent normalement de 1 à 10 ms. En général, les crépitants sont causés par l'ouverture brutale des petites voies respiratoires (Sengupta *et al.*, 2016). Ils sont généralement divisés en deux types: les crépitants fins et les gros crépitants. Pour différencier les crépitants fins des gros crépitants, les critères fréquemment utilisés sont la largeur de déflexion initiale (IDW) et la durée des deux premiers cycles (2CD). Conformément à la définition, les valeurs moyennes de l'IDW et de la 2CD des crépitements fins sont respectivement de 0.7 ms et 5 ms, et celles des gros crépitants sont respectivement de 1.5 ms et 10 ms, autrement dit, les crépitants fins sont caractérisés par une fréquence plus élevée qui occupe une durée courte, tandis que les gros crépitants sont caractérisés par une fréquence plus faible qui occupe une durée plus longue (Lu, 2008). La figure 1.3 représente un son respiratoire contenant des crépitants dans le domaine temporel ainsi que son spectrogramme où les crépitants sont délimités par des rectangles.

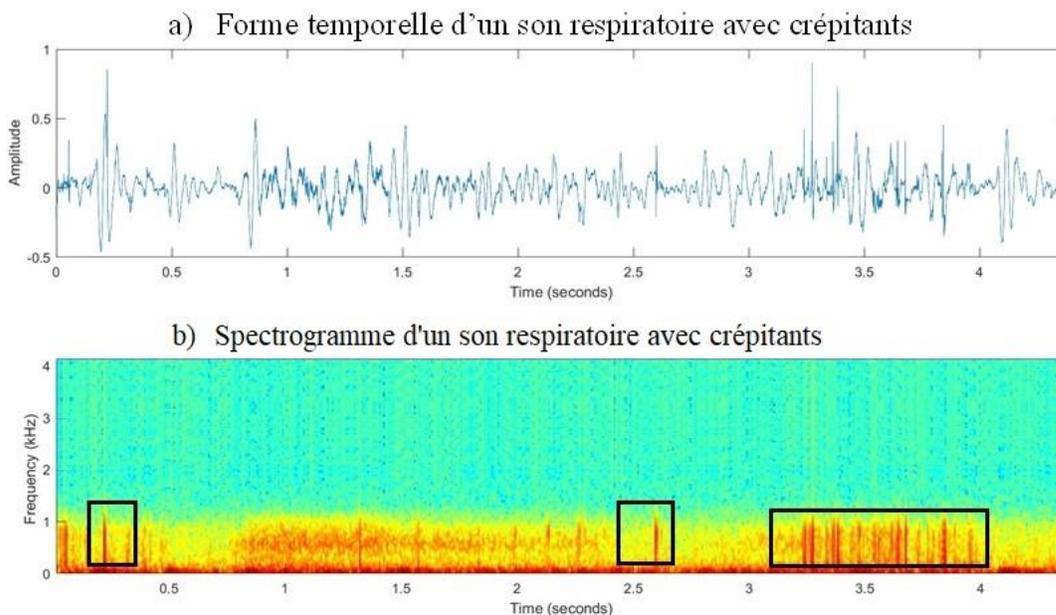


Figure 1.3 Représentation temporelle d'un son respiratoire contenant des crépitants et son spectrogramme

1.1.3 Base de données utilisée

La base de données du défi scientifique ICBHI 2017 contient des enregistrements audio, collectés par deux équipes de recherche dans deux pays différents, sur plusieurs années. L'un des objectifs du défi était la création d'un ensemble de données commun, vaste et ouvert, pour l'analyse des sons respiratoires. La base de données contient 920 enregistrements audio annotés provenant de 126 participants. Les sons ont été recueillis dans six emplacements différents (gauche/droite, antérieure, postérieure et latérale), comme illustrés dans la figure 1.4.

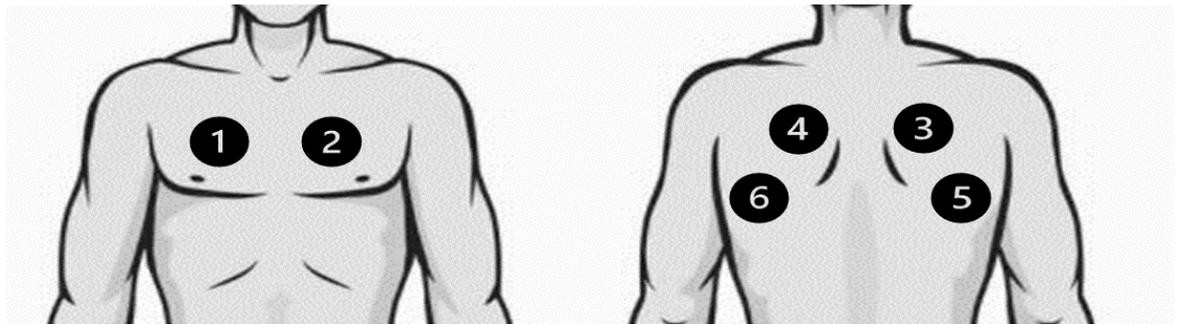


Figure 1.4 Emplacement où les sons respiratoires ont été recueillis : (1) Antérieur droit ou (Ar), (2) Antérieur gauche ou (Al), (3) Postérieur droite ou (Pr), (4) Postérieur gauche ou (Pl), (5) Latérale droite ou (Lr), (6) Latérale gauche ou (Ll)

Les cycles respiratoires dans les enregistrements ont été annotés par des experts comme contenant des crépitations, des sifflements, une combinaison des deux, ou aucun son respiratoire adventice. Les enregistrements ont été recueillis à l'aide d'équipements hétérogènes et leur durée varie de 10 s à 90 s. Chaque nom de fichier comporte 5 éléments spécifiques, comme par exemple « 112_1b1_Ar_Sc_Meditron » (Rocha *et al.*, 2019).

- Numéro du patient (101,102..., 226);
- Indice d'enregistrement (1b1, 1b2, ...);
- Emplacement de l'écoute sur le torse (comme mentionné dans la figure 1.4);
- Mode d'acquisition des données (monocanal (sc), multi-canal (mc));
- Équipement d'enregistrement.

- Microphone AKG C417L (AKGC417L);
- Stéthoscope Littmann Classic II SE de 3M (LittC2SE);
- Stéthoscope électronique 3M Litmmann 3200 (Litt3200);
- Stéthoscope électronique Meditron Master Elite de WelchAllyn (Meditron).

Le tableau 1.1 représente la signification des éléments composants le nom de fichier « 112_1b1_Ar_Sc_Meditron ».

Tableau 1.1 Exemple de nom de fichier de la base de données ICBHI 2017

Éléments	Signification
112	Numéro du patient
1b1	Indice d'enregistrement
Ar	Emplacement (1) de l'écoute sur le torse
Sc	Mode d'acquisition des données
Méditerons	Équipement d'enregistrement

Chacun des fichiers d'annotation comprend quatre colonnes :

- Début du cycle respiratoire;
- Fin du cycle respiratoire;
- Présence/absence de crépitants (présence=1, absence=0);
- Présence/absence de sifflements (présence=1, absence=0).

Donc, les quatre combinaisons (00,01,10 et 11) des deux dernières colonnes correspondent respectivement aux classes de son normal, son avec crépitants, son avec sifflements et son avec crépitants et sifflements.

1.2 PROBLÉMATIQUE DE RECHERCHE

L'auscultation par stéthoscope est une technique qui permet aux médecins de rechercher la présence de bruits pulmonaires normaux et de bruits adventices (anormaux) chez un patient. En dépit de ses avantages, cette technique (subjective) reste limitée étant donné que l'évaluation des sons captés par le stéthoscope dépend de la perception auditive et de l'expertise du médecin. Ces facteurs signifient que l'interprétation des sons perçus par le médecin est subjective et peut être non précise. De là, nous pouvons dire que nous nous retrouvons dans le besoin de réaliser un système de classification automatique de sons respiratoires. Ce dernier serait en mesure d'identifier de façon objective, et assez précise, les différents types de sons respiratoires avec une haute exactitude. Cette méthode serait dédiée à l'assistance des médecins dans la prise de décision lors de l'auscultation, afin d'établir un diagnostic précis des patients.

Aujourd'hui, plusieurs équipes œuvrent dans des projets de recherche qui visent à concevoir et réaliser des systèmes de reconnaissances des sons respiratoires, mais l'utilisation de bases de données différentes complique la comparaison des systèmes réalisés et des résultats obtenus. En 2017, dans le cadre de la conférence ICBHI 2017, une base de données de référence a été rendue publique et regroupe plusieurs classes de sons respiratoires. Elle a été utilisée par plusieurs chercheurs. (Amose *et al.*, 2022; Monaco *et al.*, 2020; Ullah *et al.*, 2021) qui ont proposé l'étude de diverse techniques d'extraction de caractéristiques et de classification automatique basé sur l'apprentissage machine, tandis que d'autres (Petmezas *et al.*, 2022) et (Demir *et al.*, 2020) ont proposé la réalisation de systèmes de classification automatique basés sur l'apprentissage profond.

Dans ce projet de recherche, nous allons utiliser cette base de données pour la réalisation d'un système de classification automatique de sons respiratoires. Ce dernier repose sur l'apprentissage machine, et est capable d'identifier quatre classes de sons (Normaux, sifflements, Crépitants, et le cas contenant les sifflements et les crépitants simultanément).

1.3 OBJECTIFS

L'objectif principal de ce projet est de réaliser un système capable de faire une extraction de caractéristiques des sons respiratoires contenus dans la base de données ICBHI 2017, dont le but est de réduire leurs tailles le plus possible et de garder seulement les composantes significatives. Ensuite, utiliser ces dernières pour faire la classification avec l'exactitude la plus élevée possible. Afin d'y parvenir, nous allons étudier plusieurs méthodes d'extraction de caractéristiques tel que l'analyse par ondelettes et l'analyse cepstrale. Nous allons les comparer en termes de meilleure caractérisation. Enfin, nous étudierons diverses architectures de classifieurs dans le but d'atteindre les résultats de classification (l'exactitude et le score ICBHI) les plus élevés possibles.

1.4 HYPOTHÈSE

Afin d'atteindre les objectifs que nous avons fixés, nous avons formulé une hypothèse basée sur la ressemblance entre les sons de parole et les sons respiratoires. Ainsi, nous allons nous baser sur les méthodes développées dans le domaine de la classification automatique de la parole concernant l'étape d'extraction de caractéristiques.

Cette hypothèse nous conduit vers l'étude de l'analyse par ondelettes, et l'analyse cepstrale, qui sont largement utilisées dans le domaine de la parole. Pour l'étape de classification automatique, nous allons également se baser sur les techniques proposées dans le domaine de la parole tels que les réseaux de neurones (MLP), les machines à vecteur de support (SVM) et le k-plus proches voisins (k-NN).

1.5 MÉTHODOLOGIE

La méthodologie adoptée dans notre projet opère en deux étapes principales. En premier lieu, l'étape d'extraction des caractéristiques, qui vise à créer un espace de caractéristiques représentant le signal original, mais avec de faibles redondances. Cette dernière doit être effectuée minutieusement, car c'est le nouvel espace qui va permettre de bien mener l'étape suivante. La deuxième étape est la classification dans laquelle les sons respiratoires sont reconnus et classés selon leurs catégories d'appartenance, elle passe par deux grandes phases qui sont la phase d'apprentissage et la phase de test comme le montre la figure 1.5. Nous avons opté pour les techniques de classifications basées sur l'apprentissage machine.

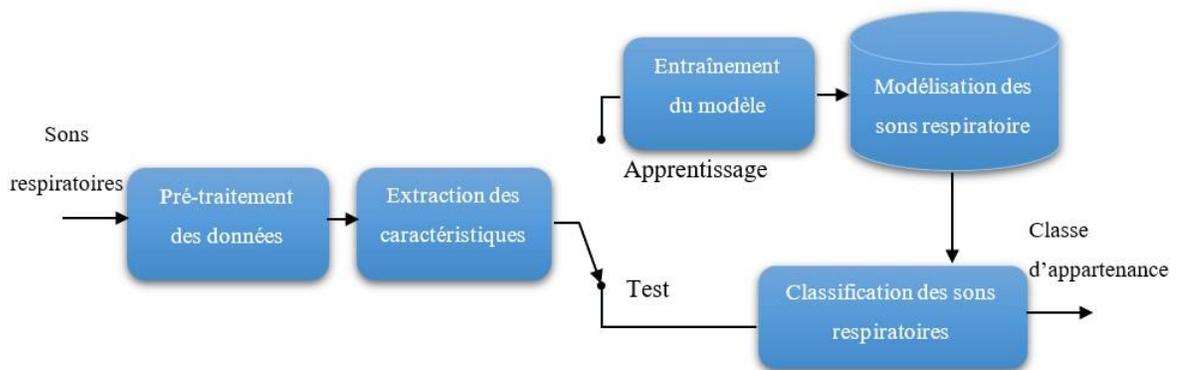


Figure 1.5 Diagramme d'un système de classification des sons respiratoires

La méthodologie adoptée sera brièvement décrite dans les deux sections suivantes :

1.5.1 Extraction de caractéristiques des sons respiratoires

Plusieurs techniques d'extraction de caractéristiques seront étudiées et testées durant la réalisation de notre projet de recherche. Parmi elles, les coefficients cepstraux en fréquences Gammatones (GFCC) qui sont conçus pour simuler le processus du système auditif humain selon Dua *et al.* (2018), les coefficients cepstraux à l'échelle de Mel (MFCC), la transformée en ondelettes (WT), les paquets d'ondelettes (WPT), ainsi que des combinaisons WPT avec les coefficients MFCC et GFCC. Généralement les coefficients

GFCC sont jugés plus robustes que les coefficients MFCC quand il s'agit de données bruitées, c'est pourquoi cette technique a été retenue pour la suite du projet avec les WPT qui se sont avérés prometteurs en ce qui concerne la caractérisation des crépitants (Quandt *et al.*, 2015).

1.5.2 Classification automatique des sons respiratoires

Trois modèles de classification seront étudiés durant la réalisation de notre projet de recherche concernant l'étape de la classification des sons respiratoires : perceptrons multicouches (MLP), machines à vecteur de support (SVM) et le k-plus proches voisins (k-NN). Le réseau MLP est probablement l'architecture la plus fréquemment utilisé pour les applications pratiques des réseaux de neurones. Dans la plupart des cas, le réseau se compose de deux couches, de poids ajustable lors de l'entraînement, avec une connectivité totale entre les entrées et les unités cachées et entre les unités cachées et les sorties (Nabney, 2002). Ce type de classifieur a la capacité d'effectuer un apprentissage à partir de données et de tester ensuite avec un nouvel ensemble de données différentes, ce qui le rend bien adapté pour le cas de classification de sons respiratoires. Le SVM est un classifieur non linéaire qui maximise la marge des échantillons dans l'hyperplan (Meng *et al.*, 2020). Étant le principal concurrent du MLP, le SVM est le deuxième classifieur proposé à être étudié et testé dans notre projet. Enfin, le troisième classifieur étudié est le k-NN, qui classe les données en fonction des caractéristiques les plus proches (Taunk *et al.*, 2019).

1.6 CONTRIBUTION

La contribution que nous avons apportée dans ce projet est située au niveau de l'étape d'extraction des caractéristiques. En effet, dans la littérature, les chercheurs utilisent souvent une ou plusieurs techniques de caractérisation des sons respiratoires dans un but comparatif, tel que l'analyse par ondelettes (Kandaswamy *et al.*, 2004), ou l'analyse cepstrale (Bahoura, 2009; Dua *et al.*, 2018; Rizal *et al.*, 2015; Semmad, 2021). Dans notre projet, nous avons proposé l'utilisation de l'analyse par ondelettes, et de l'analyse cepstrale, ainsi qu'une

combinaison de ces deux techniques d'analyse. La première proposition était d'effectuer des tests en combinant les coefficients extraits à partir des WPT et les coefficients MFCC. La deuxième proposition était de réaliser une combinaison entre les coefficients WPT et les coefficients GFCC. Cette dernière a mené à d'excellents résultats au niveau de la classification en termes d'exactitude et de score ICBHI.

CHAPITRE 2

EXTRACTION DES CARACTÉRISTIQUES

Dans ce chapitre, nous allons présenter quelques techniques d'extraction de caractéristiques rencontrées dans la littérature. Nous allons mettre l'emphase sur celles utilisées dans notre projet de recherche.

2.1 DÉFINITION ET PRINCIPE DE L'EXTRACTION DES CARACTÉRISTIQUES

L'extraction de caractéristiques est une étape du traitement du signal qui va, dans de nombreux cas, de pair avec la classification. Elle débute à partir d'un ensemble de données brutes, et produit des valeurs destinées à être utiles et non redondantes, ce qui simplifie la phase de classification qui suit. L'extraction de caractéristiques peut être qualifiée également de processus de réduction des données, qui tente de ne capturer que l'essentiel des caractéristiques du signal analysé avec un faible débit de données, afin de réduire la complexité des calculs dans les étapes suivantes (Bahoura, 2009). Le principe de l'extraction des caractéristiques consiste à diviser un signal en segments de même taille (fenêtrage). Chaque segment peut être converti en une représentation à une, à deux voir plusieurs composantes. Le nombre de composantes choisi correspondra à la dimension de l'espace des caractéristiques (Pelletier, 2006).

Dans ce chapitre, nous allons présenter les principales techniques proposées dans la littérature pour l'extraction de caractéristiques dans les sons respiratoires. Ces techniques seront globalement réparties en quatre groupes :

- Analyse de Fourier.
- Analyse par ondelettes.
- Analyse cepstrale.

- Caractéristiques statistiques.

2.2 ANALYSE DE FOURIER

L'analyse de Fourier consiste à calculer et représenter le contenu fréquentiel d'un signal. La transformée de Fourier discrète (DFT) d'un signal discret $x[n]$ est donnée par l'équation (2.1) (Pelletier, 2006) :

$$S[k] = \sum_{n=0}^{N-1} s[n] e^{-j \frac{2\pi nk}{N}} \quad k = 0 \dots N-1 \quad (2.1)$$

où n est l'indice du temps, k est l'indice de fréquence et N le nombre d'échantillons du signal $s[n]$ analysé, qui correspond également au nombre de fréquences discrètes de $S[k]$.

L'analyse de Fourier est particulièrement pertinente dans le cas des sons respiratoires. En effet, la présence de sifflements est généralement représentée par des pics de fréquences qui s'ajoutent aux spectres des sons normaux (Pelletier, 2006).

Cependant, la transformée de Fourier à court terme STFT a été proposée pour analyser les signaux non-stationnaires. Elle permet d'analyser l'évolution temporelle du spectre d'un signal. La transformée de Fourier à court terme (STFT) d'un signal à temps discret $s[n]$ est définie par l'équation (2.2)

$$S[m, k] = \sum_{n=0}^{N-1} s[n] w[n - mL] e^{-j \frac{2\pi nk}{N}} \quad (2.2)$$

où $w[n]$ est une fonction de fenêtrage de pondération de N échantillons, L est le pas d'avancement. N représente aussi le nombre de fréquences discrètes (Bahoura, 2009). Avec la transformée STFT, une fenêtre plus étroite améliorera la résolution temporelle de la représentation temps-fréquence, mais entraînera une résolution fréquentielle plus faible. Inversement, une fenêtre plus large améliorera la résolution fréquentielle mais entraînera une résolution temporelle plus faible (Yan *et al.*, 2020). Comme la transformée de Fourier est une

fonction complexe, la densité spectrale de puissance (DSP) est généralement utilisée et qui est donnée par l'équation (2.3)

$$P_s[m, k] = \frac{1}{N} |S[m, k]|^2 \quad (2.3)$$

À la fréquence d'échantillonnage f_s , chaque signal fenêtré (trame) est représenté par un spectre de puissance à N points couvrant la plage de fréquences $[-f_s/2, f_s/2]$.

Dans la recherche de Rizal *et al.* (2015), les données ont été analysées dans le domaine temporel et dans le domaine fréquentiel. La transformation du domaine temporel au domaine fréquentiel a été effectuée à l'aide d'une transformée de Fourier rapide (FFT) de 256 échantillons. Tandis que dans Rizal *et al.* (2022) la meilleure exactitude de classification de 100% a été obtenue par la méthode STFT.

2.3 ANALYSE PAR ONDELETTES

La transformée en ondelettes vise à pallier aux limitations de la STFT liées aux résolutions temporelle et fréquentielle et à faire un bien meilleur traitement. En effet, la STFT a été employée pour effectuer l'analyse en temps et en fréquence. Cependant, il s'est avéré que cette technique ne permet pas une amélioration simultanée des résolutions temporelle et fréquentielle. Une réduction de la fenêtre d'analyse de la STFT améliore la résolution temporelle au détriment de la résolution fréquentielle et vice-versa.

La transformée en ondelettes (WT) est l'une des techniques les plus utilisées pour l'extraction des caractéristiques des signaux biomédicaux. Elle définit d'abord une fonction de base spécifique nommée "ondelette mère" qui est localisée en temps et en fréquence. On peut ainsi générer une série de fonctions d'ondelettes par translation et compression/dilatation, chacune analysera une partie du spectre temps-fréquence (Rajoub, 2020). La figure 2.1 montre la principale différence entre la transformée de Fourier à court-terme et la transformée en ondelette (Wang & Karayiannis, 1998).

Dans la transformée de Fourier à court terme (STFT), une fenêtre coulissante de taille fixe est utilisée pour calculer la transformée de Fourier du signal. Il en résulte une représentation temps/fréquence uniforme du signal. Cependant, le calcul de la transformée en ondelettes (WT) implique une application unique d'un banc de filtres sur l'ensemble du signal. Ainsi, nous nous retrouvons avec une représentation temps/fréquence qui n'est pas uniforme (c'est-à-dire une résolution grossière à fine) (Rajoub, 2020).

