République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Echahid Chikh Larbi Tébessi - Tébessa

Faculté des Sciences Exactes et Sciences de la Nature et de la Vie

Département des Mathématiques et Informatique

MÉMOIRE DE FIN D'ÉTUDES

Présenté en vue de l'obtention du diplôme de

MASTER

Filière : **Informatique**

Option : **Systèmes et multimédia**

Classification Audio pour les Maladies Respiratoires

Présenté par :

DJEDIDI Abdelghani

Devant le jury composé de :

Membre	Grade	Établissement	Qualité
Dr. DAOUADI Kheir Eddine	MCA	Université Larbi Tébessi	Président
Dr. TABET khaoula	MCB	Université Larbi Tébessi	Examinateur
Mr. GAHMOUSSE Abdelatif	MAA	Université Larbi Tébessi	Encadreur
Dr. GASMI Mohammed	MCA	Université Larbi Tébessi	Co-Encadreur

Date de soutenance : 09/06/2025

Année Universitaire : 2024-2025

Résumé

Les maladies respiratoires constituent un défi sanitaire mondial majeur, nécessitant des outils diagnostiques rapides, fiables et accessibles. L'analyse audio des sons de toux offre une alternative non invasive prometteuse. Ce projet vise à développer un système basé sur l'intelligence artificielle pour la classification automatisée des maladies respiratoires à partir de l'audio de la toux, en utilisant le jeu de données COUGHVID annoté par des experts.

Face aux défis tels que la variabilité acoustique et le bruit, nous avons exploré diverses techniques de prétraitement et d'extraction de caractéristiques, en nous concentrant sur la combinaison des coefficients MFCC et des spectrogrammes Mel. Plusieurs modèles de classification ont été entraînés et évalués sur ces caractéristiques.

Notre approche méthodologique a abouti à la sélection d'un modèle XGBoost optimisé comme étant le plus performant. Les expériences menées ont rigoureusement validé son efficacité : le modèle XGBoost final a atteint une précision remarquable de 98.01% sur l'ensemble de test (et 97.51% en validation) pour la classification des conditions respiratoires basées sur les diagnostics experts. Ces résultats surpassent significativement ceux d'autres modèles testés, tels que Random Forest (96.20% test) et les approches d'Apprentissage Profond standard (94.12% test).

Ce travail contribue une méthodologie optimisée combinant des caractéristiques acoustiques pertinentes (MFCC + Mel-spectrogramme) avec un classificateur XGBoost performant, démontrant une haute précision sur des données validées cliniquement. Ces résultats soulignent le potentiel de l'analyse audio de la toux pour développer des outils de dépistage évolutifs et économiques, renforçant ainsi la télémédecine et les stratégies de santé publique.

الملخص

تُمثل أمراض الجهاز التنفسي عبئًا صحيًا عالميًا كبيرًا، مما يستلزم أدوات تشخيص سريعة وسهلة الوصول ودقيقة. يُقدم التحليل الصوتي لأصوات السعال نهجًا غير جراحي واعدًا، مستفيدًا من التقنيات المتاحة بسهولة. تتناول هذه الأطروحة التصنيف الآلي لأمراض الجهاز التنفسي باستخدام إشارات صوتية للسعال من مجموعة بيانات COUGHVID واسعة النطاق والمُجمعة من الجمهور، مع التركيز بشكل خاص على المجموعة الفرعية التي شرحها خبراء طبيون لضمان أهميتها السريرية.

تشمل التحديات في هذا المجال التباين الصوتي، وضوضاء الخلفية، والحاجة إلى استخراج فعال للخصائص ونماذج تصنيف متينة. يهدف بحثنا إلى تطوير وتقييم أطر عمل متقدمة للتعلم الآلي للتغلب على هذه التحديات. لقد بحثنا في تقنيات مُختلفة لمعالجة الإشارات مسبقًا، وركزنا على استخراج خصائص صوتية مفيدة، وخاصةً الجمع بين معاملات ميل-فريكوينسي سيبسترا (MFCCs) ومخططات ميل الطيفية.

يتضمن جوهر منهجيتنا المقترحة تدريب وتحسين العديد من نماذج التصنيف على هذه الخصائص المُجمعة المُستمدة من بيانات COUGHVID المُصنفة من قِبل الخبراء. أظهر التحليل المقارن تفوق نموذج XGBoost المُحسَّن. وقد أثبتت تجاربنا بدقة فعالية هذا النهج، وتوجت نموذج XGBoost النهائي بتحقيق دقة اختبار ملحوظة بلغت 98.01% (مع دقة تحقق بلغت 97.51%) في تصنيف حالات الجهاز التنفسي بناءً على تشخيصات الخبراء. وقد تفوق هذا بشكل ملحوظ على النماذج الأخرى المختبرة، بما في ذلك نموذج الغابة العشوائية (دقة اختبار بلغت 96.20%) ونهج التعلم العميق القياسي (دقة اختبار بلغت 94.12%).

يساهم هذا العمل في تحقيق منهجي في تصنيف أمراض الجهاز التنفسي القائمة على الصوت باستخدام بيانات مُتحققة سريريًا. وتشمل المساهمات الرئيسية خط أنابيب مُحسَّن يجمع بين ميزات MFCC وMel-spectrogram مع مُصنِّف XGBoost، مما يُظهر دقة عالية. وتُؤكد النتائج على إمكانات هذا النهج في تطوير أدوات قابلة للتطوير وفعالة من حيث التكلفة لفحص ومراقبة صحة الجهاز التنفسي، مما يُسهم بشكل كبير في تطوير استراتيجيات الطب عن بُعد والصحة العامة من خلال توفير طريقة فعّالة لتحليل صوت السعال.

Abstract

Respiratory diseases represent a significant global health burden, necessitating diagnostic tools that are rapid, accessible, and accurate. Audio-based analysis of cough sounds offers a promising non-invasive approach, leveraging readily available technology. This thesis addresses the automated classification of respiratory diseases using cough audio signals from the large-scale, crowdsourced COUGHVID dataset, focusing specifically on the subset annotated by medical experts to ensure clinical relevance.

Challenges in this domain include acoustic variability, background noise, and the need for effective feature extraction and robust classification models. Our research aims to develop and evaluate advanced machine learning frameworks to overcome these challenges. We investigated various signal preprocessing techniques and focused on extracting informative acoustic features, particularly combining Mel-Frequency Cepstral Coefficients (MFCCs) and Mel spectrograms.

The core of our proposed methodology involved training and optimizing several classification models on these combined features derived from the expert-labeled COUGHVID data. Comparative analysis demonstrated the superiority of an optimized XGBoost model. Our experiments rigorously validated the effectiveness of this approach, culminating in the final XGBoost model achieving a remarkable test accuracy of 98.01% (with a validation accuracy of 97.51%) in classifying respiratory conditions based on the expert diagnoses. This significantly outperformed other tested models, including Random Forest (96.20% test accuracy) and standard Deep Learning (Apprentissage Profond) approaches (94.12% test accuracy).

This work contributes a systematic investigation into audio-based respiratory disease classification using clinically validated data. Key contributions include an optimized pipeline combining MFCC and Mel-spectrogram features with an XGBoost classifier, demonstrating high accuracy. The findings underscore the potential of this approach for developing scalable, cost-effective tools for respiratory health screening and monitoring, significantly advancing telemedicine and public health strategies by providing a robust method for cough sound analysis.

