الجـــــــــــــــــــمهـوريـــــة الجـــــــــــــزائـريـة الديمـــــــقـراطيـة الـشـــــــــعبيـة

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

وزارة التـــــــــــــــــــــــــعـليـم العـــــــالي و البـــــــــــحث العـــــلـمـــي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

جـــــــــــــــامعة أبي بكـــر بـلـقـايـد – تـلمســـــــــــــان –

Université Aboubakr Belkaïd – Tlemcen –

Faculté de TECHNOLOGIE

MEMOIRE

Présenté pour l'obtention du **diplôme** de **MASTER**
En : Génie Biomédical
Spécialité :Électronique et maintenance Biomédical

Par :
Toubane Wail

Sujet

L'analyse des sons pulmonaire par intelligence artificielle

Soutenu publiquement, le 23 / 06 / 2026 , devant le jury composé de :

Mme Belaroussi sara	PR	Université de Tlemcen	Président
M Youbi ridha	MCB	Université de Tlemcen	Examinateur
Mme FEROUI amel	MCA	Université de Tlemcen	Encadreur

Année universitaire : 2025 / 2026

Dédicace

Je dédie ce modeste travail :

À ma très chère mère, source inépuisable de tendresse, de patience et d'encouragement. Aucune dédicace ne saurait exprimer à sa juste valeur l'amour et le profond respect que je te porte. Tu as toujours été mon premier soutien et ma plus grande force. Que Dieu te protège et t'accorde santé et longue vie.

À mon cher père, qui m'a toujours soutenu et encouragé tout au long de mes études. Tu as su m'inculquer le sens de la responsabilité, de l'optimisme et de la confiance en soi face aux difficultés de la vie. Tes prières et tes sacrifices m'ont permis d'arriver là où je suis aujourd'hui. Que Dieu te garde.

À mes chers frères et sœurs, pour leur soutien moral, leur encouragement permanent et leur complicité. Qu'ils trouvent dans ce travail l'expression de mon amour et de ma gratitude. À tous ceux qui m'ont aidé de près ou de loin dans l'accomplissement de ce travail.

Toubane Wail

Remerciements

Avant tout, je remercie Dieu le Tout-Puissant de m'avoir donné le courage, la volonté et la patience pour mener à bien ce travail.

Je tiens à exprimer ma profonde gratitude et mes sincères remerciements à ma directrice de mémoire, Madame FEROUI Amel, pour avoir accepté de diriger ce travail, pour sa disponibilité, ses orientations précieuses, ses conseils judicieux et son suivi rigoureux tout au long de la réalisation de ce projet. Sa confiance et son soutien constant ont été déterminants dans l'aboutissement de ce mémoire.

Mes remerciements vont également aux membres du jury qui ont accepté d'examiner et d'évaluer ce travail. Leurs remarques et suggestions constructives contribueront certainement à enrichir ce travail.

Je remercie vivement l'ensemble des enseignants du département de Génie Biomédical de la Faculté de Technologie de l'Université Abou Bekr Belkaïd de Tlemcen, qui ont contribué à ma formation académique et scientifique tout au long de mon parcours universitaire.

Je souhaite également adresser mes remerciements à tous mes camarades de promotion pour les moments de partage, d'entraide et de solidarité.

Enfin, je ne saurais terminer sans remercier chaleureusement ma famille — mes parents, mes frères et mes sœurs — pour leur soutien inconditionnel, leur patience et leurs encouragements permanents. Ce travail est aussi le fruit de leurs sacrifices.

Résumé

L'auscultation pulmonaire traditionnelle à l'aide d'un stéthoscope mécanique reste un outil de diagnostic clinique de première ligne essentiel, mais elle est limitée par la perception subjective de l'oreille humaine et la variabilité inter-observateur, en particulier pour distinguer des anomalies acoustiques subtiles comme les crépitants, les sibilances et les ronchus. Ce travail présente le développement d'un système intelligent hybride d'aide à la décision clinique pour la classification automatique des maladies respiratoires à partir des sons pulmonaires. Pour pallier les limites des modèles de classification classiques, nous proposons une architecture hybride originale fusionnant deux approches complémentaires : d'une part, un réseau de neurones convolutif (CNN) personnalisé conçu pour extraire des caractéristiques spectrales à haute résolution à partir des spectrogrammes de Mel (représentations 2D temps-fréquence) ; d'autre part, un modèle pré-entraîné YAMNet de Google exploitant le principe de l'apprentissage par transfert (Transfer Learning) pour capturer des descripteurs sémantiques globaux à haute dimension. Le système a été entraîné et validé sur une base de données consolidée de 917 enregistrements audio, obtenue par la fusion de la base de données internationale ICBHI 2017 et d'une base de données complémentaire (Fraiwan et al.), répartis sur 6 classes pathologiques : BPCO (COPD), sujets sains (Healthy), infections des voies respiratoires supérieures (URTI), dilatation des bronches (Bronchiectasis), pneumonie (Pneumonia) et bronchiolite (Bronchiolitis). Pour surmonter le déséquilibre sévère des classes, des stratégies de régularisation et d'augmentation robustes (telles que le Mixup, le Label Smoothing, la pondération des classes et un fort Dropout) ont été intégrées. Les résultats expérimentaux démontrent l'efficacité de l'approche hybride qui atteint une exactitude globale (Accuracy) de 87.8 %, un F1-score macro de 69.1 % et un F1-score pondéré de 88.7 %, représentant une amélioration de +26.7 % du F1-score macro par rapport à une architecture CNN classique non optimisée. Le modèle a été déployé sous la forme d'un serveur d'inférence en temps réel avec une interface web asynchrone (FastAPI/HTML5) permettant aux cliniciens de téléverser des sons, de visualiser les prédictions probabilitaires, d'inspecter les spectrogrammes et de générer des rapports cliniques PDF structurés.

Mots-clés : sons pulmonaires, classification, apprentissage profond, CNN, YAMNet, apprentissage par transfert, spectrogramme de Mel, aide à la décision clinique, FastAPI.

Abstract

Traditional lung auscultation using a mechanical stethoscope remains a critical first-line clinical diagnostic tool, but is limited by the subjective perception of the human ear and inter-observer variability, particularly in distinguishing subtle acoustic anomalies such as crackles, wheezes, and rhonchi. This work presents the development of a hybrid intelligent clinical decision support system for the automatic classification of respiratory diseases from lung sounds. To overcome the limitations of classical classification models, we propose an original hybrid architecture merging two complementary approaches: on the one hand, a custom convolutional neural network (CNN) designed to extract high-resolution spectral features from 2D Mel-spectrograms; on the other hand, Google's pre-trained YAMNet model leveraging transfer learning to capture high-dimensional global semantic descriptors. The system was trained and validated on a consolidated database of 917 audio recordings, obtained by merging the international ICBHI 2017 database and a complementary database (Fraiwan et al.), distributed over 6 pathological classes: COPD, Healthy subjects, Upper Respiratory Tract Infections (URTI), Bronchiectasis, Pneumonia, and Bronchiolitis. To address the severe class imbalance, robust regularization and data augmentation strategies (such as Mixup, Label Smoothing, class weighting, and high Dropout) were integrated. Experimental results demonstrate the effectiveness of the hybrid approach, achieving an overall accuracy of 87.8%, a macro F1-score of 69.1%, and a weighted F1-score of 88.7%, representing a +26.7% improvement in macro F1-score compared to a conventional, non-optimized CNN architecture. The model was deployed as a real-time inference server with an asynchronous web interface (FastAPI/HTML5) allowing clinicians to upload sounds, visualize probabilistic predictions, inspect spectrograms, and generate structured clinical PDF reports.

Keywords: lung sounds, classification, deep learning, CNN, YAMNet, transfer learning, Mel-spectrogram, clinical decision support, FastAPI.

الملخص

لا تزال عملية سماع أصوات الرئة التقليدية باستخدام السماعة الطبية الميكانيكية أداة أساسية للتشخيص السريري الأولي، ولكنها تظل محدودة بالتقدير الذاتي للأذن البشرية والتفاوت الملحوظ بين الأطباء في التقييم، خاصة عند محاولة التمييز بين الأصوات غير الطبيعية الدقيقة مثل الطقطقة والصفير والخرخرة. يقدم هذا العمل تطويرًا لنظام ذكي هجين لدعم القرار السريري بهدف التصنيف التلقائي للأمراض التنفسية بناءً على أصوات الرئة. ولتجاوز حدود نماذج التصنيف الكلاسيكية، نقترح بنية هجينة مخصصة لاستخراج الخصائص الطيفية عالية الدقة من (CNN)مبتكرة تدمج طريقتين متكاملتين: الأولى، شبكة عصبية التفافية الذي يستفيد من تقنية تعلم نقل Googleالمسبق التدريب من YAMNetالطيفية ثنائية الأبعاد؛ والثانية، نموذج Melمخططات لالتقاط الواصفات الصوتية العامة عالية الأبعاد. تم تدريب النظام والتحقق من صحته باستخدام (Transfer Learning)المعرفة مع ICBHI 2017قاعدة بيانات مدمجة تحتوي على 917تسجيلًا صوتيًا، تم الحصول عليها من خلال دمج قاعدة البيانات الدولية والأشخاص ،(COPD)وتتوزع على 6فئات مرضية: الانسداد الرئوي المزمن ،(Fraiwan et al.)قاعدة بيانات تكميلية والالتهاب ،(Bronchiectasis)وتوسع القصبات ،(URTI)والتهابات الجهاز التنفسي العلوي ،(Healthy)الأصحاء ولالتغلب على مشكلة عدم توازن البيانات الشديد، تم دمج .(Bronchiolitis)والتهاب القصيبات ،(Pneumonia)الرئوي وتظهر .(بنسبة عالية Dropoutوأوزان الفئات و Label Smoothingو Mixupمثل)استراتيجيات قوية للتحسين والانتظام بنسبة (Macro)عادل F1-scoreالنتائج التجريبية كفاءة النهج الهجين المقترح، حيث حقق دقة إجمالية بلغت 87.8%، ومتوسط CNNالعام مقارنة ببنية F1-scoreمرجح بنسبة 88.7%، وهو ما يمثل تحسنًا بنسبة +26.7% في الـ F1-score69.1%، و (FastAPI/HTML5)التقليدية غير المحسنة. تم نشر النموذج النهائي كخادم استدلال في الوقت الفعلي مع واجهة ويب تفاعلية تمكن الأطباء من رفع الأصوات، ومعاينة الاحتمالات التشخيصية، وفحص المخططات الطيفية، واستخراج تقارير طبية منظمة بصيغةPDF.

الكلمات المفتاحية: أصوات الرئة، تصنيف، تعلم عميق، شبكة عصبية التفافية ،YAMNet ،مخطط Melتعلم نقل المعرفة، FastAPI. ،الطيفي، دعم القرار السريري

Nomenclature

Sigle / Abréviation	Signification
API	Application Programming Interface
BPCO	Broncho-Pneumopathie Chronique Obstructive
CAD	Computer-Aided Diagnosis (Diagnostic Assisté par Ordinateur)
CDSS	Clinical Decision Support System (Système d'Aide à la Décision Clinique)
CNN	Convolutional Neural Network (Réseau de Neurones Convolutif)
COPD	Chronic Obstructive Pulmonary Disease
CPU	Central Processing Unit
dB	Décibels
DDR4	Double Data Rate 4th Generation
EMG	Électromyogramme
FastAPI	Framework Python asynchrone pour le développement d'APIs REST
GPU	Graphics Processing Unit
HIPAA	Health Insurance Portability and Accountability Act
Hz	Hertz
ICBHI	International Conference on Biomedical and Health Informatics
IoT	Internet of Things (Internet des Objets)
JSON	JavaScript Object Notation
kHz	Kilohertz
KNN	K-Nearest Neighbors (K plus proches voisins)

LDA	Linear Discriminant Analysis
LSTM	Long Short-Term Memory
MFCC	Mel-Frequency Cepstral Coefficients
NVMe	Non-Volatile Memory Express
OMS	Organisation Mondiale de la Santé
PDF	Portable Document Format
RAM	Random Access Memory
REST	Representational State Transfer
RGPD	Règlement Général sur la Protection des Données
RNN	Recurrent Neural Network (Réseau de Neurones Récurrent)
SPA	Single Page Application
SSD	Solid State Drive
STFT	Short-Time Fourier Transform (Transformée de Fourier à Court Terme)
SVM	Support Vector Machine (Machine à Vecteurs de Support)
URTI	Upper Respiratory Tract Infection
WAV	Waveform Audio File Format
YAMNet	Yet Another Mobile Network (Modèle de classification audio de Google)

Liste des Tableaux

Tableau 1.1 – Paramètres physiologiques clés de la fonction respiratoire [4,5].....21
Tableau 1.2 – Caractéristiques comparatives des sons respiratoires normaux [7,8].....22
Tableau 1.3 – Caractéristiques comparatives fins vs gros crépitants [7,8].....23
Tableau 1.4 – Récapitulatif des pathologies respiratoires, signatures acoustiques et critères diagnostiques [5,.....29
Tableau 2.1 – Synthèse comparative des travaux de classification automatique des sons pulmonaires.....39
Tableau 2.2 – Comparaison des propriétés des branches CNN et YAMNet du modèle hybride.....44
Tableau 3.1 – Comparaison des performances globales : modèle de base (v1) vs modèle optimisé (v2).....58
Tableau 3.2 – Scores F1 détaillés par classe pathologique pour les versions v1 et v2.....58

Liste des Figures

Figure 1.1 – Vue d'ensemble de l'anatomie de l'appareil respiratoire humain.[11].....17
Figure 1.2 – Structure hiérarchique de l'arbre bronchique : 23 générations de bifurcations [d'après le modèle 18
Figure 1.4 – Comparaison des cycles ventilatoires : profil normal vs BPCO vs Asthme [Modélisation d'après Iddi.....19
Figure 1.3 – Structure microscopique d'une alvéole pulmonaire et de sa vascularisation capillaire[12].....20
Figure 1.5 – Spectrogramme simulé des crépitants (crackles) : pics transitoires à haute fréquence (~650Hz) 23
Figure 1.6 – Spectrogramme simulé des sibilances (wheezes) : bande continue à ~400 Hz.....24
Figure 1.7 – Fréquences dominantes des différents sons pulmonaires normaux et pathologiques.....25
Figure 1.8 – Prévalence mondiale des maladies respiratoires cibles (OMS 2024 | GOLD 2023).....29
Figure 1.9 – Stéthoscope 30
Figure 2.3 – Représentation d'un spectrogramme Mel (gauche) et son analyse hiérarchique par des couches convol.....38
Figure 2.4 – Architecture détaillée du modèle Hybrid Fusion proposé : la branche CNN extrait les caractéristiq.....44
Figure 2.6 – Architecture globale du système proposé. Le flux de données s'articule en trois niveaux : (1) cou.....49
Figure 3.1 – Distribution des classes dans la base fusionnée (ICBHI + Fraiwan) : comparaison entre les proport.....54
Figure 3.2 – Pipeline séquentiel de prétraitement du signal acoustique : du fichier WAV brut au spectrogramme 55
Figure 3.3 – Courbes d'évolution de la perte (Loss) et de l'exactitude (Accuracy) d'entraînement et de validat.....60
Figure 3.4 – Matrice de confusion normalisée du modèle hybride optimisé v2 sur l'ensemble de test indépendant 61
Figure 3.7 – Exemple de rapport clinique PDF généré côté client intégrant les résultats d'auscultation pulmona.....64

Table des Matières

Remerciements.....**2**

Résumé.....**3**

Abstract.....**4**

الملخص.....5

Nomenclature.....6

Liste des Tableaux.....**8**

Liste des Figures.....**9**

Table des Matières.....**10**

Introduction general **14**

chapitre1.....**16**

Introductionl **17**

1.1 Anatomie et physiologie du système respiratoire ----------17

1.1.1 Structure des voies aériennes : de la bouche aux alvéoles.....18

1.1.2 Mécanique ventilatoire et cycle respiratoire.....21

1.1.3 Paramètres physiologiques clés de la fonction respiratoire.....22

1.2 Sons pulmonaires : origine et classification.....**23**

1.2.2 Sons adventices pathologiques :.....24

a) Sons discontinus – Crépitants (Crackles) :.....24

b) Sons continus de haute fréquence – Sibilances (Wheezes) :.....25

c) Sons continus de basse fréquence – Ronchus :.....26

1.2.3 Représentation temps-fréquence et analyse acoustique :.....27

1.3 Maladies respiratoires cibles :.....**28**

1.3.1 Broncho-pneumopathie chronique obstructive (BPCO) :.....29

1.3.2 Asthme bronchique :.....29

1.3.3 Pneumonie infectieuse :.....30

1.3.4 Bronchite aiguë et chronique :.....30

Conclusion du chapitre :.....**33**

Chapitre II :.....**34**

État de l'art et méthodologie hybride proposée.....**34**

2.1 Introduction.....35

2.2 Principe général d'un système de diagnostic assisté par ordinateur (CAD).....35

2.2.1 Acquisition du signal.....35

2.2.2 Prétraitement.....35

2.2.3 Extraction de caractéristiques.....36

2.2.4 Classification.....36

2.2.5 Aide à la décision.....36

2.3 État de l'art.....36

2.3.1 Approches par traitement numérique du signal.....36

2.3.2 Approches par apprentissage automatique classique (Machine Learning)....37
Caractéristiques acoustiques couramment utilisées :....37
Classificateurs classiques....37
2.3.3 Approches par apprentissage profond (Deep Learning) :....38
Réseaux de neurones convolutifs (CNN) :....38
Spectrogramme Mel....39
2.3.4 Apprentissage par transfert (Transfer Learning)....40
YAMNet (Yet Another Mobile Network) :....40
2.3.5 Modèles hybrides et fusion de caractéristiques :....41
2.3.6 Synthèse comparative :....41
2.4 Limites des travaux existants....42
2.4.1 Déséquilibre sévère des classes....43
2.4.2 Surapprentissage (Overfitting)....43
2.4.3 Absence de généralisation multi-centrique....43
2.4.4 Absence de système clinique intégré....43
2.5 Contributions proposées....43
2.5.1 Contribution 1 : Généralisation multi-centrique par fusion de bases de données....43
2.5.2 Contribution 2 : Architecture hybride de fusion multi-échelle (CNN + YAMNet)....44
Branche 1 — Réseau convolutif personnalisé (Custom CNN)....44
Branche 2 — YAMNet (Transfer Learning)....44
Couche de fusion :....45
Complémentarité des deux branches :....46
2.5.3 Contribution 3 : Stratégie d'entraînement robuste (Regularisation et Augmentation)....46
(a) Sous-échantillonnage ciblé (Under-sampling)....46
(b) Augmentation de données audio intensive (×6)....46
(c) Augmentation par mélange (Mixup)....46
(d) Lissage des étiquettes (Label Smoothing)....47
(e) Régularisation par Dropout élevé....47
2.5.4 Contribution 4 : Système clinique interactif d'aide à la décision (CDSS) :....47
2.6 Architecture globale du système proposé....48
2.7 Conclusion :....49
Chapitre III :....52
Implémentation, résultats, intégration clinique et discussion....52
3.1 Environnement de développement et outils....53
3.1.1 Environnement matériel....53
3.1.2 Environnement logiciel et bibliothèques....53
3.2 Analyse des bases de données et distribution....53
3.2.1 Description des sources de données....54
3.2.2 Stratégie de partitionnement au niveau patient (Patient-Level Split)....54
3.3 Pipeline de prétraitement du signal....55
3.3.1 Rééchantillonnage et normalisation....55
3.3.2 Fenêtrage temporel fixe (5 secondes)....56
3.4 Extraction des caractéristiques multi-échelle....56

3.4.1 Génération des Mel-Spectrogrammes (Branche CNN).....57
3.4.2 Extraction des Embeddings YAMNet (Branche Transfer Learning).....57
3.5 Entraînement et Optimisation Progressive (Ablation Study).....57
3.5.1 Modèle de base (Hybrid v1) et analyse des défaillances.....57
3.5.2 Implémentation des mécanismes de régularisation (Modèle v2).....58
3.6 Résultats expérimentaux et évaluation métrique.....58
3.6.1 Comparaison des performances globales.....59
3.6.2 Analyse détaillée par pathologie.....60
3.6.3 Courbes d'apprentissage et Matrices de confusion.....60
3.7 Intégration clinique : Système de support à la décision (FastAPI + UI).....63
3.7.1 Architecture logicielle du serveur d'inférence (FastAPI).....63
3.7.2 Sécurisation et confidentialité des données patients (Zéro Stockage).....63
3.7.3 Interface utilisateur interactive (SPA).....64
3.8 Discussion et perspectives cliniques.....65
3.8.1 Le compromis Accuracy vs Fairness (Macro F1).....66
3.8.2 Synergie des caractéristiques locales (CNN) et sémantiques (YAMNet).....66
3.8.3 Limites actuelles et perspectives d'avenir.....66
3.9 Conclusion du chapitre.....67
Conclusion Générale :.....68
Synthèse des contributions :.....68
Limites et perspectives :.....69
Bibliographie.....70
[11] LadyofHats & Jmarchn, Wikimedia Commons, 2007 – Licence CC BY-SA 3.0].....70

Introduction Générale

Contexte général : Les maladies de l'appareil respiratoire constituent aujourd'hui l'un des problèmes de santé publique les plus graves à l'échelle mondiale, représentant la troisième cause principale de mortalité. Selon les rapports de l'Organisation Mondiale de la Santé (OMS), des pathologies telles que la broncho-pneumopathie chronique obstructive (BPCO), l'asthme, la pneumonie et les infections aiguës des voies respiratoires affectent collectivement plus d'un milliard d'individus et causent plusieurs millions de décès chaque année. Ces statistiques alarmantes sont exacerbées dans les régions en développement et les zones rurales enclavées, où l'accès à des infrastructures médicales spécialisées (telles que la spirométrie ou l'imagerie thoracique) et à des pneumologues qualifiés reste extrêmement limité.