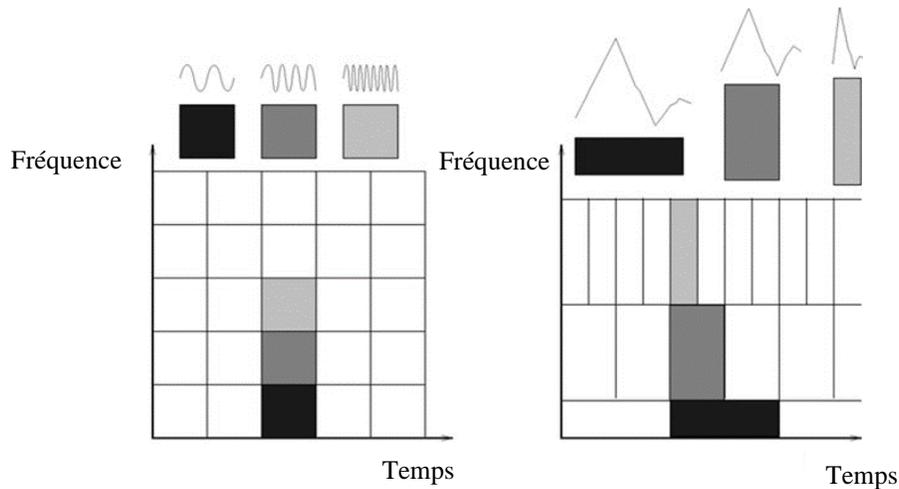


Figure 2.1 Résolutions temporelle et fréquentielle de la STFT et la WT (Wang & Karayiannis, 1998)

2.3.1 Transformée en ondelettes continue (CWT)

La transformée en ondelettes continue d'un signal $s(t)$ est définie par :

$$CWT_s^\tau(\tau, a) = \frac{1}{\sqrt{a}} \int s(t) \psi^* \left(\frac{t - \tau}{a} \right) dt \quad (2.4)$$

où a désigne le paramètre d'échelle de l'ondelette, qui peut être considéré comme représentant de l'inverse de la fréquence et τ représente le décalage temporel.

2.3.2 Transformée en ondelette discrète (DWT)

La transformée en ondelettes discrète (DWT) est une implémentation discrète de la transformée en ondelettes continue. En outre, cette méthode permet de décomposer un signal donné en un certain nombre d'ensembles, où chaque ensemble est une série temporelle de coefficients décrivant l'évolution temporelle du signal dans la bande de fréquence correspondante (Hosseinzadeh, 2020). Cela constitue la principale différence avec la transformée en ondelettes continue (CWT). Il existe plusieurs types d'implémentation de l'algorithme DWT. Le plus ancien et le plus populaire est l'algorithme pyramidal de Mallat (Mallat, 1989). Dans cet algorithme, deux filtres sont utilisés de manière récurrente pour obtenir des données pour toutes les échelles. Dans la littérature plusieurs familles d'ondelettes ont été utilisées, parmi celles-ci : Haar, Daubechies, Biorthogonal, Symlets, Coiflets etc. La procédure de décomposition multi-résolution d'un signal à temps discret $s[n]$ est représentée schématiquement par la figure 2.2.

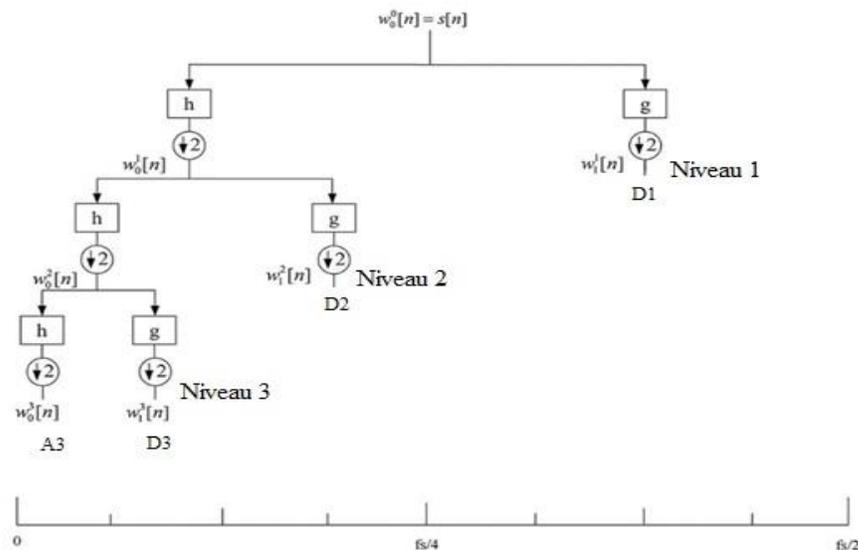


Figure 2.2 Décomposition en sous-bandes de la transformée en ondelettes discrètes; $g[n]$ est un filtre passe-haut, $h[n]$ un filtre passe-bas (Bahoura, 2009)

Chaque niveau de décomposition de ce schéma consiste en deux filtres numériques et deux sous-échantillonneurs par 2. Le premier filtre, $g[n]$ à base de l'ondelette mère, passe-

haut par nature, le second, $h[n]$ est sa version miroir, passe-bas par nature (Joshi *et al.*, 2017). Ainsi, la décomposition par ondelettes à trois niveaux ($j=3$) de la figure 2.2 fournit trois détails D1-D3 et une approximation A3.

2.3.3 Transformée par paquets d'ondelettes (WPT)

La transformée par paquets d'ondelettes (WPT) peut être considérée comme une extension de la transformée en ondelettes (Voir la figure 2.3). Alors que la transformée en ondelettes discrète offre une résolution temps-fréquence flexible, elle souffre d'une résolution relativement limitée dans la région des hautes fréquences. Cette déficience entraîne une difficulté à distinguer les composantes transitoires à haute fréquence (Gao & Yan, 2011). Contrairement à la transformée en ondelette, qui est obtenue en itérant la branche passe-bas $h[n]$, la transformée par paquets d'ondelettes est obtenue en itérant l'une ou l'autre des branches à n'importe quel niveau j de décomposition.

La décomposition par paquets d'ondelettes est équivalente à un filtrage multicanal où le nombre de filtres et leur largeur de bande sont liés au niveau j de décomposition.

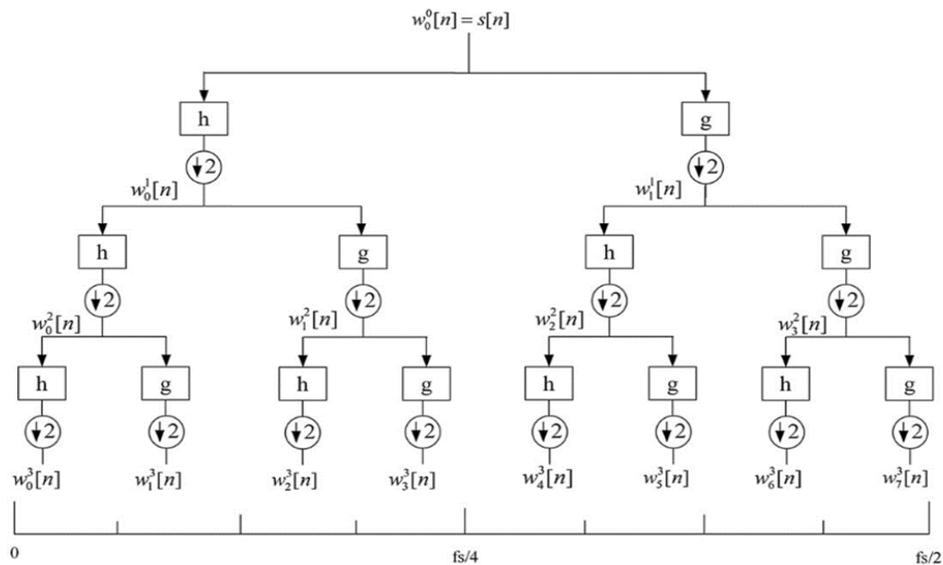


Figure 2.3 Décomposition en sous-bandes de la transformée par paquets d'ondelettes discrète; $h[n]$ est le filtre passe-haut, $g[n]$ le filtre passe-bas (Bahoura, 2009)

La figure 2.3 illustre, de manière schématique, un processus de décomposition par paquets d'ondelettes (WPT) du signal $s[n]$ à trois niveaux produisant un total de huit sous-bandes. Chaque sous-bande couvre un huitième de la bande fréquentielle du signal. Les coefficients $w_k^{j+1}[n]$ des paquets d'ondelettes représentés dans la figure 2.3 correspondant aux sous-bandes paires et impaires nommés k .

Les coefficients extraits par WPT fournissent une représentation compacte qui montre la distribution de l'énergie du signal en temps et en fréquence. Dans le système proposé par Kandaswamy *et al.* (2004), le nombre de niveaux de décomposition a été choisi en fonction des composantes de fréquence dominantes du signal. Les niveaux ont été choisis de telle sorte que les parties du signal qui sont en bonne corrélation avec les fréquences requises pour la classification du signal soient conservées dans les coefficients d'ondelette. Puisque les sons respiratoires n'ont pas de composantes fréquentielles utiles en dessous de 50 Hz, le nombre de niveaux a été choisi égale à 7. Ainsi, le signal est décomposé en détails D1-D7 et une approximation A7. Les coefficients d'ondelettes extraits fournissent une représentation compacte qui montre la distribution d'énergie du signal en temps et en fréquence. Il a été observé que les valeurs des coefficients sont très proches de zéro en D1, D2 et A7. Par conséquent, les coefficients correspondant aux bandes de fréquences D1, D2 et A7 ont été éliminés, réduisant alors le nombre de coefficients représentant le signal.

Dans le système de Kandaswamy *et al.* (2004), les caractéristiques suivantes ont été utilisées pour représenter la distribution temps-fréquence des sons respiratoires :

- Moyenne des valeurs absolues des coefficients dans chaque sous-bande;
- Puissance moyenne des coefficients d'ondelettes dans chaque sous-bande;
- Écart-type des coefficients dans chaque sous-bande;
- Rapport des valeurs moyennes absolues des sous-bandes adjacentes.

où les deux premières caractéristiques représentent la distribution de fréquence du signal, la troisième et la quatrième caractéristiques représentent l'importance des changements dans la distribution de fréquence. Dans le système proposé par Hashemi *et al.* (2011), les sifflements

ont également été décrits avec sept détails, D1 à D7, et une approximation, A7. En extrayant les caractéristiques statistiques des coefficients d'ondelettes, la dimension de l'espace du signal analysé sera réduite. Six caractéristiques statistiques ont été extraites des coefficients d'ondelettes.

- Moyenne des valeurs absolues dans chaque sous-bande;
- Puissance moyenne dans chaque sous-bande;
- Écart-type dans chaque sous-bande;
- Rapport des valeurs moyennes absolues des sous-bandes adjacentes;
- Skewness dans chaque sous-bande;
- Kurtosis dans chaque sous-bande.

Les caractéristiques de la moyenne représentent la distribution fréquentielle du signal, mais l'écart type et le rapport des moyennes des valeurs absolues des sous-bandes adjacentes montrent l'importance des changements dans la distribution fréquentielle. La différence entre (Hashemi *et al.*, 2011) et (Kandaswamy *et al.*, 2004) est l'utilisation de l'asymétrie (Skewness) et de l'aplatissement (Kurtosis) dans les sous bandes. L'asymétrie est une mesure de l'asymétrie de la distribution de probabilité d'un signal à valeurs réelles et l'aplatissement (Kurtosis) mesure la "crête" de la distribution de probabilité d'un signal à valeurs réelles. Sept familles ondelettes différentes, comprenant Haar, Symmlets d'ordre 8, Daubechies d'ordre 2, 8 et 10 et Bi-orthogonales d'ordre 1.5 et 2.8, ont été utilisées de cette façon pour extraire 46 caractéristiques de chaque signal en utilisant chacune d'elle.

2.4 ANALYSE CEPSTRALE

L'analyse cepstrale a été initialement proposée dans le domaine du traitement automatique de la parole. Les sons respiratoires sont considérés également comme le résultat de la convolution entre une source sonore (diaphragme) et les voies respiratoires (Pelletier, 2006). L'opération de convolution définissant le son respiratoire est modélisée par l'équation (2.5) suivante :

$$y[n] = x[n] * h[n] \quad (2.5)$$

où $x[n]$ est le signal d'excitation de la parole et $h[n]$ est la réponse du conduit vocal (voies respiratoires). L'utilisation de l'analyse cepstrale a été proposée afin de pouvoir séparer l'ajout apporté par les voies respiratoires du son généré par la source elle-même.

Afin de passer du domaine temporel au domaine fréquentiel, nous appliquons la transformée de Fourier, ainsi la convolution devient une multiplication comme le montre l'équation (2.6) (Pelletier, 2006) :

$$Y(e^{j\omega}) = X(e^{j\omega})H(e^{j\omega}) \quad (2.6)$$

Étant donné la difficulté de séparer les deux signaux, la transformation par la fonction logarithmique a été proposée pour séparer $x[n]$ et $h[n]$ dans le domaine fréquentiel, comme donné dans l'équation (2.7) (Pelletier, 2006):

$$\log(Y(e^{j\omega})) = \log(X(e^{j\omega})) + \log(H(e^{j\omega})) \quad (2.7)$$

Dans la majorité des applications du traitement du signal, nous n'avons besoin que des spectres d'amplitude, ainsi l'équation (2.7) devient (Pelletier, 2006):

$$\log(|Y(e^{j\omega})|) = \log(|X(e^{j\omega})|) + \log(|H(e^{j\omega})|) \quad (2.8)$$

Enfin, les coefficients cepstraux (Cc) sont obtenus par la transformée de Fourier discrète inverse (DFT^{-1}), selon l'équation (2.9) (Pelletier, 2006):

$$Cc[k] = DFT^{-1} \log(|X(e^{j\omega})|) + DFT^{-1} \log(|H(e^{j\omega})|) \quad (2.9)$$

où l'indice k correspond aux raies spectrales. Les premières composantes correspondent à la réponse en fréquence des voies respiratoires, les secondes correspondent à la réponse en fréquence de l'excitation, d'où notre intérêt. Ces coefficients ont été précédemment utilisés pour étudier le processus de génération et de transmission du son respiratoire (Bahoura, 2009). Dans la suite du chapitre, nous allons aborder deux techniques d'analyse cepstrale : les MFCC et les GFCC.

2.4.1 Coefficients cepstraux à l'échelle de Mel (MFCC)

L'une des techniques les plus utilisées dans les différents types d'applications de reconnaissances vocales est les coefficients cepstraux à l'échelle de Mel (MFCC). Cette approche répond au besoin de séparer le contenu du signal d'excitation de la parole de celui de la réponse du conduit vocal. Les MFCC reposent sur les principes du système auditif, ainsi que la propriété de décorrélation du cepstre (Bahoura, 2009). Ces caractéristiques sont obtenues à partir du spectre de puissance, en appliquant un banc de filtres sur une échelle fréquentielle modifiée, appelée l'échelle de Mel (Pelletier, 2006). L'algorithme d'extraction des caractéristiques des coefficients MFCC comprend les étapes représentées par la figure 2.4 :

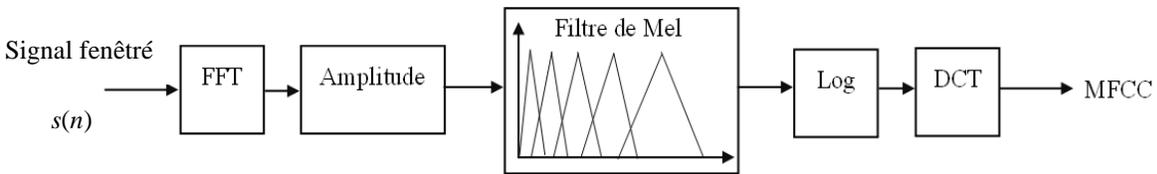


Figure 2.4 Schéma de principe de l'extraction des coefficients cepstraux à l'échelle de Mel

- Le signal à temps discret $s[n]$ est initialement divisé en trames $s[m,n]$ de N échantillons chacune, comme le montre l'équation (2.10) :

$$s[m,n] = s[n]w[n - mL] \quad (2.10)$$

où m est l'indice de trame, $w[n]$ est une fenêtre de pondération de N échantillons et L est le pas de décalage en échantillons (Bahoura, 2018). Le taux de chevauchement entre les fenêtres adjacentes est $(N-L)/N$.

- Chaque trame $s[m,n]$ est ensuite transformée dans le domaine des fréquences à l'aide de la transformée de Fourier à court terme (STFT) (Bahoura, 2018; Jeevan *et al.*, 2017), selon l'équation (2.11) :

$$S[m, k] = \sum_{n=0}^{N-1} s[m, n] e^{-j \frac{2\pi nk}{N}} \quad (2.11)$$

où N est la taille de la transformée qui représente également le nombre de fréquences discrètes et k est l'indice de fréquence ($k = 0, \dots, N - 1$)

- Ensuite, le spectre d'énergie est calculé comme suit :

$$|S[m, k]|^2 = S_r^2[m, k] + S_i^2[m, k] \quad (2.12)$$

où $S_r[m, n]$ et $S_i[m, k]$ représentent respectivement la partie réelle et la partie imaginaire.

- Le spectre d'énergie passe ensuite par M filtres triangulaires espacés dans l'échelle de Mel (voir la figure 2.5). La sortie du l ème filtre est calculée par la somme pondérée de sa réponse en fréquence $H_l(k)$ et du spectre d'énergie (équation 2.13).

$$E(m, l) = \sum_{k=0}^{N-1} |S[m, k]|^2 H_l(k) \quad (2.13)$$

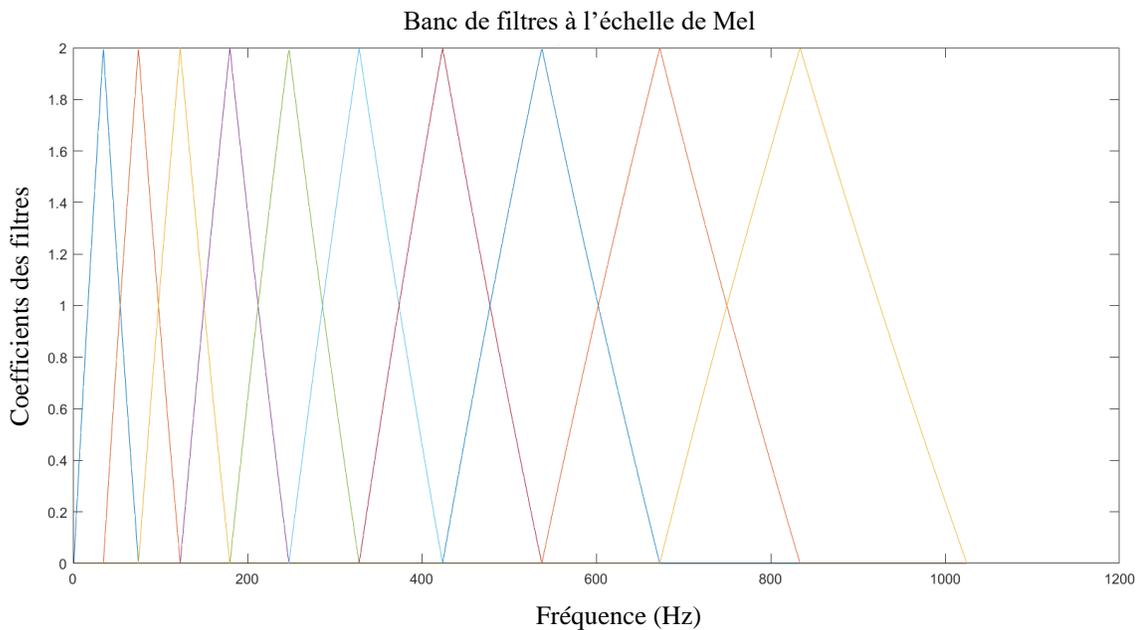


Figure 2.5 Banc de filtres à l'échelle de Mel

où $H_l(k)$ est la fonction de transfert du l ème filtre ($l = 1 \dots M$) et M est le nombre de filtres.

- L'échelle de Mel peut être approximée par l'équation (2.14) (Pelletier, 2006)

$$Mel(f) = 2595 \log_{10} \left(1 + \left(\frac{f}{700} \right) \right) \quad (2.14)$$

- L'équation (2.15) montre la façon dont les filtres triangulaires sont obtenus (Semmad, 2021) :

$$H_l(k) = \begin{cases} 0 & \text{si } k < k_c[l-1] \\ \frac{k - k_c(l-1)}{k_c(l) - k_c(l-1)} & \text{si } k_c[l-1] \leq k \leq k - k_c[l] \\ \frac{k_c(l+1) - k}{k_c(l+1) - k_c(l)} & \text{si } k_c[l] \leq k \leq k - k_c[l+1] \\ 0 & \text{si } k > k_c[l+1] \end{cases} \quad (2.15)$$

où $k_c(l)$ est l'indice correspondant à la fréquence centrale du l ème filtre.

- Ensuite, une fonction logarithmique est appliquée aux énergies filtrées par les filtres Mel, comme dans l'équation (2.16) :

$$e[m, l] = \log_{10}(E[m, l]) \quad (2.16)$$

- Enfin, le spectre d'énergie logarithmique est retransformé par la transformée en cosinus discrète (DCT) et le n ième coefficient MFCC est calculé comme suit :

$$c_m[n] = \sum_{l=1}^M e[m, l] \cos\left(n\left(l - 0.5\right) \frac{\pi}{M}\right) \quad (2.17)$$

où $n = 0 \dots, P - 1$ et $P \leq M$ est le nombre de coefficients requis.

