



UNIVERSITE LIBANAISE
FACULTE DE TECHNOLOGIE
SAIDA

**Génie Des Réseaux Informatiques Et
Télécommunications**

**RAPPORT DE PROJET FIN
D'ETUDES**

**Analyse des sons respiratoires basée sur l'apprentissage
automatique pour la détection de la
Maladie Pulmonaire Obstructive Chronique**

Préparé Par : Tamara Al Hajjar

Encadré par : Dr Nassib Abdallah

Pr Bassam Daya

Mrad Sleiman (Doctorant en

1^{ère} année de thèse)

Année Universitaire : 2022-2023

Table de matières

Remerciements	4
I. Introduction	Error! Bookmark not defined.
II. Cadre du stage	6
III. Gestion du projet	8
IV. Travail réalisé	9
V. Résultats	16
VI. Conclusion et perspectives	18
VII. Références	19

Table de figures

Figure 1 : Diagramme de la méthodologie utilisée pour la classification des sons respiratoires à partir des signaux audio	7
Figure 2 : Diagramme de Gantt	8
Figure 3 : Diagramme bête à cornes	8
Figure 4: Signal audio avant filtrage	9
Figure 5 : Signal audio après filtrage	10
Figure 6 : Décomposition du signal principal en plusieurs fenêtres et application de la FFT sur chacune d'elles	11
Figure 7: L'échelle de conversion entre Hz et Mel	11
Figure 8: Diagramme explicatif de la méthode de calcul du spectrogramme	12
Figure 9: Multiplication des filtres par les échantillons de chaque fenêtre du signal	12
Figure 10: 26 filterbanks	13
Figure 11 : Un spectrogramme.....	13
Figure 12 : Diagramme explicatif de la méthode de calcul des MFCCs.....	14
Figure 13 : MFCC	14
Figure 14 : Les segments chevauchées	15
Figure 15 : Hamming window allant jusqu'à 4000 Hz.....	15
Figure 16 : Multiplication du Hamming window par un segment du signal principal	15
Figure 17 : Matrices de confusion des classificateurs XGBoost et Random Forest respectivement.....	17

Table de tableaux

Tableau 1 : Accuracy, Precision, Recall, F1-score pour chacun des classificateurs en les appliquant sur les 920 audios de la base de données ICBHI	16
--	----

Remerciements

Chers Docteur Nassib Abdallah, Professeur Bassam Daya,

Je tiens tout d'abord à exprimer ma profonde gratitude envers chacun de vous pour la formidable expérience professionnelle que j'ai vécue lors de mon projet. Je suis extrêmement reconnaissante de la confiance que vous m'avez accordée, ainsi que de la diversité des missions que vous m'avez confiées. Votre soutien et vos conseils avisés ont été inestimables tout au long de cette période.

Je tiens également à adresser mes remerciements chaleureux à Mrad Sleiman, doctorant en 1ère année de thèse, qui a été mon maître de stage. Sa patience, sa pédagogie et son accompagnement constant ont contribué de manière significative à mon développement professionnel. Sa disponibilité et sa bienveillance ont rendu cette expérience d'apprentissage encore plus enrichissante.

Grâce à votre encadrement attentif, j'ai pu acquérir de nouvelles compétences et approfondir mes connaissances dans le domaine de l'apprentissage automatique (Machine Learning), ainsi que m'introduire à l'apprentissage profond (Deep Learning). Vos conseils éclairés, vos remarques constructives et votre ouverture d'esprit m'ont permis de me surpasser et d'élargir mes connaissances.

I. Introduction

L'apprentissage automatique (ML) et l'apprentissage profond (DL) jouent un rôle de plus en plus crucial dans le domaine biomédical. Ces approches d'apprentissage automatique permettent d'extraire des informations précieuses à partir de vastes ensembles de données biomédicales, ce qui contribue à une meilleure compréhension des maladies, à des diagnostics plus précis et à des traitements plus efficaces.

Ces techniques peuvent être utilisées pour analyser des images médicales, telles que les radiographies, les scanners et les IRM.

Cependant, analyser les images prend beaucoup de temps et de calcul, donc les signaux audio y seront un complémentaire puissant pour révéler les maladies pulmonaires de manière non invasive et rentable simplement en écoutant le son de la poitrine.

L'auscultation pulmonaire est un élément précieux de l'évaluation clinique des patients. Le traitement des signaux audio bénéficie grandement des révolutions technologiques innovantes, qui offrent aux cliniciens et aux médecins des moyens optimaux pour diagnostiquer les maladies pulmonaires, et détecter toutes sortes de crépitements, sifflement ou autres sons pulmonaires.

La maladie pulmonaire obstructive chronique (MPOC) est le terme médical utilisé pour décrire des formes d'affections pulmonaires graves qui provoquent un rétrécissement, une obstruction et une inflammation des voies respiratoires, ce qui rend la respiration difficile.