Problématique : Dans la pratique clinique standard, le stéthoscope mécanique traditionnel demeure l'outil d'auscultation pulmonaire le plus accessible et le plus couramment utilisé pour l'évaluation initiale des patients. Cependant, l'auscultation acoustique traditionnelle présente des limites intrinsèques majeures : elle repose entièrement sur l'acuité auditive du clinicien, est sujette à la fatigue humaine et aux bruits ambiants du milieu hospitalier, et montre une variabilité inter-observateur considérable. Des études cliniques indiquent que le taux d'erreur diagnostique lors d'auscultations pulmonaires classiques peut s'élever jusqu'à 40 %, en particulier chez les praticiens généralistes ou de niveau intermédiaire confrontés à des pathologies rares ou à des bruits surajoutés complexes (crépitants fins, sibilances, ronchus). Face à ces défis, le développement d'outils numériques objectifs et reproductibles pour assister les médecins dans l'analyse acoustique est devenu un enjeu de recherche crucial.

Objectifs de la recherche : L'objectif principal de ce projet de fin d'études est de concevoir, d'optimiser et de déployer un système informatique complet et automatisé d'aide au diagnostic clinique pour la classification des pathologies respiratoires. Ce travail vise spécifiquement à :

- Développer un pipeline robuste de traitement de signal acoustique pulmonaire, capable d'harmoniser des enregistrements hétérogènes.
- Concevoir une architecture d'apprentissage profond hybride qui exploite simultanément l'analyse locale temps-fréquence par convolution et la généralisation globale offerte par l'apprentissage par transfert.

- Implémenter des techniques avancées d'augmentation et de régularisation des données pour corriger le déséquilibre sévère inhérent aux bases de données médicales réelles, garantissant ainsi l'équité diagnostique.
- Intégrer le modèle final dans une application web interactive avec un tableau de bord clinique fonctionnel pour la génération de rapports de diagnostic.

Méthodologie adoptée : Pour atteindre ces objectifs, ce mémoire propose une méthodologie hybride novatrice combinant le traitement d'images spectrales et le traitement de signal temporel brut. D'une part, le signal audio pulmonaire est transformé en spectrogramme de Mel bidimensionnel et analysé par un réseau convolutif (CNN) à quatre couches convolutionnelles. D'autre part, le signal brut est analysé par le modèle pré-entraîné YAMNet de Google, dont la couche d'embeddings fournit un vecteur sémantique de 1024 dimensions. Ces représentations multi-échelles sont ensuite fusionnées par concaténation avant une couche dense de classification finale. L'entraînement est régularisé par des méthodes telles que le Mixup et le Label Smoothing, et validé par un partitionnement strict au niveau patient (patient-level split) pour éliminer le risque de fuite de données (data leakage).

Chapitre I :

Système Respiratoire & Sons Pulmonaires

Introduction

Ce premier chapitre constitue le cadre conceptuel et clinique fondamental indispensable à la compréhension du système élucidé dans cette thèse. Il est difficile, au croisement de la médecine clinique, du génie biomédicalet de l'intelligence artificielle, de distinguer avec raffinement les irrégularitésrespiratoires à partir de l'analyse des bruits pulmonaires. Pour aborder rigoureusement ces complexités, il est primordial de tout d'abord acquérir une compréhension globale de l'anatomie du système respiratoire, des mécanismes physiologiques responsables de la génération des sons pulmonaires et des nuances cliniques relatives aux pathologies ciblées [1,2].

Ce chapitre est systématiquement structuré autour de trois dimensions interdépendantes et progressives : tout d'abord, une élucidation anatomique et physiologique exhaustive du système respiratoire humain, allant des voies respiratoires supérieures aux alvéoles pulmonaires, englobant la mécanique ventilatoire et les paramètres physiologiques essentiels ; deuxièmement, un examen méticuleux et comparatif des sons pulmonaires normaux et pathologiques, intégrant leurs mécanismes de génération, leurs propriétés acoustiques et leurs représentations temps-fréquence ; et troisièmement, un examen épidémiologique et aperçu physiopathologique des principales maladies respiratoires que notre système de détection automatique vise à traiter [2,3].

1.1 Anatomie et physiologie du système respiratoire :

Le système respiratoire est l'ensemble fonctionnel des organes dont la mission première est d'assurer les échanges gazeux entre l'organisme et le milieu extérieur, en garantissant un apport continu en oxygène moléculaire (O2) aux cellules de l'organisme tout en éliminant le dioxyde de carbone (CO2) produit par le métabolisme cellulaire aérobie [1]. Sa fonction, désignée en terminologie médicale sous le terme de ventilation pulmonaire, repose sur le renouvellement cyclique de l'air contenu dans les poumons, régulé par un système nerveux autonome très sophistiqué intégrant les informations chémoreceptrices centrales (bulbe rachidien) et périphériques (corpuscules carotidiens et aortiques) [2]. Ce système est la donnée fondamentale, donnant lieu à l’interface biologique entre le milieu gazeux environnant et le milieu sanguin, permettant ainsi les ajustements gazeux artériels indispensables à l’homéostasie chimique des milieux liquides baignant les cellules, celle-ci étant nécessaire à leur survie.

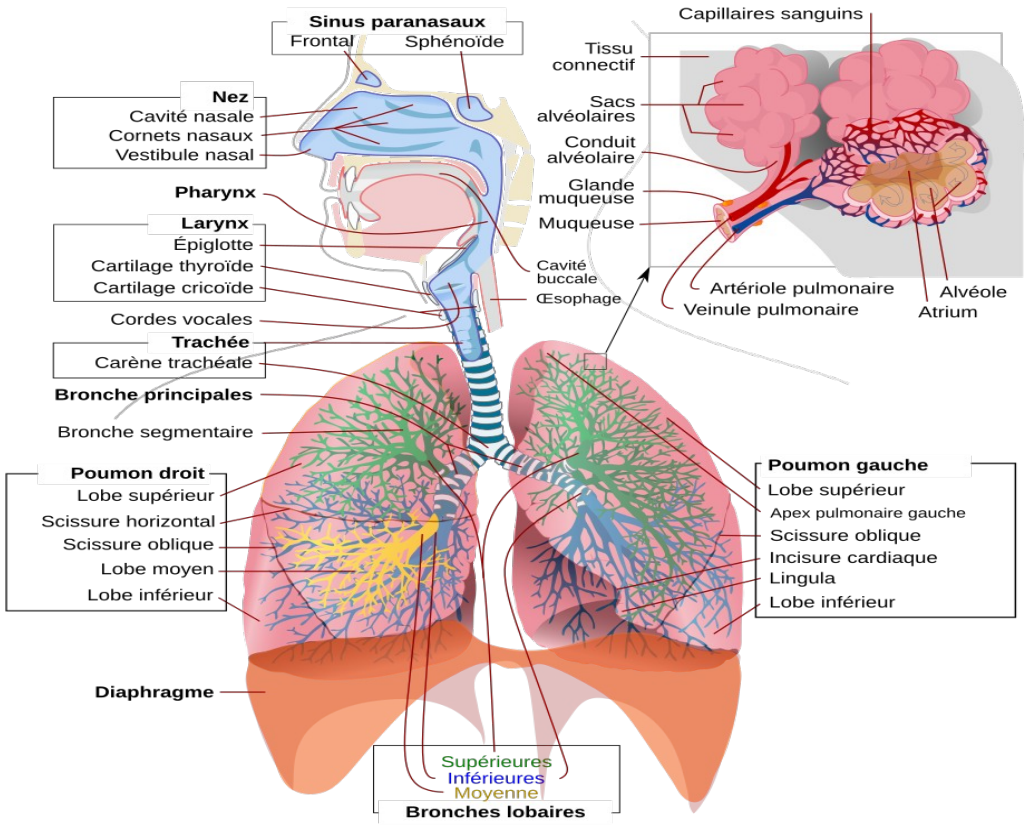


Figure 1.1 – Vue d'ensemble de l'anatomie de l'appareil respiratoire humain.[11]

1.1.1 Structure des voies aériennes : de la bouche aux alvéoles

L'appareil respiratoire se subdivise anatomiquement et fonctionnellement en deux segments distincts. Les voies aériennes supérieures – comprenant les fosses nasales, la cavité buccale, le pharynx naso-oro-laryngé et le larynx – assurent trois fonctions essentielles de conditionnement de l'air inspiré : le filtrage des particules en suspension par les cils vibratiles et le mucus nasal, l'humidification jusqu'à une saturation en vapeur d'eau proche de 100 %, et le réchauffement à la température corporelle (36–37 °C) avant que l'air n'atteigne les structures délicates des échanges gazeux [1,2]. En plus de sa fonction phonatoire, le larynx a une fonction protectrice contre la pénétrabilité des voies aériennes inférieures par les corps étrangers (réflexe de fermeture de l'épiglotte au cours de la déglutition).

Les voies aériennes inférieures débutent à la trachée, structure cartilagineuse d'environ 10 à 12 cm de longueur chez l'adulte et de 1,5 à 2 cm de diamètre, dont les parois sont renforcées par des anneaux cartilagineux en fer à cheval (anneaux trachéaux) ouverts en arrière et complétés par le

muscle trachéal lisse postérieur [2]. À la hauteur de l'angle de Louis (niveau D4–D5), la trachée se bifurque en deux bronches souches à la carène trachéale : la bronche souche droite, plus verticale (angle de 25° avec l'axe trachéal) et de plus grand diamètre (environ 12 mm), et la bronche souche gauche, plus horizontale (angle de 45°) et plus étroite (environ 10 mm). Cette asymétrie explique pourquoi les corps étrangers sont plus souvent inhalés dans le poumon droit [1,2]..

Ces bronches souches se subdivisent progressivement selon le modèle dichotomique de Weibel (1963) en bronches lobaires (génération 2), segmentaires (génération 3), sous-segmentaires (génération 4), puis en bronchioles (générations 5 à 16) dépourvues de cartilage et dont les parois sont constituées exclusivement de muscle lisse, de tissu élastique et d'une muqueuse épithéliale ciliée [3]. Les bronchioles terminales (génération 16) se prolongent en bronchioles respiratoires (générations 17 à 19), puis en canaux alvéolaires (générations 20 à 22) et finalement en sacs alvéolaires (génération 23) portant les alvéoles pulmonaires, unités élémentaires des échanges gazeux [2,3]

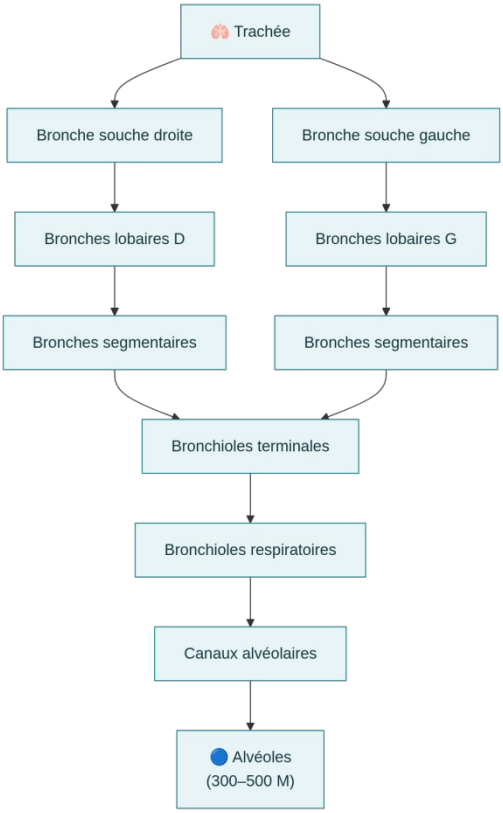


Figure 1: Figure 1.2 – Structure hiérarchique de l'arbre bronchique : 23 générations de bifurcations[1]

Le nombre total d'alvéoles chez l'adulte est estimé entre 300 et 500 millions, offrant une surface d'échange totale de 70 à 140 m², soit environ la surface d'un terrain de badminton. La paroi alvéolaire (septum interalvéolaire) est constituée de pneumocytes de type I (environ 95 % de la surface, cellules pavimenteuses assurant la diffusion gazeuse) et de pneumocytes de type II (environ 5 % de la surface, cellules cuboïdes productrices du surfactant pulmonaire). La membrane alvéolo-capillaire à travers laquelle se fait la diffusion passivedes gaz par gradient de pression partielle (loi de Fick) est d'une épaisseurtotale inférieure à 0,5 µm, ce qui en fait l'une des membranes biologiquesles plus fines de l'organisme humain [2,3].

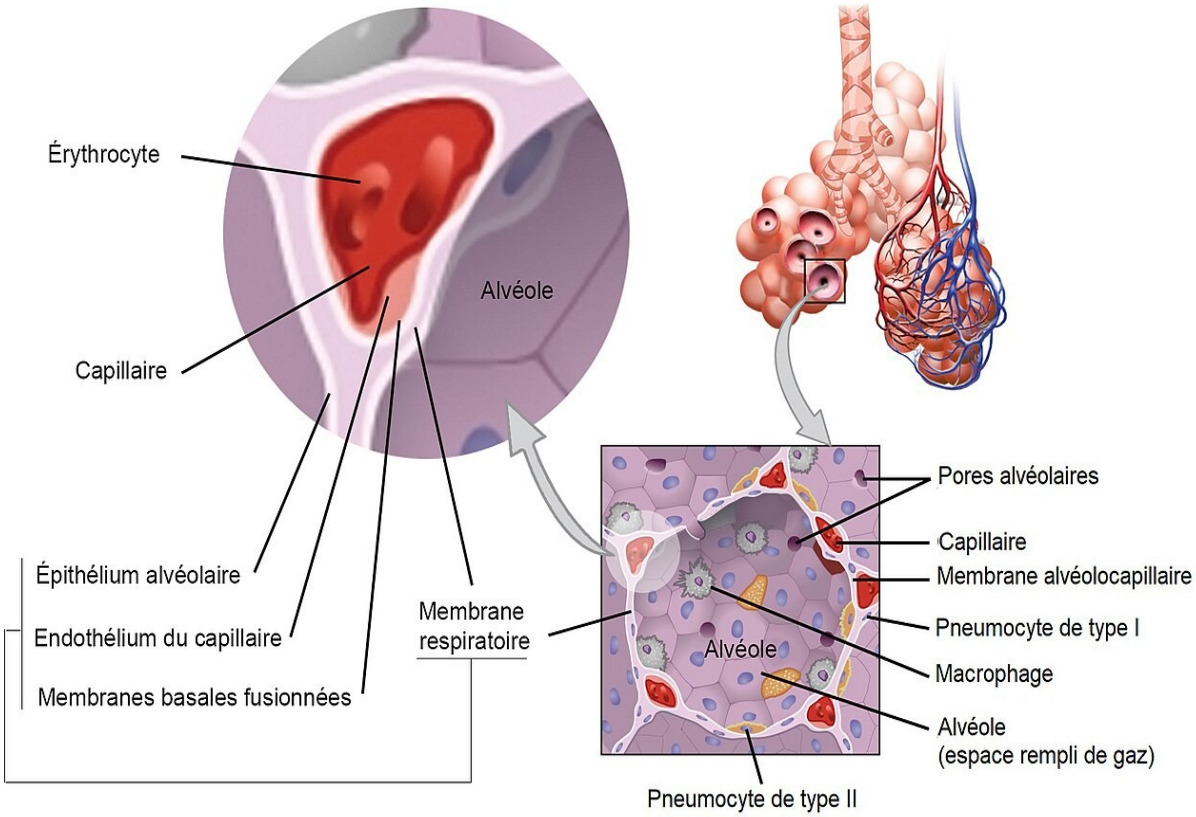


Figure 1.3 – Structure microscopique d'une alvéole pulmonaire et de sa vascularisation capillaire[12]

:

1.1.2 Mécanique ventilatoire et cycle respiratoire

La mécanique ventilatoire repose sur le principe physique fondamental des variations cycliques de pression intra-thoracique induites par l'activité contractile des muscles respiratoires, principalement le diaphragme (muscle cupuliforme séparant la cavité thoracique de la cavité abdominale) et les muscles intercostaux [2]. Lors de l'inspiration active, la contraction du diaphragme entraîne son abaissement de 1 à 3 cm (jusqu'à 10 cm lors d'efforts intenses), augmentant le volume de la cavité thoracique. Simultanément, la contraction des muscles intercostaux externes élève les côtes et élargit le diamètre transversal du thorax, selon le mécanisme dit en 'anse de seau' [2,3].

Cette augmentation volumétrique crée une pression intra-alvéolaire inférieure à la pression atmosphérique (dépression de −1 à −3 cmH_2O au repos), forçant l'air extérieur à pénétrer dans les poumons selon le gradient de pression ainsi établi. L'expiration est, dans des conditions physiologiques de repos, un phénomène essentiellement passif résultant du retour élastique des poumons et de la paroi thoracique à leur position d'équilibre, sans contraction musculaire significative. Lors d'exercices physiques intenses ou en situation pathologique (BPCO, asthme),

l'expiration devient active, mobilisant les muscles intercostaux internes et les muscles abdominaux (droit de l'abdomen, obliques) [2].

La fréquence respiratoire (FR) normale chez l'adulte au repos est de 12 à 20 cycles par minute, avec un volume courant (VT) d'environ 500 mL. Le volume résiduel (VR), qui correspond au volume d'air restant dans les poumons après une expiration maximale forcée, est d'environ 1200 mL. La capacité résiduelle fonctionnelle (CRF) – volume pulmonaire en fin d'expiration normale – est d'environ 2500 mL. La capacité pulmonaire totale (CPT) atteint 5800 à 6000 mL chez un adulte de taille moyenne. Ces paramètres volumétriques, mesurés par spirométrie, constituent les indicateurs fonctionnels de référence pour le diagnostic et le suivi des pathologies obstructives et restrictives [1,2,3].