Dans le système réalisé par Dua *et al.* (2018), après avoir effectué un pré-traitement sur le signal vocal d'entrée il a été paramétré à l'aide de diverses techniques d'extraction de

caractéristiques cepstrales, tels que les coefficients MFCC. Initialement, 12 coefficients MFCC ont été utilisés. Ensuite, les chercheurs ont proposé une caractérisation supplémentaire, connue sous le nom d'énergie de la trame, qui est également ajoutée, donc le vecteur de caractéristiques comporte 13 composantes MFCC au lieu de 12. Tandis que dans le système proposé par Valente *et al.* (2020), une représentation visuelle (image) de chaque segment audio provenant de la base de données ICBHI 2017 a été réalisée. Pour chaque fichier audio de l'ensemble de données, les coefficients MFCC ont été extraits, ce qui signifie que chaque segment audio a une représentation par une image, semblable à un spectrogramme, mais avec des détails plus distincts. La taille des images de 224x140 a été utilisée.

2.4.2 Coefficients cepstraux des fréquences Gammatone (GFCC)

Le GFCC est un modèle plus complet basé sur l'échelle ERB (*Equivalent rectangular band-width*) et un ensemble de bancs de filtres Gammatone. Il est conçu pour simuler aussi le comportement du système auditif humain. La principale différence entre les MFCC et les GFCC est leurs bancs de filtres. Le banc de filtres Gammatone est un groupe de filtres dont les réponses impulsionnelles sont similaires à la magnitude caractéristique du filtre auditif humain (Dua *et al.*, 2018). Les étapes de fenêtrage et le passage par la transformée de Fourier sont les mêmes que celle des MFCC et sont exprimées par les équations (2.10) et (2.11). Les autres étapes suivantes sont définies comme suit :

- Le signal de sortie de la transformée de Fourier passera ensuite à travers un banc de filtres Gammatone (voir la figure 2.6). Ce dernier, avec une fréquence centrale f_c , sera défini selon l'équation suivante (Semmad, 2021) :

$$g(f, t) = at^{(n-1)}e^{(-2\pi bt)}\cos(2\pi f_c t + \phi) \quad (2.18)$$

où a est une constante, b est un facteur qui détermine la largeur de bande du filtre et dépend de f_c , ϕ détermine la phase et n l'ordre du filtre. Les coefficients GFCC disposent de filtres plus rapprochés pour les basses fréquences et des filtres plus larges et plus espacés pour les hautes fréquences.

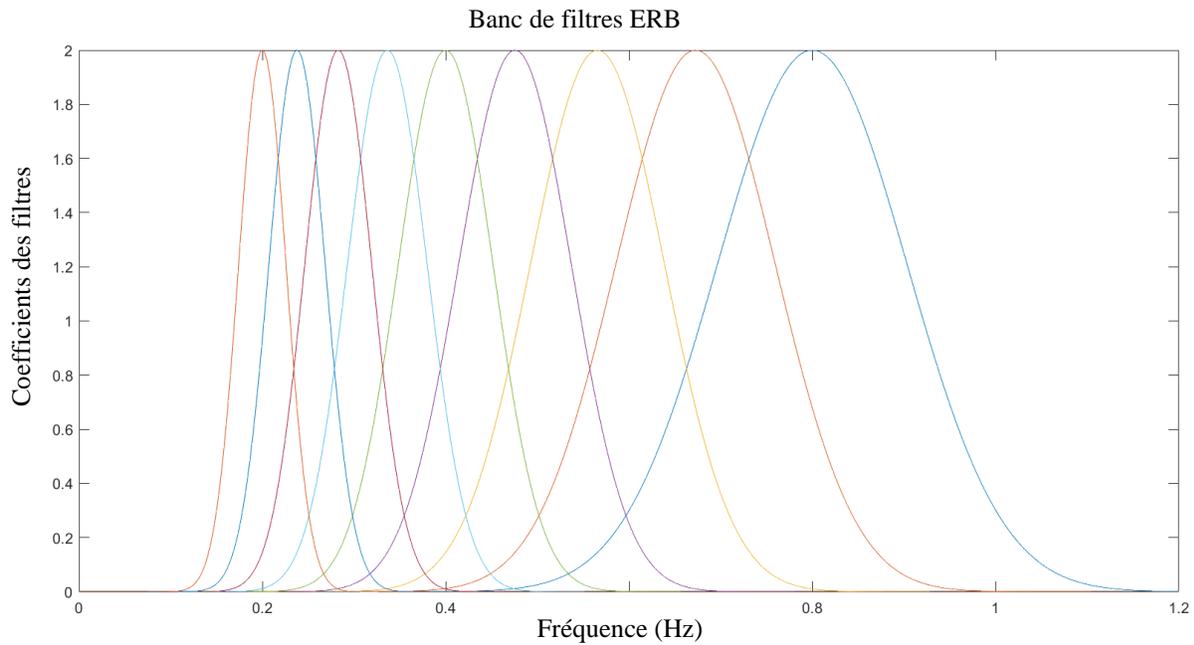


Figure 2.6 Banc de filtres Gammatone

Dua *et al.* (2018) ont fixé n à 4, car c'est l'ordre du filtre Gammatone qui est le plus proche des filtres auditifs humains. Le nombre de filtres dans le GFCC détermine la robustesse au bruit du banc de filtres en général, donc il a été fixé à 128.

- La fréquence centrale f_c et la largeur de bande b peuvent être dérivées de la largeur de bande rectangulaire équivalente (ERB) du filtre (Admuthe & Patil, 2015; Jeevan *et al.*, 2017), selon les équations (2.19) et (2.20) :

$$ERB(f_c) = 24.7(4.37 \frac{f_c}{1000} + 1) \quad (2.19)$$

$$b = 1.019ERB(f_c) = 25.17(4.37 \frac{f_c}{1000} + 1) \quad (2.20)$$

Dans l'étude proposée par Jeevan *et al.* (2017), le signal de parole a été multiplié par le banc de filtres Gammatone dans le domaine des fréquences. Par conséquent, ce signal est à nouveau ramené dans le domaine temporel à l'aide d'une transformée de Fourier inverse, il a été ensuite décimé à 100 Hz pour réduire les effets du bruit.

2.5 CARACTÉRISTIQUES STATESTIQUES

Les caractéristiques statistiques ont été employées dans de nombreuses applications de traitement des signaux et de données. Elles modélisent les données en termes de propriétés statistiques telles que le maximum, le minimum, la médiane, le mode et les moments statistiques tels que la moyenne, la variance, l'aplatissement (Kurtosis) et l'asymétrie (Skewness) (Rajoub, 2020).

Les moments statistiques permettent de recueillir des informations sur les propriétés de la distribution des caractéristiques. Sur la base d'échantillons d'une variable aléatoire x , nous sommes en mesure de déterminer les moments statistiques liés à l'origine de n'importe quel ordre α à l'aide de la série géométrique comme le montre l'équation (2.21) :

$$E[X^\alpha] = \frac{1}{N} \sum_{i=1}^N x_i^\alpha \quad (2.21)$$

où N est le nombre d'échantillons générés à partir d'une variable aléatoire x , x_i est le i ème échantillon, et $E[X^\alpha]$ est l'opérateur d'espérance. Dans Shi *et al.* (2018) Les moments statistiques suivants ont été proposés :

- Le premier moment ($\alpha = 1$) représente la moyenne (μ) des données et est exprimé par

$$\mu = E(x) = \frac{1}{N} \sum_{i=1}^N x_i \quad (2.22)$$

La moyenne d'une série de données est la somme des valeurs divisée par le nombre total de valeurs. Il existe d'autres paramètres alternatifs, tels que le mode ou la médiane, qui sont plus résistants aux valeurs aberrantes. Une solution pour détecter les valeurs aberrantes, dans le cas de variables aléatoires unimodales, consiste à comparer les valeurs de la moyenne et de la médiane. Si la différence entre la moyenne et la médiane est assez élevée, il est probable que des valeurs aberrantes soient présentes dans les données. Nous trouvons également la moyenne quadratique qui est un type de moyenne. Elle mesure la magnitude absolue d'un ensemble de nombres et se calcule comme suit :

$$\mu_{rms} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \quad (2.23)$$

- Le deuxième moment est la variance σ^2 , qui est obtenu en soustrayant la moyenne des données de chaque point et en fixant $\alpha = 2$. La racine carrée de la variance est appelée l'écart-type, σ qui indique la dispersion des données autour de la moyenne. Un faible écart-type signifie que les valeurs sont très similaires à la moyenne des données. La variance et l'écart-type sont définie par les équations (2.24) et (2.25)

$$\text{var} = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \quad (2.24)$$

$$\sigma = \sqrt{\text{var}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (2.25)$$

- Le troisième moment est appelé asymétrie (*Skewness*) et s'obtient en utilisant $\alpha = 3$, comme le montre l'équation (2.26) :

$$Skew = \frac{\frac{1}{N} \sum_{i=1}^N \left(\frac{(x_i - \mu)}{\sigma} \right)^3}{\mu_{rms}^3} \quad (2.26)$$

L'asymétrie caractérise la structure de la distribution en déterminant son asymétrie par rapport au mode. Elle est notamment utilisée pour mesurer la distance de la distribution des données vis-à-vis à une distribution normale de même variance.

- Le quatrième moment ($\alpha = 4$) s'appelle l'aplatissement (*Kurtosis*) et est donné par l'équation (2.27) :

$$Kurt = \frac{\frac{1}{N} \sum_{i=1}^N \left(\frac{(x_i - \mu)}{\sigma} \right)^4}{\mu_{rms}^4} \quad (2.27)$$

L'aplatissement permet aussi de définir la forme de la distribution, il montre l'étendue de l'aplatissement de la distribution (Rajoub, 2020). Un aplatissement positif veut dire que la distribution est pointue, tandis qu'un aplatissement négatif indique que la distribution est plus plate. Dans la recherche de Monaco *et al.* (2020), les auteurs ont utilisé les caractéristiques statistiques (moyenne, l'écart type, Skewness, max, min etc.) dans le deuxième niveau d'analyse à long terme et consistait à calculer 10 moments statistiques associés aux séries temporelles de chacune des 33 caractéristiques à court terme (ZCR, entropie spectrale, Rolloff, etc.).

2.6 RÉDUCTION DE DIMENSIONNALITÉ DES CARACTÉRISTIQUES

La réduction de dimensionnalité de données est un processus très souvent utilisé dans le domaine de l'informatique et la technologie de l'information. Il s'agit de prendre un espace de données de taille considérablement grande et d'ensuite en sortir un nouvel espace de données plus petit que le précédent, mais à condition qu'il représente bien les données originales. Parmi les méthodes de réduction de dimension les plus courantes, l'analyse en composante principale PCA (Principal component analysis) (Sehgal *et al.*, 2014). Les

composantes principales sont de nouvelles variables qui sont conçues comme étant des combinaisons linéaires des variables initiales (Alkandari & Aljaber, 2015). Cette méthode procède selon les étapes ci-dessous :

- La normalisation : Le principal objectif de cette opération est de standardiser la plage des variables initiales afin que chaque variable apporte une contribution équivalente à l'analyse.
- Calcul de la matrice de covariance : Le rôle de cette opération est de définir la manière dont les variables de l'ensemble de ces données varient de la moyenne les unes par rapport aux autres. Il peut arriver que les variables soient très corrélées de telle sorte qu'elles comportent des informations redondantes (Yang *et al.*, 2021). Ainsi, afin de déterminer ces corrélations, nous calculons la matrice de covariance.
- Calcul des vecteurs et des valeurs propres de la matrice de covariance : Les vecteurs propres et les valeurs propres sont les constituants de l'algèbre linéaire que nous devons calculer à partir de la matrice de covariance pour pouvoir déterminer les principales composantes des données (Haris Mita *et al.*, 2021).
- Le vecteur de caractéristiques : Dans cette étape, nous décidons de garder toutes ces composantes ou de supprimer celles qui sont peu pertinentes. Nous formons avec les composantes qui restent un vecteur que nous appellerons le vecteur caractéristique. Le vecteur caractéristique est alors une matrice qui a pour colonnes les vecteurs propres des éléments que nous voulons conserver.

2.7 APPROCHE PROPOSÉE

Dans ce projet de recherche, nous allons commencer par l'étude de trois techniques d'extraction de caractéristiques:

- Les moments statistiques (Moyenne, Variance, Skewness et Kurtosis) appliqués aux coefficients d'ondelettes. Nous avons opté pour la décomposition à 6 niveaux et l'utilisation des détails D1-D6. Concernant les paquets d'ondelettes, nous avons testé l'analyse sur plusieurs niveaux de décomposition ($j=1, 2, 3$ et 4) ensuite, de nombreuses simulations ont été effectuées sur l'ensemble des nœuds et des niveaux de décomposition, dans le but d'éliminer toutes les composantes non significatives et ainsi réduire le nombre de coefficients utilisés.
- Les coefficients MFCC ont également été utilisés dans ce projet. Plusieurs tests ont été effectués avec cette technique d'extraction de caractéristiques, et ce en variant le nombre de coefficients à plusieurs valeurs. Le nombre de coefficients retenu à la fin des tests est celui de 26 coefficients réduits à 12 grâce à la technique PCA.
- Les coefficients GFCC ont été testés en présence et absence des coefficients delta (Variation des coefficients d'une fenêtre d'analyse à l'autre) et delta Delta (Variation des valeurs delta). Après les tests effectués, nous avons jugé que les coefficients initiaux sont suffisants pour la caractérisation des sons respiratoires. Le nombre de coefficients retenu à la fin des tests a été fixé à 26 coefficients. La technique PCA de réduction de dimension a été ensuite appliquée aux 26 coefficients pour les réduire à 12 éliminant ainsi les informations redondantes.

Une fois les paramètres fixés pour chaque technique, nous procédons ensuite avec des combinaisons entre ces dernières:

- La combinaison WPT-MFCC regroupant ainsi 12 coefficients MFCC et 4 moments statistiques.
- La combinaison WPT-GFCC regroupant 12 coefficients GFCC et 4 moments statistiques.

CHAPITRE 3

MÉTHODES DE CLASSIFICATION AUTOMATIQUE DES SONS RESPIRATOIRES

Après avoir extrait les caractéristiques pertinentes des sons respiratoires, nous passerons ensuite à l'étape de classification de ces dernières. Dans ce chapitre, nous allons présenter quatre techniques de classification automatique des sons respiratoires (MLP, SVM, k-NN et GMM) rencontrées dans la littérature. Trois d'entre elles vont être retenues pour la suite du projet.

3.1 DÉFINITION ET PRINCIPE DE LA CLASSIFICATION AUTOMATIQUE

La classification est l'étape où nous accomplissons la détermination de la classe d'appartenance d'un nouveau son respiratoire, sur la base d'un ensemble de données d'apprentissage contenant des représentations dont les classes sont connues. Dans la littérature, il est souvent proposé d'utiliser des réseaux de neurones, pour qu'un apprentissage se fasse à partir d'un ensemble de sons respiratoires, tout en connaissant leurs classes d'appartenance. Ensuite, la phase de test suit pour évaluer le réseau mis en place précédemment, en utilisant un autre ensemble de sons respiratoires contenant des caractéristiques étrangères au réseau. Dans ce chapitre, nous allons présenter quelques techniques de classification rencontrées dans la littérature, particulièrement celles qui ont été adoptées pour la réalisation de ce projet. Ces techniques seront présentées dans quatre sections :

- Perceptron multicouche MLP (*Multi-Layer Perceptron*).
- Machines à vecteur de support SVM (*Support vector machines*).
- k-plus proches voisins k-NN (*k-nearest neighbours*).

- Modèle de mélange gaussien (*Gaussian Mixture Model*).

3.2 PERCEPTRON MULTICOUCHE MLP (MULTI-LAYER PERCEPTRON)

Inspiré des neurones biologiques, le perceptron multicouche est probablement l'architecture la plus fréquemment utilisée pour les applications pratiques des réseaux de neurones (Tocchetto *et al.*, 2014). Dans la plupart des cas, le réseau de neurones se compose de deux couches avec une connectivité totale entre les entrées et les unités cachées, et entre les unités cachées et les sorties. Les entrées sont combinées avec les poids de départ dans une somme pondérée et elles sont ensuite soumises à la fonction d'activation (Nabney, 2002). La figure 3.1 est une représentation simple d'un réseau de neurones de type perceptron multicouche (MLP) ayant une seule couche cachée.

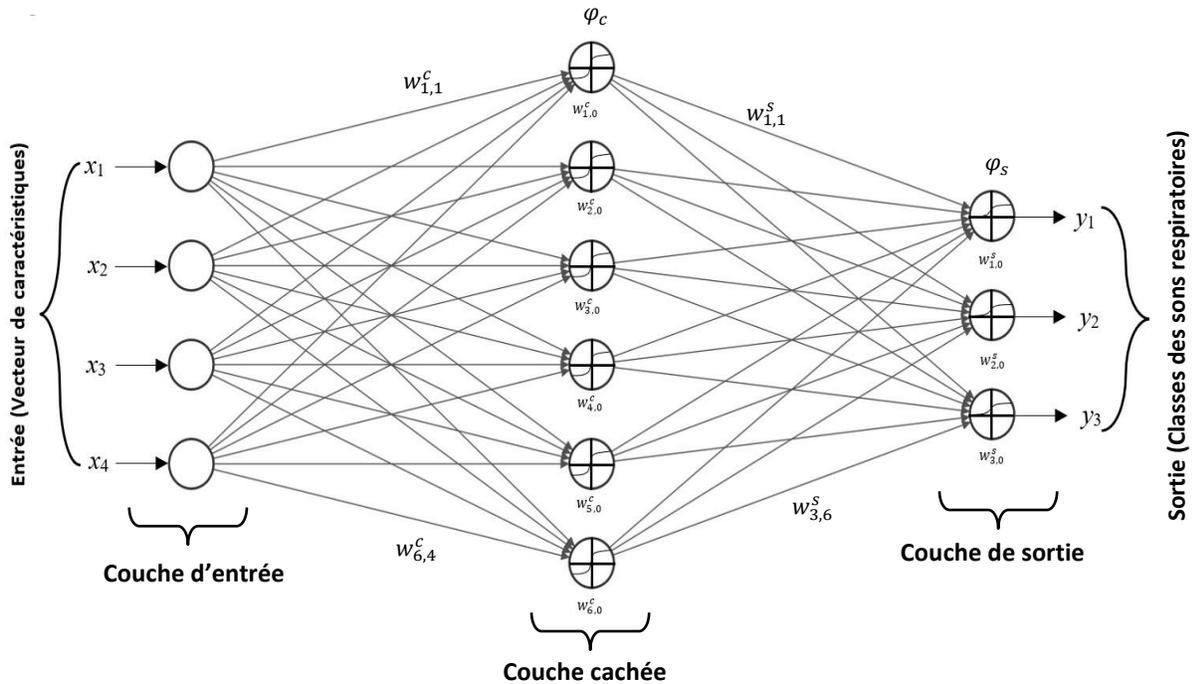


Figure 3.1 Représentation d'un réseau de neurone simple

Le vecteur de caractéristiques $x = \{x_1, x_2, x_3, x_4\}$ constitue l'entrée du réseau, alors que $y = \{y_1, y_2, y_3\}$ constitue sa sortie correspondante aux classes.

En raison de leur capacité à reproduire et modéliser des processus non linéaires, les réseaux de neurones artificiels ont été appliqués dans de nombreuses disciplines. Un réseau de neurones artificiels passe d'abord par une phase d'apprentissage où il apprend à reconnaître les formes à partir de leurs caractéristiques dans les données d'apprentissage (Semmad, 2021), qu'elles soient visuelles, auditives ou textuelles (audio contenant quatre classes de sons respiratoires dans notre cas). Ensuite, la phase de test où de nouvelles données sont présentées au réseau afin que ce dernier puisse les classer chacune dans sa catégorie d'appartenance (Pelletier, 2006).

3.2.1 Perceptron

Les réseaux de neurones sont principalement constitués de neurones artificiels, il s'agit de simples entités de traitement qui reçoivent des données d'entrée pondérées et produisent un signal de sortie limitée à l'aide d'une fonction d'activation (figure 3.2). Le perceptron peut être considéré comme un outil de classification binaire en définissant une frontière de décision linéaire (James *et al.*, 2013).

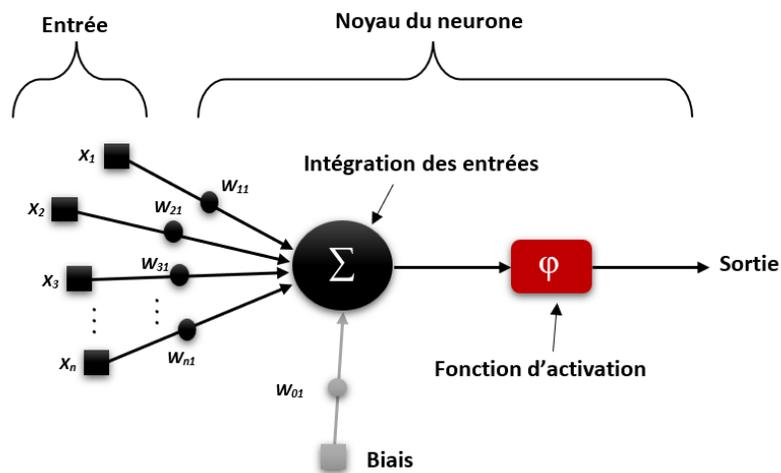


Figure 3.2 Modèle d'un neurone simple

Les modèles de perceptron utilisent habituellement la fonction tangente hyperbolique pour

les neurones de la couche cachée et la fonction sigmoïde pour les neurones de la couche de sortie comme le montre la figure 3.3.