REMERCIEMENTS

Je loue Dieu Tout-Puissant, source de toute force, de toute sagesse et de toute patience, pour m'avoir guidé et soutenu tout au long de ce long parcours, souvent semé de défis mais riche en apprentissages et en expériences. Sa présence m'a été un pilier essentiel dans chaque étape de cette aventure académique.

Je tiens à exprimer ma profonde gratitude à mes encadrants, Monsieur GAHMOUSSE Abdelatif et Monsieur GASMI Mohammed, pour avoir accepté de superviser ce travail avec tant de sérieux, de bienveillance et de compétence. Leur accompagnement constant, leur disponibilité, leurs conseils éclairés ainsi que leur exigence scientifique ont grandement contribué à l'orientation et à l'aboutissement de ce mémoire. Leur confiance en mes capacités et leur soutien, tant moral que méthodologique, m'ont permis d'avancer avec assurance.

Mes remerciements vont également aux membres du jury, pour l'honneur qu'ils me font en acceptant d'évaluer ce travail. Leur regard critique et leur expertise sont pour moi une reconnaissance précieuse et un encouragement à poursuivre sur la voie de l'excellence.

Je ne saurais oublier l'ensemble des enseignants qui m'ont transmis, au fil des années, non seulement un savoir académique solide, mais aussi des valeurs de rigueur, de persévérance et de curiosité intellectuelle. Leur dévouement et leur engagement ont profondément marqué mon parcours.

Enfin, je souhaite adresser mes remerciements les plus sincères à ma famille, qui a toujours cru en moi et m'a entouré de son amour, de sa patience et de son soutien indéfectible. Leur présence constante à mes côtés a été la clé de mon équilibre et de ma réussite.

Table des matières

1 Introduction générale **13**

1.1 Contexte et Motivation 13

1.1.1 Problématique et Défis 14

1.1.2 Objectifs et Contributions 15

1.1.3 Impact et Perspectives 16

1.1.4 Structure de la Mémoire 17

1.2 Le Fardeau Mondial des Maladies Respiratoires et les Défis Diagnostiques 17

1.2.1 Émergence des Diagnostics Basés sur l'Audio comme Fron- tière Prometteuse 18

1.3 Les Principales Affections Respiratoires et la Toux 18

1.3.1 Maladies Respiratoires Courantes 18

1.3.2 Comprendre la Toux comme Symptôme Clé 18

1.4 L'Analyse de la Toux comme Outil de Diagnostic 19

1.4.1 Importance et Limites de l'Évaluation Traditionnelle 19

1.4.2 Émergence des Diagnostics Basés sur l'Analyse Audio 19

1.5 Conclusion 19

2 Concepts et outils pour le traitement et reconnaissance du son numérique **22**

2.1 Introduction 22

2.2 Fondamentaux du Traitement Numérique du Signal pour l'Audio 23

2.2.1 Propriétés et Caractéristiques des Ondes Sonores 23

2.2.2 Conversion Analogique-Numérique 24

2.2.3 Théorie de l'Échantillonnage et Théorème de Nyquist 24

2.2.4 Représentations dans le Domaine Temporel et Fréquentiel 24

2.2.5 Techniques de Fenêtrage 25

2.2.6 Transformées de Fourier et Analyse Spectrale 25

2.2.7 Filtrage Numérique 26

2.3 Caractéristiques Acoustiques Utilisées 26

2.3.1 Caractéristiques dans le Domaine Fréquentiel 26

2.3.2 Caractéristiques Cepstrales . . . 27
2.3.3 Caractéristiques Temps-Fréquence . . . 28
2.4 Approches de Classification Audio . . . 28
2.4.1 Méthodes Traditionnelles d'Apprentissage Automatique . . . 28
2.4.2 Méthodes d'Apprentissage Profond . . . 30
2.4.3 Apprentissage par Transfert . . . 31
2.5 Outils et Bibliothèques pour l'Analyse Audio . . . 31
2.5.1 Écosystème Python pour le Traitement Audio . . . 31
2.5.2 Bibliothèques Audio Spécialisées . . . 32
2.6 Évaluation des Modèles . . . 32
2.6.1 Métriques d'Évaluation . . . 33
2.6.2 Techniques de Validation . . . 33
2.7 Défis et Directions Futures . . . 34
2.8 Applications de l'analyse des sons respiratoires dans la détection de la COVID-19 . . . 34
2.8.1 Analyse des toux . . . 35
2.8.2 Reconnaissance des motifs respiratoires . . . 35
2.8.3 Évaluation de la qualité vocale . . . 36
2.8.4 Recherches actuelles et défis . . . 36
2.8.5 Considérations éthiques . . . 37
2.9 Conclusion . . . 37

3 État De L'art **40**
3.1 Introduction . . . 40
3.1.1 Émergence des Diagnostics Basés sur l'Audio comme Frontière Prometteuse . . . 40
3.2 Travaux connexes . . . 41
3.2.1 Approches Basées Sur Machine Learning (ML) . . . 41
3.2.2 Approches Basées Sur Deep Learning (Apprentissage Profond) . . . 43
3.2.3 Approches Hybrides et Ensembles . . . 43
3.3 Le Dataset COUGHVID . . . 44
3.3.1 Origine, Échelle et Méthodologie de Collecte Initiale . . . 44
3.3.2 Évolution et Amélioration de la Qualité : Étiquetage par Experts et Approches Semi-Supervisées . . . 45
3.3.3 Importance et Défis des Datasets Médicaux Crowdsourcés . . . 46
3.4 Présentation du jeu de données COUGHVID : Introduction, Collecte et Description . . . 47
3.4.1 Démographie des Participants . . . 47
3.4.2 Distribution Géographique . . . 49

3.5 Avancées en IA/ML pour la Classification des Maladies Respiratoires utilisant COUGHVID (2022-2025) . . . 49
3.5.1 Architectures d'Apprentissage Profond (Deep Learning) : Repousser les Limites de Performance . . . 50
3.5.2 Techniques de Prétraitement et d'Extraction de Caractéristiques . 51
3.5.3 Défis et Directions Futures . . . 52
3.6 Conclusion . . . 53

4 Exploration, Nettoyage et Préparation des Données Acoustiques 55
4.1 Présentation du jeu de données COUGHVID : Introduction, Collecte et Description . . . 55
4.1.1 Analyse de l'État de Santé . . . 56
4.1.2 Caractéristiques Audio . . . 57
4.2 Analyse Exploratoire des Données (AED) . . . 58
4.2.1 Démographie des Participants . . . 58
4.2.2 Distribution Géographique . . . 60
4.2.3 Analyse de l'État de Santé . . . 60
4.2.4 Caractéristiques Audio . . . 61
4.3 Analyse Exploratoire des Données (AED) sur les Étiquettes Expertes . . 62
4.3.1 Qualité des Enregistrements . . . 62
4.3.2 Type de Toux . . . 63
4.3.3 Symptômes Audibles . . . 64
4.3.4 Sévérité de la Toux . . . 66
4.3.5 Diagnostic Expert . . . 67
4.3.6 Interprétation et Implications des Distributions . . . 67
4.4 Préparation et Filtrage des Données Expertes . . . 68
4.5 Segmentation des Événements de Toux . . . 69
4.6 Chargement et normalisation des segments audio . . . 70
4.7 Augmentation des Données pour la Réduction du Surapprentissage . . . 71
4.8 Extraction de Caractéristiques Acoustiques . . . 73