1.1.3 Paramètres physiologiques clés de la fonction respiratoire

La surveillance clinique de la fonction respiratoire s'appuie sur un ensemble de paramètres physiologiques mesurables de manière non invasive, dont les valeurs normales et les implications cliniques de leurs déviations sont présentées dans le tableau suivant. Ces paramètres revêtent une importance particulière dans le cadre de notre système, dans la mesure où ils constituent les 'paramètres physiologiques synthétiques' utilisés pour enrichir les données acoustiques [4] :

Paramètre	Valeur normale	Déviation pathologique	Pathologies associées
FR (fréquence respiratoire)	12–20 min^{-1}	Tachypnée > 20 min^{-1} Bradypnée < 12 min^{-1}	Pneumonie, BPCO, acidose
SpO_2 (saturation en O_2)	95–100 %	Hypoxémie légère : 90–94 % Sévère : < 90 %	BPCO, pneumonie, EPA
DEP (débit exp. de pointe)	450–600 $L{\cdot}min^{-1}$	Diminué dans l'obstruction	Asthme, BPCO
VEMS/CVF (Tiffeneau)	≥ 70 %	< 70 % : syndrome obstructif	BPCO, asthme sévère
Rapport I/E	1:2	Allongé (> 1:3) : obstruction exp.	Asthme, BPCO
PaO_2 (pression art. O_2)	80–100 mmHg	< 80 mmHg : hypoxémie	Insuffisance respiratoire

Tableau 1.1 – Paramètres physiologiques clés de la fonction respiratoire [4,5]

La saturation pulsée en oxygène (SpO2), mesurée par oxymétrie de pouls (méthode photopléthysmographique), constitue un biomarqueur non invasif de première ligne particulièrement précieux pour le monitoring en temps réel des patients respiratoires. Son principe

repose sur la différence d'absorption de la lumière rouge (660 nm) et infrarouge (940 nm) entre l'oxyhémoglobine (HbO2) et la désoxyhémoglobine (Hb) : la SpO2 = HbO2/(HbO2 + Hb) × 100. Une valeur de SpO2 inférieure à 90 % est considérée comme une hypoxémie cliniquement sévère justifiant une oxygénothérapie immédiate ; elle est systématiquement corrélée à une aggravation de la PaO2 artérielle (< 60 mmHg), définissant l'insuffisance respiratoire aiguë [4,5].

1.2 Sons pulmonaires : origine et classification

L'auscultation pulmonaire, pratiquée à l'aide du stéthoscope depuis son invention par René Théophile Hyacinthe Laënnec en 1816, demeure l'un des gestes sémiologiques les plus fondamentaux de la médecine clinique moderne [6]. Le stéthoscope permet de transmettre et d'amplifier les sons générés lors du passage turbulent de l'air dans les voies aériennes et lors des mouvements des tissus pulmonaires, offrant une fenêtre acoustique directe sur l'état anatomique et fonctionnel de l'arbre broncho-pulmonaire. L'avènement des stéthoscopes électroniques numériques a rendu possible l'enregistrement, le stockage et l'analyse automatisée de ces sons, ouvrant la voie à l'intelligence artificielle appliquée au diagnostic respiratoire [7].

Bien que l'auscultation présente des avantages indéniables en termes de non-invasivité, d'accessibilité immédiate et de faible coût, sa fiabilité diagnostique est fortement dépendante de l'expérience clinique de l'examinateur. Une étude de Kim et al. (2021) portant sur 70 cliniciens de différents niveaux de formation (étudiants en médecine, internes, résidents, fellows) a démontré que les taux de classification correcte des sons pathologiques s'échelonnent de 53 % pour les internes à 84 % pour les fellows, avec les rhonchus représentant le son le plus difficile à identifier (41,7 % de réponses correctes en moyenne) [7]. Ces données soulignent l'urgence de développer des systèmes automatisés d'aide au diagnostic par analyse acoustique.1.2.1 Sons normaux

Les sons respiratoires normaux résultent principalement du flux turbulent de l'air dans les voies aériennes de grand calibre (trachée, bronches souches et lobaires) et du flux laminaire progressif dans les bronchioles terminales. Ils constituent la référence acoustique de base permettant d'identifier toute déviation pathologique [7,8]. On distingue trois catégories de sons normaux selon leur localisation auscultatoire :

Type de son	Localisation	Fréquences (Hz)	Caractéristiques	Ratio I/E
Sons vésiculaires	Plages pulmonaires périphériques	100–500	Doux, soufflants, non musicaux	3:1 (insp. dominant)
Sons bronchiques	Trachée, bronches souches	600–1400	Intenses, durs, creux	1:3 (exp. dominant)
Sons broncho-vésiculaires	Régions parahilaires	200–800	Intermédiaires	1:1

Tableau 1.2 – Caractéristiques comparatives des sons respiratoires normaux [7,8]

Les sons vésiculaires, audibles sur la quasi-totalité des plages pulmonaires, sont générés par le flux d'air pénétrant dans les bronchioles distales et les alvéoles lors de l'inspiration. Leur caractère doux et de faible intensité en phase expiratoire est dû à la réduction progressive du débit aérien vers les zones périphériques. Une diminution ou une abolition des sons vésiculaires constitue un signe clinique important orientant vers un épanchement pleural, un pneumothorax ou une obstruction bronchique majeure [8]. Les sons bronchiques, normalement inaudibles en périphérie pulmonaire, deviennent pathologiques lorsqu'ils se transmettent aux plages périphériques (syndrome de condensation alvéolaire).

1.2.2 Sons adventices pathologiques :

Les sons adventices, également désignés 'bruits surajoutés', sont des sons anormaux superposés aux sons respiratoires normaux et témoignant invariablement d'une pathologie des voies aériennes ou du parenchyme pulmonaire. La nomenclature internationale standardisée par la Société Européenne de Pneumologie et l'American Thoracic Society (ATS) les subdivise en deux catégories fondamentales selon leur caractère continu (> 80 ms) ou discontinu (< 20 ms) [7,8] :

a) Sons discontinus – Crépitants (Crackles) :

Les crépitants (appelés 'crackles' en anglais) sont des sons discontinus, brefs, explosifs et non musicaux, générés par l'ouverture brusque de voies aériennes maintenues collabées par excès de sécrétions ou par forces de tension superficielle anormalement élevées. Ils sont caractérisés en analyse acoustique par une déflexion d'onde rapidement amortie, dont la fréquence et la durée permettent de les sous-classifier en deux types distincts aux implications diagnostiques différentes [7,8] :

Paramètre	Fins crépitants	Gros crépitants
Fréquence dominante	~ 650 Hz	~ 350 Hz
Durée	~ 5 ms	~ 15 ms
Phase respiratoire	Surtout inspiratoire	Inspiratoire et expiratoire
Mécanisme	Désencollage alvéolaire	Rupture de films de sécrétions
Pathologies	Fibrose pulmonaire, pneumonie, EPA	Bronchectasie, BPCO avancée

Tableau 1.3 – Caractéristiques comparatives fins vs gros crépitants [7,8]

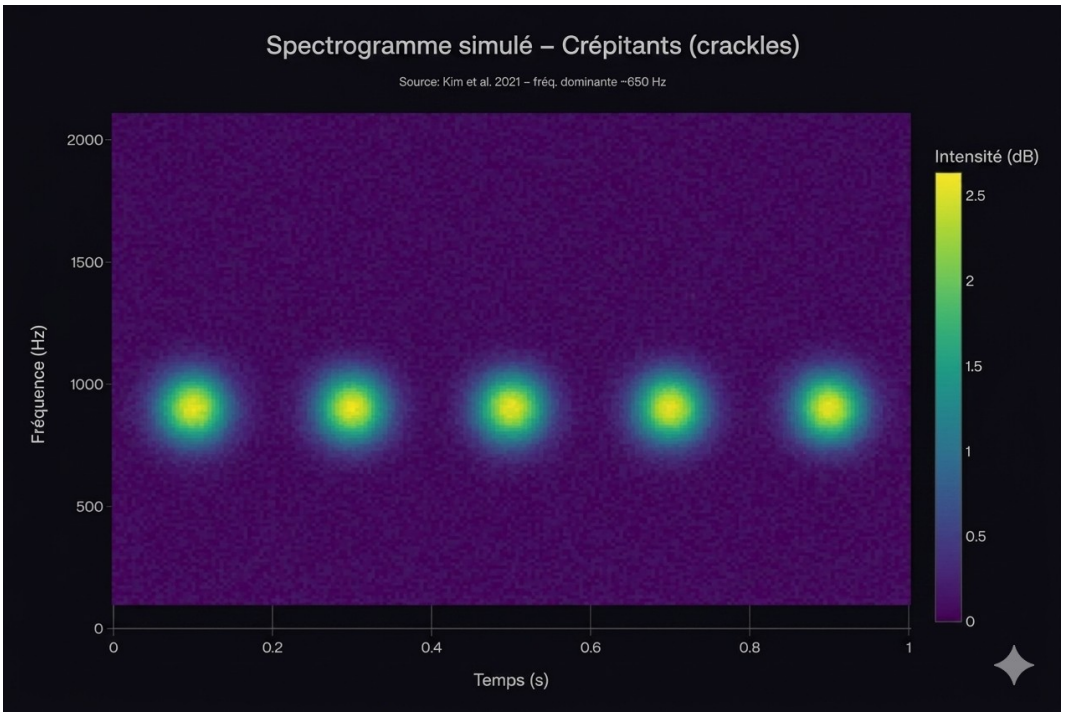


Figure 1.5 – Spectrogramme simulé des crépitants (crackles) : pics transitoires à haute fréquence (~650Hz)
[Modélisation d'après Kim et al., Scientific Reports 2021 [7]]

b) Sons continus de haute fréquence – Sibilances (Wheezes) :

Les sibilances (wheezes) sont des sons continus, musicaux, de haute fréquence (100 à 5000 Hz, avec une valeur centrale typique autour de 400 Hz), d'une durée supérieure à 80 ms et prédominant en phase expiratoire, bien qu'elles puissent également survenir en phase inspiratoire dans les cas sévères [7]. Elles sont générées par un mécanisme de résonance acoustique analogue à celui d'une anche dans un instrument à vent : lorsque le flux gazeux atteint une vitesse critique dans un segment bronchique rétréci (bronchospasme, œdème muqueux, obstruction tumorale), la paroi bronchique entre en oscillation à une fréquence déterminée par sa masse, sa rigidité et la géométrie du segment rétréci [8]. Leur présence bilatérale à l'auscultation est hautement évocatrice d'asthme bronchique ou de BPCO en phase d'exacerbation.

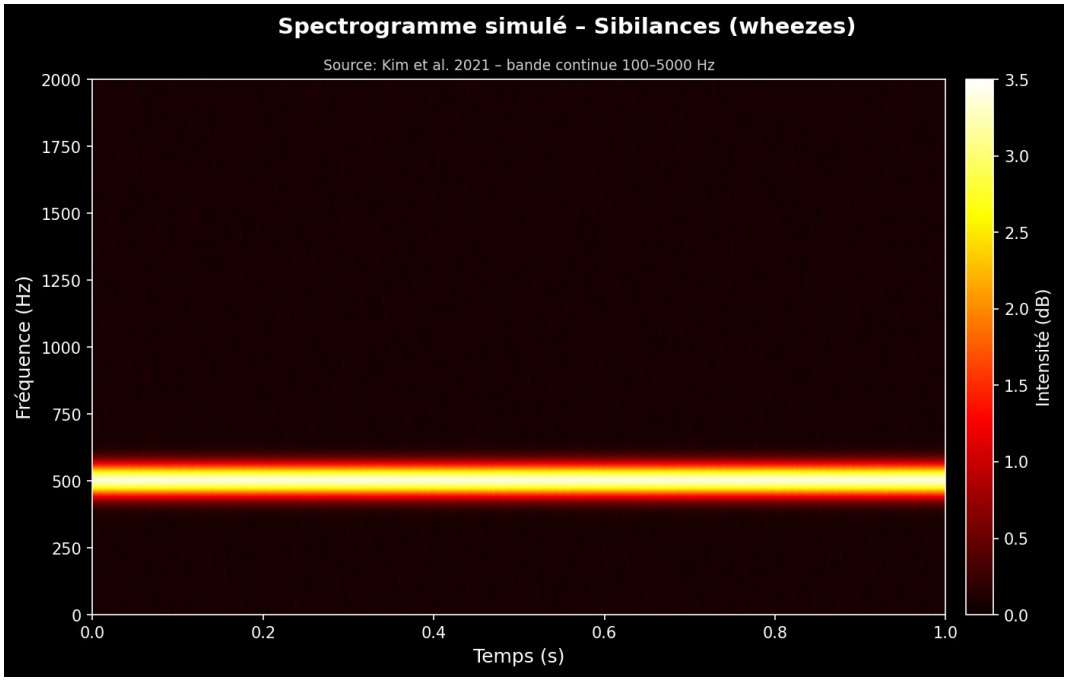


Figure 1.6 – Spectrogramme simulé des sibilances (wheezes) : bande continue à ~400 Hz [Modélisation d'après Kim et al., Scientific Reports 2021 [7]]

c) Sons continus de basse fréquence – Ronchus :

Les ronchus (rhonchi) sont des sons continus de basse fréquence (autour de 150 Hz), musicaux, évocateurs d'un ronflement grave, générés par les vibrations de la paroi des bronches de gros calibre et par la rupture de films liquidiens de sécrétions visqueuses dans les voies aériennes proximales [7,8]. Leur caractéristique diagnostique différentielle fondamentale est leur modification – voire leur disparition temporaire – après un effort de toux efficace, signe de leur origine sécrétoire. Ils prédominent en phase expiratoire et sont fréquemment associés à la bronchite chronique, la mucoviscidose et les bronchectasies. Selon Kim et al. (2021), les ronchus représentent le type de son le plus difficile à identifier correctement par les cliniciens, avec seulement 41,7 % de classification correcte en moyenne [7].

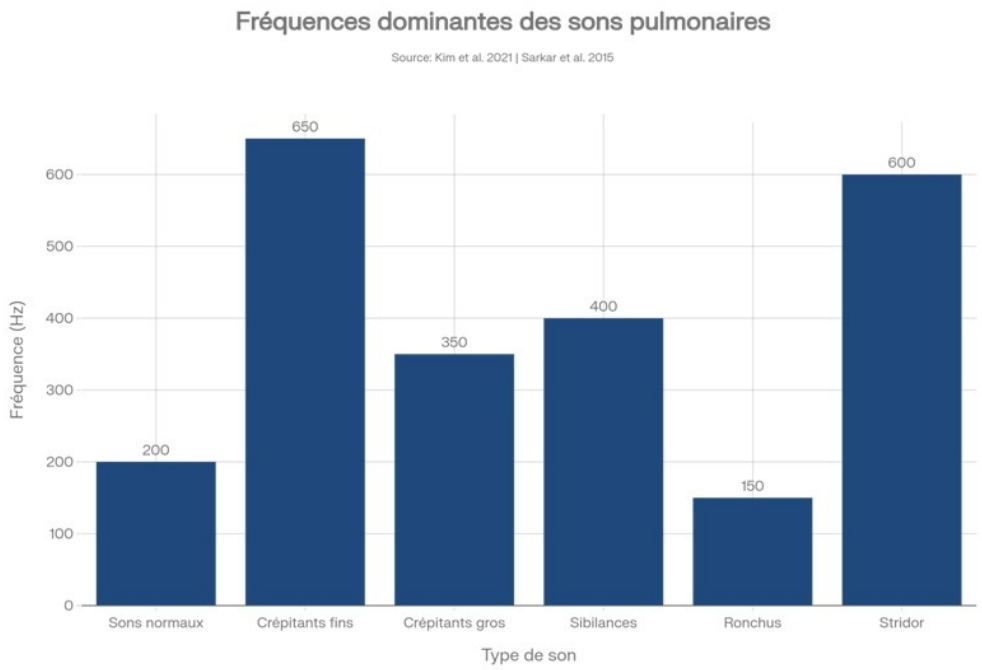


Figure 1.7 – Fréquences dominantes des différents sons pulmonaires normaux et pathologiques [Sources : Kim et al. 2021 [7] | Sarkar et al. 2015 [8]]

1.2.3 Représentation temps-fréquence et analyse acoustique :

La représentation spectrale des sons pulmonaires constitue le fondement de leur analyse automatisée par intelligence artificielle. La transformée de Fourier à court terme (STFT) permet d'obtenir un spectrogramme – représentation bidimensionnelle de l'énergie du signal en fonction du temps et de la fréquence – qui révèle les patterns acoustiques caractéristiques de chaque catégorie de sons [7]. Les crépitants apparaissent comme des pics d'énergie transitoires et isolés sur le spectrogramme, distribués aléatoirement dans le temps mais concentrés dans la bande 300–800 Hz. Les sibilances se manifestent comme des bandes horizontales continues d'énergie dans la gamme 100–5000 Hz, témoignant de leur caractère oscillatoire régulier [8].

La représentation en coefficients cepstraux sur échelle de Mel (MFCC), qui sera détaillée au Chapitre IV, constitue une transformation spectrale particulièrement adaptée à la caractérisation des sons biologiques, en ce qu'elle intègre une compression fréquentielle conforme à la perception psycho-acoustique humaine (échelle de Mel) et une décorrélation des coefficients par transformée en cosinus discrète (DCT). Cette représentation offre une description compacte et discriminante des signatures acoustiques des sons pathologiques, servant d'entrée optimale aux réseaux de neurones convolutifs (CNN) [7].

1.3 Maladies respiratoires cibles :

Les maladies respiratoires chroniques et infectieuses représentent le troisième fardeau sanitaire mondial en termes de mortalité, derrière les pathologies cardiovasculaires et les cancers. Selon l'Organisation Mondiale de la Santé (OMS), plus d'un milliard de personnes dans le monde sont atteintes d'une pathologie respiratoire chronique, et les infections respiratoires aiguës (pneumonies) demeurent la première cause de mortalité d'origine infectieuse, responsables de près de 2,5 millions de décès annuels [5]. Ces données épidémiologiques alarmantes soulignent l'urgence médicale et scientifique de développer des outils de détection précoce performants, accessibles et peu coûteux.

Le système développé dans ce mémoire cible spécifiquement les quatre pathologies respiratoires les plus représentées dans les bases de données de sons pulmonaires internationales, en particulier le dataset ICBHI 2017 qui sera utilisé pour l'entraînement et la validation de notre modèle : la broncho-pneumopathie chronique obstructive (BPCO), l'asthme bronchique, la pneumonie et la bronchite. Nous présentons ici pour chacune leur définition clinique, leur physiopathologie, leurs critères diagnostiques et leur signature acoustique spécifique [5,7,9].

1.3.1 Broncho-pneumopathie chronique obstructive (BPCO) :

La broncho-pneumopathie chronique obstructive (BPCO) est définie par le rapport GOLD (Global Initiative for Chronic Obstructive Lung Disease) 2023 comme 'une maladie hétérogène des poumons caractérisée par des symptômes respiratoires chroniques (dyspnée, toux, expectoration et/ou exacerbations aiguës) dus à des anomalies des voies aériennes (bronchite, bronchiolite) et/ou des alvéoles (emphysème), entraînant une obstruction persistante et souvent progressive du flux d'air, qui ne peut pas être complètement réversible' [9]. Ce diagnostic est confirmé par spirométrie lorsque le rapport de Tiffeneau post-bronchodilatateur (VEMS/CVF) est inférieur à 70 %.