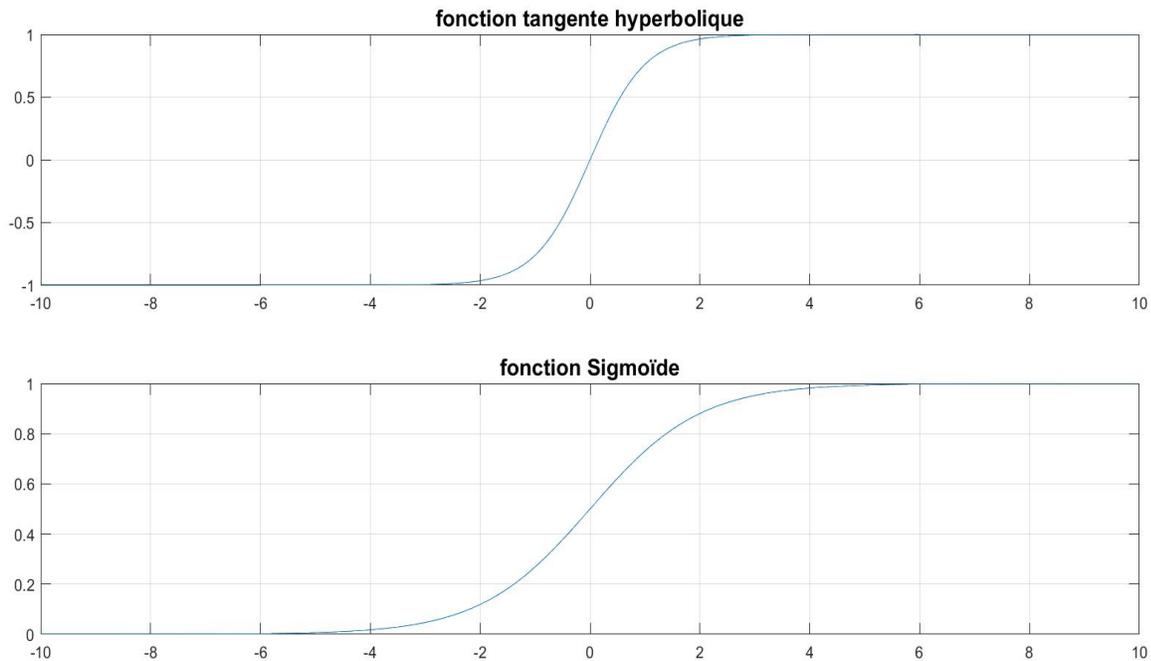


Figure 3.3 Fonction tangente hyperbolique et fonction sigmoïde

Chaque nœud dans un réseau effectue une somme pondérée qui est transférée par une fonction d'activation non-linéaire φ_c . La sortie d'un neurone (nœud) de la couche cachée est donnée par :

$$y_j^c = \varphi_c \left(\sum_{i=0}^D w_{j,i}^c x_i^c \right) \quad (3.1)$$

où D est le nombre d'entrées, les i sont les nœuds de la couche d'entrée, les j sont les nœuds de la couche de cachée. $w_{j,i}^c$ est le poids de connexion entre l'entrée i et le nœud j (Bahoura, 2009).

De la même façon, la sortie du neurone de la couche de sortie est donnée par :

$$y_k^s = \varphi_s \left(\sum_{j=0}^D w_{k,j}^s y_j^c \right) \quad (3.2)$$

où φ_s est la fonction de transfert, $w_{k,j}^s$ est le poids de connexion, et ($y_0^c=1$) est le biais de la couche de sortie. Lorsque le i ème élément (x_i, d_i) est présenté, l'erreur de l'apprentissage est définie comme suite (Bahoura, 2009) :

$$E(w) = \frac{1}{K} \sum_{k=1}^K (d_k - y_k^s)^2 \quad (3.3)$$

où d_k et y_k sont respectivement les réponses souhaitées et les réponses réelles du neurone k. K est le nombre de classes.

3.2.2 Principe de rétro-propagation (back propagation)

L'algorithme de la rétropropagation (*Back propagation*) est un mode d'apprentissage qui permet au perceptron multicouche d'ajuster de manière itérative ses poids (Nabney, 2002). Il peut être implémenté de deux manières différentes : en mode séquentiel et en mode batch. En mode séquentiel, les poids sont mis-à-jour après que chaque donnée d'apprentissage ait été appliquée au réseau. En mode batch, toutes les données d'apprentissage, qui constituent une époque, sont appliquées au réseau avant que les poids ne soient mis-à-jour. Il est évident que pour une application en temps-réel, l'apprentissage séquentiel doit être choisi (Bahoura & Park, 2012). À la n ème itération, c'est-à-dire la présentation du n ème exemple d'apprentissage, le signal d'erreur à la sortie du neurone k est défini comme suit :

$$e_k(n) = d_k(n) - y_k^s(n) \quad (3.4)$$

où $d_k(n)$ et $y_k^s(n)$ sont respectivement la sortie désirée et la sortie réelle de ce neurone.

La valeur instantanée de l'énergie d'erreur totale au niveau de la couche de sortie est définie comme suit (Haykin, 1999) :

$$E(n) = \frac{1}{K} \sum_{k=1}^K e_k^2(n) \quad (3.5)$$

Cette erreur peut être réduite en mettant à jour les poids à l'aide de la méthode de descente du gradient (Haykin, 1999) :

$$\Delta w_{k,j}(n) = -\eta \frac{\partial E(n)}{\partial w_{k,j}(n)} \quad (3.6)$$

où ($0 < \eta < 1$) est le taux d'apprentissage. La valeur suivante de $w_{k,j}(n)$ est alors donnée par (Bahoura & Park, 2012) :

$$w_{k,j}(n+1) = w_{k,j}(n) + \Delta w_{k,j}(n) \quad (3.7)$$

3.2.2.1 Couche de sortie

les poids de connexion dans la couche de sortie sont mis à jour en utilisant (Bahoura & Park, 2012) :

$$\Delta w_{k,j}^s(n) = \eta \delta_k^s(n) y_j^c(n) \quad (3.8)$$

où le gradient local $\delta_k^s(n)$ est défini par (Bahoura & Park, 2012) :

$$\delta_k^s(n) = e_k(n) \varphi'_s(v_k^s(n)) \quad (3.9)$$

3.2.2.2 Couche cachée

Les poids de connexion dans la couche cachée sont mis à jour en utilisant l'équation suivante (Bahoura & Park, 2012) :

$$\Delta w_{j,i}^c(n) = \eta \delta_j^c(n) x_i(n) \quad (3.10)$$

où le gradient local $\delta_j^c(n)$ est défini par

$$\delta_j^c(n) = \varphi'_h(v_j^c(n)) \sum_{k=1}^K \delta_k^s(n) w_{k,j}^s(n) \quad (3.11)$$

comme une fonction tangente hyperbolique est utilisée $\varphi'_c(v_j^c(n)) = 1 - \varphi_c^2(v_j^c(n))$ l'équation (3.11) peut être simplifiée a (Bahoura & Park, 2012) :

$$\delta_j^c(n) = (1 - \varphi_c^2(v_j^c(n))) \sum_{k=1}^k \delta_k^s(n) w_{k,j}^s(n) \quad (3.12)$$

3.3 MACHINE À VECTEUR DE SUPPORT SVM (*SUPPORT VECTOR MACHINE*)

3.3.1 Introduction aux SVM

Le SVM est un classificateur qui maximise la marge entre les échantillons de différentes classes séparées par un hyperplan (Meng *et al.*, 2020). Les machines à vecteur de support ont été largement utilisées dans diverses tâches de classification, car elles peuvent résoudre des problèmes de haute dimension et non-linéairement séparables. L'algorithme SVM a été développé à la base pour la classification binaire, mais il peut être utilisé pour classifier des données multi classes (Boujelben, 2017).

Comme le montre la figure 3.4. Nous allons appeler la marge maximale dans l'hyperplan par MMH. Les points de l'échantillon situés sur le MMH sont appelés vecteur de support (Ertekin, 2009).

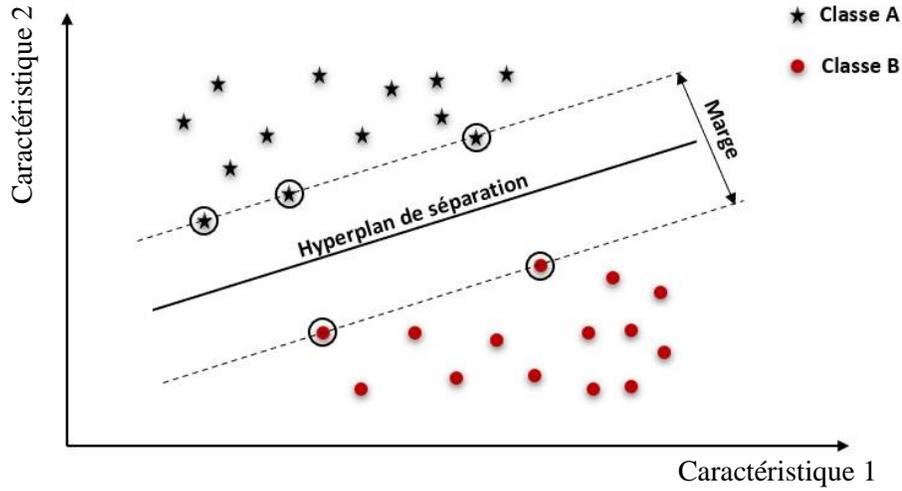


Figure 3.4 Description de la Marge

Pour la classification binaire basée sur les SVM, un ensemble d'apprentissage étiqueté de n observations est défini par la formule suivante :

$$\phi = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \quad (3.13)$$

où x_n sont les entrées à d -dimensions (vecteurs de caractéristiques) et $y_n \in \{1, -1\}$ leurs identifications associées, tandis que n est le nombre d'échantillons, c'est-à-dire la taille de l'ensemble de données d'observation.

3.3.2 Espace caractéristique linéairement séparable

Comme le montre la figure 3.4, le principal intérêt d'un SVM est de déterminer un hyperplan tel que les indices de classe des données (± 1) soient situés de part et d'autre de l'hyperplan et que la distance du vecteur le plus proche de l'hyperplan soit maximale. Le vecteur de support des données d'apprentissage est défini par $w^T x + b = 1$ pour la classe A, et $w^T x + b = -1$ pour la classe B, où w est un vecteur de poids à d -dimensions dans l'espace des caractéristiques et b est un terme de biais (Meng *et al.*, 2020).

L'hyperplan de séparation à marge maximale peut être construit en résolvant le problème d'optimisation primaire (Vapnik, 1999) présenté par les deux équations suivantes (Boujelben & Bahoura, 2018) :

$$\begin{aligned}
 \text{Maximise} \quad & J_p(w, \xi_i) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \\
 \text{Relativement à} \quad & y_i(w^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0
 \end{aligned} \tag{3.14}$$

où $\|w\|$ est la norme euclidienne de w , C est le paramètre de mauvaise classification (c'est-à-dire un compromis entre la maximisation de la marge et la minimisation de l'erreur) et ξ_i sont les variables qui représentent la distance des points x_i , par rapport à leurs marges associées. Le problème d'optimisation initial donné par (3.14) concerne la minimisation de deux quantités, le premier terme permet de contrôler la marge et le second terme limite le nombre de points mal classés (Boujelben & Bahoura, 2018). Pour résoudre (3.14), le problème d'optimisation a été transformé en un problème de Lagrange comme suit (Boujelben & Bahoura, 2018) :

$$\begin{aligned}
 \text{Maximise} \quad & L_d(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j \\
 \text{Relativement à} \quad & \sum_{i=1}^n \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C
 \end{aligned} \tag{3.15}$$

où n est le nombre d'échantillons d'apprentissage, et α_i sont les multiplicateurs de Lagrange. Par conséquent, le vecteur de poids w et le biais b peuvent être obtenus en utilisant l'équation suivante (Boujelben & Bahoura, 2018) :

$$w = \sum_{i=1}^S \alpha_i y_i x_i \quad (3.16)$$

$$b = y_i - w^T x_i \quad (3.17)$$

où x_i représente les vecteurs de support, avec $i=1, \dots, S$, et S est le nombre de vecteur de support (Boujelben & Bahoura, 2016). Le nombre S représente le nombre d'instances en apprentissage qui résolvent le problème d'optimisation initial donné par (3.14). Géométriquement, les vecteurs de support sont le point de l'ensemble de données d'apprentissage le plus proche de l'hyperplan optimal (Boujelben & Bahoura, 2018). La décision de classification pour le classifieur SVM linéaire est présentée comme suit :

$$d(x) = \text{sign} \left(\sum_{i=1}^S \alpha_i y_i x_i^T x + b \right) \quad (3.18)$$

3.3.3 Espace caractéristique non linéairement séparable

Dans le cas de données non linéairement séparables, le SVM risque de ne pas fonctionner correctement. C'est pourquoi la méthode des noyaux a été introduite pour traiter les données non linéairement séparables à l'aide d'un classificateur SVM linéaire (Boujelben & Bahoura, 2018). Le principe de la méthode du noyau est de convertir des données d'entrée non linéairement séparables en données linéairement séparables par un processus non linéaire qui les transforme de leur espace original de basse dimension (\mathbb{R}^d) en un nouvel espace de haute dimension (\mathbb{R}^q), où $q \geq d$. Dans le cas présent, le vecteur x est transformé en $\varphi(x)$ (Boujelben & Bahoura, 2018).

$$\begin{aligned} \varphi: \mathbb{R}^d &\rightarrow \mathbb{R}^q \\ x &\rightarrow \varphi(x) \end{aligned} \quad (3.19)$$

Une fonction noyau est alors définie $K(x_i, x_j) = \varphi^T(x_i)\varphi(x_j)$. Dans ce cas, la fonction optimale (3.15) devient (Boujelben & Bahoura, 2018) :

$$\begin{aligned}
\text{Maximise} \quad & L_d(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\
\text{Relativement à} \quad & \sum_{i=1}^n \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C
\end{aligned} \tag{3.20}$$

Le vecteur de poids w et le biais b peuvent être obtenus comme suit (Boujelben & Bahoura, 2018) :

$$w = \sum_{i=1}^S \alpha_i y_i \varphi(x_i) \tag{3.21}$$

$$b = y_i - w^T \varphi(x_i) \tag{3.22}$$

La fonction de décision dans le contexte de données non linéaires est définie par l'équation suivante (Boujelben & Bahoura, 2018) :

$$d(x) = \text{sign} \left(\sum_{i=1}^S \alpha_i y_i K(x_i, x) + b \right) \tag{3.23}$$

3.4 K-PLUS PROCHES VOISINS K-NN (*K-NEAREST NEIGHBORS*)

La méthode k-plus proches voisins (k-NN) est un algorithme simple et facile à implémenter afin de résoudre des problèmes de classification et de régression (Jivani, 2013). La méthode k-NN est principalement utilisée comme classifieur (Taunk *et al.*, 2019) pour classer les données en fonction des caractéristiques les plus proches ou les plus éloignées (voir la figure 3.5).

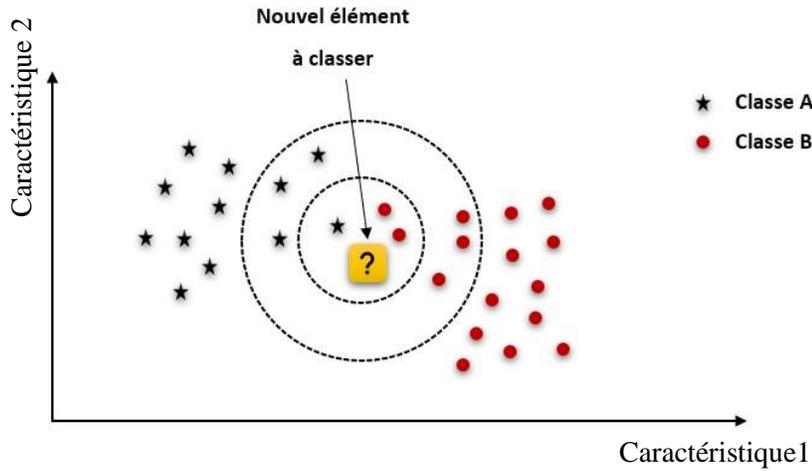


Figure 3.5 Règle de décision k-NN pour les valeurs $k = 3$ et $k = 9$

Dans ce concept de classification, la méthode k-NN peut être considérée comme un outil de classification supervisée. La classification supervisée a pour but de faire correspondre une nouvelle observation x de l'espace des vecteurs de caractéristiques à d -dimensions à la classe désirée de $\{\omega_1, \dots, \omega_K\}$, où K est le nombre de classes. Considérons un ensemble d'apprentissage étiqueté de N observations $O_N = \{(x_1, \omega_1), \dots, (x_N, \omega_N)\}$, où x_i et ω_i sont respectivement les vecteurs de caractéristiques et leurs identifiants scalaires correspondants (Bahoura & Simard, 2012).

Pour classer une nouvelle observation x de classe inconnue, une cellule de volume V est placée autour de x pour capturer k prototypes. Soit k_j le nombre d'échantillons identifiés ω_j capturés par la cellule (de sorte que $\sum_{j=1}^K k_j = k$), la probabilité conjointe peut être estimée comme suit (Bahoura & Simard, 2012) :

$$p_N(x, \omega_j) = \frac{k_j / V}{k} \quad (3.24)$$

et peut être utilisé pour fournir une estimation acceptable de la probabilité postérieure

$$p_N(\omega_j|x) = \frac{p_N(x, \omega_j)}{\sum_{j=1}^M p_N(x, \omega)} = \frac{k_j}{k} \quad (3.25)$$

La nouvelle observation x est prédite dans la classe ω par un vote majoritaire comme suit (Bahoura & Simard, 2012) :

$$\omega = \arg \max_{1 \leq j \leq k} \{k_j\} \quad (3.26)$$

La phase de classification par k-NN est beaucoup plus rapide par rapport aux autres algorithmes de classification. Il n'est pas nécessaire d'entraîner un modèle pour la généralisation, c'est pourquoi k-NN est connu comme l'algorithme d'apprentissage le plus simple car il est basé sur les instances. Il peut être utile dans le cas de données non linéaires (Navlani, 2018). Bien que cette méthode de classification soit aussi simple et rapide, mais il arrive qu'elle devienne particulièrement lente à mesure que le nombre de points de données croît, car ce modèle doit enregistrer tous les points de données. Elle n'est également pas efficace en termes de mémoire et elle est aussi sensible aux valeurs aberrantes.

3.5 MODÈLES DE MÉLANGE GAUSSIEN GMM (GAUSSIAN MIXTURE MODELS)

Une méthode statistique performante basé sur le modèle de mélange gaussien (GMM) est largement utilisée pour l'identification/vérification des locuteurs (Bahoura, 2009). Elle est considérée comme une technique de modélisation d'une densité de probabilité inconditionnelle $p_j(x)$ pour un ensemble fini de points de données non identifiés $X = \{x_1, \dots, x_M\}$, cet algorithme peut être utilisé pour construire un système de classification (Jeevan *et al.*, 2017). Un modèle à M composantes s'écrit sous la forme (Bahoura, 2009) :

$$p(x|\lambda) = \sum_{j=1}^M w_j p_j(x) \quad (3.27)$$

où x est un vecteur caractéristique à d-dimensions, λ est le modèle de la classe de son, les paramètres w_j sont appelés les coefficients de pondération de mélange des fonctions de

densité de probabilité $p_j(x)$ varient généralement avec j . Le modèle de mélange est un modèle génératif et il est utile de prendre en compte le processus de génération d'échantillons à partir de la densité qu'il représente. Dans un premier temps, l'une des composantes j est choisie au hasard avec la probabilité $p(j)$ (Jeevan *et al.*, 2017). Nous pouvons donc considérer $p(j)$ comme la probabilité antérieure de la j ième composante. Les probabilités postérieures correspondantes peuvent être écrites comme suit :

$$P(j|x) = \frac{p_j(x)w_j}{p(x|\lambda)} \quad (3.28)$$

où $p(x|\lambda)$ est donné par (3.27).

Les poids du mélange w_j vérifie la contrainte suivante : $\sum_{j=1}^M w_j$ chaque densité de composant est une fonction gaussienne à D-variables (Nabney, 2002) de la forme :

$$p_j(x) = \frac{1}{(2\pi)^{d/2} |\sum_j|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu_j)' \sum_j^{-1} (x - \mu_j) \right\} \quad (3.29)$$

Le modèle complet de mélange gaussien est paramétré par les vecteurs moyens, les matrices de covariance et les poids de mélange de toutes les densités de composantes (Reynolds, 2000). Ces paramètres sont représentés collectivement par la relation (3.30).

$$\lambda = \left\{ w_j, \mu_j, \sum_j \right\} \quad j = 1, \dots, M \quad (3.30)$$

Étant donné les vecteurs d'apprentissage et une configuration GMM, nous souhaitons estimer le paramètre du GMM, λ , qui, dans un certain sens, correspond le mieux à la distribution des vecteurs de caractéristiques d'apprentissage (Nabney, 2002). Dans le contexte des sons respiratoires, chaque classe de son est modélisée par un GMM spécifique qui est entraîné en utilisant un ensemble de données avec l'algorithme Expectation Maximization (EM), pendant la phase de test, un son inconnu est comparé à tous les modèles GMM et la décision de classification est basée sur le critère du maximum de vraisemblance (ML) (Bahoura, 2009). L'objectif de l'estimation ML (Maximum Likelihood) est de trouver

les paramètres du modèle qui maximisent la probabilité pour une séquence de vecteurs d'apprentissage $X = \{x_1, \dots, x_M\}$.