La maladie pulmonaire obstructive chronique est une affection courante. En fait, selon l'Organisation mondiale de la santé (OMS), il y a environ 251 millions de cas de MPOC dans le monde. Les chiffres suggèrent que d'ici 2030, la MPOC pourrait devenir la troisième cause de décès dans le monde.[1]

Les principales causes d'inflammation du poumon sont le tabagisme, plus grand facteur de risque de la MPOC et l'inhalation de toxines (chimiques, poussière, polluants) lors d'une exposition professionnelle ou par la pollution atmosphérique. Le facteur génétique peut aussi être l'une de ses causes.[2]

L'importance et la nécessité du diagnostic proactif des maladies pulmonaires, ont poussé les chercheurs à réaliser des études dans ce domaine. Des articles de recherches précédents, ont traité le sujet de détection des maladies pulmonaires, et la classification des anomalies respiratoires.

Siddiqui, H.U.R et al [3], introduisant une nouvelle méthode de détection des maladies pulmonaires, le radar UWB ayant des capacités non intrusives et sans tacle, afin de détecter les sons respiratoires à une distance 1.5m du patient. Les caractéristiques calculées sont ensuite fournies à des modèles ML et DL. Le modèle DL LSTM donne l'accuracy la plus élevée de 93 %.

Lam Pham et al [4] ont proposé un modèle CNN-MoE pour effectuer 4 tâches sur la base de données ICBHI, nommées tâche 1-1, tâche 1-2, tâche 2-1, tâche 2-2, et ont obtenu 79%, 84%, 91% et 92% respectivement pour chacune des tâches, surpassant les modèles de l'état de l'art.

II. Cadre du stage

Etat de l'art :

Le but principal de ce stage est d'obtenir un modèle de classification des sons respiratoires en se basant sur les signaux audio et l'apprentissage automatique.

Pour ce faire, plusieurs outils et différentes méthodes ont été appliqués :

-L'utilisation de la distribution **Anaconda** et du logiciel **Spyder** pour la programmation en Python. La librairie « librosa » a été utilisée principalement dans ce projet.

-La base de données sujet de ce projet : la base de données ICBHI [5] formée de signaux audios, et créée par deux équipes de recherche au Portugal et en Grèce.

1- Description de la base de données ICBHI

L'ensemble de données ICBHI fournit une grande base de sons respiratoires étiquetés comprenant 920 enregistrements audio d'une durée combinée de 5,5 heures. Les durées d'enregistrement sont inégales, allant de 10 à 90 secondes, et ont été enregistrés avec une large gamme de fréquences d'échantillonnage à partir de 4 kHz à 44,1 kHz : 90 enregistrements à 4 KHz, 6 à 10 KHz et 824 à 44.1 KHz. Au total, l'ensemble de données contient des enregistrements de 128 patients, qui sont identifiés comme étant en bonne santé ou présentant l'une des maladies ou affections respiratoires suivantes : MPOC, Bronchiectasie, Asthme, infection des voies respiratoires supérieures et inférieures, Pneumonie, Bronchiolite. Ces étiquettes d'état respiratoire sont liées à des fichiers d'enregistrement audio. Dans chaque enregistrement audio, quatre types différents de cycle respiratoire sont présentés – appelés 'Crackle', 'Wheeze', 'Both' (Crackle & Wheeze) et 'Normal'. Ces cycles, étiquetés par des experts, incluent le début identifié et temps de décalage. Les cycles ont différentes longueurs d'enregistrement allant de 0,2 à 16,2 secondes, le nombre de cycles étant déséquilibré (soit 1864, 886, 506 et 3642 cycles respectivement pour 'Crackle', 'Wheeze', 'Both' et 'Normal').

2- Modèle de classification des sons respiratoires :