5 Modélisation Acoustique et Évaluation des Performances 76
5.1 Division Stratégique des Données : Entraînement, Validation et Test . . . 76
5.2 Correction du déséquilibre des classes par pondération des classes . . . 77
5.3 Modélisation et Optimisation . . . 78
5.3.1 Choix du Modèle : XGBoost . . . 78
5.3.2 Optimisation des Hyperparamètres par Optimisation Bayésienne . 79
5.3.3 Résultats et Évaluation du Modèle Final . . . 80
5.4 Analyse de l'Importance des Caractéristiques et Sélection . . . 81

5.4.1 Analyse Initiale de l'Importance 81
5.4.2 Contributions par Type de Caractéristique 82
5.4.3 Expérimentations avec Combinaisons de Caractéristiques 82
5.4.4 Conclusion sur la Sélection des Caractéristiques 83
5.5 Évaluation par Validation Croisée 83
5.6 Analyse Visuelle de la Performance du Modèle Final 84
5.6.1 Matrices de Confusion 84
5.6.2 Courbes de Perte d'Entraînement et de Validation 87
5.6.3 Conclusion de l'Analyse Visuelle 88
5.7 Comparaison avec d'Autres Modèles de Classification 88
5.8 Évaluation avec les Étiquettes Auto-Déclarées 90
Conclusion générale 92

Table des figures

2.1 Processus de conversion analogique-numérique (ADC) et numérique-analogique (DAC). La partie supérieure montre l'échantillonnage, la quantification et la sérialisation pour la conversion analogique-numérique. La partie inférieure illustre le processus inverse pour la conversion numérique-analogique. 38

3.1 Distribution de l'Âge des Participants. 48
3.2 Distribution du Genre des Participants. 48
3.3 Distribution Géographique des Échantillons (colorée par statut auto-déclaré). 49

4.1 Distribution de l'État COVID-19 Auto-Déclaré. 57
4.2 Distribution des Scores de Détection de Toux. 58
4.3 Distribution de l'Âge des Participants. 59
4.4 Distribution du Genre des Participants. 59
4.5 Distribution Géographique des Échantillons (colorée par statut auto-déclaré). 60
4.6 Distribution de l'État COVID-19 Auto-Déclaré. 61
4.7 Distribution des Scores de Détection de Toux. 62
4.8 Distribution de la Qualité des Enregistrements (Étiquettes Expertes). . . 63
4.9 Distribution du Type de Toux (Étiquettes Expertes). 63
4.10 Présence de Dyspnée Audible (Étiquettes Expertes). 64
4.11 Présence de Sifflements Audibles (Étiquettes Expertes). 64
4.12 Présence de Stridor Audible (Étiquettes Expertes). 65
4.13 Présence d'États d'Étouffement Audibles (Étiquettes Expertes). 65
4.14 Présence de Congestion Nasale Audible (Étiquettes Expertes). 65
4.15 Absence d'Autres Symptômes Audibles ('Nothing', Étiquettes Expertes). 66
4.16 Distribution de la Sévérité de la Toux (Étiquettes Expertes). 66
4.17 Distribution du Diagnostic Expert (Étiquettes Expertes). 67
4.18 Illustration du processus de segmentation basé sur l'énergie avec seuils adaptatifs. 70
4.19 Illustration conceptuelle des techniques däugmentation de données audio appliquées. 72
4.20 Illustration de l'extraction de caractéristiques : Forme d'onde, Spectrogramme Mel et MFCCs. 73

5.1 Précision Moyenne par Type de Caractéristique (Basée sur les Top 20 Caractéristiques Individuelles). 82
5.2 Matrice de Confusion pour l'Ensemble d'Entraînement (Modèle Final). . 85
5.3 Matrice de Confusion pour l'Ensemble de Validation (Modèle Final). . . 86
5.4 Matrice de Confusion pour l'Ensemble de Test (Modèle Final). 87
5.5 Courbes de Perte d'Entraînement et de Validation au fil des Itérations (Modèle Final). 88

Liste des tableaux

3.1 Comparaison des approches basées sur Machine Learning pour la détection des maladies respiratoires 42
3.2 Comparaison des approches basées sur Deep Learning (Apprentissage Profond) pour la détection des maladies respiratoires 44
3.3 Comparaison des approches hybrides et d'ensemble pour la détection des maladies respiratoires 45

5.1 Résultats des Expérimentations avec Combinaisons de Caractéristiques . 83
5.2 Comparaison des Performances de Différents Modèles de Classification . . 89

Chapitre 1
Contexte, Enjeux et Diagnostic Audio des Maladies Respiratoires

Chapitre 1

Introduction générale

Les maladies respiratoires constituent un défi majeur pour la santé publique mondiale, affectant des centaines de millions de personnes et représentant une cause significative de morbidité et de mortalité [1]. Ces affections, qui englobent un large spectre allant des infections aiguës comme la pneumonie aux maladies chroniques telles que l'asthme et la bronchopneumopathie chronique obstructive (BPCO), imposent un fardeau considérable sur les systèmes de santé et la qualité de vie des patients. La pandémie de COVID-19 a davantage mis en lumière l'importance cruciale d'un diagnostic précoce et précis des maladies respiratoires, ainsi que la nécessité de développer des méthodes de dépistage accessibles, non invasives et évolutives [2, 3].

Dans ce contexte, l'analyse des sons respiratoires, et particulièrement des sons de toux, émerge comme une approche prometteuse pour la détection et la surveillance des pathologies respiratoires [4]. La toux, symptôme cardinal de nombreuses affections respiratoires, possède des caractéristiques acoustiques distinctives qui peuvent être analysées pour identifier des signatures spécifiques à différentes pathologies. Les avancées récentes dans les domaines du traitement du signal numérique, de l'apprentissage automatique et de l'intelligence artificielle ont ouvert de nouvelles perspectives pour l'automatisation de cette analyse, permettant potentiellement un diagnostic plus rapide, plus précis et plus accessible [5, 6].

1.1 Contexte et Motivation

Le diagnostic traditionnel des maladies respiratoires repose principalement sur l'examen clinique, l'imagerie médicale (radiographies, tomodensitométries), les tests de fonction pulmonaire et les analyses de laboratoire. Ces méthodes, bien qu'efficaces, présentent plusieurs limitations : elles nécessitent souvent un équipement spécialisé, un personnel qualifié, et impliquent des coûts significatifs, les rendant inaccessibles dans de nombreuses régions du monde, particulièrement dans les zones à ressources limitées [7]. De plus, ces

approches conventionnelles peuvent être chronophages, et difficiles à déployer à grande échelle lors d'épidémies ou de pandémies.

L'analyse audio des sons respiratoires offre une alternative prometteuse, exploitant la nature non invasive et la facilité d'acquisition des enregistrements sonores [8]. Les sons respiratoires, notamment la toux, contiennent des informations riches sur l'état physiologique du système respiratoire. Les caractéristiques acoustiques comme la durée, la fréquence, l'intensité et le timbre peuvent révéler des anomalies spécifiques associées à différentes pathologies [9]. Par exemple, la toux sèche et persistante caractéristique de la COVID-19 présente des signatures acoustiques distinctes de celles observées dans l'asthme ou la BPCO [10].