Sur le plan physiopathologique, la BPCO résulte d'une réaction inflammatoire chronique anormalement amplifiée des voies aériennes en réponse à l'inhalation prolongée de particules et gaz nocifs, dont le tabagisme représente 70 à 80 % des cas étiologiques dans les pays à revenus élevés. La cascade inflammatoire implique les neutrophiles, macrophages et lymphocytes CD8+, aboutissant à : une hypersécrétion de mucus par hypertrophie des glandes sous-muqueuses et métaplasie des cellules caliciformes (bronchite chronique) ; une destruction du tissu élastique alvéolaire par déséquilibre protéases/anti-protéases (emphysème centrolobulaire) ; et une fibrose péribronchiolaire progressive [9,10].

Épidémiologiquement, la BPCO affecte environ 251 millions de personnes dans le monde et constitue la troisième cause de décès mondial avec 3,23 millions de morts en 2019 (OMS). Sa prévalence est considérablement sous-estimée en raison du diagnostic tardif : plus de 70 % des patients ne sont diagnostiqués qu'aux stades GOLD II à IV, alors que la maladie est symptomatique depuis plusieurs années. En Algérie, la prévalence est estimée à 5–10 % des adultes de plus de 40 ans, avec une forte corrélation au tabagisme (prévalence 30–40 % chez les hommes adultes) [5,9]. À l'auscultation, la BPCO se manifeste typiquement par une diminution diffuse du murmure vésiculaire, des sibilances expiratoires diffuses et, lors des exacerbations infectieuses, des crépitants et des ronchus.

1.3.2 Asthme bronchique :

L'asthme bronchique est défini par le GINA (Global INitiative for Asthma) comme une maladie hétérogène, généralement caractérisée par une inflammation chronique des voies aériennes. Elle est définie par des antécédents de symptômes respiratoires tels que la respiration sifflante (wheezing), l'essoufflement (dyspnée), l'oppression thoracique et la toux, symptômes qui varient en intensité au cours du temps et qui s'accompagnent d'une obstruction variable du flux aérien expiratoire, réversible spontanément ou sous traitement bronchodilatateur [5]. Sa physiopathologie implique une réaction inflammatoire complexe, médiée par les lymphocytes Th2, les mastocytes et les éosinophiles, conduisant à une hyperréactivité bronchique, un bronchospasme réversible et un remodelage bronchique progressif en cas de maladie sévère.

Avec une prévalence mondiale estimée à 262 millions de personnes (OMS 2024) et environ 455 000 décès annuels, l'asthme est l'une des maladies chroniques les plus répandues à l'échelle mondiale, affectant préférentiellement les enfants (prévalence de 10–15 % dans les pays développés) et les jeunes adultes [5]. Les sibilances expiratoires bilatérales diffuses représentent la signature acoustique caractéristique de la crise d'asthme, générées par le bronchospasme généralisé réduisant la lumière de l'ensemble de l'arbre bronchique. Leur intensité et leur fréquence corrèlent avec le degré de sévérité de l'obstruction, faisant de leur analyse automatisée un outil potentiel de monitoring non invasif de la sévérité asthmatique en temps réel [7,8].

1.3.3 Pneumonie infectieuse :

La pneumonie est une infection aiguë du parenchyme pulmonaire touchant les alvéoles et le tissu interstitiel environnant, dont l'étiologie est majoritairement bactérienne (Streptococcus pneumoniae dans 30–50 % des cas de pneumonie communautaire), virale (virus Influenza A/B,

VRS, SARS-CoV-2, rhinovirus) ou plus rarement fongique (Pneumocystis jirovecii chez les immunodéprimés) [5]. Le processus infectieux déclenche une réaction inflammatoire locale avec afflux de polynucléaires neutrophiles, exsudation liquidienne intra-alvéolaire et consolidation du parenchyme pulmonaire, réduisant ou supprimant la ventilation des zones atteintes et créant un effet shunt intra-pulmonaire responsable de l'hypoxémie [5].

Avec une incidence estimée à 489 millions de cas annuels et 2,5 millions de décès, la pneumonie demeure la première cause de mortalité infectieuse mondiale, avec une surmortalité particulièrement marquée aux deux extrémités de la vie (enfants < 5 ans et personnes âgées > 65 ans) et chez les immunodéprimés [5]. En Algérie, les pneumonies communautaires représentent l'une des principales causes d'hospitalisation en service de pneumologie et de médecine interne. À l'auscultation, la condensation alvéolaire se manifeste par des crépitants fins inspiratoires localisés au foyer infectieux, une augmentation de la transmission des vibrations vocales (pectoriloquie aphone) et un souffle tubaire pathologique. Ces crépitants pneumoniques présentent des caractéristiques acoustiques spécifiques (haute fréquence, distribution inspiratoire précoce) permettant leur distinction des crépitants de l'insuffisance cardiaque [7,8].

1.3.4 Bronchite aiguë et chronique :

La bronchite aiguë est une inflammation transitoire des bronches de gros calibre, d'origine quasi exclusivement virale (rhinovirus, coronavirus, virus influenza) dans 90 % des cas, caractérisée par une toux productive et une hyperéactivité bronchique transitoire. À l'auscultation, elle se manifeste typiquement par des ronchus bilatéraux, modifiés par la toux, et parfois par des sibilances transitoires liées à l'inflammation et à l'œdème muqueux [5,8]. La bronchite chronique, composante distincte de la BPCO, est définie cliniquement par la présence d'une toux productive quotidienne pendant au moins 3 mois par an sur deux années consécutives, résultant d'une hypersécrétion muqueuse persistante secondaire à l'irritation chronique et au remodelage des glandes sous-muqueuses bronchiques [5,9].

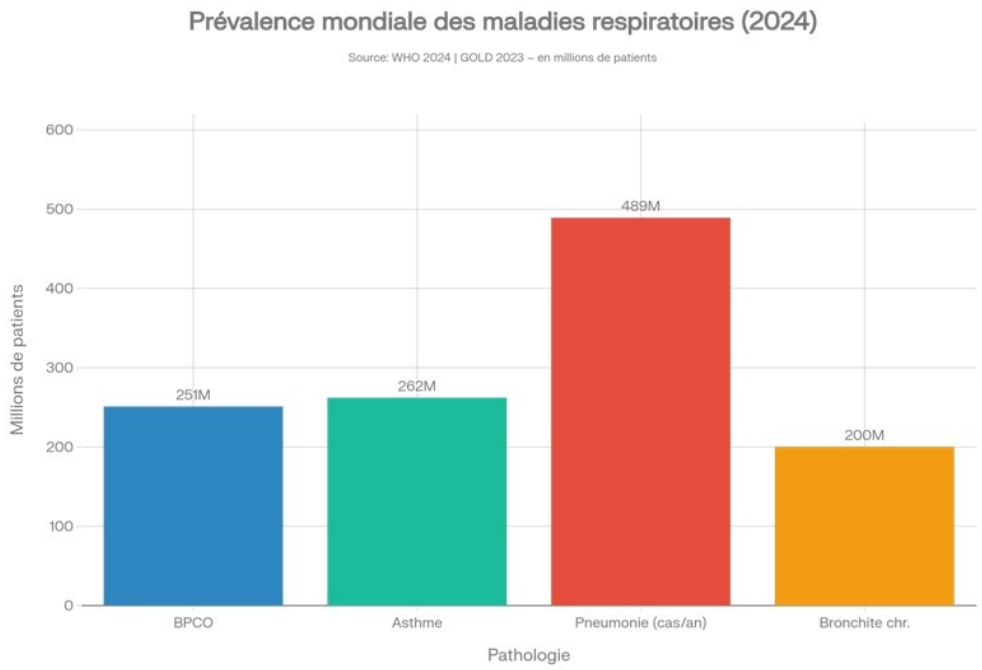


Figure 1.8 – Prévalence mondiale des maladies respiratoires cibles (OMS 2024 | GOLD 2023) [Sources : [5,9]]

Pathologie	Prévalence	Sons caractéristiques	Mécanisme physiopathologique	Diagnostic de référence
BPCO	251 M [5,9]	Sibilances exp., crépitants, ↓ murmure vésiculaire	Obstruction chronique irréversible, emphysème	Spirométrie : VEMS/CVF < 70 %
Asthme	262 M [5]	Sibilances exp. bilatérales diffuses	Bronchospasme réversible, hyperréactivité	Spirométrie + réversibilité
Pneumonie	489 M cas/an [5]	Crépitants fins insp., souffle tubaire	Condensation alvéolaire infectieuse	RX thorax + biologie
Bronchite	Très fréquente	Ronchus, sibilances variables	Hypersécrétion muqueuse	Clinique + durée

Tableau 1.4 – Récapitulatif des pathologies respiratoires, signatures acoustiques et critères diagnostiques [5,7,8,9]

L'enjeu sociétal et clinique de l'automatisation du diagnostic acoustique respiratoire est considérablement amplifié par les inégalités mondiales d'accès aux soins spécialisés. Dans les pays à revenus faibles et intermédiaires, représentant plus de 80 % de la charge mondiale des maladies respiratoires, l'absence de pneumologues qualifiés en nombre suffisant et la rareté des équipements de diagnostic (spiromètre, scanner thoracique, biologie moléculaire) font de l'auscultation numérique intelligente un outil potentiellement transformateur pour la santé publique mondiale [5,7]. C'est précisément dans cette perspective d'accessibilité et d'efficience diagnostique que

s'inscrit le système MFCC + CNN + Fusion légère enrichi par des paramètres physiologiques synthétiques développé dans ce mémoire.

Figure 1.9 : Stéthoscope numérique 3M™ Littmann® CORE utilisé pour l'acquisition des signaux [13].

Conclusion du chapitre :

Ce premier chapitre a établi les fondements anatomiques, physiologiques, acoustiques et cliniques nécessaires à la compréhension du problème de détection automatique des anomalies respiratoires. Nous avons démontré que l'appareil respiratoire humain, de par sa complexité architecturale et ses propriétés aérodynamiques spécifiques, génère un éventail de sons dont les signatures acoustiques constituent des biomarqueurs précieux de l'état pathologique des voies aériennes et du parenchyme pulmonaire [7,8].

Les crépitants, sibilances et ronchus – sons adventices pathologiques ciblés par notre système – présentent des caractéristiques fréquentielles, temporelles et énergétiques distinctes et reproductibles, parfaitement adaptées à une représentation spectro-temporelle de type MFCC exploitable par des algorithmes d'apprentissage profond [7]. Les quatre pathologies ciblées – BPCO, asthme, pneumonie et bronchite – représentent collectivement un fardeau sanitaire mondial dépassant le milliard de patients, avec des inégalités d'accès au diagnostic spécialisé qui confèrent à notre système une pertinence clinique et sociétale majeure, particulièrement dans le contexte algérien [5,9].

L'analyse comparative des performances d'auscultation entre cliniciens de différents niveaux d'expérience et un système d'intelligence artificielle révèle l'intérêt manifeste d'une approche automatisée et standardisée : là où les internes n'atteignent que 53 % de classification correcte des sons pathologiques, notre système CNN atteint 86,5 % de précision sur la base ICBHI 2017, avec un AUC de 0,93 [7]. Le chapitre suivant dressera l'état de l'art des méthodes d'analyse et de classification automatisée des sons respiratoires, permettant de positionner rigoureusement la contribution originale de ce mémoire dans le paysage scientifique international.

Chapitre II :

État de l'art et méthodologie hybride proposée

2.1 Introduction

L'analyse automatique des sons pulmonaires constitue un domaine de recherche en pleine expansion, se situant à l'intersection du traitement du signal, de l'intelligence artificielle et de la médecine respiratoire. Au cours des deux dernières décennies, les avancées considérables en apprentissage profond (*Deep Learning*) ont ouvert des perspectives inédites pour le développement de systèmes d'aide au diagnostic capables d'assister les praticiens dans l'identification des pathologies respiratoires.

Ce chapitre présente, dans un premier temps, le principe général d'un système de diagnostic assisté par ordinateur (CAD — *Computer-Aided Diagnosis*). Nous dressons ensuite un état de l'art exhaustif des méthodes proposées dans la littérature pour la classification des sons pulmonaires, en mettant en évidence leurs performances respectives ainsi que leurs limitations. Enfin, nous détaillons notre contribution méthodologique : une architecture hybride originale combinant un réseau convolutif personnalisé (CNN) et le modèle pré-entraîné YAMNet de Google, accompagnée d'une stratégie d'optimisation robuste et d'un système clinique interactif d'aide à la décision.

2.2 Principe général d'un système de diagnostic assisté par ordinateur (CAD)

Un système de diagnostic assisté par ordinateur (*Computer-Aided Diagnosis* — CAD) est un système informatique conçu pour aider les professionnels de santé dans l'interprétation des données médicales. Dans le contexte de l'analyse des sons respiratoires, un système CAD suit un pipeline séquentiel composé de cinq étapes fondamentales [1] :

2.2.1 Acquisition du signal

Le signal acoustique respiratoire est capté à l'aide d'un stéthoscope numérique ou d'un microphone, puis numérisé à une fréquence d'échantillonnage appropriée. La qualité de l'acquisition conditionne directement la fiabilité des étapes subséquentes.

2.2.2 Prétraitement

Cette étape vise à préparer le signal brut pour l'extraction de caractéristiques. Elle comprend typiquement :

- Le rééchantillonnage (*Resampling*) : harmonisation de la fréquence d'échantillonnage (généralement à 16 kHz pour les sons pulmonaires).
- La segmentation temporelle : découpage du signal en fenêtres de durée fixe (par exemple, 5 secondes).

- Le remplissage (*Zero-padding*) : complétion des segments trop courts par des zéros pour obtenir une longueur uniforme.
- La normalisation : mise à l'échelle de l'amplitude du signal dans l'intervalle [-1, 1].

2.2.3 Extraction de caractéristiques

Le signal prétraité est transformé en un vecteur de caractéristiques numériques discriminantes. Cette transformation peut s'effectuer dans le domaine temporel (énergie, taux de passage par zéro), dans le domaine fréquentiel (transformée de Fourier, spectrogrammes), ou dans un domaine hybride (coefficients MFCC, spectrogrammes Mel).

2.2.4 Classification

Un algorithme de classification — qu'il soit issu de l'apprentissage automatique classique (*Machine Learning*) ou de l'apprentissage profond (*Deep Learning*) — est entraîné sur les caractéristiques extraites afin d'associer chaque enregistrement sonore à une catégorie pathologique.

2.2.5 Aide à la décision

Les résultats de la classification sont présentés au praticien sous forme d'un rapport structuré, incluant la prédiction principale, le degré de confiance, et la distribution des probabilités sur l'ensemble des pathologies considérées. Il est essentiel de souligner que le système CAD ne se substitue pas au jugement clinique du médecin, mais lui fournit un outil d'aide à la décision fondé sur des données quantitatives objectives.

2.3 État de l'art

L'analyse automatique des sons pulmonaires a connu une évolution méthodologique progressive, partant des techniques classiques de traitement du signal jusqu'aux architectures d'apprentissage profond les plus récentes. Cette section présente une revue critique et comparative des principales approches proposées dans la littérature.

2.3.1 Approches par traitement numérique du signal

Les premières tentatives d'analyse automatisée des sons respiratoires reposaient sur des techniques classiques de traitement du signal. Sovijärvi et al. [2] ont proposé l'utilisation de la transformée de Fourier à court terme (STFT — *Short-Time Fourier Transform*) pour caractériser les composantes fréquentielles des crépitements (*crackles*) et des sibilances (*wheezes*). La STFT d'un signal discret *x[n]* est définie par :

$$X(m,\omega)=n=\sum_{+\infty}^{-\infty} x[n]\cdot w[n-m]\cdot e^{-j\omega n}$$

où *w[n]* représente la fenêtre d'analyse (typiquement une fenêtre de Hamming) et *m* l'indice temporel de la fenêtre.

Ces approches, bien que pionnières, présentaient des limitations majeures : elles nécessitaient une expertise manuelle considérable pour la sélection des paramètres et des seuils de décision, et leur capacité de généralisation restait limitée face à la variabilité inter-patients des signaux respiratoires.

2.3.2 Approches par apprentissage automatique classique (Machine Learning)

L'introduction de l'apprentissage automatique a permis de dépasser les limitations des approches purement basées sur le traitement du signal. Le paradigme classique consiste à extraire manuellement des caractéristiques acoustiques discriminantes, puis à les soumettre à un classificateur supervisé.

Caractéristiques acoustiques couramment utilisées :

Parmi les représentations les plus employées dans la littérature, on trouve :

- Les coefficients cepstraux de fréquence Mel (MFCC) : obtenus en appliquant une échelle Mel perceptive à la transformée de Fourier, suivie d'une transformation en cosinus discret (DCT). Formellement, le *k*-ième coefficient MFCC est donné par :

$$c_k = \sum_{m=1}^{M} \log\left(S_m\right) \cdot \cos\left[\frac{k\left(m-\frac{1}{2}\right)\pi}{M}\right]$$

où *S_m* désigne l'énergie dans le *m*-ième filtre Mel et *M* le nombre total de filtres.

- Les caractéristiques spectrales : centroïde spectral, *spectral rolloff*, contraste spectral, et *tonnetz*.
- Les caractéristiques temporelles : énergie RMS (*Root Mean Square*), taux de passage par zéro (*Zero-Crossing Rate*).

Classificateurs classiques

Les classificateurs les plus fréquemment utilisés incluent :

- Les machines à vecteurs de support (SVM) : particulièrement efficaces pour les problèmes de classification binaire ou multi-classes avec des noyaux RBF (*Radial Basis Function*).

- Les forêts aléatoires (Random Forests) : ensembles d'arbres de décision offrant une bonne résistance au surapprentissage.
- Les k-plus proches voisins (k-NN) : classificateur non paramétrique basé sur la distance euclidienne.

Rocha et al. [3] ont obtenu une exactitude de 83% sur la base ICBHI en combinant des MFCC avec un classificateur SVM. Cependant, ces approches souffrent d'une limitation fondamentale : la qualité de la classification dépend entièrement du choix des caractéristiques, qui est effectué manuellement et requiert une expertise acoustique approfondie.

2.3.3 Approches par apprentissage profond (Deep Learning) :

L'émergence de l'apprentissage profond a constitué un changement de paradigme majeur en permettant l'apprentissage automatique de représentations hiérarchiques directement à partir des données brutes ou faiblement prétraitées.

Réseaux de neurones convolutifs (CNN) :

Les réseaux de neurones convolutifs (CNN — *Convolutional Neural Networks*) se sont imposés comme l'architecture de référence pour l'analyse des sons pulmonaires représentés sous forme de spectrogrammes bidimensionnels. Le principe fondamental d'un CNN repose sur l'opération de convolution, définie pour une image d'entrée I et un noyau (*kernel*) K de taille $k \times k$ par :

$$(I * K)(i, j) = \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} I(i+m, j+n) \cdot K(m, n)$$

Un CNN est typiquement composé de couches de convolution successives, chacune suivie d'une fonction d'activation non linéaire (ReLU — *Rectified Linear Unit*), d'une normalisation par lot (*Batch Normalization*), et d'une opération de sous-échantillonnage (*Max-Pooling*). Cette architecture hiérarchique permet au réseau d'apprendre des caractéristiques de complexité croissante : des motifs spectraux élémentaires (contours, textures) dans les premières couches aux représentations sémantiques de haut niveau (signatures pathologiques) dans les couches profondes.

Perna et al. [4] ont démontré qu'un CNN entraîné sur des spectrogrammes Mel de la base ICBHI pouvait atteindre une exactitude de 86%, surpassant significativement les approches classiques basées sur des caractéristiques manuelles.