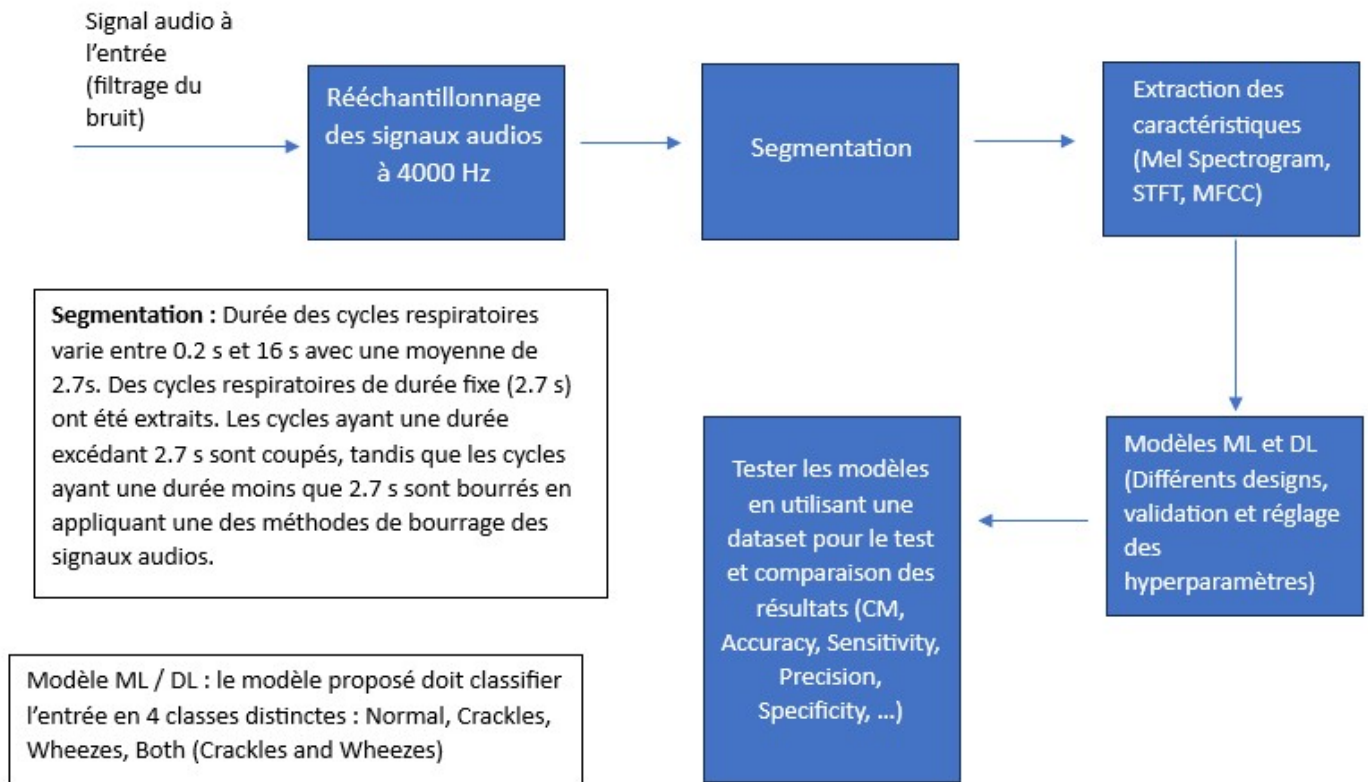


Figure 1 : Diagramme de la méthodologie utilisée pour la classification des sons respiratoires à partir des signaux audio

III. Gestion du projet



Figure 2 : Diagramme de Gantt

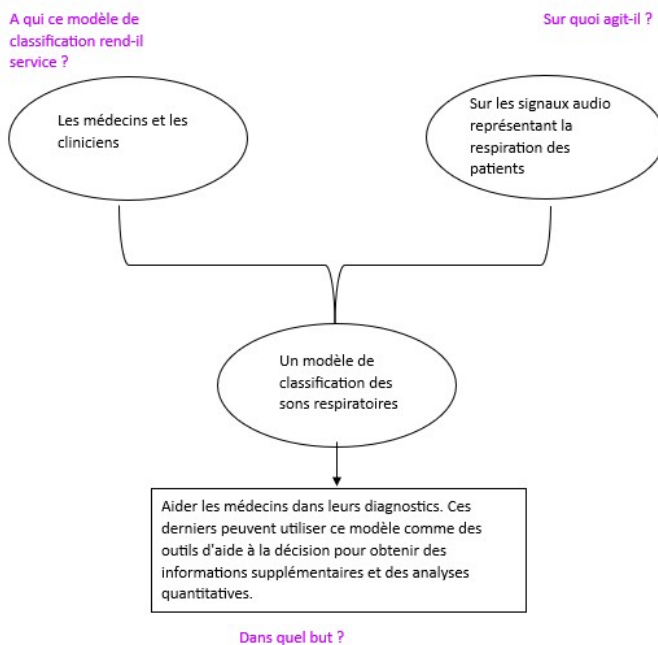


Figure 3 : Diagramme bête à cornes

IV. Travail réalisé

Afin de construire notre modèle de classification, nous avons suivi les étapes du pipeline expliqué dans la [figure 1](#).

1-1- Acquisition des signaux :

Nous allons utiliser les 920 audios de la base de données ICBHI dans les étapes suivantes de notre projet.

1-2- Prétraitement :

La base de données représente des cas de la vie réelle, donc les enregistrements capturés contiennent des bruits de l'arrière-plan, et le bruit d'autres organes tel que le battement du cœur. Pour cela il est nécessaire de débruité ces signaux. Cela est effectué en créant un filtre Butterworth d'ordre 10 et de fréquence de coupure 80 Hz. Comme déjà expliqué dans le [chapitre II-2](#), la fréquence des signaux varie entre 44,1 KHz, 4 KHz et 10 KHz.

Pour unifier ces fréquences, on choisit de rééchantillonner les signaux à la fréquence **4 KHz**. Ce choix de fréquence est justifié par le fait que, suite aux travaux précédents sur cette base de données, on a trouvé que les informations utiles s'étendent jusqu'à 2 KHz. D'après le théorème de Shannon, la fréquence d'échantillonnage est supérieure ou égal au double de la fréquence de coupure ; soit $f_c \geq 2 \cdot f_c \geq 2 \cdot 2000 = 4000$ Hz.