La démocratisation des technologies mobiles et l'omniprésence des smartphones équipés de microphones de haute qualité ont créé un environnement propice au développement d'applications de santé basées sur l'analyse audio [11]. Ces dispositifs permettent l'enregistrement, le traitement et même l'analyse en temps réel des sons respiratoires, ouvrant la voie à des solutions de télémédecine et de surveillance à distance particulièrement pertinentes dans le contexte actuel de distanciation sociale et de surcharge des systèmes de santé [12].

1.1.1 Problématique et Défis

Malgré son potentiel considérable, la classification audio automatisée des maladies respiratoires présente de nombreux défis techniques et méthodologiques qui doivent être adressés pour développer des systèmes fiables et cliniquement pertinents.

Premièrement, la variabilité intrinsèque des sons respiratoires constitue un obstacle majeur. Les caractéristiques acoustiques de la toux peuvent varier significativement entre les individus, même ceux atteints de la même pathologie, en fonction de facteurs comme l'âge, le sexe, la morphologie, la sévérité de la maladie et les comorbidités [13]. Cette hétérogénéité complique la tâche d'identification de signatures acoustiques universelles pour chaque maladie.

Deuxièmement, la qualité des enregistrements audio représente un défi important. Les conditions d'enregistrement non contrôlées, typiques des applications réelles, introduisent du bruit ambiant, des distorsions et des variations dans la distance et l'orientation du microphone, affectant la fidélité du signal [14]. Ces facteurs peuvent masquer ou altérer les caractéristiques acoustiques pertinentes pour le diagnostic.

Aussi, l'extraction de caractéristiques discriminantes à partir des signaux audio bruts nécessite des techniques sophistiquées de traitement du signal et d'apprentissage automatique. Le choix des représentations appropriées (domaine temporel, fréquentiel, ou temps-fréquence) et des descripteurs acoustiques (MFCC, spectrogrammes, coefficients LPC, etc.) est crucial pour capturer l'information diagnostique pertinente [15].

Autrement, la sélection et l'optimisation des algorithmes de classification représentent un défi considérable. L'éventail des approches s'étend des méthodes traditionnelles d'apprentissage automatique (SVM, Random Forests, XGBoost) aux architectures avancées de deep learning (Apprentissage Profond, CNN, RNN, transformers), chacune présentant des avantages et des limitations spécifiques [16, 17]. Le choix de l'architecture optimale dépend de multiples facteurs, incluant la taille et la qualité du dataset, la complexité du problème de classification, et les contraintes computationnelles.

Enfin, la disponibilité de datasets étiquetés de haute qualité constitue un obstacle majeur. L'acquisition d'enregistrements audio de patients atteints de diverses pathologies respiratoires, accompagnés de diagnostics cliniques fiables, est complexe et coûteuse [3]. Cette limitation est particulièrement prononcée pour les maladies rares ou émergentes, pour lesquelles les données sont naturellement plus restreintes.

1.1.2 Objectifs et Contributions

Face à ces défis, cette mémoire vise à développer et évaluer des méthodes avancées pour la classification audio automatisée des maladies respiratoires, avec un accent particulier sur l'analyse des sons de toux. Les objectifs spécifiques de ce travail sont les suivants :

1. Explorer et comparer différentes techniques de prétraitement et d'extraction de caractéristiques pour les signaux audio respiratoires, afin d'identifier les représentations les plus informatives et robustes pour la classification des pathologies.
2. Développer et évaluer des modèles d'apprentissage automatique et de deep learning (Apprentissage Profond) pour la classification des maladies respiratoires à partir d'enregistrements audio, en mettant l'accent sur l'optimisation des performances en termes de précision, sensibilité et spécificité.
3. Investiguer l'impact de différentes architectures de réseaux de neurones (CNN, RNN, transformers) et de techniques d'ensemble sur les performances de classification, et proposer des approches hybrides combinant les forces de différents paradigmes.
4. Évaluer la robustesse des modèles développés face aux variations dans les conditions d'enregistrement, les caractéristiques démographiques des patients, et la sévérité des pathologies.
5. Proposer des méthodes pour améliorer l'interprétabilité des modèles de classification, permettant d'identifier les caractéristiques acoustiques les plus discriminantes pour chaque pathologie et facilitant ainsi l'adoption clinique de ces technologies.

Les contributions attendues de cette mémoire incluent :

1. Une analyse comparative approfondie des méthodes existantes pour la classification audio des maladies respiratoires, identifiant leurs forces, limitations et domaines d'application optimaux.
2. Des techniques innovantes de prétraitement et d'extraction de caractéristiques spécifiquement adaptées aux sons respiratoires, améliorant la robustesse face au bruit et aux variations dans les conditions d'enregistrement.
3. Des architectures de modèles optimisées pour la classification des pathologies respiratoires à partir d'enregistrements audio, démontrant des performances supérieures à l'état de l'art sur des benchmarks standardisés.
4. Des insights sur les signatures acoustiques caractéristiques de différentes maladies respiratoires, contribuant à une meilleure compréhension de leurs manifestations sonores et potentiellement à l'amélioration des pratiques diagnostiques cliniques.
5. Des recommandations pratiques pour l'implémentation de systèmes de classification audio dans des applications réelles de télémédecine et de surveillance à distance des patients atteints de maladies respiratoires.

1.1.3 Impact et Perspectives

Cette recherche s'inscrit dans un contexte où les technologies numériques transforment rapidement le paysage des soins de santé. L'analyse audio automatisée des maladies respiratoires présente un potentiel considérable pour démocratiser l'accès au diagnostic, particulièrement dans les régions à ressources limitées où l'accès aux spécialistes et aux équipements médicaux avancés est restreint [7].

Les applications potentielles sont nombreuses et diversifiées : outils de triage pour identifier les cas nécessitant une attention médicale urgente, systèmes de télésurveillance pour le suivi à domicile des patients chroniques, applications mobiles pour l'auto-évaluation préliminaire, ou encore plateformes de dépistage à grande échelle lors d'épidémies [11].

Au-delà des applications immédiates en diagnostic, cette recherche contribue à une meilleure compréhension des manifestations acoustiques des pathologies respiratoires, potentiellement ouvrant de nouvelles perspectives pour la phénotypage des maladies et la médecine personnalisée [18].

Dans un monde où les maladies respiratoires continuent de représenter un défi majeur de santé publique, exacerbé par des facteurs comme la pollution atmosphérique, le vieillissement de la population et l'émergence de nouveaux agents pathogènes, le développement de méthodes de diagnostic innovantes, accessibles et évolutives est plus crucial que jamais. Cette mémoire aspire à contribuer significativement à cet effort, en explorant le potentiel de l'intelligence artificielle et du traitement du signal pour transformer notre approche de la détection et de la gestion des maladies respiratoires.