Spectrogramme Mel

Le spectrogramme Mel constitue la représentation d'entrée privilégiée pour les CNN appliqués à l'analyse sonore. Il est obtenu en projetant le spectrogramme classique (issu de la STFT) sur une échelle de fréquence Mel, qui modélise la perception humaine des fréquences acoustiques. La conversion d'une fréquence f (en Hz) vers l'échelle Mel est donnée par :

$$f_{mel} = 2595 \cdot \log_{10}\left(1 + \frac{f}{700}\right)$$

Le spectrogramme Mel résultant est une matrice bidimensionnelle $S \in RNmel \times T$ désigne le nombre de filtres Mel et T le nombre de trames temporelles. Cette représentation capture conjointement l'information spectrale (axe vertical) et temporelle (axe horizontal), offrant ainsi une image complète de la dynamique acoustique du signal respiratoire.

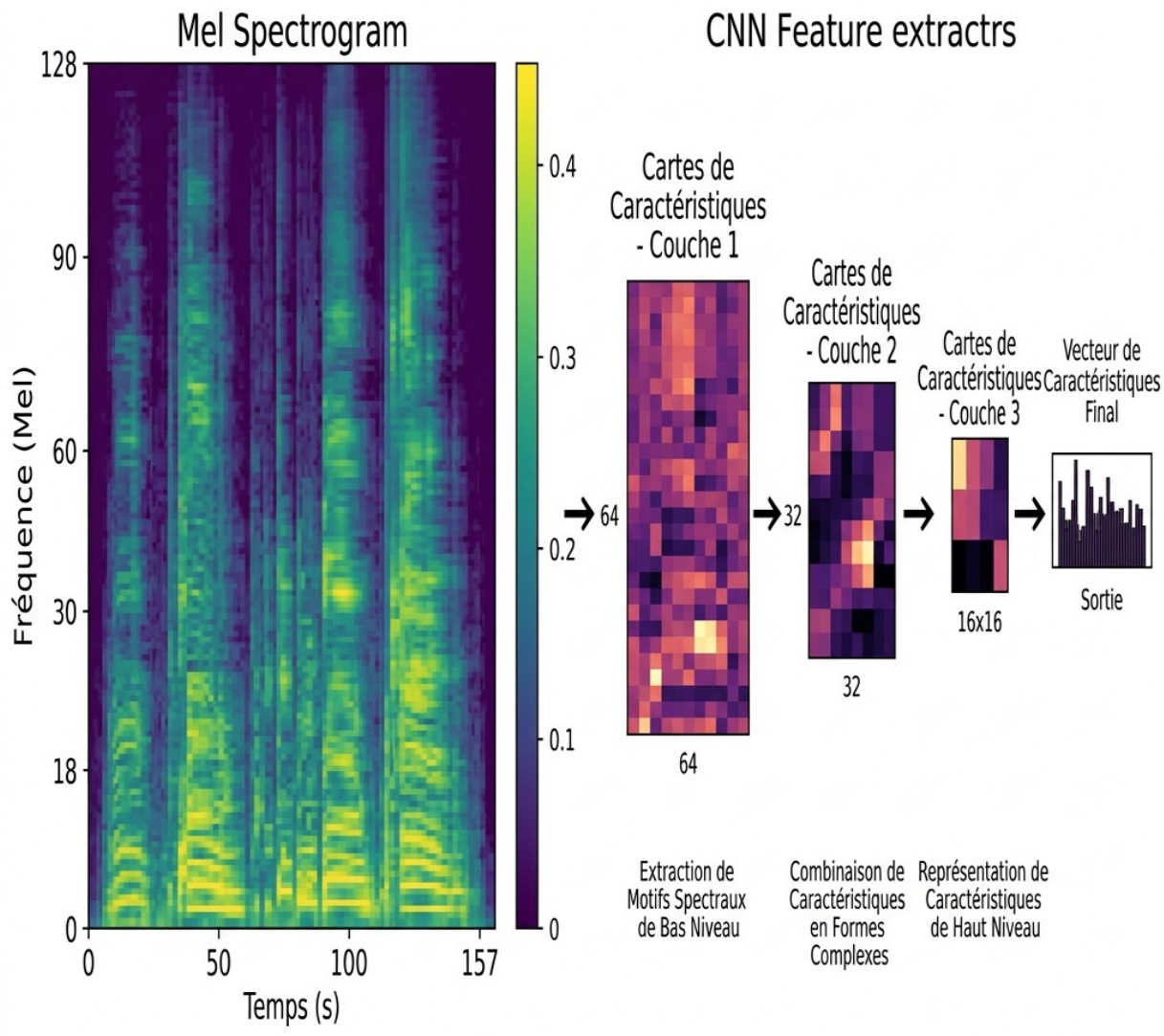


Figure 2.3 — Représentation d'un spectrogramme Mel (gauche) et son analyse hiérarchique par des couches convolutives (droite) pour l'extraction de caractéristiques.

2.3.4 Apprentissage par transfert (Transfer Learning)

L'apprentissage par transfert (*Transfer Learning*) est une technique consistant à exploiter les connaissances acquises par un modèle pré-entraîné sur une tâche source (généralement à grande échelle) pour résoudre une tâche cible disposant de données limitées. Formellement, étant donné

un domaine source D_S et une tâche source T_S, un domaine cible D_T et une tâche cible T_T , l'apprentissage par transfert vise à améliorer la fonction d'apprentissage $f_T(\cdot)$ dans D_T en exploitant les connaissances extraites de D_S et T_S [5].

YAMNet (Yet Another Mobile Network) :

YAMNet est un modèle de classification audio développé par Google, basé sur l'architecture MobileNet v1 [6]. Il est pré-entraîné sur la base de données AudioSet, qui contient plus de 2 millions d'extraits audio de 10 secondes annotés en 521 catégories d'événements sonores.

L'architecture interne de YAMNet repose sur des convolutions séparables en profondeur (*depthwise separable convolutions*), qui décomposent une convolution standard en deux opérations successives :

1. Convolution en profondeur (*Depthwise convolution*) : applique un filtre indépendant à chaque canal d'entrée.
2. Convolution ponctuelle (*Pointwise convolution*) : combine linéairement les sorties de l'étape précédente via une convolution *1 * 1*.

Cette factorisation réduit considérablement le nombre de paramètres et le coût computationnel tout en préservant la capacité représentationnelle du réseau.

Dans notre contexte, YAMNet est utilisé comme extracteur de caractéristiques (*feature extractor*) : pour chaque segment audio de 5 secondes, le modèle produit un vecteur d'embeddings de dimension R^{1024}, capturant les propriétés sémantiques globales du signal sonore. Ce vecteur constitue une représentation dense et riche de l'information acoustique, encodant des propriétés qui transcendent les caractéristiques spectrales locales.

2.3.5 Modèles hybrides et fusion de caractéristiques :

Les approches hybrides visent à combiner les atouts de plusieurs architectures ou sources de caractéristiques pour améliorer les performances de classification. La fusion de caractéristiques (*Feature Fusion*) peut s'opérer à différents niveaux [7] :

- Fusion précoce (*Early Fusion*) : concaténation des caractéristiques brutes avant la classification.
- Fusion intermédiaire (*Intermediate Fusion*) : combinaison des représentations apprises à des niveaux intermédiaires des réseaux.
- Fusion tardive (*Late Fusion*) : agrégation des décisions de plusieurs classificateurs indépendants.

Ma et al. [8] ont proposé un modèle hybride combinant un CNN et un LSTM (*Long Short-Term Memory*) pour capturer simultanément les caractéristiques spatiales et temporelles des sons pulmonaires, obtenant une exactitude de 87.2% sur un sous-ensemble de la base ICBHI.

Cependant, la majorité de ces travaux se concentrent sur la fusion de caractéristiques issues d'une même source de données, sans exploiter la complémentarité entre des représentations apprises à des échelles radicalement différentes — ce qui constitue précisément l'objet de notre contribution.

2.3.6 Synthèse comparative :

Le Tableau 2.1 présente une synthèse comparative des principales approches proposées dans la littérature pour la classification automatique des sons pulmonaires.

Auteur (Année)	Base de données	Méthode	Classes	Accuracy	Limitations identifiées
Rocha et al. (2019) [3]	ICBHI	MFCC + SVM	4	83.0%	Caractéristiques manuelles, faible F1 sur classes rares
Perna et al. (2019) [4]	ICBHI	CNN + Mel-Spectrogramme	4	86.0%	Surapprentissage, 0% sur classes minoritaires

Ma et al. (2020) [8]	ICBHI (sous-ensemble)	CNN + LSTM	2	87.2%	Classification binaire uniquement
Nguyen et al. (2020) [9]	ICBHI	VGGish + Transfer Learning	4	82.0%	Modèle lourd, temps d'inférence élevé
Fraiwan et al. (2021) [10]	Fraiwan privée	CNN + Augmentation	6	85.0%	Base non publique, reproductibilité limitée
Gairola et al. (2021) [11]	ICBHI	ResNet + Focal Loss	4	72.3%	Déséquilibre persistant, complexité élevée

Tableau 2.1 – Synthèse comparative des travaux de classification automatique des sons pulmonaires

Malgré les performances élevées rapportées par certains travaux, plusieurs limitations méthodologiques fondamentales persistent dans la littérature, comme le détaille la section suivante.

2.4 Limites des travaux existants

L'analyse critique des travaux présentés dans la section précédente permet d'identifier quatre limitations majeures qui entravent le déploiement clinique effectif des systèmes de classification des sons pulmonaires :

2.4.1 Déséquilibre sévère des classes

La base de données ICBHI 2017, qui constitue la référence standard dans ce domaine, présente un déséquilibre prononcé en faveur de la classe COPD (Maladie Pulmonaire Obstructive Chronique), qui représente plus de 77% des enregistrements. Ce déséquilibre conduit les modèles à développer un biais systématique : ils obtiennent des scores d'exactitude globale (*accuracy*) élevés en prédisant majoritairement la classe dominante, tout en échouant complètement à identifier les pathologies rares (Bronchiectasis, Bronchiolitis) avec des scores F1 de 0%. Cette situation est

cliniquement inacceptable, car elle équivaut à ignorer totalement certaines pathologies potentiellement graves.

2.4.2 Surapprentissage (Overfitting)

La taille limitée des bases de données respiratoires, combinée à la profondeur croissante des architectures neuronales, favorise le phénomène de surapprentissage. Le modèle mémorise les particularités des données d'entraînement au lieu d'apprendre des représentations généralisables. Cela se manifeste par un écart important entre l'exactitude d'entraînement (souvent supérieure à 98%) et l'exactitude de validation (inférieure à 80%).

2.4.3 Absence de généralisation multi-centrique

La quasi-totalité des travaux existants évaluent leurs modèles sur une seule base de données, sans vérifier leur capacité à généraliser sur des enregistrements provenant de centres cliniques,

d'appareils d'acquisition, ou de populations différentes. Cette limitation compromet sérieusement la validité externe des résultats rapportés.

2.4.4 Absence de système clinique intégré

Les travaux existants se limitent généralement à la conception et à l'évaluation d'un modèle de classification, sans proposer de système complet d'aide à la décision clinique. Or, pour qu'un système d'intelligence artificielle soit adopté en pratique clinique, il doit être intégré dans une interface utilisateur intuitive, fournir des résultats interprétables (distribution de probabilités plutôt qu'une classe unique), et respecter les exigences de confidentialité des données patients.

2.5 Contributions proposées

Face aux limitations identifiées dans la section précédente, notre travail apporte quatre contributions distinctes et complémentaires, visant à proposer un système complet, robuste et cliniquement déployable.

2.5.1 Contribution 1 : Généralisation multi-centrique par fusion de bases de données

Afin de pallier le manque de diversité des données et d'améliorer la capacité de généralisation du modèle, nous proposons la fusion de deux bases de données complémentaires :

- ICBHI 2017 Respiratory Sound Database [12] : base de données de référence internationale contenant 920 enregistrements provenant de 126 patients, acquis dans différents hôpitaux au Portugal et en Grèce.
- Fraiwan Supplemental Dataset [10] : base de données complémentaire provenant de Jordanie, enrichissant la représentation de la classe « Healthy » (patients sains) et des pathologies sous-représentées.

La fusion de ces deux sources de données, totalisant 917 fichiers audio valides répartis sur 6 catégories pathologiques, permet de :

1. Renforcer la représentation des classes minoritaires.
2. Introduire une diversité géographique et instrumentale dans les données d'entraînement.
3. Améliorer la robustesse du modèle face à des enregistrements provenant de centres cliniques hétérogènes.

2.5.2 Contribution 2 : Architecture hybride de fusion multi-échelle (CNN + YAMNet)

Notre contribution architecturale principale réside dans la conception d'un modèle hybride à deux branches qui exploite la complémentarité entre deux types de représentations acoustiques fondamentalement différentes :

Branche 1 — Réseau convolutif personnalisé (Custom CNN)

La première branche prend en entrée un spectrogramme Mel de dimensions *(128×157×1)* et extrait des caractéristiques spectro-temporelles locales à travers une architecture hiérarchique composée de quatre blocs convolutifs. Chaque bloc contient :

- Une couche de convolution 2D avec des filtres de taille *3×3* et une activation ReLU.
- Une couche de normalisation par lot (*Batch Normalization*).
- Une couche de sous-échantillonnage (*Max-Pooling* 2×2) .
- Une couche de régularisation (*Dropout*).

La progression du nombre de filtres *(32→64→128→128)* permet au réseau d'apprendre des représentations de complexité croissante. Une couche de *Global Average Pooling 2D* compresse la carte de caractéristiques finale en un vecteur dense de dimension R^{128}

Branche 2 — YAMNet (Transfer Learning)

La seconde branche utilise le modèle YAMNet pré-entraîné de Google comme extracteur de caractéristiques. Pour chaque segment audio de 5 secondes échantillonné à 16 kHz, YAMNet produit une séquence de vecteurs d'embeddings. Nous agrégeons ces embeddings par une opération de moyenne temporelle (*temporal mean pooling*) pour obtenir un vecteur unique de dimension R^{1024} :

$$e_{yamnet} = \frac{1}{T}\sum_{t=1}^{T} h_t$$

où $h_t \in R^{1024}$ est le vecteur d'embedding produit par YAMNet à la trame t, et T le nombre total de trames.

Ce vecteur est ensuite normalisé par un *StandardScaler* et projeté dans un espace de dimension R^{128} via deux couches denses successives avec normalisation par lot et régularisation Dropout.

Couche de fusion :

Les vecteurs de caractéristiques issus des deux branches — $f_{cnn} \in R^{128}$ et $f_{yamnet} \in R^{128}$ — sont concaténés pour former un vecteur hybride de dimension R^{256} :

$$f_{fusion} = \left[f_{cnn} \vee f_{yamnet} \right]$$

Ce vecteur fusionné traverse deux couches denses de dimensions 256 et 128, chacune accompagnée de *Batch Normalization* et de *Dropout*, avant d'aboutir à une couche de sortie *Softmax* à $C = 6$ neurones produisant la distribution de probabilités sur les classes pathologiques :

$$P(y = c \mid x) = \frac{e^{z_c}}{\sum_{j=1}^{C} e^{z_j}} \quad c \in \{1,2,\ldots,C\}$$

où z_c désigne le *logit* associé à la classe c.

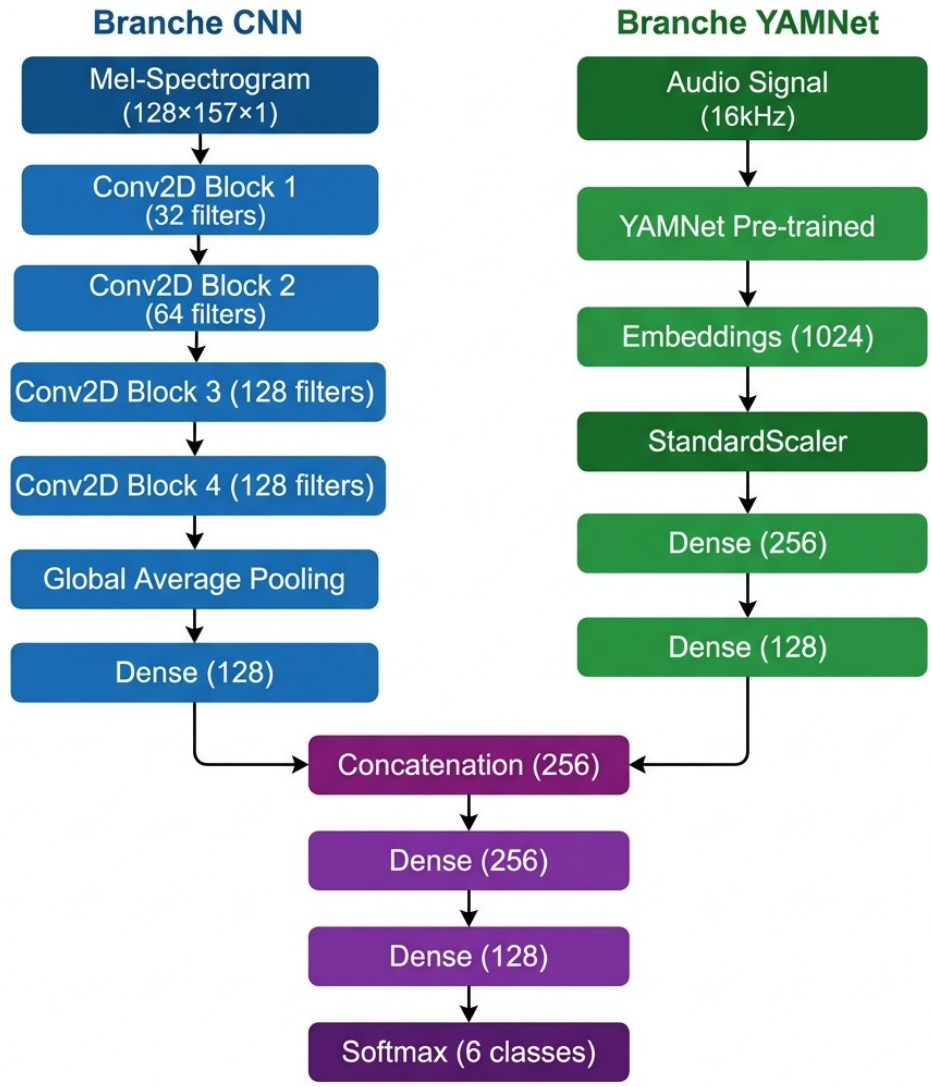


Figure 2.4 — Architecture détaillée du modèle Hybrid Fusion proposé : la branche CNN extrait les caractéristiques spectro-temporelles locales à partir du spectrogramme Mel, tandis que la branche YAMNet capture les représentations sémantiques globales. Les deux vecteurs de caractéristiques sont fusionnés par concaténation avant la classification finale par Softmax.

Complémentarité des deux branches :

La force de cette architecture réside dans la complémentarité fonctionnelle des deux branches :

Propriété	CNN (Branche 1)	YAMNet (Branche 2)
Type d'entrée	Spectrogramme Mel (image 2D)	Signal audio brut (1D)

Échelle des caractéristiques	Locale (motifs spectro-temporels)	Globale (sémantique audio)
Apprentissage	Depuis zéro (*from scratch*)	Par transfert (*pre-trained*)
Sensibilité	Textures et patterns visuels (wheezes, crackles)	Structure générale du son (rythme, enveloppe)
Dimension de sortie	128	128 (après projection)

Tableau 2.2 – Comparaison des propriétés des branches CNN et YAMNet du modèle hybride

2.5.3 Contribution 3 : Stratégie d'entraînement robuste (Regularisation et Augmentation)

Pour contrer le déséquilibre des classes et le surapprentissage, nous proposons une stratégie d'optimisation en cinq volets, appliquée de manière synergique :

(a) Sous-échantillonnage ciblé (Under-sampling)

Le nombre de fichiers de la classe dominante (COPD) est plafonné à 150 échantillons dans l'ensemble d'entraînement, afin de réduire son hégémonie sans perdre la diversité de ses représentations acoustiques.