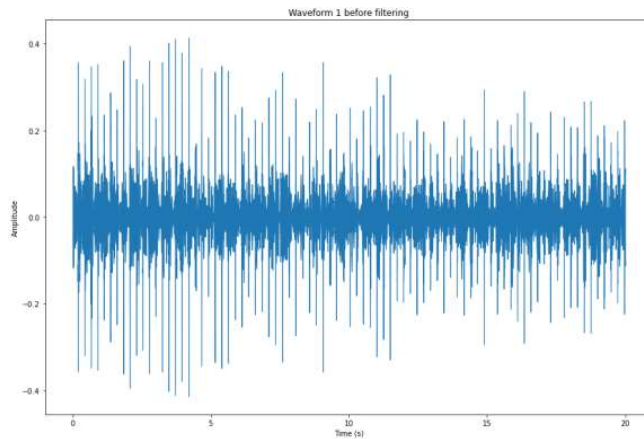


Figure 4: Signal audio avant filtrage

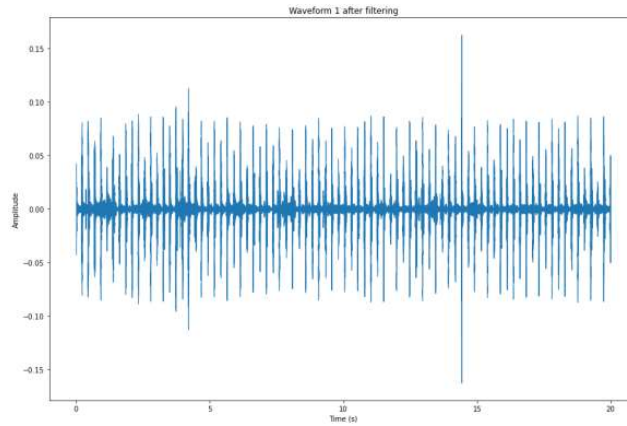


Figure 5 : Signal audio après filtrage

Après ce premier prétraitement, on passe au deuxième prétraitement.

En revenant aussi au pipeline de la [figure 1](#), nous devons effectuer une segmentation. Ce processus peut prendre trois formes, selon le signal audio :

- Si la durée du signal est inférieure à 2.7 s : on le complète jusqu'à 2.7 s à l'aide d'un bourrage par répétition. Le bourrage par répétition est une méthode qui consiste à répéter le signal un certain nombre de fois.
- Si la durée du signal est égale à 2.7 s : le signal est pris complètement tel qu'il est.
- Si la durée du signal est supérieure à 2.7 s : on coupe le signal jusqu'à 2.7 s.

1-3- Extraction des caractéristiques :

Nous avons extrait 3 types de caractéristiques :

1- Log Mel Filterbanks energy feature (ou simplement Mel Filterbanks ou Mel Spectrogram): [6]

Comme on le sait, la fréquence, et donc l'amplitude du signal audio varie au cours du temps. Ce sont les signaux **apériodiques**. L'application du Fast Fourier Transform (FFT), est efficace pour analyser un signal, mais cela seulement pour les signaux stationnaires. Donc que faire pour obtenir, à partir de notre signal audio apériodique de départ, un ensemble de signaux stationnaires ? On décompose le signal en un ensemble de fenêtres, et on applique sur chacune d'elles la FFT. Ce processus produit les *spectrogrammes*, qui sont affichés en Hz, et que nous voulons les transformer en Mel qui est une unité de mesure du son. L'être humain ne perçoit pas le son d'une façon linéaire mais d'une façon logarithmique, d'où il était nécessaire de convertir les fréquences exprimées en Hz en leur équivalent en Mel, c-à-d transformer l'intensité du son du Hz en Mel. Donc les points consécutifs seront séparés équitablement.

Les *filterbanks*, qui sont des filtres triangulaires seront utilisés. Nous multiplions chaque fenêtre par ces filterbanks, autrement dit, on donne un poids à chaque échantillon du signal. Les résultats de multiplication sont ensuite additionnés. Même processus pour les filtres restants. Dans notre travail, nous avons calculer 26 filterbanks.