1.1.4 Structure de la Mémoire

Cette mémoire est structurée en quatre chapitres principaux, précédés de cette introduction générale et suivis d'une conclusion générale. Le premier chapitre présente une vue d'ensemble des maladies respiratoires courantes, en mettant l'accent sur la toux comme symptôme clé et son importance diagnostique. Le deuxième chapitre aborde les concepts fondamentaux et les outils nécessaires à la reconnaissance numérique du son, incluant le traitement du signal, l'extraction de caractéristiques et les bases de l'apprentissage automatique appliqué à l'audio. Le troisième chapitre dresse un état de l'art détaillé des travaux existants sur la classification audio des maladies respiratoires, en analysant les différentes approches méthodologiques (Machine Learning, Deep Learning-Apprentissage Profond, hybrides) et les jeux de données utilisés, notamment le dataset COUGHVID. Le quatrième chapitre expose notre approche proposée, détaillant la méthodologie adoptée pour le prétraitement des données, l'analyse exploratoire, l'extraction de caractéristiques et la construction des modèles de classification basés sur les enregistrements de toux du jeu de données COUGHVID annoté par des experts. Enfin, la conclusion générale synthétise les principaux résultats de notre recherche, discute leurs implications, reconnaît les limites de l'étude et propose des pistes pour les travaux futurs.

1.2 Le Fardeau Mondial des Maladies Respiratoires et les Défis Diagnostiques

Les maladies respiratoires représentent un défi majeur pour la santé mondiale, englobant un large spectre de conditions allant des infections courantes comme la grippe aux affections chroniques telles que l'asthme et la bronchopneumopathie chronique obstructive (BPCO), sans oublier les pandémies émergentes comme la COVID-19 [3]. Leur nature omniprésente, leur potentiel de transmission rapide et les complexités associées à un diagnostic précis et opportun contribuent à des charges substantielles pour la santé publique. Les méthodologies diagnostiques traditionnelles, notamment les tests de laboratoire comme la réaction en chaîne par polymérase après transcription inverse (RT-PCR), l'imagerie médicale comme les radiographies et les tomodensitométries (TDM), ainsi que les examens cliniques complets, nécessitent souvent des ressources considérables, sont chronophages ou requièrent une infrastructure spécialisée et du personnel formé.

L'impératif d'outils de dépistage rapides, non invasifs et rentables est devenu particulièrement aigu pendant les périodes d'épidémies généralisées, comme la pandémie de COVID-19. Cette urgence a servi de puissant catalyseur pour l'innovation, stimulant l'exploration et le développement de paradigmes diagnostiques alternatifs capables de surmonter les limitations des approches conventionnelles.

1.2.1 Émergence des Diagnostics Basés sur l'Audio comme Frontière Prometteuse

Les signaux sonores respiratoires humains, spécifiquement la toux, les schémas respiratoires et les vocalisations, sont intrinsèquement liés à la fonction pulmonaire et peuvent manifester des signatures acoustiques distinctes indicatives de pathologies respiratoires sous-jacentes [7]. L'analyse de ces sons offre une avenue non invasive et hautement accessible pour la détection des maladies et la surveillance continue. Cette approche exploite la connexion directe de ces signaux avec les poumons, fournissant une fenêtre unique sur la santé respiratoire. L'adoption généralisée de la technologie des smartphones et la sophistication croissante des dispositifs portables ont encore amplifié le potentiel de collecte à grande échelle et à distance de ces sons physiologiques. Cette omniprésence technologique ouvre la voie au développement et au déploiement d'applications diagnostiques pouvant atteindre de larges populations, particulièrement dans les contextes à ressources limitées où l'infrastructure diagnostique traditionnelle peut être restreinte. La capacité à conduire des évaluations préliminaires et une stratification des risques à distance représente une avancée significative en santé publique.

1.3 Les Principales Affections Respiratoires et la Toux

Cette section présente un aperçu des affections respiratoires les plus prévalentes qui affectent la santé mondiale. Elle se concentre ensuite spécifiquement sur la toux, symptôme principal et souvent révélateur, commun à nombre de ces pathologies. En identifiant d'abord les principales pathologies puis en examinant les différentes facettes de la toux, nous posons les bases cliniques de son utilisation en tant qu'indicateur diagnostique.

1.3.1 Maladies Respiratoires Courantes

Les maladies respiratoires courantes regroupent un ensemble de pathologies affectant les voies respiratoires et les poumons, représentant une part significative de la morbidité et de la mortalité à l'échelle mondiale. Cette section présente un aperçu des affections les plus répandues, notamment la bronchopneumopathie chronique obstructive (BPCO), l'asthme, les infections respiratoires (comme la pneumonie ou la bronchite), ainsi que la COVID-19. Chacune de ces maladies présente des mécanismes physiopathologiques spécifiques, des symptômes caractéristiques et des implications cliniques importantes.

1.3.2 Comprendre la Toux comme Symptôme Clé

La toux est l'un des symptômes les plus fréquents en pathologie respiratoire, jouant un rôle essentiel dans la protection des voies aériennes. Les toux peuvent être classées selon

leur durée (aiguë, subaiguë, chronique) ou leurs caractéristiques (sèche, grasse/productive). Les causes sont variées, incluant les infections, l'asthme, le reflux gastro-œsophagien (RGO) ou des irritants environnementaux. Les caractéristiques de la toux varient selon la maladie ; par exemple, la BPCO est souvent associée à une toux productive chronique, tandis que la COVID-19 provoque généralement une toux sèche et persistante.

1.4 L'Analyse de la Toux comme Outil de Diagnostic

La toux étant établie comme un symptôme clinique clé, cette section explore son application pratique en tant qu'outil de diagnostic. Elle aborde dans un premier temps la valeur et les limites inhérentes à l'évaluation traditionnelle et subjective de la toux, avant d'introduire le passage à l'analyse automatisée basée sur l'audio comme solution de diagnostic moderne.

1.4.1 Importance et Limites de l'Évaluation Traditionnelle

La toux est un outil diagnostique clé, car ses caractéristiques peuvent fournir des informations précieuses sur la cause sous-jacente. Cependant, l'évaluation traditionnelle présente des limites : elle est subjective, sujette aux biais de rappel du patient, et ne permet pas une quantification facile de la fréquence ou de l'intensité.

1.4.2 Émergence des Diagnostics Basés sur l'Analyse Audio

Les signaux sonores respiratoires humains, comme la toux, sont liés à la fonction pulmonaire et peuvent présenter des signatures acoustiques distinctes indicatives de pathologies. L'analyse de ces sons offre une approche non invasive et très accessible pour la détection des maladies. L'adoption généralisée des smartphones a amplifié le potentiel de collecte de ces sons à grande échelle, ouvrant la voie à des applications de diagnostic, en particulier dans les contextes à ressources limitées. L'analyse automatisée des sons de toux offre plusieurs avantages, tels que l'objectivité, la surveillance continue, la détection précoce et l'évaluation à distance, ce qui est particulièrement pertinent pour le dépistage de masse lors d'épidémies.

1.5 Conclusion

Les maladies respiratoires affectent des millions de personnes dans le monde et causent d'importants problèmes de santé. La toux est un symptôme courant et important de nombreuses affections respiratoires. Comprendre les différents types de toux et leurs causes est essentiel pour un diagnostic et un traitement appropriés.

Les méthodes traditionnelles d'évaluation de la toux présentent des limites. Les nouvelles technologies numériques d'analyse des sons de toux offrent des moyens prometteurs d'améliorer le diagnostic, la surveillance et la gestion des maladies respiratoires. Ces technologies pourraient être particulièrement précieuses pour des affections comme la COVID-19, où la détection précoce et la surveillance à distance sont importantes.

Les chapitres suivants exploreront les aspects techniques de la reconnaissance numérique des sons et comment elle peut être appliquée pour analyser les sons de toux pour la détection des maladies respiratoires.