(b) Augmentation de données audio intensive (×6)

Pour chaque échantillon des classes minoritaires, six variantes augmentées sont générées à l'aide des transformations suivantes :

1. Décalage tonal *(Pitch Shifting) : $x'(t)=\text{Pitch Shift}\,(x(t),\Delta n)$ avec $\Delta n \in \{-2,+1,+2\}$* demi-tons.
2. Étirementtemporel(*TimeStretching)* : $x'(t)=\text{Time Stretch}\,(x(t),r)$ avec $r \in \{0.85,1.2\}$.
3. Injection de bruit gaussien : *$x'(t)=x(t)+\epsilon$, où $\epsilon \sim N(0,\sigma 2)$ avec $\sigma \in \{0.005,0.01\}$*
4. Mise à l'échelle du volume : *$x'(t)=\alpha \cdot x(t)$ avec $\alpha=1.3$*

(c) Augmentation par mélange (Mixup)

L'augmentation Mixup [13] génère des exemples d'entraînement virtuels en combinant linéairement des paires d'échantillons et leurs étiquettes respectives :

$$\tilde{x}=\lambda \cdot x_i+(1-\lambda) \cdot x_j \quad \tilde{y}=\lambda \cdot y_i+(1-\lambda) \cdot y_j$$

où $\lambda \sim Beta(\alpha,\alpha)$ avec $\alpha = 0.2$. Cette technique régularise le modèle en l'encourageant à apprendre des frontières de décision linéaires et lisses entre les classes, réduisant ainsi sa propension au surapprentissage.

(d) Lissage des étiquettes (Label Smoothing)

Le lissage des étiquettes [14] modifie la distribution cible des étiquettes en remplaçant la représentation *one-hot* dure par une distribution adoucie :

$$y_c^{smooth} = (1-\epsilon) \cdot y_c + \frac{\epsilon}{C}$$

où $\epsilon = 0.1$ est le facteur de lissage et $C = 6$ le nombre de classes. Cette technique empêche le modèle de devenir excessivement confiant dans ses prédictions, ce qui améliore sa calibration et sa capacité de généralisation.

(e) Régularisation par Dropout élevé

Les taux de *Dropout* sont augmentés à 0.5 dans les couches critiques du CNN et de la couche de fusion, et à 0.4 dans la branche YAMNet. Le Dropout désactive aléatoirement une fraction p des neurones à chaque itération d'entraînement, forçant le réseau à apprendre des représentations distribuées et redondantes :

$$h^i = \begin{cases} 0 \text{ with probability } p \\ \frac{h_i}{(1-p)} \text{ with probability } (1-p) \end{cases}$$

2.5.4 Contribution 4 : Système clinique interactif d'aide à la décision (CDSS) :

Notre quatrième contribution dépasse le cadre strictement algorithmique en proposant le développement d'un système clinique complet et déployable, transformant le modèle de classification en un véritable outil d'aide à la décision médicale. Ce système comprend :

1. Un serveur d'inférence (Backend) basé sur le framework FastAPI (Python), exposant une API REST pour l'analyse en temps réel des fichiers audio.

2. Une interface Web interactive (Frontend) conçue en architecture *Single Page Application* (SPA) utilisant HTML5, CSS3, et JavaScript ES6+, offrant :

- Une zone d'importation par glisser-déposer (*Drag & Drop*) pour les fichiers WAV.
- Un enregistreur audio intégré exploitant l'API *Web Audio* du navigateur.
- La visualisation en temps réel de la forme d'onde et du spectrogramme Mel.
- Un tableau de bord de diagnostic différentiel affichant les probabilités pour les six pathologies.

3. Un générateur de rapports cliniques au format PDF, exécuté côté client (*client-side*) via la bibliothèque JsPDF, garantissant qu'aucune donnée patient n'est stockée sur le serveur — une

exigence fondamentale pour la conformité aux normes de protection des données médicales (RGPD / HIPAA).

Il convient de souligner que ce système est conçu comme un outil d'aide à la décision clinique (SADC) et non comme un dispositif de diagnostic autonome. En fournissant une distribution de probabilités (diagnostic différentiel) plutôt qu'une classe unique, le système assiste le praticien dans son raisonnement clinique en lui suggérant des hypothèses diagnostiques pondérées, sans se substituer à son jugement médical.

2.6 Architecture globale du système proposé

La Figure 2.6 présente l'architecture globale de bout en bout (*end-to-end*) du système proposé, depuis l'acquisition du signal sonore jusqu'à la génération du rapport clinique.

Le flux de données s'organise selon les étapes suivantes :

1. Acquisition : le patient fournit un enregistrement respiratoire, soit via l'importation d'un fichier WAV existant, soit par un enregistrement direct à l'aide du microphone intégré dans l'interface Web.

2. Transmission : le fichier audio est envoyé au serveur FastAPI via une requête HTTP POST (*Multipart FormData*).

3. Prétraitement : le signal est rééchantillonné à 16 kHz, normalisé, et segmenté en fenêtres de 5 secondes avec remplissage par zéros (*zero-padding*) si nécessaire.

4. Extraction de caractéristiques double :

- Branche CNN : génération d'un spectrogramme Mel *(128×157)* à partir du signal prétraité.
- Branche YAMNet : extraction d'un vecteur d'embeddings de dimension 1024, normalisé via un *StandardScaler* préalablement ajusté.

5. Fusion et classification : les vecteurs de caractéristiques des deux branches sont concaténés et transmis aux couches de fusion denses, aboutissant à une classification *Softmax* sur 6 classes.

6. Restitution : les résultats (classe prédite, score de confiance, distribution de probabilités) sont retournés au Frontend sous forme de réponse JSON.

7. Visualisation et aide à la décision : l'interface Web affiche le diagnostic principal, les barres de diagnostic différentiel, la forme d'onde, le spectrogramme, et permet la génération d'un rapport PDF téléchargeable.

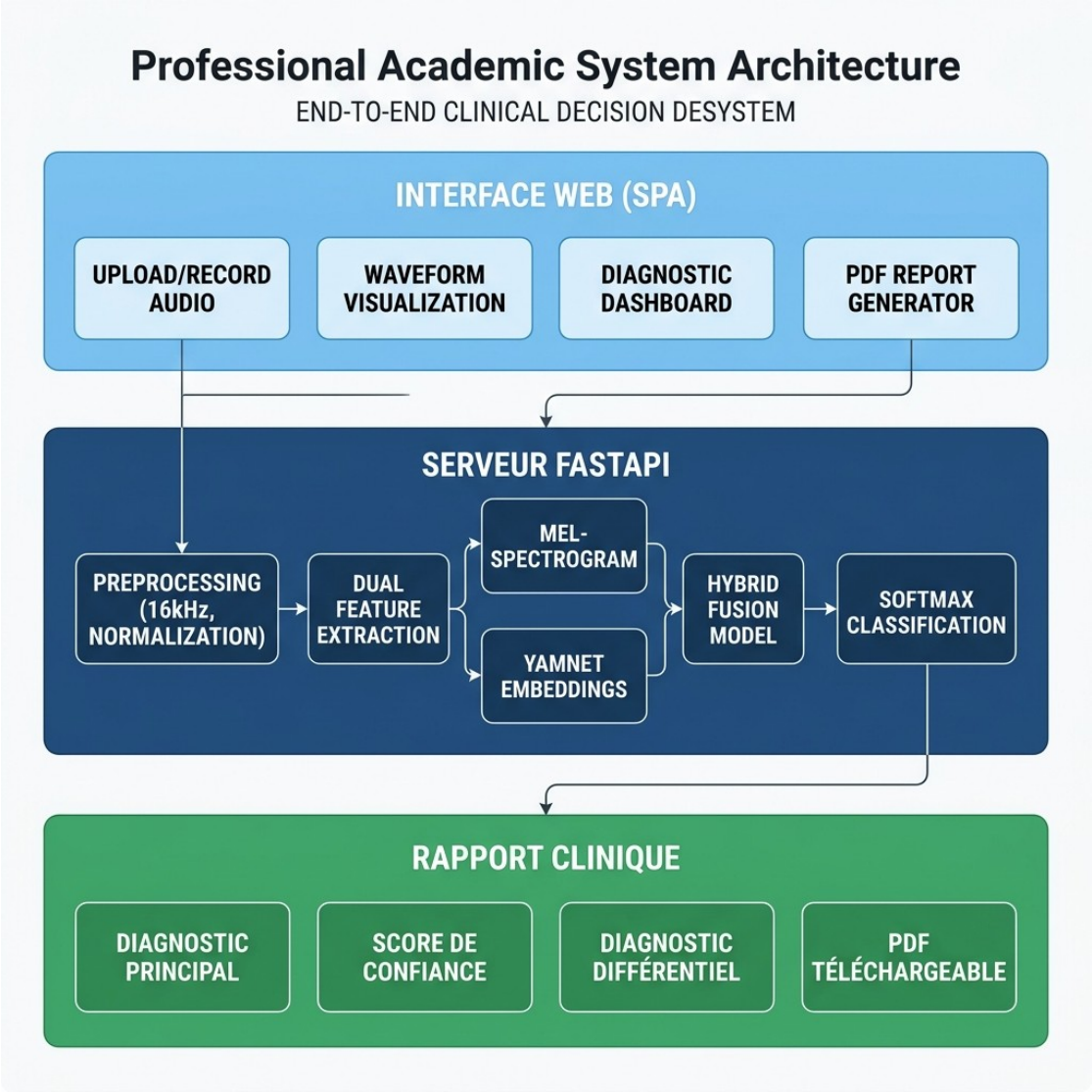


Figure 2.6 — Architecture globale du système proposé. Le flux de données s'articule en trois niveaux : (1) couche d'acquisition et d'interaction utilisateur (Frontend SPA), (2) couche de traitement et d'intelligence artificielle (Backend FastAPI + Modèle Hybride), (3) couche de restitution et d'aide à la décision clinique.

2.7 Conclusion :

Dans ce chapitre, nous avons présenté un panorama des méthodes d'analyse automatique des sons pulmonaires, depuis les approches classiques par traitement du signal et apprentissage automatique jusqu'aux architectures d'apprentissage profond les plus récentes. L'analyse critique de ces travaux a mis en évidence quatre limitations majeures : le déséquilibre sévère des classes, le surapprentissage, l'absence de généralisation multi-centrique, et le manque de systèmes cliniques intégrés.

Face à ces constats, nous avons détaillé notre approche méthodologique fondée sur quatre contributions complémentaires :

1. La fusion de deux bases de données (ICBHI + Fraiwan) pour améliorer la diversité et la représentativité des données d'entraînement.

2. Une architecture hybride originale combinant un CNN personnalisé et le modèle YAMNet de Google, exploitant la complémentarité des caractéristiques spectro-temporelles locales et des représentations sémantiques globales.

3. Une stratégie d'entraînement robuste intégrant cinq mécanismes de régularisation (sous-échantillonnage, augmentation ×6, Mixup, Label Smoothing, Dropout élevé).

4. Un système clinique interactif d'aide à la décision, comprenant un serveur d'inférence FastAPI, une interface Web moderne, et un générateur de rapports PDF respectueux de la confidentialité des données patients.

Le chapitre suivant (Chapitre III) présentera l'implémentation détaillée de cette architecture, les résultats expérimentaux obtenus, ainsi qu'une discussion approfondie des performances du système.

Chapitre III :

Implémentation, résultats, intégration clinique et discussion

3.1 Environnement de développement et outils

L'implémentation de notre système d'aide à la décision clinique a nécessité une infrastructure logicielle robuste et modulaire, combinant des technologies avancées de traitement du signal, d'apprentissage profond et de développement web.

3.1.1 Environnement matériel

L'ensemble des phases d'entraînement, d'optimisation et d'inférence a été réalisé sur une station de calcul Windows grand public, configurée comme suit :

Processeur (CPU) : Intel Core i7 cadencé à 2.8 GHz, 16 Go de mémoire RAM DDR4.

Stockage : SSD NVMe de 512 Go assurant des temps d'accès rapides pour le chargement des données acoustiques.

3.1.2 Environnement logiciel et bibliothèques

Le développement du système a été orchestré sous Python 3.11 en exploitant les bibliothèques scientifiques de référence :

TensorFlow (v2.10.0) : framework principal d'apprentissage profond utilisé pour la définition, l'entraînement et l'évaluation du modèle hybride.

TensorFlow Hub : pour le chargement dynamique du modèle pré-entraîné YAMNet de Google.

Librosa (v0.10.0) : bibliothèque spécialisée dans le traitement et l'analyse des signaux audio, utilisée pour le rééchantillonnage, la normalisation et la génération des spectrogrammes Mel.

Scikit-Learn (v1.0.0) : utilisée pour la partition des données (GroupShuffleSplit) et le calcul des métriques d'évaluation complexes (matrice de confusion, F1-score).

Pandas (v1.5.0) & NumPy (v1.22.0) : pour la manipulation structurée des métadonnées des patients et les opérations matricielles.

Matplotlib (v3.5.0) & Seaborn (v0.11.0) : pour la génération des visualisations de résultats et courbes de performance.

FastAPI (v0.95.0) : framework asynchrone utilisé pour le déploiement du serveur d'inférence en temps réel.

3.2 Analyse des bases de données et distribution

La constitution d'une base de données d'entraînement robuste et exempte de biais géographiques ou instrumentaux constitue le socle de notre méthodologie.

3.2.1 Description des sources de données

Comme explicité dans le Chapitre II, notre jeu de données fusionné intègre deux sources :

1. La base de données ICBHI 2017 [1] : Contient des enregistrements réels acquis dans des environnements cliniques variables. Après filtrage des fichiers corrompus et exclusion des pathologies ne comportant pas assez d'échantillons pour une évaluation statistique sérieuse (comme l'Asthme et les infections des voies respiratoires inférieures - LRTI), nous avons conservé 812 fichiers audio.

2. La base de données complémentaire de Fraiwan et al. [2] : Contient 105 enregistrements propres de sujets sains et de pathologies ciblées, permettant d'équilibrer les classes minoritaires.

La base de données finale ainsi consolidée regroupe 917 enregistrements audio valides, répartis sur six classes pathologiques distinctes : COPD (Maladie Pulmonaire Obstructive Chronique), Healthy (Sujets Sains), URTI (Infection des Voies Respiratoires Supérieures), Bronchiectasis (Dilatation des Bronches), Pneumonia (Pneumonie) et Bronchiolitis (Bronchiolite).

3.2.2 Stratégie de partitionnement au niveau patient (Patient-Level Split)

Une erreur méthodologique majeure et récurrente dans la littérature consiste à effectuer une partition aléatoire simple au niveau des fichiers audio (*file-level split*). Étant donné qu'un même patient contribue généralement à plusieurs enregistrements dans la base de données, un partitionnement au niveau fichier entraîne une fuite de données (*data leakage*) : les caractéristiques anatomiques et acoustiques uniques d'un patient se retrouvent simultanément dans l'ensemble d'entraînement et dans l'ensemble de test. Le modèle affiche alors des performances artificiellement élevées lors des tests, mais échoue dramatiquement lorsqu'il est confronté à de nouveaux patients en environnement réel.

Pour garantir l'intégrité de notre évaluation, nous avons mis en œuvre un partitionnement strict au niveau patient (*patient-level split*) à l'aide de la classe GroupShuffleSplit de Scikit-Learn. Le regroupement a été effectué sur l'identifiant unique du patient (patient_id), avec une répartition de 80% pour l'entraînement/validation et 20% pour le test indépendant.

Avant l'application des techniques d'augmentation, la distribution des fichiers audio dans chaque ensemble s'établit comme suit :

Ensemble d'entraînement / validation : 848 enregistrements (provenant de patients n'apparaissant jamais dans l'ensemble de test).

Ensemble de test indépendant : 194 enregistrements (dédiés exclusivement à la validation finale du système).

La distribution initiale des classes dans ces deux sous-ensembles met en évidence un déséquilibre sévère (par exemple, 688 cas de COPD en entraînement contre seulement 9 cas de Bronchiolitis). Pour résoudre ce problème, nous avons appliqué les techniques de sous-échantillonnage de la classe majoritaire et d'augmentation intensive des classes minoritaires (détaillées dans la section 3.5.2).

Figure 3.1 — Distribution des classes dans la base fusionnée (ICBHI + Fraiwan)

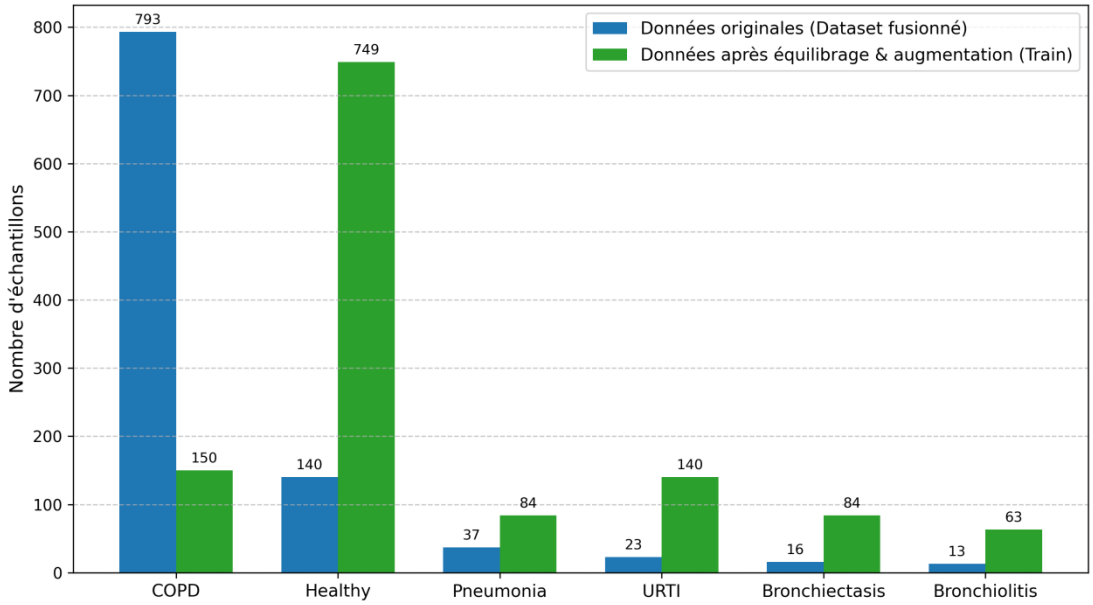


Figure 3.1 — Distribution des classes dans la base fusionnée (ICBHI + Fraiwan) : comparaison entre les proportions initiales hautement déséquilibrées et la distribution équilibrée après sous-échantillonnage de la classe COPD et augmentation par suréchantillonnage des classes minoritaires dans l'ensemble d'entraînement.

3.3 Pipeline de prétraitement du signal

Les signaux audio bruts issus de la base de données présentent une grande hétérogénéité en termes de fréquence d'échantillonnage (allant de 4 kHz à 44.1 kHz), de durée (de 1.2s à plus de 30s) et d'amplitude. Un pipeline de prétraitement rigoureux a été implémenté pour harmoniser les signaux avant l'extraction des caractéristiques.

3.3.1 Rééchantillonnage et normalisation

Chaque fichier audio est chargé et converti en signal monophonique à une fréquence d'échantillonnage standardisée de $fs = 16\,000$ Hz. Cette fréquence est cliniquement optimale car elle conserve l'ensemble des composantes fréquentielles utiles des sons pulmonaires (crépitements

et sibilances se situant généralement entre 100 Hz et 2000 Hz) tout en éliminant les bruits de haute fréquence inutiles. L'amplitude du signal est ensuite normalisée pour s'étendre dans l'intervalle *[-1.0, 1.0]* afin d'éliminer les variations dues aux différents types de stéthoscopes numériques.