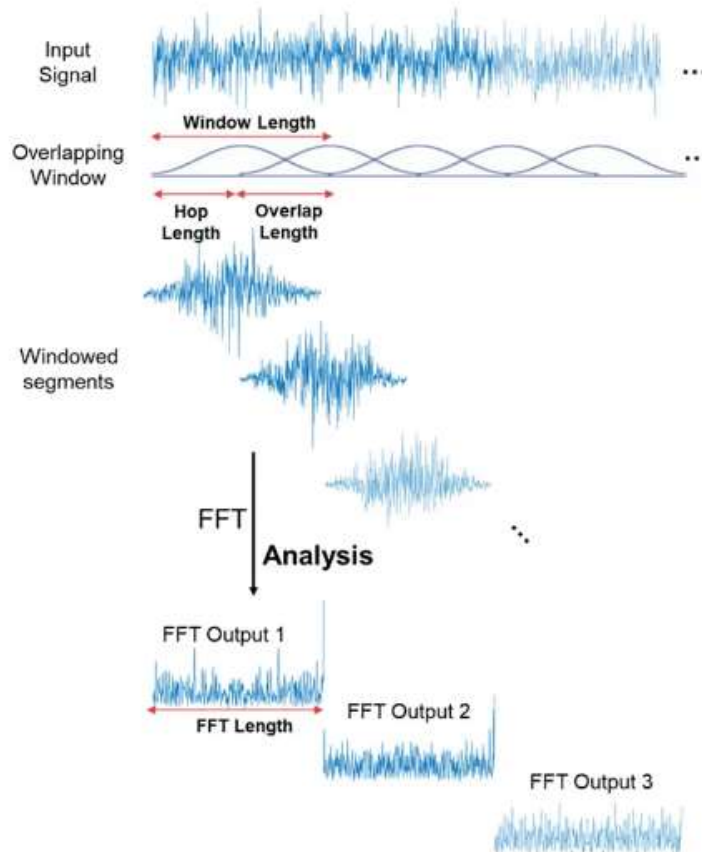


Figure 6 : Décomposition du signal principal en plusieurs fenêtres et application de la FFT sur chacune d'elles

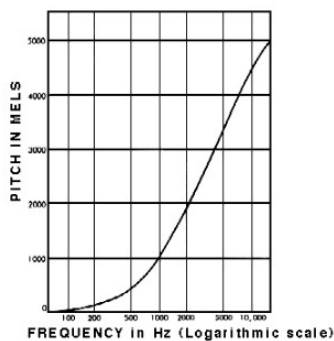


Figure 7: L'échelle de conversion entre Hz et Mel

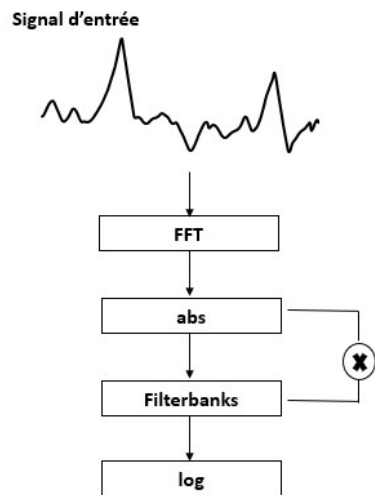


Figure 8: Diagramme explicatif de la méthode de calcul du spectrogramme

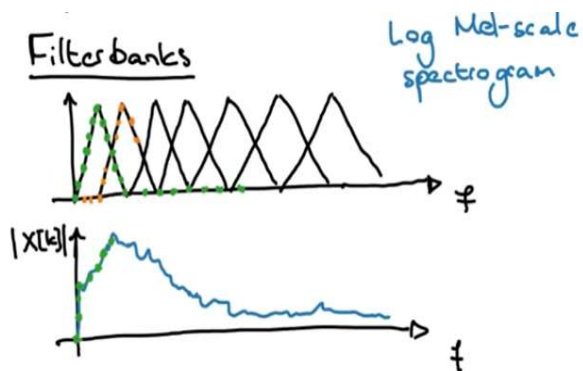


Figure 9: Multiplication des filtres par les échantillons de chaque fenêtre du signal

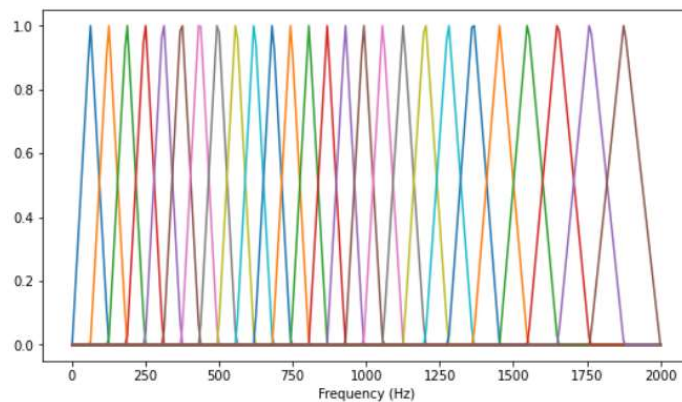


Figure 10: 26 filterbanks

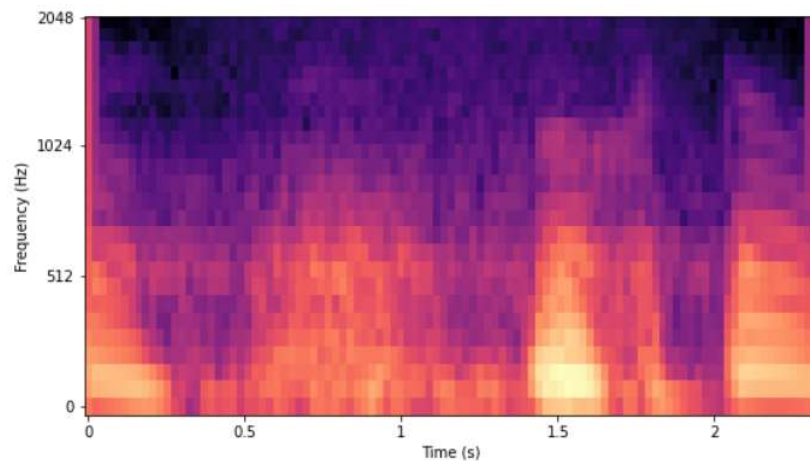


Figure 11 : Un spectrogramme

2- Mel Frequency Cepstral Coefficients (MFCC): [7][8]