Chapitre 2
Concepts et outils pour le traitement et reconnaissance du son numérique

Chapitre 2

Concepts et outils pour le traitement et reconnaissance du son numérique

2.1 Introduction

La reconnaissance numérique du son est devenue un outil puissant dans divers domaines, des systèmes de reconnaissance vocale à la classification des sons environnementaux. Ces dernières années, son application dans le domaine de la santé a suscité une attention considérable, notamment pour le diagnostic et le suivi des maladies respiratoires [19]. La capacité d'analyser et de classifier les signaux audio des sons respiratoires offre une approche non invasive, accessible et potentiellement rentable pour la détection et le suivi des maladies [20].

La pandémie de COVID-19 a accéléré la recherche dans ce domaine, avec de nombreuses études explorant comment les sons respiratoires – tels que la toux, les schémas respiratoires et les caractéristiques vocales – peuvent fournir des informations diagnostiques précieuses [10, 21]. Ces biomarqueurs acoustiques peuvent potentiellement révéler des changements subtils dans le système respiratoire qui se produisent lors d'une infection, même avant que d'autres symptômes ne deviennent apparents [22].

Les systèmes de reconnaissance numérique du son pour l'analyse respiratoire impliquent généralement plusieurs composants clés : l'acquisition du signal audio, le prétraitement, l'extraction de caractéristiques et la classification à l'aide d'algorithmes d'apprentissage automatique. Chaque composant présente des défis uniques et nécessite des connaissances spécialisées couvrant le traitement du signal, l'ingénierie des caractéristiques et l'apprentissage automatique [23].

Ce chapitre fournit un aperçu complet des concepts fondamentaux, des méthodologies et des outils essentiels à la reconnaissance numérique du son, avec un accent particulier sur les applications dans la détection des maladies respiratoires, notamment la COVID-19. Nous commençons par explorer les bases du traitement numérique du signal pour l'audio,

y compris les représentations dans les domaines temporel et fréquentiel. Nous examinons ensuite diverses méthodes d'extraction de caractéristiques qui transforment les signaux audio bruts en représentations significatives adaptées aux algorithmes d'apprentissage automatique. Ensuite, nous examinons les approches traditionnelles et d'apprentissage profond pour les tâches de classification audio. Enfin, nous discutons des outils logiciels et des bibliothèques couramment utilisés dans ce domaine, fournissant des informations pratiques pour la mise en œuvre.

En établissant une base solide dans ces concepts et outils, ce chapitre vise à fournir le contexte nécessaire à la compréhension des chapitres suivants, qui se concentreront sur les méthodologies spécifiques et les résultats expérimentaux de notre recherche sur la classification audio pour les maladies respiratoires.

2.2 Fondamentaux du Traitement Numérique du Signal pour l'Audio

Le Traitement Numérique du Signal (DSP) constitue le fondement des systèmes d'analyse audio, fournissant le cadre mathématique et computationnel nécessaire pour extraire des informations significatives des signaux sonores [24]. Cette section explore les concepts fondamentaux du DSP tels qu'ils s'appliquent à l'analyse des sons respiratoires, en particulier dans le contexte de la détection de la COVID-19.

2.2.1 Propriétés et Caractéristiques des Ondes Sonores

Le son est une onde mécanique qui se propage à travers un milieu en créant des variations de pression. Dans le contexte des sons respiratoires, ces ondes sont générées par le mouvement de l'air à travers les voies respiratoires, la vibration des cordes vocales et d'autres mécanismes physiologiques [25]. Les principales caractéristiques des ondes sonores comprennent :

- **Amplitude** : Représente la magnitude de la variation de pression, perçue comme l'intensité sonore. La COVID-19 peut affecter l'amplitude des sons respiratoires en raison de changements dans la dynamique du flux d'air [26].
- **Fréquence** : Mesurée en Hertz (Hz), la fréquence détermine la hauteur d'un son. Les sons respiratoires couvrent généralement des fréquences allant de 50 Hz à 2500 Hz, les conditions pathologiques modifiant souvent la distribution fréquentielle [27].
- **Phase** : Décrit la position d'un point dans un cycle d'onde, cruciale pour comprendre les formes d'onde complexes.
- **Timbre** : La qualité ou la "couleur" du son, déterminée par le contenu harmonique et l'enveloppe, qui peut être particulièrement informative pour distinguer les sons respiratoires sains des sons pathologiques [19].

2.2.2 Conversion Analogique-Numérique

Pour traiter les signaux sonores de manière computationnelle, les ondes sonores analogiques doivent être converties en représentations numériques par un processus appelé conversion analogique-numérique (CAN) [28]. Ce processus implique :

- **Échantillonnage** : Conversion d'un signal en temps continu en un signal en temps discret en prenant des mesures à intervalles réguliers.
- **Quantification** : Approximation des valeurs d'amplitude continues à un ensemble fini de valeurs discrètes.
- **Codage** : Représentation des valeurs quantifiées en format binaire pour le stockage et le traitement numériques.

La qualité de la représentation numérique est principalement déterminée par deux paramètres :

- **Taux d'échantillonnage** : Le nombre d'échantillons pris par seconde, mesuré en Hz. Pour l'analyse des sons respiratoires, des taux d'échantillonnage entre 8 kHz et 44,1 kHz sont couramment utilisés, les taux plus élevés capturant des informations plus détaillées [3].
- **Profondeur de bits** : Le nombre de bits utilisés pour représenter chaque échantillon, déterminant la plage dynamique. Les valeurs typiques varient de 16 à 24 bits pour les applications audio médicales [29].

2.2.3 Théorie de l'Échantillonnage et Théorème de Nyquist

Le théorème d'échantillonnage de Nyquist-Shannon est un principe fondamental du DSP qui établit le taux d'échantillonnage minimum requis pour représenter avec précision un signal [30]. Le théorème stipule que pour reconstruire parfaitement un signal à bande limitée, le taux d'échantillonnage doit être au moins deux fois la composante de fréquence la plus élevée du signal :

$$f_s \geq 2f_{max} \tag{2.1}$$

où f_s est le taux d'échantillonnage et f_{max} est la fréquence maximale du signal.

Pour les sons respiratoires, qui contiennent généralement des fréquences allant jusqu'à 2,5 kHz, un taux d'échantillonnage minimum de 5 kHz serait théoriquement suffisant. Cependant, en pratique, des taux d'échantillonnage plus élevés (par exemple, 8 kHz ou 16 kHz) sont souvent utilisés pour capturer des caractéristiques acoustiques subtiles qui pourraient être pertinentes pour la détection de la COVID-19 [10].

2.2.4 Représentations dans le Domaine Temporel et Fréquentiel

Les signaux audio peuvent être analysés dans deux domaines complémentaires :

— **Domaine Temporel** : Représente comment l'amplitude du signal change au fil du temps. L'analyse dans le domaine temporel est utile pour identifier des motifs temporels, tels que la durée de la toux ou des cycles respiratoires, qui peuvent être altérés chez les patients atteints de COVID-19 [22].
— **Domaine Fréquentiel** : Représente la distribution de l'énergie du signal à travers différentes fréquences. L'analyse dans le domaine fréquentiel est particulièrement précieuse pour l'analyse des sons respiratoires, car différentes conditions respiratoires, y compris la COVID-19, peuvent se manifester par des changements caractéristiques dans le spectre fréquentiel [21].