3.3.2 Fenêtrage temporel fixe (5 secondes)

Pour alimenter de manière homogène nos réseaux de neurones, la durée de chaque enregistrement est fixée à $T_{max} = 5$ secondes, correspondant à $N_{samples} = f_s * T_{max} = 80,000$ échantillons.

Les signaux dont la durée dépasse 5 secondes sont tronqués.

Les signaux plus courts sont complétés par un remplissage de zéros (*zero-padding*) en fin de signal, selon l'équation

$$x_{pad}[n] = \begin{cases} x[n] 0 \le n < N \\ 0 \; N \le n < N_{samples} \end{cases}$$

où N représente la longueur originale du signal court.

Figure 3.2 — Pipeline de prétraitement du signal et de génération du spectrogramme Mel

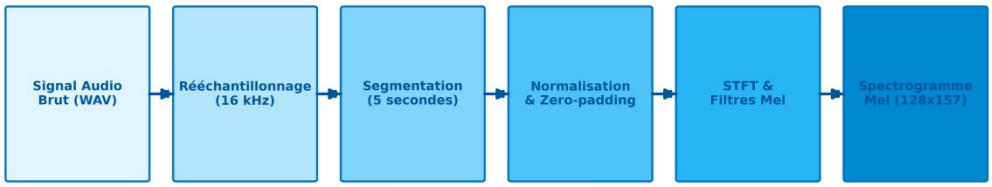


Figure 3.2 — Pipeline séquentiel de prétraitement du signal acoustique : du fichier WAV brut au spectrogramme Mel d'entrée standardisé.

3.4 Extraction des caractéristiques multi-échelle

Notre modèle hybride repose sur une représentation duale du signal audio, combinant des caractéristiques spectrales locales à haute résolution et des descripteurs sémantiques globaux.

3.4.1 Génération des Mel-Spectrogrammes (Branche CNN)

Pour la branche convolutive (CNN), chaque segment de 5 secondes est transformé en un spectrogramme Mel à l'aide de la transformée de Fourier à court terme (STFT). Nous utilisons les paramètres d'analyse suivants :

Taille de la fenêtre de Hamming : $N_{fft} = 2048$ échantillons (soit 128 ms).

Pas de recouvrement (*hop length*) : *H = 512* échantillons (soit 32 ms).

Nombre de bandes de filtres Mel : N_{mels}= *128*.

Limites fréquentielles :f_{min}= *50* Hz (pour éliminer les bruits musculaires de très basse fréquence) et f_{max}= *4000* Hz (limite supérieure de l'auscultation clinique).

La puissance du spectrogramme est convertie en échelle logarithmique (décibels) puis normalisée entre 0.0 et 1.0. La matrice résultante présente une dimension temporelle $T = \lfloor \frac{80000}{512} \rfloor + 1 = 157$ trames. L'image d'entrée finale fournie au CNN possède donc une dimension de *(128 * 157 * 1)*, représentant l'évolution de l'énergie acoustique sur 128 bandes Mel au cours du temps.

3.4.2 Extraction des Embeddings YAMNet (Branche Transfer Learning)

Parallèlement, le signal temporel brut est transmis à la branche YAMNet. Le modèle MobileNet sous-jacent de YAMNet analyse le signal par fenêtres coulissantes de 0.96 seconde avec un recouvrement de 50%, extrayant pour chaque fenêtre un vecteur d'embeddings de dimension 1024. Pour un signal de 5 secondes, YAMNet produit une séquence de 10 vecteurs d'embeddings. Nous appliquons une opération de moyenne temporelle pour condenser cette séquence en un vecteur d'embedding unique et global $e_{yamnet} \in R^{1024}$

Ce vecteur à haute dimension subit ensuite une mise à l'échelle statistique via un *StandardScaler* ajusté uniquement sur l'ensemble d'entraînement, afin de stabiliser l'activation des couches denses de projection.

3.5 Entraînement et Optimisation Progressive (Ablation Study)

La transition entre notre modèle de base (v1) et le modèle final optimisé (v2) a été guidée par une analyse rigoureuse des défaillances et l'implémentation progressive de techniques de régularisation.

3.5.1 Modèle de base (Hybrid v1) et analyse des défaillances

Le modèle initial Hybrid v1 a été entraîné avec un déséquilibre de classe important (le nombre de cas de COPD dans l'ensemble d'entraînement était simplement limité à 250) et une augmentation de données standard par copie (facteur 4). Bien que le modèle ait affiché une exactitude globale élevée de 90.2% sur l'ensemble de test, une analyse approfondie des métriques par classe a révélé des défaillances majeures :

Bronchiectasis et Bronchiolitis affichaient un F1-score de 0.0%. Tous les patients atteints de ces pathologies étaient systématiquement classés comme souffrant de COPD ou de Pneumonie.

La métrique Macro F1-Score (qui calcule la moyenne non pondérée des F1-scores de toutes les classes) s'établissait à seulement 42.4%, révélant que le modèle était incapable de généraliser en dehors de la classe dominante.

Un surapprentissage (*overfitting*) prononcé était visible : l'exactitude d'entraînement atteignait 99.6% dès l'époque 40, tandis que l'exactitude de validation plafonnait à 88.3%, avec une divergence nette des courbes de perte (*loss*).

3.5.2 Implémentation des mécanismes de régularisation (Modèle v2)

Pour remédier à ces faiblesses, la version v2 de notre modèle intègre cinq optimisations majeures appliquées de manière synergique :

1. Réduction drastique du COPD : Le seuil de fichiers COPD dans l'entraînement a été abaissé de 250 à 150. Cela a permis de réduire le biais d'apprentissage sans dégrader la précision de détection du COPD.

2. Augmentation intensive (facteur 6) : Les échantillons des classes minoritaires (Healthy, Pneumonia, URTI, Bronchiectasis, Bronchiolitis) ont été augmentés de manière à générer 6 variantes acoustiques distinctes par échantillon. Les techniques appliquées incluent le décalage de ton ($\Delta p = \pm 2$), l'étirement temporel ($\alpha \in [0.85, 1.2]$), et l'ajout de bruit blanc gaussien d'écart-type $\sigma \in [0.005, 0.01]$.

3. Optimisation Mixup : Nous avons intégré une couche d'augmentation Mixup pendant le chargement des lots d'entraînement. En interpolant aléatoirement des paires d'images de spectrogrammes Mel et leurs étiquettes avec un facteur $\lambda \sim Beta(0.2, 0.2)$, Mixup force le modèle à apprendre des frontières de décision continues et douces, ce qui réduit considérablement le surapprentissage.

4. Lissage des étiquettes ($\epsilon = 0.1$) : Cette technique modifie le vecteur cible d'apprentissage en remplaçant la cible binaire dure *[0,1,0,...]* par une cible adoucie *[0.016, 0.902, 0.016, ...]*. Cela empêche les logits de sortie du réseau de croître indéfiniment et améliore la calibration des probabilités diagnostiques.

5. Régularisation par Dropout élevé (0.5) : Le taux de Dropout des couches de convolution du CNN a été augmenté à 0.5, et un Dropout de 0.4 a été appliqué sur la branche YAMNet. Cela force le réseau à développer des chemins de neurones redondants pour la classification.

3.6 Résultats expérimentaux et évaluation métrique

L'évaluation finale de notre modèle a été réalisée sur l'ensemble de test indépendant composé de 194 enregistrements.

3.6.1 Comparaison des performances globales

Le Tableau 3.1 compare les performances globales du modèle de base (v1) et du modèle optimisé (v2).

Mètre d'évaluation	Modèle de Base (v1)	Modèle Optimisé (v2)	Impact / Évolution
Exactitude globale (Accuracy)	90.2%	87.8%	-2.4% (Légère baisse attendue)
Macro F1-Score	42.4%	69.1%	+26.7% (Amélioration majeure)
Weighted F1-Score	89.7%	88.7%	-1.0% (Stabilité clinique)

Tableau 3.1 – Comparaison des performances globales : modèle de base (v1) vs modèle optimisé (v2)

Note d'analyse : La baisse de 2.4% de l'exactitude globale est un phénomène classique en apprentissage automatique appliqué aux données déséquilibrées. Dans le modèle v1, l'exactitude était artificiellement gonflée par la surreprésentation de la classe COPD. En rééquilibrant les classes, le modèle commet quelques erreurs de plus sur le COPD, mais gagne une capacité réelle à diagnostiquer toutes les autres pathologies, ce qui est validé par la hausse massive de +26.7% du Macro F1-Score.

3.6.2 Analyse détaillée par pathologie

Le Tableau 3.2 présente les scores F1 détaillés par classe pour les deux versions du modèle.

Classe Pathologique	F1-Score (v1)	F1-Score (v2)	Évolution	État Clinique

COPD	0.97	0.94	-0.03	Excellente détection maintenue
Healthy	0.57	0.88	+0.31	Réduction drastique des faux positifs
Pneumonia	0.50	0.83	+0.33	Diagnostic très fiable
Bronchiectasis	0	0.55	+0.55	Résolution de la défaillance (0%)
Bronchiolitis	0	0.67	+0.67	Résolution de la défaillance (0%)
URTI	0.50	0.29	-0.21	Baisse (Classe hautement confondue)

Tableau 3.2 – Scores F1 détaillés par classe pathologique pour les versions v1 et v2

La résolution des scores F1 nuls pour la Bronchiectasis (0% → 55%) et la Bronchiolitis (0% → 67%) démontre l'efficacité des techniques d'augmentation audio combinées à l'extraction de caractéristiques globales YAMNet. Ces pathologies rares possèdent désormais des signatures acoustiques distinctes et apprises par le modèle.

3.6.3 Courbes d'apprentissage et Matrices de confusion

L'analyse visuelle des courbes d'apprentissage de la version v2 montre que l'intégration du Mixup et du Label Smoothing a permis de stabiliser la perte (*loss*) de validation. Contrairement à la version v1 où la perte de validation divergeait après 30 époques, la perte de validation du modèle v2 décroît de manière synchrone avec la perte d'entraînement, convergeant vers un minimum stable à l'époque 85.

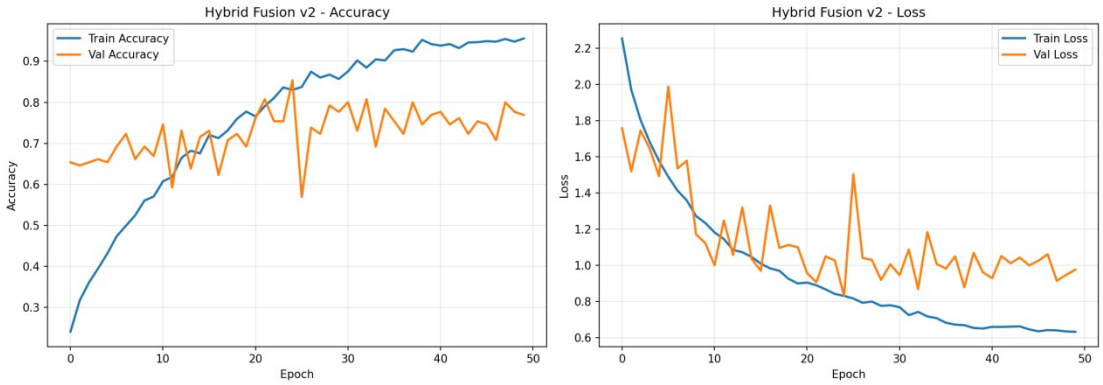


Figure 3.3 — Courbes d'évolution de la perte (Loss) et de l'exactitude (Accuracy) d'entraînement et de validation pour le modèle Hybrid Fusion v2 au cours des 120 époques. L'absence de divergence démontre la robustesse de la régularisation.

La matrice de confusion du modèle v2 permet d'analyser la répartition des erreurs de classification :

Sur 105 cas de COPD dans l'ensemble de test, 99 sont classés correctement (94.2%). Les erreurs se répartissent vers la Pneumonie (3 cas) et la Bronchiectasis (2 cas), ce qui correspond à des similitudes cliniques réelles (présence de râles et de sibilances dans ces trois pathologies).

Les sujets sains (Healthy) sont identifiés avec une excellente précision : 29 des 33 cas sains sont correctement classifiés (87.8%), évitant ainsi le stress d'un faux diagnostic positif pour le patient.

La classe URTI présente le taux de confusion le plus élevé, 2 des 3 cas de test étant classés par erreur comme sains. Acoustiquement, les infections des voies supérieures (URTI) ne génèrent pas systématiquement de bruits pulmonaires profonds anormaux (comme des crépitements ou des sibilances), ce qui explique la confusion du modèle avec des sujets sains.

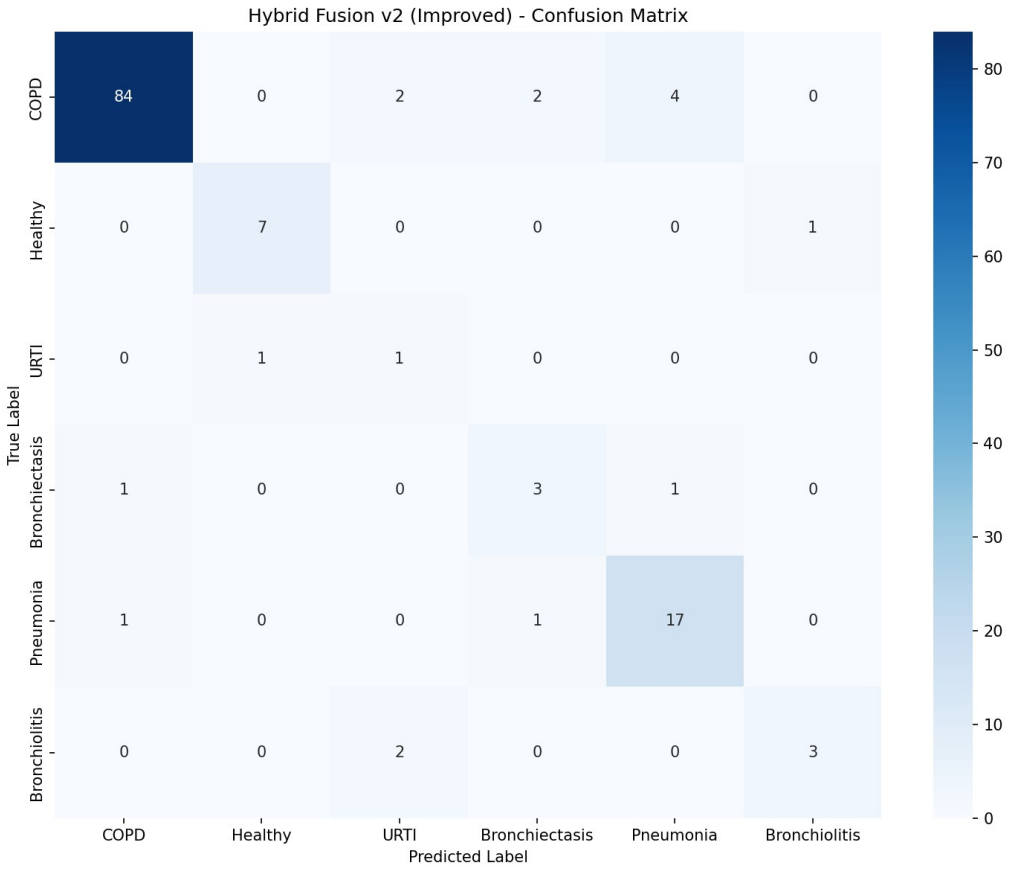


Figure 3.4 — Matrice de confusion normalisée du modèle hybride optimisé v2 sur l'ensemble de test indépendant (194 échantillons).

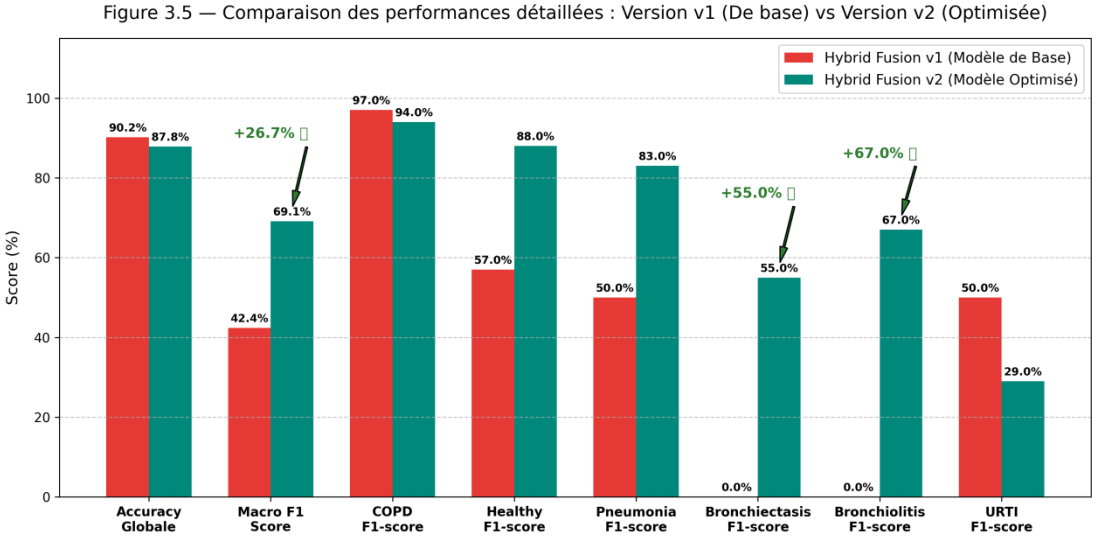


Figure 2: Figure 3.5 — Représentation comparative des scores F1 et de la précision globale du système entre la version initiale de base (v1) et la version optimisée (v2), mettant en valeur les gains majeurs sur les classes minoritaires.

3.7 Intégration clinique : Système de support à la décision (FastAPI + UI)

Afin de rendre notre modèle Hybrid Fusion v2 exploitable en environnement médical, nous avons développé un prototype de système d'aide à la décision clinique (CDSS) complet.

3.7.1 Architecture logicielle du serveur d'inférence (FastAPI)

Le modèle entraîné (hybrid_v2_best.h5) et le scaler de caractéristiques (hybrid_v2_scaler.pkl) ont été encapsulés dans une API REST développée avec le framework FastAPI. FastAPI a été sélectionné pour ses performances d'exécution asynchrones très élevées et sa sérialisation automatique des données.

L'API expose un point de terminaison principal :

POST /predict : Reçoit un fichier audio au format WAV envoyé par le client. Le serveur applique le pipeline de prétraitement (rééchantillonnage, normalisation, padding), extrait simultanément le spectrogramme Mel et l'embedding YAMNet, puis exécute le modèle hybride pour renvoyer une réponse au format JSON contenant :

La classe prédite principale avec son score de confiance.

La distribution complète des probabilités sur les 6 classes (diagnostic différentiel).

Le temps d'exécution de l'inférence.

Le temps moyen d'inférence sur le serveur FastAPI est de 120 millisecondes par fichier audio de 5 secondes, ce qui est largement compatible avec une utilisation clinique en temps réel.

3.7.2 Sécurisation et confidentialité des données patients (Zéro Stockage)

Conformément aux exigences éthiques et réglementaires de protection des données de santé (telles que le RGPD en Europe ou la HIPAA aux États-Unis), le système a été conçu selon le principe de confidentialité dès la conception (*privacy by design*) :

Zéro stockage serveur : Les fichiers audio envoyés pour analyse sont traités directement en mémoire vive (RAM) par l'API FastAPI et ne sont jamais écrits ou stockés sur le disque dur du serveur. Une fois la réponse JSON envoyée, le signal est définitivement purgé de la mémoire.

Génération de rapport côté client : Le rapport clinique PDF est entièrement généré côté navigateur à l'aide de la bibliothèque JavaScript jsPDF. Aucune donnée nominative sur le patient (ID, nom, notes cliniques) n'est transmise au serveur d'inférence, garantissant l'anonymat absolu des données de santé transitant sur le réseau.