Supposons que nous avons K classes de sons λ_k représentées par les modèles GMM $\lambda_1, \lambda_2, \dots, \lambda_K$. Le but est de trouver le modèle k qui maximise la probabilité a posteriori pour une séquence d'observation donnée $X = \{x_1, x_2, \dots, x_M\}$ (Bahoura, 2009) :

$$\hat{k} = \arg \max_{1 \leq k \leq K} p(\lambda_k | X) \quad (3.31)$$

En utilisant les règles de Bayes, l'équation (3.31) peut être exprimée comme suit (Bahoura, 2009) :

$$\hat{k} = \arg \max_{1 \leq k \leq K} \frac{p(X | \lambda_k) p(\lambda_k)}{p(X)} \quad (3.32)$$

En supposant que les classes sont également probables et en notant que $p(X)$ est le même pour toutes les classes, la classification peut être simplifiée pour trouver (Bahoura, 2009) :

$$\hat{k} = \arg \max_{1 \leq k \leq K} p(X | \lambda_k) \quad (3.33)$$

En utilisant la fonction logarithmique et en supposant l'indépendance entre les observations, la formule de classification est la suivante (Bahoura, 2009) :

$$\hat{k} = \arg \max_{1 \leq k \leq K} \sum_{i=1}^M \log p(x_i | \lambda_k) \quad (3.34)$$

3.6 TECHNIQUES DE VALIDATION DES DONNÉES

La validation croisée est une méthode classique d'évaluation et de comparaison des algorithmes d'apprentissage qui consiste à répartir les données en deux parties : l'une utilisée pour entraîner le modèle et l'autre pour tester le modèle. Dans une validation croisée typique,

les ensembles d'apprentissages et de test devront se croiser dans des tours successifs de sorte que chaque élément de données ait une possibilité d'être validé (Refaeilzadeh *et al.*, 2009). Il existe de multiples versions de la validation mais les plus courantes sont les suivantes :

3.6.1 La méthode de retenue (*Holdout*)

Le *hold-out* consiste à diviser l'ensemble de données en un ensemble "*d'apprentissages*" et un ensemble de "*test*". L'ensemble d'apprentissage est celui sur lequel le système est entraîné, tandis que l'ensemble de test est utilisé pour voir comment ce système se comporte sur des données inconnues (Pal & Patel, 2020). Une division habituelle lors de l'utilisation de cette méthode est de consacrer 80% des données pour l'apprentissage et les 20% de données qui restent pour le test.

3.6.2 La validation croisée à k-blocs (*k-fold cross validation*)

Dans la validation croisée k-bloc, les données sont réparties en k sous-ensembles.

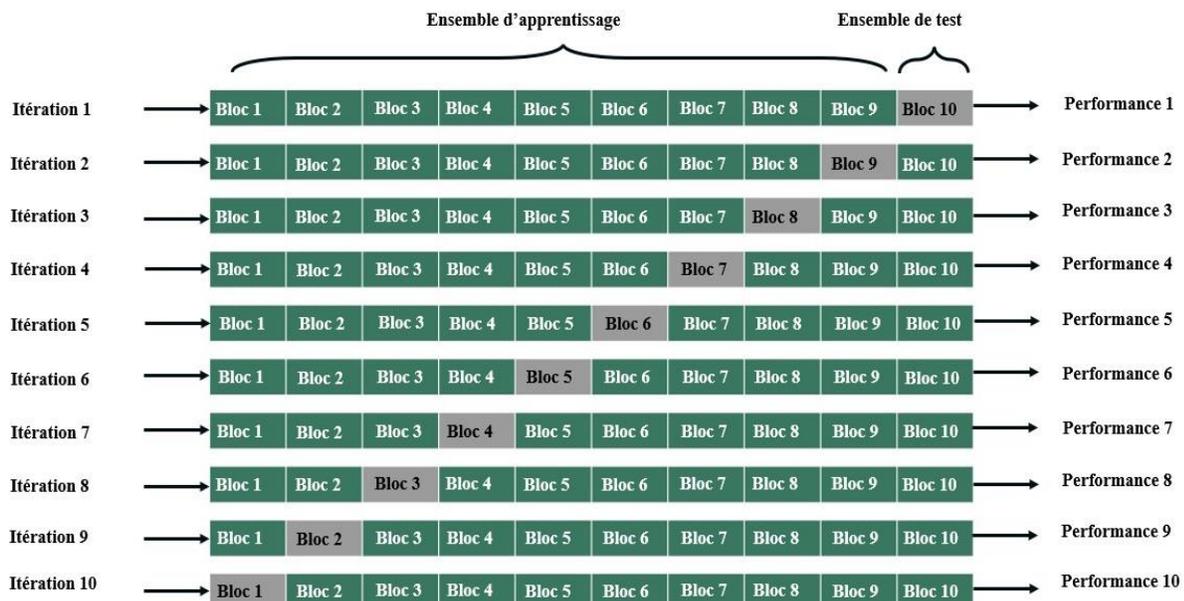


Figure 3.6 Exemple de 10-fold cross validation

La méthode de retenue est ensuite répétée k fois, de façon que, chaque fois, l'un des k sous-ensembles de données soit considéré comme l'ensemble de tests et que les $k-1$ autres sous-ensembles soient réunis pour former un ensemble d'apprentissages (Yadav & Shukla, 2016), comme le montre la figure 3.6. La validation croisée k -bloc peut effectivement être utilisée pour une classification de qualité. Cela permet de réduire fortement le biais, car la majeure partie des données est utilisée pour l'apprentissage (Yadav & Shukla, 2016).

Dans l'exemple de la figure 3.6, pour la validation croisée 10-blocs (*10-fold cross validation*), l'ensemble de données est réparti en 10 sous-groupes, et le modèle est entraîné et testé 10 fois de façon que chaque groupe ait la possibilité de servir d'ensemble de test.

3.6.3 La validation croisée avec exclusion (*Leave-One-Out Cross-Validation*)

La procédure de validation croisée avec exclusion (*Leave-One-Out Cross-Validation*) est utilisée pour estimer la performance des algorithmes d'apprentissage automatique lorsqu'ils sont utilisés pour faire des prédictions sur des données qui n'ont pas été utilisées pour entraîner le modèle (Cawley, 2006). La validation croisée avec exclusion est un cas particulier de la validation croisée k -bloc dans lequel k correspond au nombre total d'exemples dans l'ensemble de données.

La technique de validation croisée avec exclusion est une version extrême de la validation croisée k -bloc qui a le coût de calcul le plus élevé. Elle nécessite la création et l'évaluation d'un modèle pour chaque exemple de l'ensemble de données d'apprentissage (Lv *et al.*, 2020). L'avantage d'un si grand nombre de modèles adaptés et évalués est une estimation plus robuste de la performance du modèle, car chaque série de données a la possibilité de représenter la totalité de l'ensemble de données de test.

Compte tenu du coût de calcul, cette technique n'est pas appropriée pour les très grands ensembles de données, tels que les dizaines ou centaines de milliers d'exemples, utilisées pour l'apprentissage des modèles tels que les réseaux de neurones.

3.7 APPROCHE PROPOSÉE POUR LA CLASSIFICATION

Pour l'étape de classification automatique des sons respiratoires, nous allons réaliser trois modèles de classification basés sur le réseau MLP, le SVM et le k-NN.

Le modèle à base de réseau MLP va être optimisé en accomplissant une étude paramétrique concernant le nombre de neurones et de couches cachées utilisés par le réseau. Le deuxième modèle proposé est le SVM qui va être réalisé en utilisant l'algorithme *one-vs-one* afin de pouvoir l'utiliser pour la classification de 4 classes. Le troisième modèle à tester est le k-NN qui utilise la distance euclidienne pour la mesure du voisinage. Nous allons effectuer également des tests pour différentes valeurs de k (k=1, k=3 et k=5).

CHAPITRE 4

RÉSULTATS EXPÉRIMENTAUX

Dans ce chapitre, les techniques d'extraction de caractéristique et de classification, abordées dans les deuxième et troisième chapitres, ont été testées sous différentes combinaisons. Nous allons commencer par présenter le prétraitement accompli sur la base de données. Nous donnerons ensuite les différentes étapes réalisées pour l'optimisation de nos systèmes réaliser ainsi que leurs performances atteintes.

4.1 OUTILS ET MATÉRIEL UTILISÉ

Les expériences décrites dans les sections suivantes, ont été réalisées sur un ordinateur doté d'un processeur AMD Ryzen 5 3500U, d'une carte graphique Radeon Vega Mobile Gfx 2.10 GHz et d'une mémoire vive de 8 Go.

Tous les traitements de données et les simulations ont été réalisés sur le logiciel MATLAB R2021a, incluant divers toolbox tel que Audio Toolbox, Wavelet Toolbox, DeepLearning Toolbox etc.

4.2 PRÉTRAITEMENT DE LA BASE DE DONNÉES

Comme mentionné dans le premier chapitre, la base de données ICBHI 2017 contient 920 enregistrements de sons respiratoires allant de 10 à 90 secondes. Le premier défi de cette base de données est que chaque enregistrement contient lui-même deux, et même plusieurs cycles de classes différentes. De plus, les enregistrements ont trois fréquences d'échantillonnage différentes : 44.1 kHz, 10kHz et 4 kHz. La première étape était de mettre toute la base de données à la même fréquence d'échantillonnage (8 kHz dans notre cas). La figure 4.1 est le schéma descriptif d'un son respiratoire contenant les quatre classes (normal, sifflement, crépitation, sifflement et crépitation) de manière que chaque classe est représentée par une couleur.



Figure 4.1 Un enregistrement audio contenant six cycles respiratoires : le premier, le troisième et le cinquième sont normaux (Bleu). le deuxième contient des crépitations (Orange), le quatrième contient des sifflements (Vert), le sixième contient à la fois des crépitations et des sifflements (Rouge). La flèche noire indique le début de l'enregistrement. Les lignes grises signifient la fin d'un cycle respiratoire et le début d'un autre cycle.

L'idée de base est de séparer les cycles, et de les regrouper ensuite dans des enregistrements homogènes, de manière qu'à la fin, nous obtenons au total quatre enregistrements de sons respiratoires où chacun regroupe tous les cycles d'une même classe de toute la base de données. Ainsi, nous pourrions traiter chaque classe séparément des autres. La figure 4.2 est un schéma descriptif qui illustre la manière dont nous séparons et regroupons les classes de la base de données.

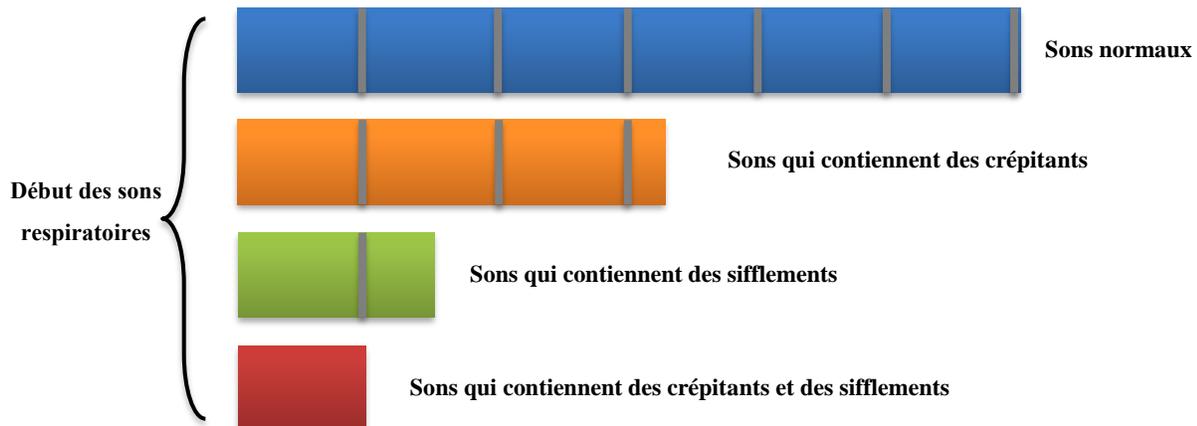


Figure 4.2 Regroupement des sons de mêmes classes dans les mêmes enregistrements (Bleu pour une respiration normal, orange pour une respiration qui contient des crépitations, vert pour une respiration qui contient des sifflements et rouge pour une respiration qui contient les deux classes précédentes)

Après avoir séparé et regroupé les classes de sons respiratoires, nous avons obtenu quatre nouveaux enregistrements de durées différentes. Nous avons proposé de faire en sorte que les quatre nouveaux enregistrements soient de même taille afin de ne pas provoquer un surapprentissage lors de la phase d'apprentissage durant le processus de classification. En prenant dans un premier temps des tranches de mêmes tailles à partir des quatre classes. Cette

approche a été utilisée pour l'optimisation de nos systèmes afin d'atteindre les meilleures performances. Une fois les paramètres du modèle de classification fixés, nous avons proposé de faire plusieurs combinaisons (voir la figure 4.3) dans le but d'utiliser quasiment tous les enregistrements de la base de données.

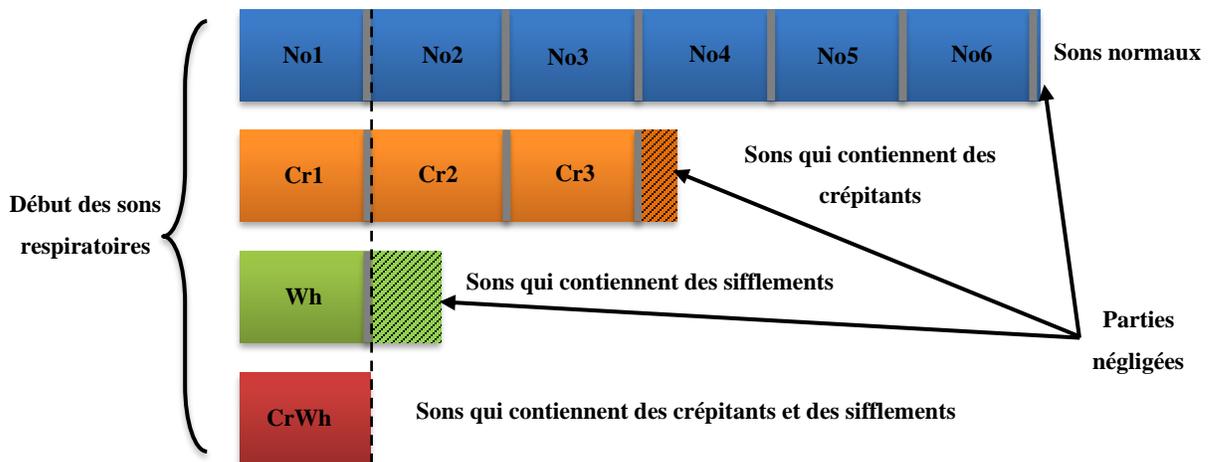


Figure 4.3 Différentes combinaisons de segments de la base de données utilisées

L'idée est d'effectuer plusieurs combinaisons de telle façon qu'à chaque test une nouvelle tranche de données est utilisée. Ce qui conduit à la fin à obtenir $6 \times 3 \times 1 \times 1 = 18$ combinaisons. Pour évaluer les performances totales, nous avons proposé de faire la moyenne des 18 résultats de simulation (Spécificité, Sensitivité, Score ICBHI, Score harmonique et l'exactitude) obtenus par les tests des 18 combinaisons. La combinaison utilisée pour l'optimisation de nos systèmes est constituée des quatre sons respiratoires suivants : No1, Cr1, Wh1 et CrWh (voir la figure 4.3). Chaque segment représentant une classe dure 24 minutes et 48 secondes environ. Le nombre d'échantillon correspondant cette durée est obtenu par le produit de la durée total et la fréquence d'échantillonnage ce qui va donner 12 386 664 échantillons. En tenant compte de la largeur des fenêtres de 2048, nous devrions avoir un nombre total de fenêtres non-chevauchantes égale à 6 048. Notre traitement devrait se faire avec un taux de chevauchement a 50% ce qui conduit à avoir 12095 fenêtres chevauchantes pour chaque classe à traiter.

4.3 CRITÈRES D'ÉVALUATION

L'évaluation d'un classifieur consiste en général à entraîner un modèle sur un ensemble de données et à utiliser le modèle pour faire des prédictions sur un ensemble de données qui n'a pas été utilisé pendant l'apprentissage. Ensuite, à comparer les prédictions aux valeurs prévues dans les données retenu pour le test (Brownlee, 2020). Parmi les critères rencontrés dans les recherches utilisant la base de données ICBHI 2017 l'exactitude et le score ICBHI.

Pour mesurer les performances de classification d'un système trois grandeurs sont souvent utilisées : la sensibilité, la spécificité et l'exactitude. Dans le cas de l'utilisation de la base de données ICBHI 2017, la sensibilité et la spécificité sont données comme suit (Rocha et al., 2019):

$$\text{Sensibilité}(S_E) = \frac{C_v + W_v + CW_v}{C + W + CW} \quad (4.1)$$

$$\text{Spécificité}(S_P) = \frac{N_v}{N} \quad (4.2)$$

où C , W , CW et N désigne respectivement le nombre total de segments de crépitant, de sibilants, de crépitant & sibilants et normaux, alors que C_v , W_v , CW_v et N_v représente respectivement la prédictions positives de chacune des classes. L'exactitude (Acc) est définie par :

$$Acc = \frac{C_v + W_v + CW_v + N_v}{C + W + CW + N} \quad (4.3)$$

Il y a deux autres mesures proposées dans la littérature qui sont dédiées pour l'évaluation de la base de données ICBHI 2017, le score ICBHI (S_{ICBHI}) et le score harmonique (S_H) données par les deux équations suivantes :

$$S_{ICBHI} = \frac{S_E + S_P}{2} \quad (4.4)$$

$$S_H = \frac{2 \times S_E \times S_P}{S_E + S_P} \quad (4.5)$$

- La ROC

La fonction d'efficacité du récepteur ROC (Receiver Operating Characteristic) est une représentation graphique de la performance d'un classifieur. La mesure ROC est habituellement illustrée par une courbe qui exprime le taux de vrais positifs (segments des positifs correctement détectés) par rapport au taux de faux positifs (segments des négatifs incorrectement détectés).

- La matrice de confusion

La matrice de confusion peut-être également utilisée pour déterminer l'exactitude d'un système de classification. Il s'agit d'une mesure de performance où la sortie peut être deux classes ou plus.

4.4 ÉTUDE PARAMÉTRIQUE DES TECHNIQUES D'EXTRACTION DE CARACTÉRISTIQUES

Dans cette section, nous allons effectuer une étude paramétrique sur les différentes techniques d'extraction de caractéristiques proposée. Les paramètres étudiés vont être: la largeur de la fenêtre utilisée, le nombre de coefficients et l'utilisation de la technique PCA pour la réduction de la taille des caractéristiques.

Les techniques d'extraction de caractéristiques proposées sont:

- Un vecteur de 26 coefficients GFCC pour chaque fenêtre;
- Un vecteur de 26 coefficients MFCC pour chaque fenêtre;
- Un vecteur de 24 coefficients des caractéristiques statistiques appliqués aux coefficients de la transformée en ondelettes à six niveaux;
- Un vecteur de 4 caractéristiques statistiques appliqués aux coefficients de la transformée par paquets d'ondelettes;

- Un vecteur de 12 coefficients GFCC avec les 4 coefficients statistiques appliqués aux coefficients de la transformé en paquet d'ondelettes;
- Un vecteur de 12 coefficients MFCC avec les 4 coefficients statistiques appliqués aux coefficients de la transformé en paquet d'ondelettes.

Pour chacune des techniques, nous avons calculer la spécificité, la sensibilité, l'exactitude, le score harmonique et le score ICBHI de la base de données. Le tableau 4.1 représente les résultats de classification en utilisant un réseau de neurone MLP d'une seule couche cachée de 15 neurones. Les différentes techniques ont été évaluées seulement sur la première combinaison de la base de données.

Tableau 4.1 Effet de la technique de caractérisation sur les performances de classification

Technique de caractérisation	Nbr de coefficients	Se (%)	Sp (%)	Acc (%)	S _H (%)	S _{ICBHI} (%)
GFCC	26	45.66	76.83	53.46	57.29	61.25
MFCC	26	40.93	74.25	49.26	52.77	57.59
WT	24	23.32	54.04	38.68	32.58	38.68
WPT	4	19.18	75.39	33.23	47.28	30.58
GFCC-PCA	12	59.24	90.64	67.08	71.65	74.94
MFCC-PCA	12	44.63	82.85	54.19	58.02	63.74
WT-PCA	12	24.05	42.62	28.70	33.34	30.75
WPT-GFCC	16	59.11	88.82	66.54	73.97	70.99
WPT-MFCC	16	41.55	76.08	50.19	58.81	53.75

Après analyse des résultats de classification, nous pouvons conclure que l'utilisation des techniques à base de coefficients cepstraux a révélé de bons résultats par rapport au ceux des ondelettes. La technique de réduction de dimension PCA a également permis d'augmenter le taux d'exactitude et des scores de classification. Ainsi, pour la suite des tests nous proposons de retenir les techniques d'extraction de caractéristique MFCC, GFCC, GFCC-PCA, MFCC-PCA, WPT-GFCC et WPT-MFCC.

4.5 RÉSULTATS DE CLASSIFICATION PAR RÉSEAU MLP

Pour commencer, nous allons analyser plusieurs combinaisons de techniques d'extraction de caractéristiques en vue de la classification automatique des sons respiratoires. En utilisant un réseau de neurones MLP de taille de fenêtre d'analyse et nombre de neurones non optimisés. Nous allons ensuite procéder par une étude paramétrique, dans le but d'optimiser le réseau de neurones afin d'obtenir les performances les plus élevées possible. Dans le tableau 4.2 nous avons listé les paramètres des techniques de caractérisation proposées en association avec un réseau MLP de 15 neurones cachés afin d'obtenir les résultats initiaux.