La transformation entre ces domaines est généralement réalisée à l'aide de transformées de Fourier, qui décomposent un signal en ses fréquences constitutives.

2.2.5 Techniques de Fenêtrage

En pratique, les signaux audio sont souvent analysés par segments courts ou "fenêtres" pour capturer les caractéristiques variables dans le temps [31]. Le fenêtrage implique :

— Diviser le signal en segments chevauchants ou non chevauchants.
— Appliquer une fonction de fenêtre (par exemple, Hamming, Hanning ou Blackman) à chaque segment pour réduire les fuites spectrales.
— Traiter chaque segment fenêtré indépendamment.

Le choix de la taille de la fenêtre représente un compromis entre la résolution temporelle et fréquentielle. Pour l'analyse des sons respiratoires dans la détection de la COVID-19, des tailles de fenêtre comprises entre 20 et 50 ms avec un chevauchement de 50% sont couramment utilisées, équilibrant la nécessité de capturer des événements transitoires avec l'exigence d'une résolution fréquentielle adéquate [23].

2.2.6 Transformées de Fourier et Analyse Spectrale

La transformée de Fourier est une opération mathématique qui décompose un signal en ses fréquences constitutives, permettant l'analyse spectrale [24]. Plusieurs variantes sont particulièrement pertinentes pour l'analyse des sons respiratoires :

— **Transformée de Fourier Discrète (DFT)** : Convertit une séquence finie d'échantillons équidistants en une séquence de même longueur de composantes fréquentielles équidistantes.
— **Transformée de Fourier Rapide (FFT)** : Un algorithme efficace pour calculer la DFT, réduisant la complexité computationnelle de $O(N^2)$ à $O(N \log N)$.
— **Transformée de Fourier à Court Terme (STFT)** : Applique la transformée de Fourier à de courtes fenêtres chevauchantes du signal, capturant comment le contenu fréquentiel change au fil du temps.

L'analyse spectrale des sons respiratoires a révélé des motifs distinctifs chez les patients atteints de COVID-19, y compris des changements dans la distribution de puissance à travers les bandes de fréquences et des structures harmoniques altérées dans les sons de toux [10, 21].

2.2.7 Filtrage Numérique

Les filtres numériques sont des outils essentiels pour le prétraitement des sons respiratoires, améliorant les caractéristiques pertinentes tout en supprimant le bruit [28]. Les types de filtres courants comprennent :

- **Filtres passe-bas** : Conservent les fréquences inférieures à une coupure, utiles pour supprimer le bruit haute fréquence.
- **Filtres passe-haut** : Conservent les fréquences supérieures à une coupure, utiles pour éliminer la dérive de la ligne de base.
- **Filtres passe-bande** : Conservent les fréquences dans une plage spécifiée, souvent utilisés pour se concentrer sur les bandes de fréquences les plus informatives pour la détection de la COVID-19 (généralement 100-2000 Hz) [3].
- **Filtres coupe-bande (notch)** : Suppriment des fréquences spécifiques, telles que les interférences du réseau électrique à 50/60 Hz.

La conception des filtres implique des considérations sur l'ordre du filtre, les fréquences de coupure et le type de filtre (par exemple, Butterworth, Chebyshev ou elliptique), le choix optimal dépendant des caractéristiques spécifiques des sons respiratoires analysés et de l'environnement d'enregistrement [27].

Les concepts fondamentaux du traitement numérique du signal abordés dans cette section fournissent la base nécessaire à l'extraction de caractéristiques significatives des sons respiratoires, qui est l'objet de la section suivante.

2.3 Caractéristiques Acoustiques Utilisées

L'extraction de caractéristiques est une étape critique dans les systèmes de classification audio, transformant les signaux audio bruts en représentations significatives qui capturent les caractéristiques essentielles du son tout en réduisant la dimensionnalité [32]. Cette section décrit les caractéristiques acoustiques spécifiques qui ont été extraites et utilisées dans le cadre de cette mémoire.

2.3.1 Caractéristiques dans le Domaine Fréquentiel

Les caractéristiques spectrales suivantes ont été utilisées pour capturer des informations sur la distribution de l'énergie du signal à travers différentes fréquences [33].

Centroïde Spectral

Le centroïde spectral représente le "centre de masse" du spectre, indiquant la fréquence autour de laquelle la majeure partie de l'énergie est concentrée :

$$\text{Centroïde} = \frac{\sum_{k=1}^{K} f(k) \cdot A(k)}{\sum_{k=1}^{K} A(k)} \tag{2.2}$$

où $f(k)$ est la fréquence de la k-ième bande de fréquence et $A(k)$ est son amplitude. Cette caractéristique reflète la "brillance" perçue du son. Des changements dans le centroïde spectral des sons de toux peuvent refléter des altérations dans les voies respiratoires [21].

Contraste Spectral

Le contraste spectral mesure la différence relative d'amplitude entre les pics et les vallées dans le spectre pour plusieurs bandes de fréquences. Il capture la distribution de l'énergie et peut aider à distinguer les sons texturés des sons tonals [33]. Des changements dans le contraste spectral peuvent indiquer des modifications dans la résonance et l'amortissement des voies respiratoires.

2.3.2 Caractéristiques Cepstrales

Les caractéristiques cepstrales sont dérivées en appliquant une transformée au spectre logarithmique d'un signal et sont efficaces pour représenter les caractéristiques spectrales liées à la perception humaine [34].

Coefficients Cepstraux à Fréquence Mel (MFCC)

Les MFCC sont des caractéristiques fondamentales dans l'analyse audio, imitant la réponse non linéaire de l'oreille humaine aux fréquences via l'échelle de Mel [32]. Leur calcul implique typiquement :

1. Calculer le spectre de puissance du signal (via STFT).
2. Appliquer un banc de filtres Mel au spectre de puissance.
3. Prendre le logarithme des énergies des bandes Mel.
4. Appliquer la Transformée Cosinus Discrète (DCT) pour obtenir les MFCC.

Les MFCC se sont avérés très efficaces pour capturer les caractéristiques spectrales pertinentes pour distinguer différents types de sons respiratoires [10, 22] et ont été un élément clé de notre approche.

2.3.3 Caractéristiques Temps-Fréquence

Les caractéristiques temps-fréquence représentent l'évolution du contenu spectral d'un signal au fil du temps [35].

Spectrogrammes Mel

Les spectrogrammes Mel sont des représentations temps-fréquence qui utilisent l'échelle de fréquence Mel, accordant plus de résolution aux basses fréquences conformément à la perception humaine [32]. Ils sont calculés en appliquant le banc de filtres Mel au spectre de puissance obtenu à partir de la STFT pour chaque trame temporelle. Les spectrogrammes Mel fournissent une représentation riche et perceptivement pertinente des caractéristiques spectrales variables dans le temps et ont été largement utilisés comme entrée pour les modèles d'apprentissage profond dans les tâches de classification audio [23]. Ils constituent une autre caractéristique essentielle utilisée dans cette étude.

Ces quatre types de caractéristiques – Centroïde Spectral, Contraste Spectral, MFCC et Spectrogrammes Mel – ont été sélectionnés et combinés dans nos expérimentations pour construire un modèle de classification robuste pour les sons de toux, comme détaillé au Chapitre **??**.