Les MFCCs visent à capturer les caractéristiques de la perception du son par le système auditif humain, et sont dérivées de l'échelle de fréquences de Mel. Le chevauchement des filtres utilisés dans la méthode précédente cause une corrélation très forte entre ces filtres. D'où le besoin de les séparer. On obtient les MFCC en suivant les mêmes étapes pour calculer 'Log Mel Frequency Spectrogramme', mais en ajoutant 2 étapes :

Tout d'abord on applique DCT (Discrete Cosine Transform = Transformée de Cosinus Discrète) qui est une sorte de compression et qui conserve seulement les premiers coefficients. Nous nous intéressons aux premiers coefficients, car ce sont ceux qui représentent des aspects plus simples de la forme spectrale du signal, et donc seront plus pertinents dans l'apprentissage du

modèle que les coefficients d'ordre plus élevé, et qui représentent le bruit dans le signal.

Pour des raisons de simplicité, nous avons calculer 3 MFCC.

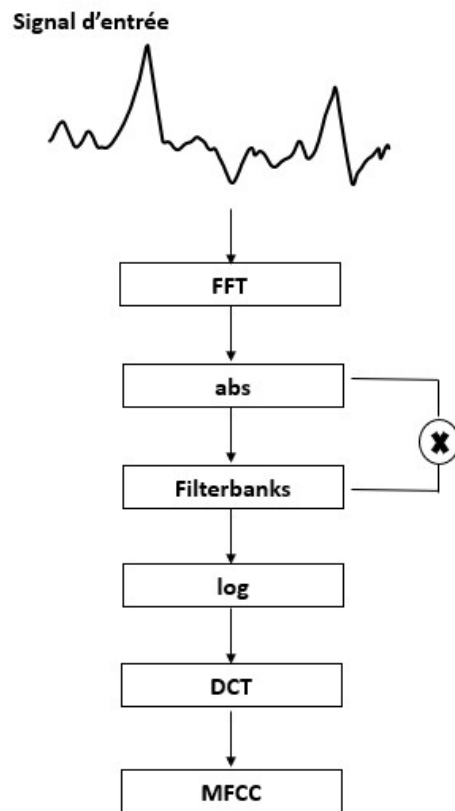


Figure 12 : Diagramme explicatif de la méthode de calcul des MFCCs

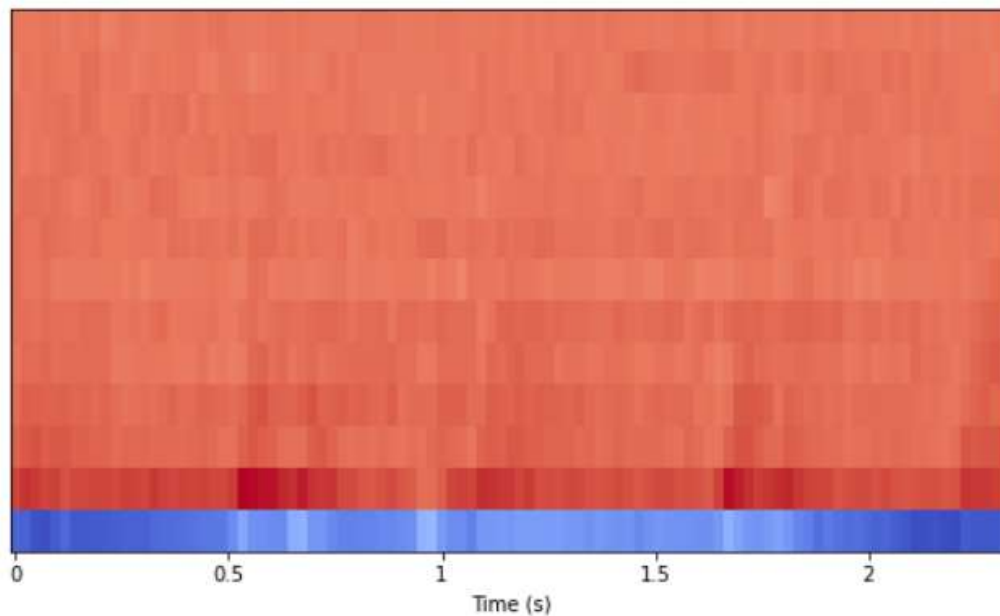


Figure 13 : MFCC

3- Short Time Fourier Transform (STFT):[9][10]

La STFT représente un signal audio dans le domaine temps-fréquence. Elle divise le signal en courts segments temporels chevauchants et calcule la transformée de Fourier pour chaque segment, tel que chaque segment représente un signal *stationnaire*. STFT fournit des informations sur les composantes fréquentielles présentes dans le signal et leur évolution dans le temps.