Tableau 4.2 Paramètres de la première expérience

	GFCC	MFCC	GFCC	MFCC	WPT-MFCC	WPT-GFCC
Nb de Coefficients	26	26	12	12	16	16
Taille des fenêtres	1024	1024	1024	1024	1024	1024
Recouvrement	50%	50%	50%	50%	50%	50%
PCA	-	-	✓	✓	-	✓

La figure 4.4 représente les résultats de simulation obtenus, en utilisant les paramètres listés dans le tableau 4.2. La sensibilité, la spécificité, l'exactitude, le score ICBHI et le score harmonique sont représentés par des barres verticales, en pourcentage, pour chaque technique d'extraction de caractéristiques proposées. Nous pouvons constater à première vue, que les meilleurs résultats ont été obtenus avec la méthode d'extraction de caractéristiques utilisant les GFCC réduit avec la technique PCA ($Acc = 67.08\%$, $S_{ICBHI} = 74.94\%$ et $S_H = 71.65\%$).

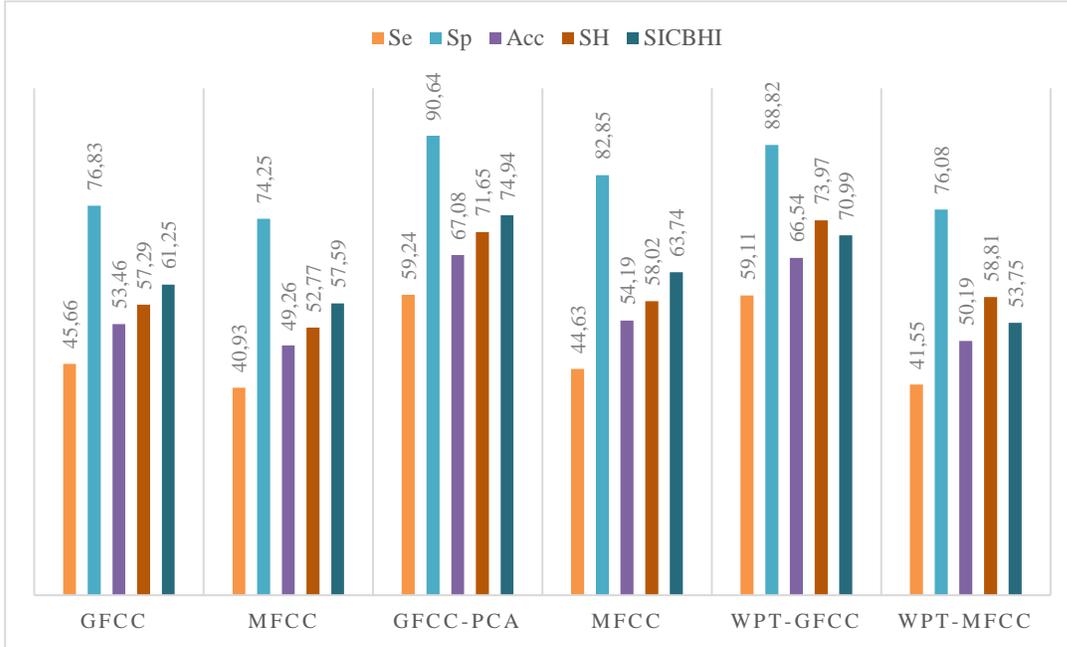


Figure 4.4 Résultats de simulation initial (la sensibilité, la spécificité, l’exactitude, le score ICBHI et le score harmonique)

4.5.1 Étude paramétrique

L’étape suivante à faire est une étude paramétrique concernant la configuration du réseau de neurones.

4.5.1.1 Effets de la largeur des fenêtres

Pour commencer, nous avons proposé d’étudier l’effet de la largeur de fenêtre, en variant la valeur initiale qui est de 1024 échantillons vers une fenêtre moins large de 512 échantillons et de prendre une fenêtre plus large de 2048 échantillons. Les résultats de simulation obtenus sont donnés par les deux figures 4.5 et 4.6 suivantes :

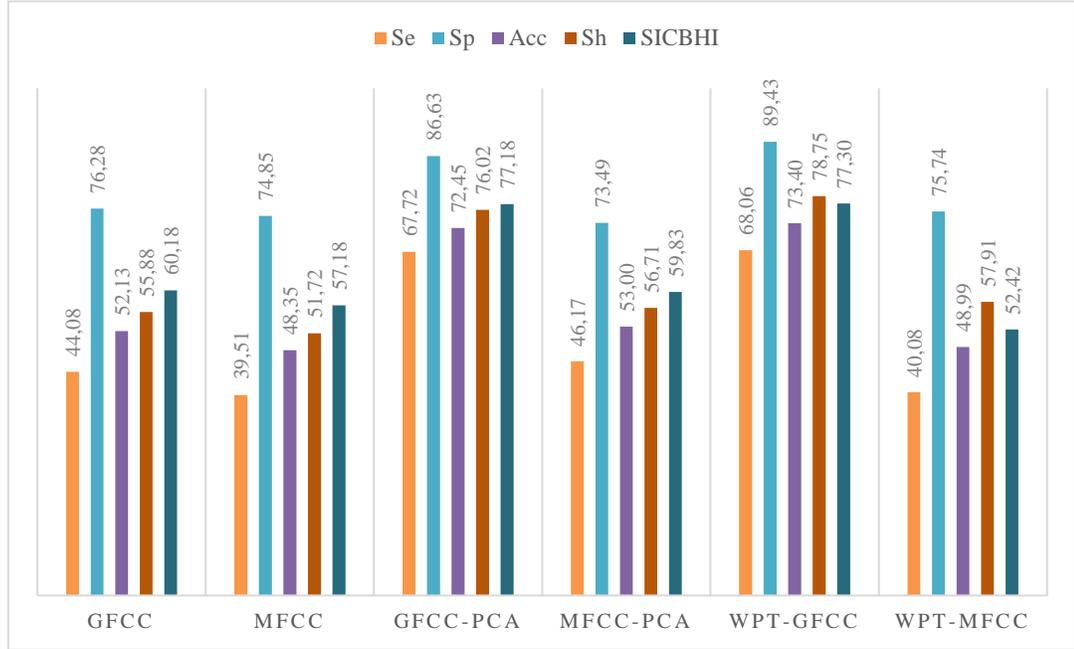


Figure 4.5 Résultats de simulation pour une fenêtre de taille 512 échantillons

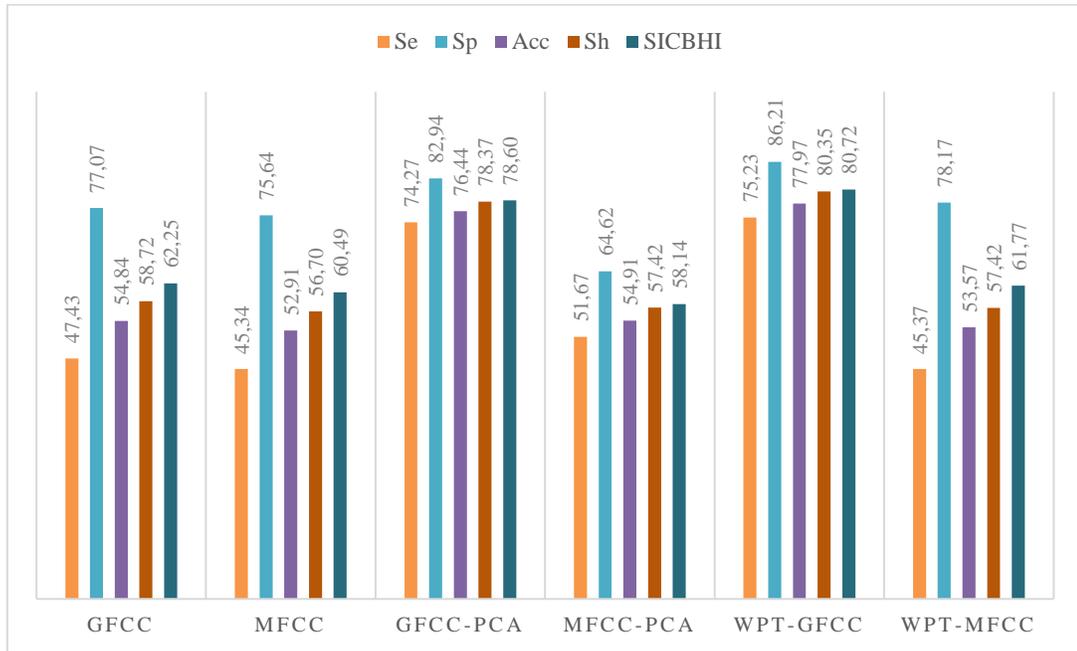


Figure 4.6 Résultats de simulation pour une fenêtre de taille 2048 échantillons

Après analyse des résultats des deux simulations, nous concluons que l'utilisation d'une fenêtre de 2048 échantillons donne environ 4% d'exactitude et de score ICBHI de plus que lors de l'utilisation d'une fenêtre de 512 échantillons. À partir des résultats précédents, nous allons garder la largeur de fenêtre de 2048 échantillons pour les tests suivants. Nous allons se passer des deux techniques GFCC et MFCC utilisant 26 coefficients vu qu'ils ne fournissent pas des résultats assez élevés que les autres techniques.

4.5.1.2 Effets du nombre de neurones

La présente étude consiste à varier le nombre de neurones cachés (initialement 15 neurones) et le nombre de couches cachées afin de voir l'effet de la variation de ce paramètre. Ensuite, nous allons fixer les valeurs qui vont nous permettre d'avoir les meilleurs résultats, dans le but de finaliser notre système de classification. Pour effectuer cela, nous avons proposé d'effectuer des tests en variant le nombre de neurones cachés de 20, 30 et 40 afin de constater le sens de variation des résultats, ensuite nous avons tester notre réseau avec deux couches cachées avec plusieurs valeurs de neurones cachés également.

Le tableau 4.3 montre les résultats de classification obtenus avec les différentes techniques de caractérisation testées en utilisant un réseau MLP d'une seule couche cachée de 15 neurones.

Tableau 4.3 Résultats de classification avec 15 neurones cachés

	Nbr de coefficients	Se (%)	Sp (%)	Acc (%)	S _H (%)	S _{ICBHI} (%)
<i>GFCC</i>	26	47,43	77,07	54,84	58,72	62,25
<i>MFCC</i>	26	45,34	75,64	52,91	56,70	60,49
<i>GFCC-PCA</i>	12	74,27	82,94	76,44	78,37	78,60
<i>MFCC-PCA</i>	12	51,67	64,62	54,91	57,42	58,14
<i>WPT-GFCC</i>	16	75,23	86,21	77,97	80,35	80,72
<i>WPT-MFCC</i>	16	45,37	78,17	53,57	57,42	61,77

Nous remarquons que les technique GFCC-PCA et WPT-GFCC ont fournis les résultats les plus élevés. Tandis que les techniques GFCC, MFCC, MFCC-PCA et WPT-MFCC ne montrent pas des résultats de classification prometteurs. Ainsi, pour la suite des tests ces dernières ne seront pas prises en compte.

Ensuite, les techniques GFCC-PCA et WPT-GFCC ont été évaluées avec des réseaux MLP de 20, 30 et 40 neurones, dont les résultats sont respectivement représentés aux tableaux 4.4, 4.5 et 4.6.

Tableau 4.4 Résultats de classification avec 20 neurones

	Nbr de coefficients	Se (%)	Sp (%)	Acc (%)	S_H (%)	S_{ICBHI} (%)
<i>GFCC-PCA</i>	12	78,21	84,27	79,73	81,13	81,24
<i>WPT-GFCC</i>	16	79,89	88,26	81,98	83,87	84,08

Tableau 4.5 Résultats de classification avec 30 neurones

	Nbr de coefficients	Se (%)	Sp (%)	Acc (%)	S_H (%)	S_{ICBHI} (%)
<i>GFCC-PCA</i>	12	82,08	89,23	83,87	85,51	85,66
<i>WPT-GFCC</i>	16	82,44	89,67	84,25	85,90	86,05

Tableau 4.6 Résultats de classification avec 40 neurones

	Nbr de coefficients	Se (%)	Sp (%)	Acc (%)	S_H (%)	S_{ICBHI} (%)
<i>GFCC-PCA</i>	12	84,51	90,68	86,05	87,49	87,60
<i>WPT-GFCC</i>	16	85,26	92,27	87,02	88,63	88,77

À partir des résultats obtenus avec différents nombres de neurones à la couche cachée, nous apercevons une amélioration des performances entre la simulation avec 15 neurones et celle de 20 et 30. Nous allons ensuite faire des essais avec deux couches cachées pour voir

l'effet sur le taux de classification. Les tableaux 4.7 et 4.8 montrent les résultats de classification pour deux couches cachées de 10 et 20 neurones, ainsi que de 20 et 30 neurones.

Tableau 4.7 Résultats de classification avec deux couches cachées de 10 et 20 neurones

	Nbr de coefficients	Se (%)	Sp (%)	Acc (%)	S_H (%)	S_{ICBHI} (%)
<i>GFCC-PCA</i>	12	78,40	83,79	79,75	81,01	81,09
<i>WPT-GFCC</i>	16	77.60	87.06	79.99	82.32	82.33

Tableau 4.8 Résultats de classification avec deux couches cachées de 20 et 30 neurones

	Nbr de coefficients	Se (%)	Sp (%)	Acc (%)	S_H (%)	S_{ICBHI} (%)
<i>GFCC-PCA</i>	12	84.45	89.61	85.74	86.95	87.03
<i>WPT-GFCC</i>	16	85.30	91.66	86.89	88.36	88.48

Après un grand nombre de tests nous avons abouti à l'architecture de deux couches cachées comportant 45 et 55 neurones. Cette dernière a donné les meilleurs résultats de classification jusqu'à présent comme montre le tableau 4.9.

Tableau 4.9 Résultats de classification avec deux couches cachées de 45 et 55 neurones

	Nbr de coefficients	Se (%)	Sp (%)	Acc (%)	S_H (%)	S_{ICBHI} (%)
<i>GFCC-PCA</i>	12	89.52	93.74	90.58	91.58	91.63
<i>WPT-GFCC</i>	16	90.84	94.78	91.83	92.77	92.81

Il faut noter qu'un grand nombre de neurones ne peut pas forcément donner un meilleur résultat, car cela peut causer un surapprentissage du système qui conduit à une détérioration des performances. La combinaison de la méthode d'extraction des caractéristiques utilisant les WPT et GFCC avec le MLP à deux couches cachées de 55 et 45 neurones a permis d'avoir les meilleurs résultats d'environ 91.83% d'exactitude, de 92.81% de score ICBHI et 92.77% de score harmonique. Tandis que la méthode GFCC-PCA, et pour la même architecture du

réseau a permis d'avoir les résultats : 90.58 % d'exactitude, de 91.58% de score ICBHI et 91.63 % de score harmonique.

Matrice de Confusion

Classe de sortie	1	11464 23.7%	121 0.3%	353 0.7%	202 0.4%	94.4% 5.6%
	2	115 0.2%	11041 22.8%	191 0.4%	969 2.0%	89.6% 10.4%
	3	362 0.7%	127 0.3%	11225 23.2%	228 0.5%	94.0% 6.0%
	4	154 0.3%	806 1.7%	326 0.7%	10696 22.1%	89.3% 10.7%
		94.8% 5.2%	91.3% 8.7%	92.8% 7.2%	88.4% 11.6%	91.8% 8.2%
		Classe cible				
		1	2	3	4	

Figure 4.7 Matrice de confusion pour la classification des quatre classes obtenues avec la méthode d'extraction des caractéristiques WPT-GFCC en combinaison avec le classifieur basés sur MLP

La figure 4.7 représente la matrice de confusion pour les quatre classes et qui démontre une exactitude totale d'environ 91.8% soit 94.8% pour 11464 prédictions de la classe de sons normaux, 91.3 % pour 11041 prédictions de la classe des crépitants, 92.8 % pour 11225 prédictions de la classe des sifflements et 88.4 % pour 10696 prédictions de la classe qui regroupe les crépitants et les sifflements pour un total de 12095 prédictions pour chaque classe. Globalement, 91.8 % des prédictions sont correctes et 8.2 % sont fausses.

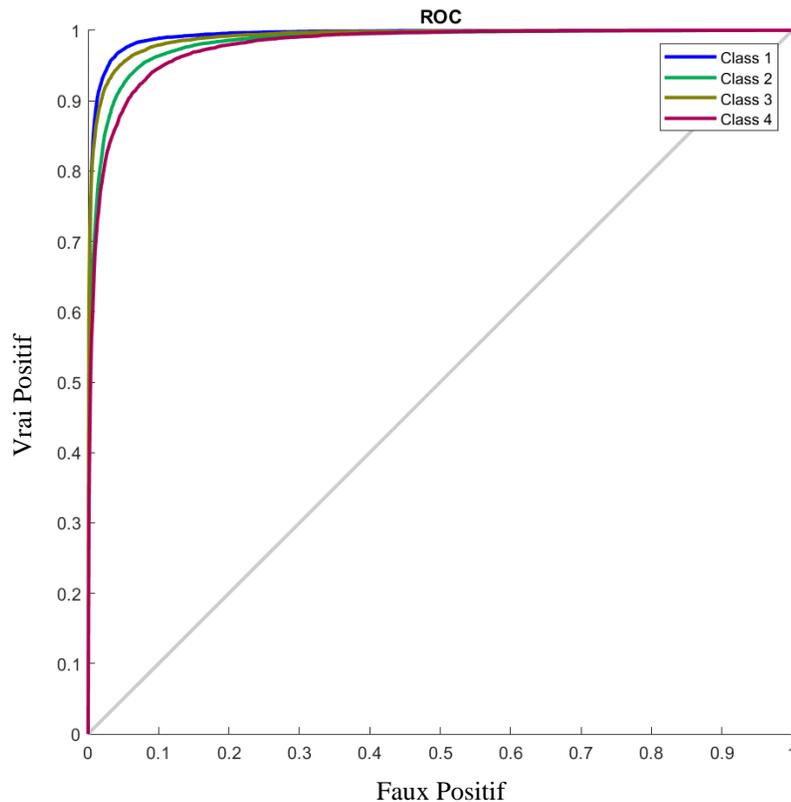


Figure 4.8 La Courbe ROC pour les 4 classes

À partir des courbes ROC de la figure 4.8, nous pouvons voir les performances de classification de chaque classe par rapport aux autres. La courbe bleue représente la classe normale suivie des sifflements et des crépitants, la courbe en rouge représente la classe qui contient en même temps les crépitants et les sifflements et qui est la moins bien classées parmi les autres. Une classification parfaite correspond au vrai positif égal à 1 et un faux positif égal à 0.

4.5.2 Utilisation de la totalité la base de données

Afin d'utiliser tous les enregistrements de la base de données, nous allons procéder comme mentionné dans la section de prétraitement de la base de données. Pour cela, nous allons effectuer les 18 combinaisons de tranches de la base de données, mais à chaque fois avec une combinaison différente. Au total nous avons obtenu six tranches de sons normaux

(No1 – No6), trois tranches de son contenant des crépitants (Cr1 – Cr3), une tranche de sons contenant les sifflements (Wh), et une tranche contenant les crépitants et les sifflements (CrWh). Les différentes combinaisons sont listées dans le tableau 4.10 :

Tableau 4.10 Combinaisons prises en compte pour l'utilisation de toutes les ressources de la base de données

Combinaison	Tranches de données correspondantes
01	No1,Cr1,Wh,CrWh
02	No2,Cr1,Wh,CrWh
03	No3,Cr1,Wh,CrWh
04	No4,Cr1,Wh,CrWh
05	No5,Cr1,Wh,CrWh
06	No6,Cr1,Wh,CrWh
07	No1,Cr2,Wh,CrWh
08	No2,Cr2,Wh,CrWh
09	No3,Cr2,Wh,CrWh
10	No4,Cr2,Wh,CrWh
11	No5,Cr2,Wh,CrWh
12	No6,Cr2,Wh,CrWh
13	No1,Cr3,Wh,CrWh
14	No2,Cr3,Wh,CrWh
15	No3,Cr3,Wh,CrWh
16	No4,Cr3,Wh,CrWh
17	No5,Cr3,Wh,CrWh
18	No6,Cr3,Wh,CrWh

Les 18 combinaisons ont été utilisées pour évaluer la méthode d'extraction de caractéristiques GFCC-PCA et WPT-GFCC et le classifieur MLP contenant 45 et 55 neurones dans la couche cachée. Les tableaux 4.11 et 4.13 représente les résultats de simulation de chaque combinaison de données.

Tableau 4.11 Résultats de simulation pour les 18 combinaisons de données avec la méthode GFCC-PCA et un réseau MLP 45-55

Combinaison	Sensitivité (%)	Spécificité (%)	Acc (%)	S _H (%)	S _{ICBHI} (%)
01	89.52	93.74	90.58	91.58	91.63
02	88.20	88.10	88.17	88.15	88.15
03	89.36	88.14	89.06	88.75	88.75
04	88.70	91.44	89.39	90.05	90.07
05	89.55	91.84	90.12	90.68	90.70
06	89.27	91.01	89.70	90.13	90.14
07	88.17	94.40	89.73	91.18	91.28
08	86.58	88.58	87.06	87.56	87.57
09	87.81	88.28	87.93	88.05	88.04
10	86.91	91.82	88.13	89.29	89.36
11	87.57	91.82	88.63	89.64	89.69
12	87.82	91.25	88.68	89.51	89.54
13	89.72	93.81	90.74	91.72	91.77
14	88.98	88.37	88.83	88.68	88.68
15	89.57	89.18	89.40	89.33	89.33
16	89.23	90.86	89.63	90.03	90.04
17	89.57	91.73	90.11	90.64	90.65
18	89.60	90.94	89.93	90.26	90.27

Après avoir effectué les différentes simulations, nous remarquons que les performances varient entre 86.58 % et 94.40 %. Pour chacune des 5 performances, nous avons calculé la moyenne pour les 18 combinaisons afin d'obtenir les résultats de l'utilisation de toute la base donnée, comme présenté dans le tableau 4.12.