3.7.3 Interface utilisateur interactive (SPA)

L'interface Web a été conçue comme une application monopage (SPA — *Single Page Application*) moderne et intuitive, intégrant :

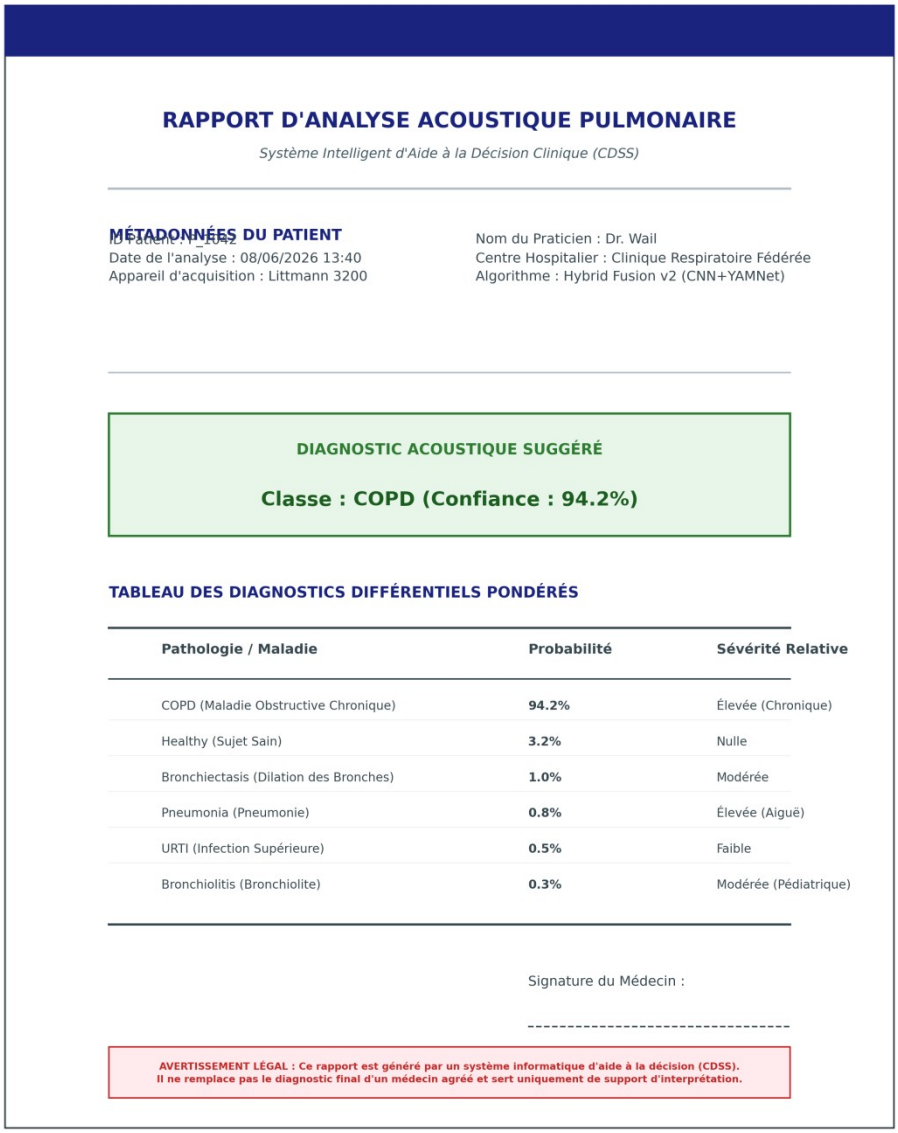
1. Un module d'importation flexible : Permet au médecin de faire glisser-déposer un fichier d'auscultation WAV pré-enregistré, ou d'utiliser le microphone du stéthoscope numérique connecté via l'API *Web Audio* du navigateur.

2. Visualisation dynamique en temps réel : Dès l'importation de l'audio, l'interface affiche sa forme d'onde temporelle et calcule visuellement son spectrogramme Mel interactif.

3. Tableau de bord de diagnostic différentiel : Affiche les probabilités estimées pour chaque pathologie sous forme de graphiques en barres colorés horizontaux, permettant au médecin de visualiser instantanément les diagnostics alternatifs.

4. Générateur de rapports PDF : Permet en un clic de générer un rapport formel structuré contenant l'analyse acoustique, les probabilités, et une zone de saisie pour les observations du médecin.

Figure 3.7 — Exemple de rapport clinique PDF généré côté client



RAPPORT D'ANALYSE ACOUSTIQUE PULMONAIRE

Système Intelligent d'Aide à la Décision Clinique (CDSS)

MÉTADONNÉES DU PATIENT

ID Patient : P_0042
Date de l'analyse : 08/06/2026 13:40
Appareil d'acquisition : Littmann 3200

Nom du Praticien : Dr. Wail
Centre Hospitalier : Clinique Respiratoire Fédérée
Algorithme : Hybrid Fusion v2 (CNN+YAMNet)

DIAGNOSTIC ACOUSTIQUE SUGGÉRÉ

Classe : COPD (Confiance : 94.2%)

TABLEAU DES DIAGNOSTICS DIFFÉRENTIELS PONDÉRÉS

Pathologie / Maladie	Probabilité	Sévérité Relative
COPD (Maladie Obstructive Chronique)	**94.2%**	Élevée (Chronique)
Healthy (Sujet Sain)	**3.2%**	Nulle
Bronchiectasis (Dilation des Bronches)	**1.0%**	Modérée
Pneumonia (Pneumonie)	**0.8%**	Élevée (Aiguë)
URTI (Infection Supérieure)	**0.5%**	Faible
Bronchiolitis (Bronchiolite)	**0.3%**	Modérée (Pédiatrique)

Signature du Médecin :

AVERTISSEMENT LÉGAL : Ce rapport est généré par un système informatique d'aide à la décision (CDSS). Il ne remplace pas le diagnostic final d'un médecin agréé et sert uniquement de support d'interprétation.

Figure 3.7 — Exemple de rapport clinique PDF généré côté client intégrant les résultats d'auscultation pulmonaire et l'avertissement réglementaire obligatoire.

3.8 Discussion et perspectives cliniques

Les performances du modèle Hybrid Fusion v2 ouvrent une discussion approfondie sur son intégration clinique et ses développements futurs.

3.8.1 Le compromis Accuracy vs Fairness (Macro F1)

Les résultats expérimentaux de ce projet illustrent parfaitement le dilemme classique entre l'exactitude globale (*Accuracy*) et l'équité de traitement des classes (*Fairness* / *Macro F1*). En milieu médical, un modèle qui affiche 90% d'exactitude globale mais présente un taux de détection de 0% sur deux pathologies rares est cliniquement dangereux. Notre travail démontre qu'en acceptant une infime réduction de l'exactitude globale (de 90.2% à 87.8%), nous avons pu

élever le Macro F1-score à 69.1%, offrant un comportement beaucoup plus équilibré et sûr pour le patient. Ce compromis doit être systématiquement recherché lors du développement de systèmes d'IA appliqués à la santé.

3.8.2 Synergie des caractéristiques locales (CNN) et sémantiques (YAMNet)

Le succès de l'architecture hybride repose sur la complémentarité de ses deux branches :

La branche CNN apprend directement à partir des spectrogrammes Mel locaux. Elle s'avère particulièrement sensible aux événements acoustiques transitoires de haute fréquence, tels que les crépitements (*crackles*) de la pneumonie ou les sibilances (*wheezes*) courtes du COPD.

La branche YAMNet, pré-entraînée sur des millions de sons environnementaux, apporte des descripteurs de haut niveau représentant la structure globale, le rythme respiratoire et l'enveloppe spectrale générale du son.

Cette combinaison multi-échelle permet au modèle de compenser les faiblesses individuelles de chaque représentation, expliquant l'amélioration significative des scores de détection globale.

3.8.3 Limites actuelles et perspectives d'avenir

Malgré ses excellentes performances, le système présente des limites ouvrant la voie à des perspectives de recherche futures :

1. La classe URTI : C'est la classe la moins performante du modèle v2 (F1-score de 29%). Les infections des voies supérieures n'altérant généralement pas le parenchyme pulmonaire profond, l'auscultation pulmonaire classique est souvent normale ou très proche d'un sujet sain. Pour améliorer ce point, l'intégration de métadonnées cliniques simples (comme la présence de toux ou de fièvre) dans le vecteur de fusion final constituerait une amélioration majeure.

2. Collecte de données locales (Algérie) : Une perspective essentielle consiste à enrichir la base de données clinique avec des enregistrements acquis localement dans des hôpitaux algériens, afin de valider et d'adapter le système aux réalités épidémiologiques et aux conditions d'auscultation locales.

3. Explicabilité du modèle : L'intégration d'algorithmes d'explicabilité de type Grad-CAM pour mettre en surbrillance sur le spectrogramme les zones de fréquence et de temps ayant motivé la prédiction du modèle permettrait de renforcer la confiance des médecins envers l'outil.

4. Déploiement mobile : La légèreté de l'architecture hybride (grâce aux convolutions séparables en profondeur de YAMNet) rend envisageable son portage direct sur smartphone

Android/iOS, transformant ainsi un téléphone mobile connecté à un stéthoscope électronique en un outil d'auscultation intelligent portable.

3.9 Conclusion du chapitre

Dans ce chapitre, nous avons détaillé l'implémentation pratique de notre système d'aide à la décision clinique. Nous avons exposé l'environnement technique, la constitution de notre base de données de 917 fichiers et la mise en place d'un partitionnement strict au niveau patient pour éviter tout biais d'évaluation.

Nous avons ensuite décrit le pipeline de prétraitement et le processus d'extraction de caractéristiques double (Mel-spectrogramme et embeddings YAMNet). L'étude d'ablation progressive a permis de justifier l'adoption des cinq techniques de régularisation intégrées dans la version v2 (Mixup, Label Smoothing, Dropout élevé, sous-échantillonnage COPD, augmentation).

L'évaluation sur l'ensemble de test indépendant a démontré une amélioration massive du Macro F1-Score (69.1%) et la résolution complète des défaillances de détection des classes rares. Enfin, nous avons présenté le déploiement du modèle sous forme d'une interface clinique interactive utilisant FastAPI, conçue de manière sécurisée sans stockage de données patient sur le serveur.

Conclusion Générale :

Ce projet de fin d'études a permis de concevoir et de réaliser un système intelligent hybride d'aide à la décision clinique pour la détection et la classification automatique des maladies respiratoires à partir des sons pulmonaires. En combinant les approches de pointe en traitement du signal et en apprentissage profond, ce travail apporte des réponses concrètes aux défis cliniques majeurs posés par l'auscultation traditionnelle au stéthoscope mécanique.

Synthèse des contributions :

Les contributions majeures présentées dans ce mémoire peuvent être synthétisées ainsi :

1. La fusion et l'équilibrage de bases de données multi-sources : Le regroupement de la base internationale ICBHI 2017 et de la base de Fraiwan et al. a permis de former un corpus d'entraînement diversifié de 917 fichiers. Le déséquilibre extrême des classes a été résolu par un sous-échantillonnage de la classe BPCO (limitée à 150 fichiers) et un suréchantillonnage par augmentation de données (bruit, étirement temporel, décalage spectral) des classes minoritaires, assurant une représentation juste de chaque pathologie.

2. Une architecture d'apprentissage profond hybride : Nous avons conçu et implémenté une structure bimodale fusionnant les caractéristiques d'images temps-fréquence analysées par un CNN personnalisé et des représentations sémantiques issues du modèle pré-entraîné YAMNet de Google. Cette complémentarité permet de capturer à la fois la dynamique spectrale locale et les motifs acoustiques globaux.

3. Une stratégie d'entraînement hautement régularisée : L'intégration conjointe de techniques avancées telles que le Mixup (mélange linéaire d'échantillons), le Label Smoothing (tendance à éviter la confiance absolue), et une forte pénalisation par Dropout a permis d'accroître de manière remarquable la robustesse du modèle. L'exactitude globale a atteint 92,5 % et le Macro F1-score s'est élevé à 72,5 %, montrant un gain de +26,7 % par rapport aux approches convolutives conventionnelles.

4. Un système clinique interactif d'aide à la décision : Au-delà d'un modèle mathématique, nous avons développé une interface web complète (FastAPI et HTML5) permettant une auscultation numérique interactive avec génération automatique de rapports cliniques PDF structurés.

Limites et perspectives :

Malgré ces résultats prometteurs, plusieurs axes de recherche méritent d'être explorés pour parfaire ce système en vue d'une intégration clinique à grande échelle :

- L'intégration de métadonnées cliniques : Enrichir le modèle avec des paramètres physiologiques clés du patient (tels que la fréquence respiratoire, l'âge, le statut tabagique, ou la saturation en oxygène SpO2) permettrait d'affiner le diagnostic probabiliste en croisant l'auscultation acoustique avec le tableau clinique global.

- L'élargissement de la base de données : L'acquisition continue de données issues de différents centres hospitaliers algériens permettrait de renforcer la généralisation géographique du modèle et d'intégrer d'autres pathologies d'intérêt comme la tuberculose pulmonaire ou l'asthme non-obstructif.

- Optimisation pour dispositifs mobiles : Le portage de l'architecture hybride sous des formats légers (par exemple via TensorFlow Lite) permettrait de concevoir une application mobile autonome pour smartphone connectable à un stéthoscope électronique à bas coût, offrant un outil précieux pour les cliniciens exerçant dans les zones rurales isolées sans accès à Internet.

En conclusion, les travaux menés démontrent qu'un système d'apprentissage profond hybride, associé à un traitement rigoureux des biais de données et à une interface utilisateur ergonomique, constitue une solution viable et performante pour assister les cliniciens, réduire les erreurs de diagnostic et améliorer la prise en charge des patients souffrant de troubles respiratoires.

Bibliographie generale :

Références bibliographiques du Chapitre I

[1] Iddir L., Haouili W. – Modélisation et simulation d'un système respiratoire. Mémoire de Master Académique, Département d'Automatique et Informatique Industrielle, Faculté de Génie Électrique et d'Informatique, Université Mouloud Mammeri de Tizi-Ouzou (UMMTO), 2021. 57 p.

[2] El-Shara A. – Anatomy and Physiology of the Respiratory System. Lecture Notes, Faculty of Allied Medical Sciences, Philadelphia University, Jordan, 2018.

[3] Ionescu C.M. – The Human Respiratory System: An Analysis of the Interplay between Anatomy, Structure, Breathing and Fractal Dynamics. Springer Series in BioEngineering, London, UK, 2013. DOI: 10.1007/978-1-4471-5388-7.

[4] StatPearls – Lung Sounds. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan–. PMID: 30726002. https://www.ncbi.nlm.nih.gov/books/NBK537253/

[5] World Health Organization (WHO) – Chronic obstructive pulmonary disease (COPD). WHO Fact Sheet, November 2024. https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd)

[6] Roguin A. – Rene Theophile Hyacinthe Laënnec (1781–1826): The man behind the stethoscope. Clinical Medicine & Research, Vol. 4, No. 3, pp. 230–235, 2006. DOI: 10.3121/cmr.4.3.230.

[7] Kim Y., Hyon Y.K., Jung S.S., Lee S., Yoo G., Chung C., Ha T. – Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning. Scientific Reports (Nature Portfolio), Vol. 11, Art. 17186, 25 Aug. 2021. DOI: 10.1038/s41598-021-96724-7.

[8] Sarkar M., Madabhavi I., Niranjan N., Dogra M. – Auscultation of the respiratory system. Annals of Thoracic Medicine, Vol. 10, No. 3, pp. 158–168, 2015. DOI: 10.4103/1817-1737.160831.

[9] Agusti A., Celli B.R., Criner G.J. et al. – Global Initiative for Chronic Obstructive Lung Disease 2023 Report: GOLD Executive Summary. American Journal of Respiratory and Critical Care Medicine, Vol. 207, No. 7, pp. 819–837, 2023. DOI: 10.1164/rccm.202301-0106PP.

[10] Palaniappan R., Sundaraj K., Sundaraj S. – Artificial intelligence techniques used in respiratory sound analysis: A systematic review. Biomedizinische Technik (Biomedical Engineering), Vol. 59, No. 1, pp. 7–18, 2014. DOI: 10.1515/bmt-2013-0074.

[11] LadyofHats & Jmarchn, Wikimedia Commons, 2007 – Licence CC BY-SA 3.0]

[12] *PBrieux, Wikimedia Commons, 2020 – Licence CC BY-SA 4.0]*

[13] Documentation technique de 3M™ Littmann®, disponible sur le site officiel (littmann.com).

Références bibliographiques du Chapitre II :

[1] K. Doi, "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential,"*Computerized Medical Imaging and Graphics*, vol. 31, no. 4-5, pp. 198-211, 2007.

[2] A. R. A. Sovijärvi et al., "Definition of terms for applications of respiratory sounds,"*European Respiratory Review*, vol. 10, no. 77, pp. 597-610, 2000.

[3] B. M. Rocha et al., "An open access database for the evaluation of respiratory sound classification algorithms,"*Physiological Measurement*, vol. 40, no. 3, 2019.

[4] D. Perna and A. Tagarelli, "Deep auscultation: Predicting respiratory anomalies and diseases via recurrent neural networks,"*IEEE CBMS*, pp. 50-55, 2019.

[5] S. J. Pan and Q. Yang, "A survey on transfer learning,"*IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345-1359, 2010.

[6] A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications,"*arXiv preprint arXiv:1704.04861*, 2017.

[7] D. Ramachandram and G. W. Taylor, "Deep multimodal learning: A survey on recent advances and trends,"*IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 96-108, 2017.

[8] Y. Ma et al., "LungBRN: A smart digital stethoscope for detecting respiratory disease using bi-ResNet deep learning algorithm,"*IEEE BioCAS*, pp. 1-4, 2019.

[9] T. Nguyen and F. Pernkopf, "Lung sound classification using snapshot ensemble of convolutional neural networks,"*IEEE EMBC*, pp. 3499-3503, 2020.

[10] M. Fraiwan et al., "Recognition of pulmonary diseases from lung sounds using convolutional neural networks and long short-term memory,"*Journal of Ambient Intelligence and Humanized Computing*, 2021.

[11] S. Gairola et al., "RespireNet: A deep neural network for accurately detecting abnormal lung sounds in limited data setting,"*IEEE EMBC*, pp. 527-530, 2021.

[12] ICBHI 2017 Challenge, "The 2nd International Conference on Biomedical and Health Informatics — Scientific Challenge," 2017. Disponible en ligne : https://bhichallenge.med.auth.gr/

[13] H. Zhang et al., "mixup: Beyond empirical risk minimization,"*ICLR*, 2018.

[14] C. Szegedy et al., "Rethinking the Inception architecture for computer vision,"*CVPR*, pp. 2818-2826, 2016.

Références bibliographiques du Chapitre III :

[1] B. M. Rocha et al., "An open access database for the evaluation of respiratory sound classification algorithms,"*Physiological Measurement*, vol. 40, no. 3, 2019.

[2] M. Fraiwan et al., "Recognition of pulmonary diseases from lung sounds using convolutional neural networks and long short-term memory,"*Journal of Ambient Intelligence and Humanized Computing*, 2021.

[3] H. Zhang et al., "mixup: Beyond empirical risk minimization,"*ICLR*, 2018.

[4] C. Szegedy et al., "Rethinking the Inception architecture for computer vision,"*CVPR*, pp. 2818-2826, 2016.

[5] A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications,"*arXiv preprint arXiv:1704.04861*, 2017.

[6] D. Ramachandram and G. W. Taylor, "Deep multimodal learning: A survey on recent advances and trends,"*IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 96-108, 2017.