Chaque segment, a une durée de 25 ms et le chevauchement est de 10 ms.

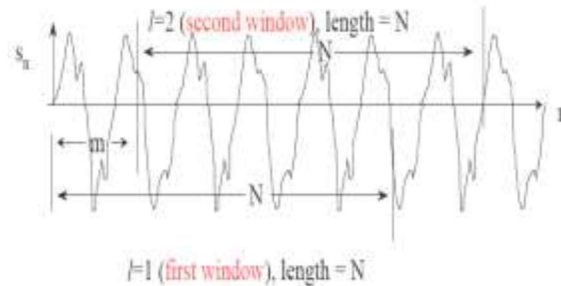


Figure 14 : Les segments chevauchés

Pour minimiser les distorsions spectrales lors de la segmentation, on multiplie chaque segment par une *fenêtre*, le 'Hamming Window'.

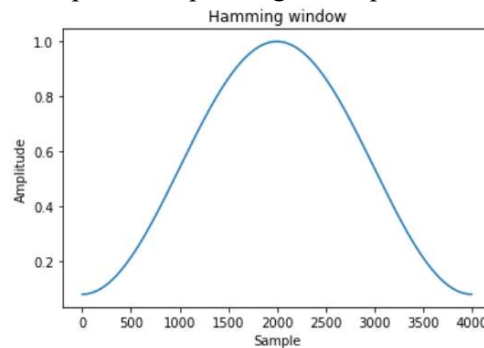


Figure 15 : Hamming window allant jusqu'à 4000 Hz

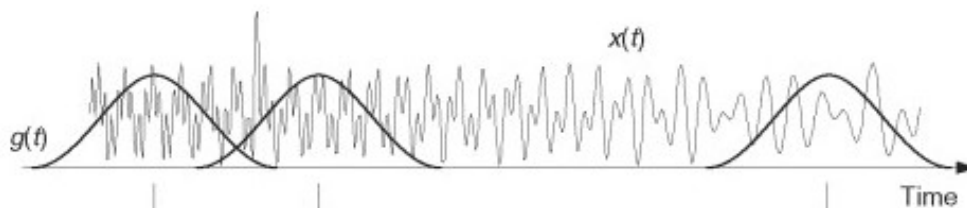


Figure 16 : Multiplication du Hamming window par un segment du signal principal

Après ces étapes, on extrait les caractéristiques STFT du signal.

V. Résultats

En revenant au pipeline de la [figure 1](#), on voit qu'après la phase d'extraction de caractéristiques, on doit construire un modèle de classification.

Tout d'abord, nous avons essayé 2 algorithmes ML : Extreme Gradient Boost (XGBoost) - Random Forest (RF)

Random Forest :

Cet algorithme de classification consiste à construire des arbres de décision (decision trees) en parallèle, et chacun prend un sous-échantillon de la base de données initiale pour effectuer les prédictions. La prédiction finale du modèle sera le vote majoritaire de ces arbres, ce qui réduit l'overfitting et augmente l'exactitude des prédictions.

Extreme Gradient Boost :

L'idée principale du gradient boost est que chaque prédicteur corrige l'erreur de son prédécesseur. Donc dans XGBoost, les arbres sont construits séquentiellement, et vont faire les prédictions d'une façon itérative, chacun se basant sur les erreurs résiduelles du modèle précédent pour adapter le modèle suivant. Les résultats prédits correctement reçoivent un poids inférieur et des poids supérieurs sont associés aux résultats mal prédits. La prédiction finale est une somme pondérée de toutes les prédictions des arbres. Le XGBoost améliore la performance du modèle et réduit l'overfitting.

L'overfitting signifie que le modèle ne donne pas des prédictions précises (surtout sur les données de test). Dans ce cas l'erreur de l'entraînement est peu significative tandis que celle du test est large.

Les résultats obtenus sont affichés dans le tableau ci-dessous :

Classifieur	Taux de test (%)	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
XGBoost	20	65.28	64	65	63
RF	20	66.45	68	66	63

Tableau 2 : Accuracy, Precision, Recall, F1-score pour chacun des classificateurs en les appliquant sur les 920 audios de la base de données ICBHI.