Tableau 4.12 Moyennes des performances obtenus pour les 18 combinaisons de données avec la méthode GFCC-PCA

Se (%)	Sp (%)	Acc (%)	S _H (%)	S _{ICBHI} (%)
88.67	90.85	89.21	89.73	89.76

Tableau 4.13 Résultats de simulation pour les 18 combinaisons de données avec la méthode WPT-GFCC et un réseau MLP 45-55

Combinaison	Sensitivité (%)	Spécificité (%)	Acc (%)	S _H (%)	S _{ICBHI} (%)
01	90.84	94.78	91.83	92.77	92.81
02	89.18	88.34	88.97	88.76	88.76
03	81.27	89.35	83.29	85.12	85.31
04	89.77	93.33	90.66	91.51	91.55
05	84.48	82.92	84.09	83.69	83.70
06	81.95	92.47	84.58	86.89	87.21
07	79.49	94.78	83.31	86.46	87.13
08	72.27	80.40	74.30	76.12	76.34
09	79.05	88.44	81.40	83.48	83.75
10	87.33	92.99	88.74	90.07	90.16
11	87.42	92.27	88.63	89.78	89.84
12	74.31	82.66	76.40	78.26	78.48
13	91.21	94.59	92.05	92.87	92.90
14	88.61	87.86	88.42	88.23	88.24
15	90.14	88.98	89.85	89.56	89.56
16	90.18	92.52	90.76	91.33	91.35
17	71.91	93.34	77.27	81.24	82.63
18	90.45	92.84	91.05	91.64	91.66

Après avoir effectué les différentes simulations, nous remarquons que les résultats varient entre 71.91 % et 94.78 %. Nous avons ensuite calculé la moyenne des 5 performances obtenue avec les 18 combinaisons afin d'obtenir les résultats de l'utilisation de toute la base donnée, comme le montre le tableau 4.14.

Tableau 4.14 Moyennes des performances obtenus pour les 18 combinaisons de données avec la méthode WPT-GFCC

Se (%)	Sp (%)	Acc (%)	S _H (%)	S _{ICBHI} (%)
84.44	90.16	85.87	87.10	87.30

Nous concluons de la précédente expérience, que pour les 18 combinaisons effectuées sur la base de données, les paramètres du système (fenêtre de 2048 échantillons, 12 caractéristiques pour la combinaison GFCC-PCA, un réseau MLP de 45 et 55 neurones

cachés) sont optimaux pour l'extraction des caractéristiques et la classification des quatre classes de sons respiratoires avec une exactitude de 89.21 % pour une erreur de classification de 10.79 %.

4.5.3 Reproductibilité des résultats

Dans les essais précédents, nous avons fixé les poids synoptiques initiaux au début de la phase d'apprentissage du réseau de neurones dans le but d'avoir les mêmes résultats lorsque nous refaisons les mêmes tests. À présent, nous allons faire en sorte que les poids initiaux changent à chaque itération de test des tranches de la base de données, car ce n'était pas forcément ceux qui donnent les meilleurs résultats. Pour ce faire, nous allons effectuer 50 itérations pour chacune des 18 combinaisons et calculer la moyenne pour chacune des performances (Sp, Se, Acc, S_H, S_{ICBHI}) avec une marge en \pm , et ce pour les méthodes d'extraction de caractéristiques GFCC-PCA et WPT-GFCC. À la fin nous aurons effectué un total de 1800 simulations (900 pour chaque méthode d'extraction). Cela correspond à un total d'environ 720 heures de simulation en continu, afin d'englober toutes les possibilités. Les résultats des 1800 simulations sont représentées par les graphiques de la figure 4.9.

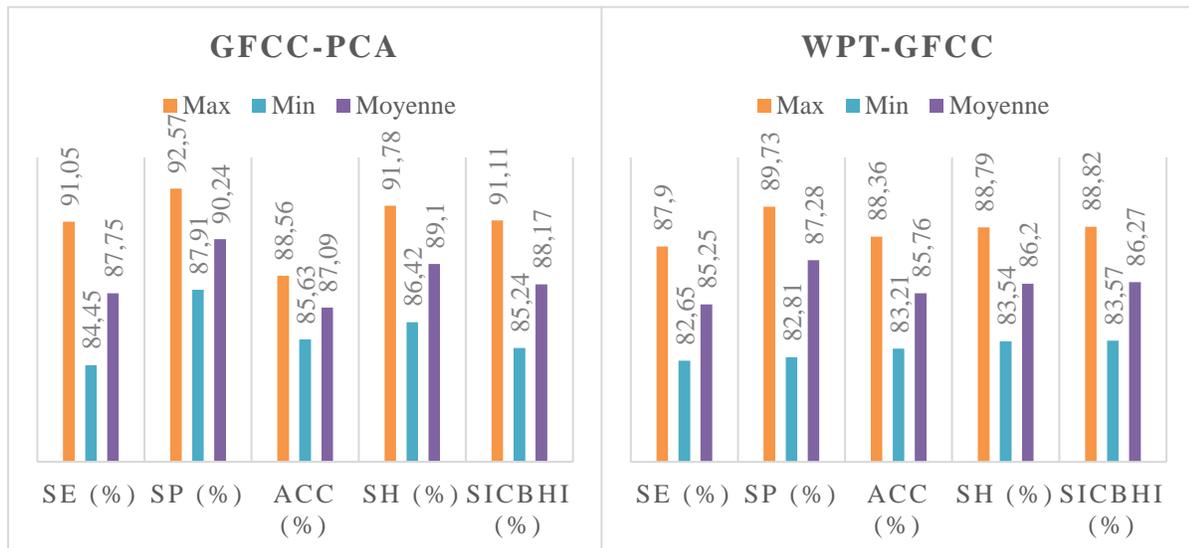


Figure 4.9 Représentation du Max, du Min et de la moyenne des 1800 résultats

Le tableau 4.15 représente les résultats précédents plus une marge autour de la valeur moyenne obtenue.

Tableau 4.15 Les valeurs du Max, min et la moyenne des 1800 résultats

		<i>Se (%)</i>	<i>Sp (%)</i>	<i>Acc (%)</i>	<i>S_H (%)</i>	<i>S_{ICBHI} (%)</i>
GFCC- PCA	Max	91.05	92.57	88.56	91.78	91.11
	Min	84.45	87.91	85.63	86.42	85.24
	Moyenne	87.75	90.24	87.09	89.10	88.17
	Marge	±3.3	±2.33	±1.46	±2.68	±2.93
WPT- GFCC	Max	87.90	89.73	88.36	88.79	88.82
	Min	82.65	82.81	83.21	83.54	83.57
	Moyenne	85.25	87.28	85.76	86.20	86.27
	Marge	±2.60	±4.47	±2.60	±2.66	±2.70

Après avoir effectué toutes les itérations possibles, nous avons abouti aux résultats représentés dans le tableau 4.15. Nous pouvons voir que l'exactitude de notre système est passée de 89.21% à 87.09% avec une marge de $\pm 1.46\%$ pour la méthode GFCC-PCA, et de 85.87 % à 85.76% avec une marge de $\pm 2.60\%$ pour la méthode WPT-GFCC, tandis que le score ICBHI est passé de 87.30% à 86.27 % avec une marge d'environ $\pm 2.70\%$.

4.6 RÉSULTATS DE LA CLASSIFICATION PAR SVM

Après avoir effectué les tests sur le classifieur à base de réseau MLP, nous proposons maintenant de réaliser une classification utilisant le classifieur SVM, en passant par les techniques d'extraction de caractéristiques qui ont donnés de bons résultats (GFCC-PCA, WPT-GFCC) et par toutes les combinaisons permettant l'utilisation de la totalité de la base de données.

Nous avons proposé d'utiliser la méthode *one-vs-one* qui permet d'utiliser des algorithmes de classification binaire pour la classification multi-classe. L'approche *one-vs-one* divise l'ensemble total de données en un autre ensemble de données pour chaque classe, par rapport à chaque autre classe. Cela va nous permettre d'utiliser le classifieur SVM qui

est à la base binaire pour effectuer la classification des quatre classes de notre base de données.

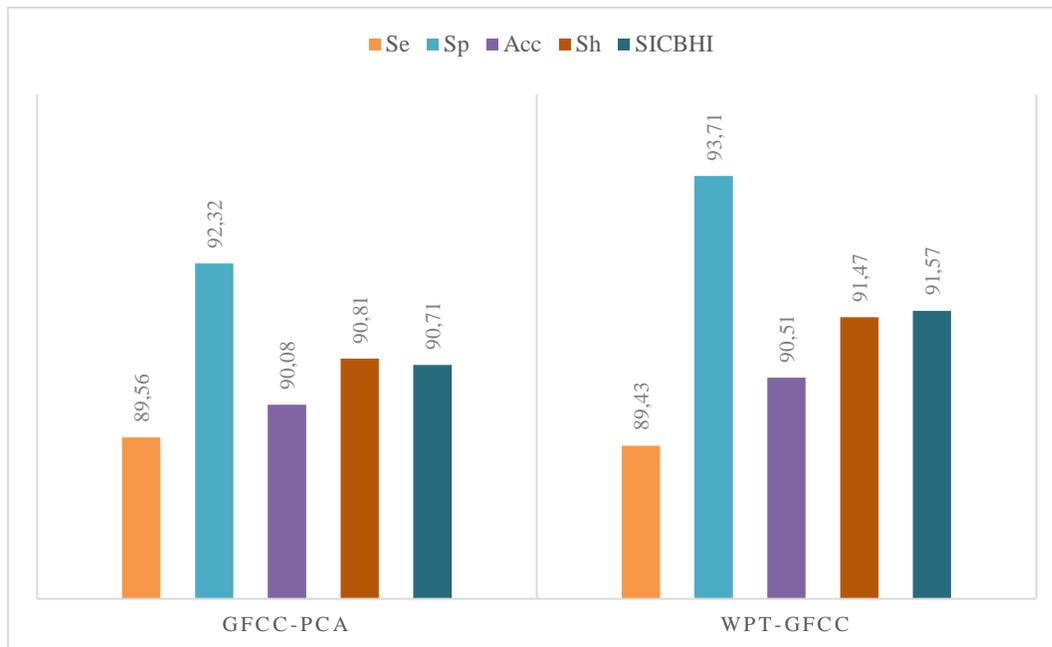


Figure 4.10 Résultats de simulation du classifieur SVM pour 18 combinaisons

Contrairement pour le classifieur MLP, la technique d'extraction de caractéristiques WPT combinée avec les GFCC a révélé le meilleur résultat pour le SVM également, pour un score ICBHI de plus de 91% et d'une exactitude de plus de 90% comme le montre la figure 4.10. Le tableau 4.16 résume les résultats de simulation de la classification par SVM.

Tableau 4.16 Résultats de classification par SVM pour 18 combinaisons

	Se (%)	Sp (%)	Acc (%)	S_H (%)	S_{ICBHI} (%)
GFCC-PCA	89.56	92.32	90.08	90.71	90.81
WPT-GFCC	89.43	93.71	90.51	91.57	91.47

Dans ce cas, nous pouvons conclure que le SVM a donné des résultats plus prometteurs que son concurrent le MLP. Pour l'étape de comparaison, nous allons nous limiter aux résultats basés sur la technique WPT-GFCC.

4.7 RÉSULTATS DE LA CLASSIFICATION PAR K-NN

Nous terminons les tests des classifieurs par la technique k-NN, qui est considérée comme la plus simple basé sur la mesure de distance, pour résoudre les problèmes de classification. Dans le cas de cette étude, nous allons tester notre classifieur pour trois valeurs des plus proches voisins $k = 1, 3$ et 5 et nous garderons le meilleur résultat pour la comparaison d'après. Nous avons adopté la méthode de mesure de distance euclidienne pour la recherche de k plus proches voisins.

Les 18 combinaisons citées précédemment ont été utilisées pour parcourir toute la base de données, et cela pour toutes les techniques d'extraction de caractéristiques. Les figures 4.11, 4.12 et 4.13 montrent les performances obtenues par les trois modèles k-NN.

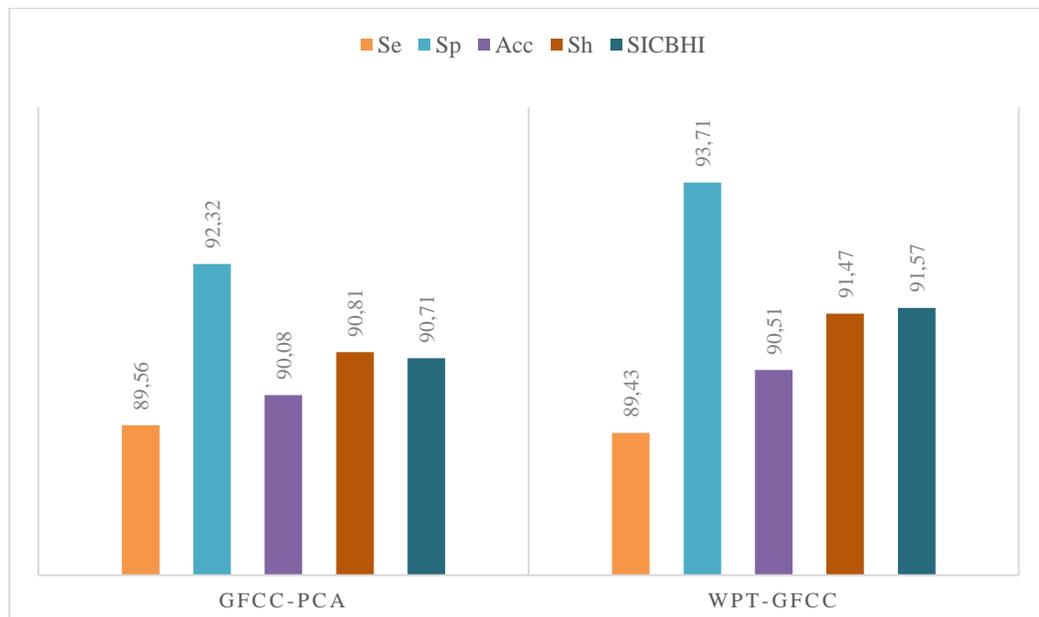


Figure 4.11 Résultats de simulation du classifieur k-NN pour $k = 1$

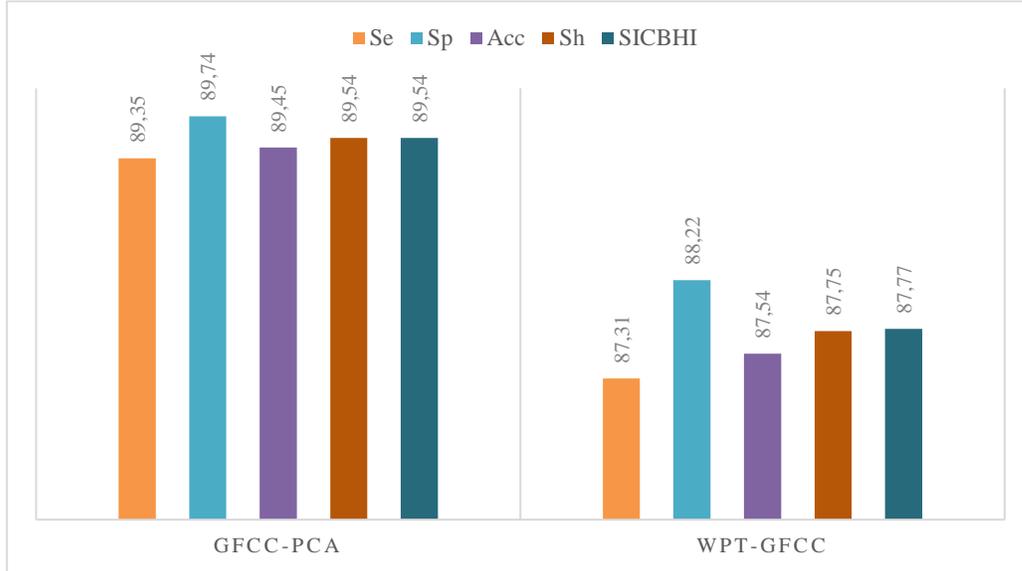


Figure 4.12 Résultats de simulation du classifieur k-NN pour k = 3

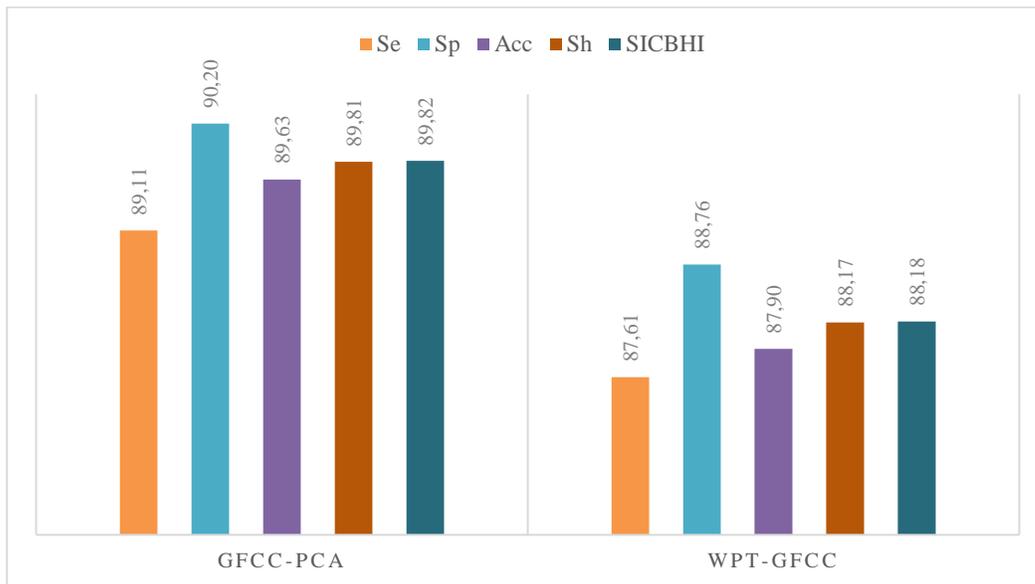


Figure 4.13 Résultats de simulation du classifieur k-NN pour k = 5

Nous remarquons que le score ICBHI et l'exactitude, entre k = 1 et k = 3, s'améliorent et grimpe d'environ 1 % pour les deux techniques d'extraction de caractéristiques, tandis que pour k = 5 leurs valeurs varient moins. Le tableau 4.17 résume les résultats pour les trois modèles k-NN.

Tableau 4.17 Résultats de classification par k-NN pour 18 combinaisons

	Se (%)	Sp (%)	Acc (%)	S _H (%)	S _{ICBHI} (%)
WPT-GFCC et k=1	85.86	88.19	86.44	86.99	87.02
GFCC-PCA et k=1	88.34	89.76	88.68	89.02	89.04
WPT-GFCC et k=3	87.31	88.22	87.54	87.75	87.77
GFCC-PCA et k=3	89.35	89.74	89.45	89.54	89.54
WPT-GFCC et k=5	87.61	88.76	87.90	88.17	88.18
GFCC-PCA et k=5	89.11	90.20	89.63	89.82	89.81

Pour la section de comparaison suivante, nous allons garder seulement le modèle qui utilise le k=5 plus proches voisins puisqu'il donne les meilleurs résultats.

4.8 COMPARAISON DES RÉSULTATS DES TROIS MODÈLES DE CLASSIFICATION RÉALISER

À présent, nous pouvons comparer les résultats de classification obtenus par chacun des trois classifieurs utilisés dans ce projet. Pour cela, nous prendrons en compte seulement les combinaisons (technique d'extraction de caractéristique et classifieur) qui ont fourni de bons résultats. Le tableau 4.18 résume les trois meilleurs résultats obtenus.

Tableau 4.18 Comparaisons des résultats de classification

	Se (%)	Sp (%)	Acc (%)	S _H (%)	S _{ICBHI} (%)
GFCC-PCA & MLP	87.75 ±3.3	90.24 ±2.3	87.09 ±1.46	89.10 ±2.68	88.17 ±2.93
WPT-GFCC & SVM	89.43	93.72	90.51	91.47	91.57
GFCC-PCA & k-NN	89.11	90.20	89.63	89.82	89.81

Durant cette étude, nous avons évalué l'exactitude, le score ICBHI et le score harmonique. Ces trois paramètres définissent le mieux la classification des quatre classes de sons respiratoires de la base de données ICBHI 2017. Après de nombreux essais et tests, il s'est avéré que le modèle de classification à base de SVM donne la meilleur exactitude et

score ICBHI, mais il reste que le MLP et le k-NN fournissent également des résultats assez élevés, et comparativement meilleurs que les résultats rencontrés dans la littérature comme la montre de tableau 4.19.

4.9 COMPARAISON AVEC DES APPROCHES SIMILAIRES

Dans cette section, nous comparons les performances de nos modèles de classification réalisés et qui sont basés sur l'extraction de caractéristiques WPT-GFCC et GFCC-PCA en combinaison avec les classifieurs à base de réseau MLP, SVM et k-NN. Avec des recherches bibliographiques sur les systèmes proposés, qui ont déjà été effectués, nous nous limiterons aux solutions qui utilisent les classifieurs plutôt similaires, afin que les résultats soient comparables.