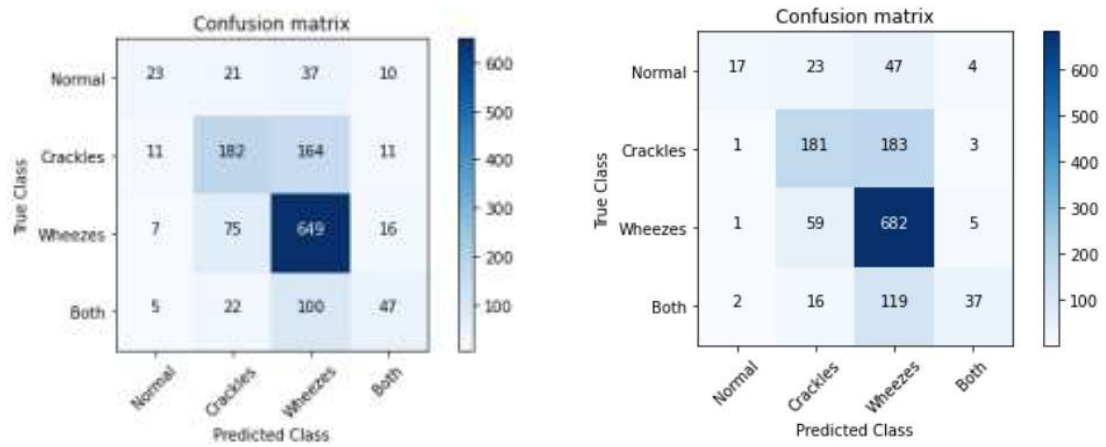


Figure 17 : Matrices de confusion des classificateurs XGBoost et Random Forest respectivement

Le tableau 1 montre que l'accuracy de Random Forest surpasse légèrement celle de XGBoost, les deux étant aux alentours de 66%. Ces résultats sont affichés dans les matrices de confusion de la figure 19.

VI. Conclusion et perspectives

Dans ce projet, nous avons effectué la classification des sons respiratoires conduisant à la détection des maladies pulmonaires, à travers l'analyse des signaux audio représentant les 4 classes (Normal, Crackles, Wheezes, Both) de la base de données ICBHI.

Le diagnostic précoce des maladies pulmonaires est crucial afin de traiter les affections pulmonaires le plus tôt possible, avant que les symptômes ne deviennent plus sévères ou que la maladie ne progresse pas trop loin. L'imagerie médicale est sans doute utile et efficace dans certaines situations, pour détecter et classer les maladies pulmonaires.

Cependant, l'utilisation des signaux audio dans la classification des maladies pulmonaires peut offrir certains avantages complémentaires à l'imagerie médicale traditionnelle. En effet, certains problèmes pulmonaires peuvent produire des sons caractéristiques, tels que des crépitements et les sifflements. Ces sons peuvent être enregistrés et analysés pour identifier certaines affections, comme l'asthme et la MPOC. De plus, les signaux audio sont faciles à utiliser et à collecter comparativement à l'imagerie médicale.

Le rééchantillonnage des signaux et leur segmentation permettent une extraction plus nette des caractéristiques.

L'ensemble de caractéristiques extraites est ensuite fourni à deux classificateurs, le Random Forest étant meilleur avec une accuracy de 66.45 % et une précision de 68%.

Dans de prochains travaux, nous appliquerons un modèle Deep Learning pour tester sa performance de classification en comparaison avec celle des algorithmes du Machine Learning.

Et puisque la base de données ICBHI présente un déséquilibre de données (le nombre de cycles des 4 classes ne sont pas égaux), nous voulons adopter une approche pour traiter les ensembles de données déséquilibrés consistant à suréchantillonner la classe minoritaire avant de les fournir au modèle. Le défi de travailler avec des ensembles de données déséquilibrés est que la plupart des techniques d'apprentissage profond ignorent et, à leur tour, ont de mauvaises performances sur la classe minoritaire, bien que ce soit généralement la performance sur la classe minoritaire qui soit la plus importante. L'une des méthodes qui pourrait être utilisée est la technique SMOTE, ou Synthetic Minority Oversampling Technique. Il s'agit donc d'un type d'augmentation des données pour la classe minoritaire.

VII. Références

- [1] <https://cnfs.ca/pathologies/maladie-pulmonaire-obstructive-chronique-mpoc>
- [2] <https://fr.gaapp.org/diseases/copd>
- [3] Hafeez-Ur-Rehman Siddiqui, Ali Raza, Adil Ali Saleem, Furqan Rustam, Isabel de la Torre Diez, Daniel Gavilanes Aray, Vivian Lipari, Imran Ashraf, Sandra Dudley: An Approach to Detect Chronic Obstructive Pulmonary Disease Using UWB Radar- Based Temporal and Spectral Features (2023). <https://doi.org/10.3390/diagnostics13061096>
- [4] Lam Pham, Huy Phan, Ramaswamy Palaniappan, Alfred Mertins, Ian McLoughlin: CNN-MoE Based Framework for Classification of Respiratory Anomalies and Lung Disease Detection (2023). [10.1109/JBHI.2021.3064237](https://doi.org/10.1109/JBHI.2021.3064237)
- [5] <https://www.kaggle.com/datasets/vbookshelf/respiratory-sound-database>
- [6] <https://medium.com/analytics-vidhya/understanding-the-mel-spectrogram-fca2afa2ce53>
- [7] <https://medium.com/@tanveer9812/mfccs-made-easy-7ef383006040>
- [8] <https://librosa.org/doc/main/generated/librosa.feature.mfcc.html>
- [9] https://www.researchgate.net/publication/331635757_Frame_Blocking_and_Windowing_Speech_Signal
- [10] <https://librosa.org/doc/main/generated/librosa.stft.html>
- [11] <https://machinelearningmastery.com/multi-class-imbalanced-classification/>