Même si les performances de classification des systèmes proposés dans la littérature sont difficiles à comparer entre elles, elles atteignent ou dépassent 80 % d'exactitude, ce qui se compare bien aux résultats obtenus par nos modèles (~ 90 %). L'utilisation de données distinctes ou l'adoption de concepts d'analyse spécifiques, peuvent également rendre difficile la comparaison des résultats. Le tableau 4.19 donne un aperçu des études récentes publiées et utilisant la base de données ICBHI 2017.

Tableau 4.19 Performances de classification présentées par des études utilisant la base de données ICBHI 2017

Étude	Se (%)	Sp (%)	Acc (%)	S_{ICBHI} (%)	Sh (%)
(Serbes et al., 2018)	55.29	83.25	57.88	69.27	-
(Rocha et al., 2019)	-	-	-	52.5	-
(Wu & Li, 2020)	-	91	88	81	-
(Monaco et al., 2020)	-	-	85 ± 3	-	-
(Ren et al., 2021)	27.78	72.96	-	50.37	-
Notre modèle GFCC-PCA / MLP	87.75 ±3.3	90.24 ±2.3	87.09 ±1.46	88.17 ±2.93	89.10 ±2.68
Notre modèle WPT-GFCC / SVM	89.43	93.72	90.51	91.57	91.47
Notre modèle GFCC-PCA / k-NN	89.11	90.20	89.63	89.81	89.82

En général, les techniques de classification exigent un nombre élevé de données d'apprentissage pour chaque classe et les classes non équilibrées peuvent réduire les performances de classification. Afin de remédier à ces problèmes, nous avons adopté dans nos modèles une validation croisée à 10 blocs (*10-fold Cross validation*), afin d'utiliser toutes les données des tranches dans chaque itération. De plus, nous avons proposé une méthode de combinaison de tranches de données, puisque les classes de sons respiratoires n'étaient pas de même taille, dans le but d'exploiter tous les enregistrements de la base de données. Au final six systèmes ont été retenus en se basant sur les deux techniques d'extraction de caractéristiques, la première qui utilise en même temps les WPT et les GFCC et la deuxième qui applique les PCA aux coefficients GFCC. Chacune de ces techniques est combinée séparément à trois classifieurs assez optimisés: un réseau MLP, un SVM et un k-NN.

CONCLUSION GÉNÉRALE

Le présent travail consiste à étudier et mettre en place un système d'extraction de caractéristiques et de classification des sons respiratoires, dans le but d'assistance ou de suivi à distance, de patients présentant des maladies respiratoires. Ce travail s'inscrit dans un projet ayant cours au Département de Mathématiques d'Informatique et de Génie (DMIG) de l'Université du Québec À Rimouski (UQAR).

Le processus de classification que nous avons proposé dans ce projet passe par deux phases principales. La phase d'apprentissage dans laquelle nous donnons en entrée à un système de classification des données d'apprentissage, sous forme de vecteurs de caractéristiques extrais depuis les données originales. Ensuite, la phase de test sert à évaluer le système entraîné avec de nouvelles données sous forme de vecteurs de caractéristiques, mais qui n'ont évidemment pas étaient déjà présentés en phase d'apprentissage. L'étude effectué dans ce projet vise la classification de quatre classes de sons respiratoires : normaux, crépitants, sifflements, et les deux dernières mélangées.

Les expériences menées durant nos travaux de recherche étaient l'expérimentation de plusieurs combinaisons de techniques d'extraction de caractéristiques. Parmi les techniques testées, l'analyse cepstrale qui regroupe les coefficients MFCC (Coefficients cepstraux à l'échelle de Mel) et les coefficients GFCC (Coefficients cepstraux des fréquences Gammatone), la transformée en ondelettes et les paquets d'ondelettes. Nous avons également eu recours à l'utilisation de la méthode de réduction de dimensionnalité dite PCA (Principal component analysis) dans le but de réduire la taille du vecteur de caractéristiques.

Cependant, pour la deuxième étape qui consiste à classifier, nous avons opté pour l'utilisation des réseaux de neurones. Nous avons commencé les tests avec un réseau MLP (Multi Layer Perceptron) en ajustant quelques paramètres tels que le nombre de couches et

le nombre de neurones cachés, etc., selon les valeurs rencontrées dans la littérature. La validation croisée à k-blocs (*k-fold cross validation*) a été proposée et tester sous la forme 10-Bloc (*10-fold cross validation*) telle que mentionnée dans le quatrième chapitre. Par la suite, nous avons procédé à une étude paramétrique pour optimiser le système de classification. Nous avons également proposé l'utilisation d'un classifieur à base de SVM et de k-NN, afin de pouvoir comparer ces différents modèles de classification.

Les expériences ont été complétées en utilisant quatre classes de sons respiratoires à la fois. Vu que la base de données n'était pas équilibrée, nous avons eu recours à faire plusieurs simulations mais à chaque fois avec une nouvelle combinaison de sons respiratoires pour utiliser pratiquement toutes les données disponibles dans la base.

À la fin de tous les tests, nous avons conclu que la combinaison GFCC-WPT avec le SVM a fourni les meilleurs résultats, soit une exactitude de plus de 90 % et un score ICBHI de plus de 91 % (Tableau 4.19).

Comme travaux futur, nous proposons d'effectuer un pré-traitement différent de la base de données tel que l'utilisation du traitement d'images (les spectrogrammes ou les cochleogrammes) comme dans certaines recherches récentes. Pour la partie caractérisation/classification, nous proposons également l'étude de nouvelles techniques pour l'extraction de caractéristiques tel que l'extraction de caractéristiques à plusieurs niveaux (d'une représentation à échelle réduite à une représentation globale par le biais de généralisation progressives), ainsi que d'autres architectures de réseaux de neurones plus avancées ou l'utilisation de l'apprentissage profond. Ainsi, les performances de classification utilisant la base de données ICBHI 2017 peuvent être améliorées.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Admuthe, S. S., & Patil, P. H. (2015). Feature extraction method-MFCC and GFCC used for Speaker Identification Miss. *IJSRD - International Journal for Scientific Research & Development*, Vol. 3(04, 2015). <https://ijsrd.com/Article.php?manuscript=IJSRDV3I40661>
- Alkandari, A., & Aljaber, S. J. (2015). Principle Component Analysis algorithm (PCA) for image recognition. *2015 Second International Conference on Computing Technology and Information Management (ICCTIM)*, 76-80. <https://doi.org/10.1109/ICCTIM.2015.7224596>
- Amose, J., Manimegalai, P., Narmatha, C., & Raj, P. (2022). Comparative Performance Analysis of Kernel Functions in Support Vector Machines in the Diagnosis of Pneumonia using Lung Sounds. *2022 2nd International Conference on Computing and Information Technology (ICCIT)*, 320-324. <https://doi.org/10.1109/ICCIT52419.2022.9711608>
- Bahoura, M. (2009). Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes. *Computers in Biology and Medicine*, 39(9), 824-843. <https://doi.org/10.1016/j.combiomed.2009.06.011>
- Bahoura, M. (2018). FPGA implementation of an automatic wheezing detection system. *Biomedical Signal Processing and Control*, 46, 76-85. <https://doi.org/10.1016/j.bspc.2018.05.017>
- Bahoura, M., & Park, C.-W. (2012). FPGA-implementation of dynamic time delay neural network for power amplifier behavioral modeling. *Analog Integrated Circuits and Signal Processing*, 73(3), 819-828. <https://doi.org/10.1007/s10470-012-9857-0>
- Bahoura, M., & Simard, Y. (2012). Serial combination of multiple classifiers for automatic blue whale calls recognition. *Expert Systems with Applications*, 39(11), 9986-9993. <https://doi.org/10.1016/j.eswa.2012.01.156>
- Boujelben, O. (2017). *Étude des techniques de détection des sibilants dans les sons respiratoires en vue d'un traitement temps-réel sur FPGA*. [Mémoire de maîtrise, Université du Québec à Rimouski]. Rimouski, Québec, Canada. <https://semaphore.uqar.ca/id/eprint/1274/>

- Boujelben, O., & Bahoura, M. (2016). FPGA implementation of an automatic wheezes detector based on MFCC and SVM. *2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, 647-650. <https://doi.org/10.1109/ATSIP.2016.7523173>
- Boujelben, O., & Bahoura, M. (2018). Efficient FPGA-based architecture of an automatic wheeze detector using a combination of MFCC and SVM algorithms. *Journal of Systems Architecture*, 88, 54-64. <https://doi.org/10.1016/j.sysarc.2018.05.010>
- Brownlee, J. (2020). *Tour of Evaluation Metrics for Imbalanced Classification*. <https://machinelearningmastery.com/tour-of-evaluation-metrics-for-imbalanced-classification/>
- Cawley, G. C. (2006). Leave-One-Out Cross-Validation Based Model Selection Criteria for Weighted LS-SVMs. *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, pp 1661-1668. <https://doi.org/10.1109/IJCNN.2006.246634>
- Demir, F., Ismael, A. M., & Sengur, A. (2020). Classification of Lung Sounds With CNN Model Using Parallel Pooling Structure. *IEEE Access*, 8, 105376-105383. <https://doi.org/10.1109/ACCESS.2020.3000111>
- Dua, M., Aggarwal, R. K., & Biswas, M. (2018). Performance evaluation of Hindi speech recognition system using optimized filterbanks. *Engineering Science and Technology, an International Journal*, 21(3), 389-398. <https://doi.org/10.1016/j.jestch.2018.04.005>
- Elissa, A., Geert, t. J., Connie, Y., Society., C. P., Section., A., Committee., D. T. a. H. S., & Section., R. H. (2020). *Paediatric asthma and COVID-19*. Société canadienne de pédiatrie. Retrieved 05-02 from www.cps.ca/en/documents/position/paediatric-asthma-and-covid-19
- Ertekin, S. (2009). *Learning in extreme conditions: Online and active learning with massive, imbalanced and noisy data*. [Thèse de doctorat, The Pennsylvania State University Park]. Pennsylvania, USA.

- Gao, R. X., & Yan, R. (2011). Wavelet Packet Transform. In R. X. Gao & R. Yan (Eds.), *Wavelets: Theory and Applications for Manufacturing* (pp. 69-81). Springer US. https://doi.org/10.1007/978-1-4419-1545-0_5
- Haris Mita, J., Ganesh Babu, C., & Gowri Shankar, M. (2021). Performance Analysis of Dimensionality Reduction using PCA, KPCA and LLE for ECG Signals. *IOP Conference Series: Materials Science and Engineering*, 1084(1), 012005. <https://doi.org/10.1088/1757-899x/1084/1/012005>
- Hashemi, A., Arabalibiek, H., & Agin, K. (2011). Classification of wheeze sounds using wavelets and neural networks. *International Conference on Biomedical Engineering and Technology*, 11(2011), 127-131. https://www.researchgate.net/publication/266569975_Classification_of_Wheeze_Sounds_Using_Wavelets_and_Neural_Networks/stats
- Haykin, S. S. (1999). *Neural networks : a comprehensive foundation*. Prentice Hall.
- Hosseinzadeh, M. (2020). 4 - Robust control applications in biomedical engineering: Control of depth of hypnosis. *Control Applications for Biomedical Engineering Systems*, 89-125. <https://doi.org/10.1016/B978-0-12-817461-6.00004-4>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.
- Jeevan, M., Dhingra, A., Hanmandlu, M., & Panigrahi, B. K. (2017). Robust Speaker Verification Using GFCC Based i-Vectors. *Proceedings of the International Conference on Signal, Networks, Computing, and Systems*, 85-91.
- Jivani, A. G. (2013). The Novel k Nearest Neighbor Algorithm. *2013 International Conference on Computer Communication and Informatics*, 1-4. <https://doi.org/10.1109/ICCCI.2013.6466287>
- Joshi, D., Khajuria, A., & Joshi, P. (2017). An Automatic Non-Invasive Method for Parkinson's disease Classification. *Computer Methods and Programs in Biomedicine*, 145. <https://doi.org/10.1016/j.cmpb.2017.04.007>

- Kandaswamy, A., Kumar, C. S., Ramanathan, R. P., Jayaraman, S., & Malmurugan, N. (2004). Neural classification of lung sounds using wavelet coefficients. *Computers in Biology and Medicine*, 34(6), 523-537. [https://doi.org/10.1016/S0010-4825\(03\)00092-1](https://doi.org/10.1016/S0010-4825(03)00092-1)
- Lu, X. G. (2008). *Reconnaissance automatique des crépitants et des sibilants dans les sons acoustiques respiratoires*. [Mémoire de maîtrise, Université du Québec à Rimouski]. Rimouski, Québec, Canada. <https://semaphore.uqar.ca/id/eprint/298/>
- Lv, L., Song, X., & Sun, W. (2020). Modify Leave-One-Out Cross Validation by Moving Validation Samples around Random Normal Distributions: Move-One-Away Cross Validation. *Applied Sciences*, 10(7), 2448. <https://www.mdpi.com/2076-3417/10/7/2448>
- Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7), 674-693. <https://doi.org/10.1109/34.192463>
- Meng, F., Shi, Y., Wang, N., Cai, M., & Luo, Z. (2020). Detection of Respiratory Sounds Based on Wavelet Coefficients and Machine Learning. *IEEE Access*, 8, 155710-155720. <https://doi.org/10.1109/ACCESS.2020.3016748>
- Monaco, A., Amoroso, N., Bellantuono, L., Pantaleo, E., Tangaro, S., & Bellotti, R. (2020). Multi-Time-Scale Features for Accurate Respiratory Sound Classification. *Applied Sciences*, 10(23), 8606. <https://www.mdpi.com/2076-3417/10/23/8606>
- Nabney, I. T. (2002). *NETLAB. Algorithms for Pattern Recognition*. (Nabney2002)
- Navlani, A. (2018). *Learn K-Nearest Neighbor(KNN) Classification and build KNN classifier using Python Scikit-learn package*. <https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn>
- OMS. (2020). *Les 10 principales causes de mortalité*. <https://www.who.int/fr/news-room/fact-sheets/detail/the-top-10-causes-of-death>

- Pal, K., & Patel, B. V. (2020). Data Classification with k-fold Cross Validation and Holdout Accuracy Estimation Methods with 5 Different Machine Learning Techniques. *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, 83-87. <https://doi.org/10.1109/ICCMC48092.2020.ICCMC-00016>
- Pelletier, C. (2006). *Classification des sons respiratoires en vue d'une détection automatique des sibilants*. [Mémoire de maîtrise, Université du Québec à Rimouski]. Rimouski, Québec, Canada. <https://semaphore.uqar.ca/id/eprint/36/>
- Petmezas, G., Cheimariotis, G.-A., Stefanopoulos, L., Rocha, B., Paiva, R. P., Katsaggelos, A. K., & Maglaveras, N. (2022). Automated Lung Sound Classification Using a Hybrid CNN-LSTM Network and Focal Loss Function. *Sensors (Basel, Switzerland)*, 22(3), 1232. <https://doi.org/10.3390/s22031232>
- Quandt, V., Pacola, E., Pichorim, S., Gamba, H., & Sovierzoski, M. (2015). Pulmonary crackle characterization: Approaches in the use of discrete wavelet transform regarding border effect, mother-wavelet selection, and subband reduction. *Research on Biomedical Engineering*, 31, 148-159. <https://doi.org/10.1590/2446-4740.0639>
- Rajoub, B. (2020). Chapter 2 - Characterization of biomedical signals: Feature engineering and extraction. *Biomedical Signal Processing and Artificial Intelligence in Healthcare*, 29-50. <https://doi.org/10.1016/B978-0-12-818946-7.00002-0>
- Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-Validation. *Encyclopedia of Database Systems*, 532-538. https://doi.org/10.1007/978-0-387-39940-9_565
- Ren, Z., Nguyen, T. T., & Nejdil, W. (2021). Prototype Learning for Interpretable Respiratory Sound Analysis. *arXiv*. <https://doi.org/10.48550/arXiv.2110.03536>
- Reynolds, D. A., Quatieri, T.F., Dunn, R.B. (2000). Speaker Verification Using Adapted Gaussian Mixture Models. *Digital Signal Processing*, 10(1–3). <https://doi.org/10.1006/dspr.1999.0361>
- Rizal, A., Hidayat, R., & Nugroho, H. A. (2015). Determining lung sound characterization using Hjorth descriptor. *2015 International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC)*, 54-57. <https://doi.org/10.1109/ICCEREC.2015.7337053>

- Rizal, A., Priharti, W., Rahmawati, D., & Mukhtar, H. (2022). Classification of Pulmonary Crackle and Normal Lung Sound Using Spectrogram and Support Vector Machine. *Journal of Biomimetics, Biomaterials and Biomedical Engineering*, 55, 143-153. <https://doi.org/10.4028/p-tf63b7>
- Rocha, B. M., Filos, D., Mendes, L., Serbes, G., Ulukaya, S., Kahya, Y. P., Jakovljevic, N., Turukalo, T. L., Vogiatzis, I. M., Perantoni, E., Kaimakamis, E., Natsiavas, P., Oliveira, A., Jácome, C., Marques, A., Maglaveras, N., Pedro Paiva, R., Chouvarda, I., & de Carvalho, P. (2019). An open access database for the evaluation of respiratory sound classification algorithms. *Physiological Measurement*, 40(3), 035001. <https://doi.org/10.1088/1361-6579/ab03ea>
- Sehgal, S., Singh, H., Agarwal, M., Bhasker, V., & Shantanu. (2014). Data analysis using principal component analysis. *2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom)*, 45-48. <https://doi.org/10.1109/MedCom.2014.7005973>
- Semmad, A. (2021). *Réseaux de neurones pour la détection des sibilants dans les sons acoustiques respiratoires dans le but d'une implantation efficace temps-réel sur circuit FPGA* [Mémoire de maîtrise, Université du Québec à Rimouski]. Rimouski, Québec, Canada. <https://semaphore.uqar.ca/id/eprint/1994/>
- Sengupta, N., Sahidullah, M., & Saha, G. (2016). Lung sound classification using cepstral-based statistical features. *Comput Biol Med*, 75, 118-129. <https://doi.org/10.1016/j.combiomed.2016.05.013>
- Serbes, G., Ulukaya, S., & Kahya, Y. P. (2018). An Automated Lung Sound Preprocessing and Classification System Based On Spectral Analysis Methods. *Precision Medicine Powered by pHealth and Connected Health*, 45-49. https://doi.org/10.1007/978-981-10-7419-6_8
- Shi, Y., Wang, G., Niu, J., Zhang, Q., Cai, M., Sun, B., Wang, D., Xue, M., & Zhang, X. D. (2018). Classification of Sputum Sounds Using Artificial Neural Network and Wavelet Transform. *International journal of biological sciences*, 14(8), 938-945. <https://doi.org/10.7150/ijbs.23855>

- Statistique Canada. (2022). *Décès, selon la cause, Chapitre X : Maladies de l'appareil respiratoire (J00 à J99)*. Gouvernement du Canada. <https://www150.statcan.gc.ca/t1/tbl1/fr/tv.action?pid=1310078201>
- Taunk, K., De, S., Verma, S., & Swetapadma, A. (2019). A Brief Review of Nearest Neighbor Algorithm for Learning and Classification. *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, 1255-1260. <https://doi.org/10.1109/ICCS45141.2019.9065747>
- Tocchetto, M. A., Bazanella, A. S., Guimaraes, L. V., Fragoso, J. L., & Parraga, A. (2014). An Embedded Classifier of Lung Sounds based on the Wavelet Packet Transform and ANN. *IFAC Proceedings Volumes*, 47, 2975-2980. <https://doi.org/10.3182/20140824-6-ZA-1003.01638>
- Ullah, A., Khan, M. S., Khan, M. U., & Mujahid, F. (2021). Automatic Classification of Lung Sounds Using Machine Learning Algorithms. *2021 International Conference on Frontiers of Information Technology (FIT)*, 131-136. <https://doi.org/10.1109/FIT53504.2021.00033>
- Valente, A., Soares, S., Ferreira, N., Sousa, J., Francisco, A., Santos, D., & Saraiva, A. (2020). Classification of Respiratory Sounds with Convolutional Neural Network [CONF]. *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies*, 138-144. <https://doi.org/10.5220/0008965101380144>
- Vapnik, V. (1999). *The nature of statistical learning theory*. Springer science & business media.
- Wang, T., & Karayiannis, N. B. (1998). Detection of Microcalcifications in Digital Mammograms Using Wavelets. *IEEE transactions on medical imaging*, 17, 498-509. <https://doi.org/10.1109/42.730395>
- Wu, L., & Li, L. (2020). Investigating into segmentation methods for diagnosis of respiratory diseases using adventitious respiratory sounds. *Annu Int Conf IEEE Eng Med Biol Soc*, 2020, 768-771. <https://doi.org/10.1109/embc44109.2020.9175783>
- Yadav, S., & Shukla, S. (2016). Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification. *2016 IEEE 6th*

International Conference on Advanced Computing (IACC), 78-83.
<https://doi.org/10.1109/IACC.2016.25>

Yan, J., Laflamme, S., Singh, P., Sadhu, A., & Dodson, J. (2020). A Comparison of Time-Frequency Methods for Real-Time Application to High-Rate Dynamic Systems. *Vibration*, 3(3), 204-216. <https://www.mdpi.com/2571-631X/3/3/16>

Yang, F., Liu, S., Dobriban, E., & Woodruff, D. P. (2021). How to Reduce Dimension With PCA and Random Projections. *IEEE Transactions on Information Theory*, 67(12), 8154-8189. <https://doi.org/10.1109/TIT.2021.3112821>