2.4 Approches de Classification Audio

Une fois les caractéristiques pertinentes extraites des signaux audio, l'étape suivante consiste à utiliser des algorithmes de classification pour attribuer une étiquette (par exemple, sain, COVID-19, autre maladie respiratoire) à chaque enregistrement. Cette section donne un aperçu des approches de classification traditionnelles et basées sur l'apprentissage profond couramment utilisées dans l'analyse audio [36, 37].

2.4.1 Méthodes Traditionnelles d'Apprentissage Automatique

Les méthodes traditionnelles d'apprentissage automatique reposent souvent sur des caractéristiques conçues manuellement (comme celles décrites dans la section précédente) comme entrée pour divers algorithmes de classification.

XGBoost

XGBoost (eXtreme Gradient Boosting) est une implémentation hautement efficace et évolutive du gradient boosting qui a gagné une popularité significative ces dernières années [38]. Il construit un ensemble de modèles de prédiction faibles, généralement des arbres de décision, de manière séquentielle où chaque nouveau modèle corrige les erreurs de ses prédécesseurs :

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathcal{F} \tag{2.3}$$

où $\hat{y}_i$ est la prédiction pour la i-ème instance, K est le nombre d'arbres, f_k représente le k-ème arbre, et $\mathcal{F}$ est l'espace de tous les arbres de régression possibles.

XGBoost optimise la fonction objectif régularisée suivante :

$$\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{2.4}$$

où l est une fonction de perte convexe différentiable qui mesure la différence entre la prédiction $\hat{y}_i$ et la cible y_i, et Ω est un terme de régularisation qui pénalise la complexité du modèle pour prévenir le surapprentissage :

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \|w\|^2 \tag{2.5}$$

où T est le nombre de feuilles dans l'arbre, w est le vecteur des poids des feuilles, et γ et λ sont des paramètres de régularisation.

XGBoost introduit plusieurs innovations clés qui contribuent à ses performances supérieures dans les tâches de classification des sons respiratoires [38] :

- Un nouvel algorithme conscient de la sparsité pour traiter efficacement les données éparses
- Une esquisse quantile pondérée théoriquement justifiée pour gérer les poids des instances
- Une structure de blocs consciente du cache pour l'apprentissage d'arbres hors mémoire
- Des capacités de calcul parallèle et distribué pour l'évolutivité

Dans le contexte de la classification des sons respiratoires, XGBoost a démontré des performances exceptionnelles pour distinguer les toux de COVID-19 des autres affections respiratoires, surpassant souvent les algorithmes traditionnels comme SVM et les Forêts Aléatoires [3, 39]. Sa capacité à gérer des caractéristiques acoustiques complexes et de haute dimension tout en maintenant une efficacité computationnelle le rend particulièrement adapté aux applications de détection en temps réel des maladies respiratoires [5].

Machines à Vecteurs de Support (SVM)

Les SVM sont des classificateurs puissants qui visent à trouver l'hyperplan optimal séparant les points de données appartenant à différentes classes dans un espace de caractéristiques de haute dimension [40]. Ils sont efficaces dans les espaces de haute dimension et peuvent gérer les relations non linéaires à l'aide de fonctions noyau.

Random Forests (RF)

Les forêts aléatoires sont des méthodes d'ensemble qui construisent plusieurs arbres de décision pendant l'entraînement et produisent la classe qui est le mode des classes (classification) ou la prédiction moyenne (régression) des arbres individuels [41]. Elles sont robustes au surapprentissage et peuvent gérer un grand nombre de caractéristiques.

K Plus Proches Voisins (KNN)

KNN est un classificateur non paramétrique basé sur l'instance qui classe un nouveau point de données en fonction de la classe majoritaire parmi ses k plus proches voisins dans l'espace des caractéristiques [42]. Sa simplicité est un avantage, mais il peut être coûteux en calcul pour de grands ensembles de données.

2.4.2 Méthodes d'Apprentissage Profond

Les méthodes d'apprentissage profond, en particulier les réseaux neuronaux, ont obtenu des résultats de pointe dans de nombreuses tâches de classification audio. Elles peuvent apprendre automatiquement des représentations hiérarchiques des caractéristiques directement à partir des données brutes ou de représentations temps-fréquence comme les spectrogrammes [43].

Réseaux Neuronaux Convolutifs (CNN)

Les CNN sont particulièrement bien adaptés au traitement de données de type grille, telles que les images ou les spectrogrammes [44]. Ils utilisent des couches convolutives pour apprendre des motifs locaux, suivies de couches de pooling pour réduire la dimensionnalité et de couches entièrement connectées pour la classification. Les CNN se sont avérés très efficaces pour la classification des sons environnementaux et l'analyse des sons respiratoires à l'aide d'entrées de spectrogrammes [23].

Réseaux Neuronaux Récurrents (RNN)

Les RNN sont conçus pour traiter des données séquentielles, ce qui les rend adaptés à l'analyse des signaux audio au fil du temps [45]. Cependant, les RNN simples peuvent avoir du mal à apprendre les dépendances à long terme [46]. Les variantes comme la Long Short-Term Memory (LSTM) [47] et les Gated Recurrent Units (GRU) résolvent ce problème et sont couramment utilisées pour la modélisation audio.

Transformers

Initialement développés pour le traitement du langage naturel, les Transformers ont également été appliqués avec succès aux tâches audio [48]. Ils utilisent des mécanismes

d'auto-attention pour pondérer l'importance des différentes parties de la séquence d'entrée, leur permettant de capturer efficacement les dépendances à longue portée.

2.4.3 Apprentissage par Transfert

L'apprentissage par transfert est une technique où un modèle pré-entraîné sur une tâche à grande échelle (par exemple, la classification d'images sur ImageNet ou la classification audio sur AudioSet) est affiné sur une tâche cible spécifique avec un ensemble de données plus petit [49]. Cela peut être particulièrement utile dans l'analyse des sons médicaux où les ensembles de données étiquetés sont souvent limités. Les modèles pré-entraînés peuvent fournir des extracteurs de caractéristiques puissants qui capturent des motifs audio généraux.

2.5 Outils et Bibliothèques pour l'Analyse Audio

La mise en œuvre de systèmes de classification audio pour la détection des maladies respiratoires nécessite des outils logiciels et des bibliothèques spécialisés. Cette section fournit un aperçu des outils et bibliothèques clés utilisés dans le pipeline d'analyse audio, de l'acquisition des données au déploiement du modèle, en mettant l'accent sur leur application à l'analyse des sons respiratoires liés à la COVID-19.

2.5.1 Écosystème Python pour le Traitement Audio

Python est devenu le langage de programmation dominant pour l'analyse audio et l'apprentissage automatique en raison de sa simplicité, de sa lisibilité et de son vaste écosystème de bibliothèques [50]. Les composants principaux de cet écosystème comprennent :

NumPy

NumPy fournit des opérations efficaces sur les tableaux et des capacités de calcul numérique, servant de base à la plupart des calculs scientifiques en Python [50].

Pour l'analyse des sons respiratoires, NumPy permet une manipulation efficace des signaux audio, la mise en œuvre d'algorithmes de traitement numérique du signal et des opérations numériques sur les caractéristiques extraites.

Pandas

Pandas fournit des structures de données et des fonctions pour une manipulation et une analyse efficaces des données [51].

Dans l'analyse des sons respiratoires liés à la COVID-19, Pandas facilite l'organisation des métadonnées audio, la gestion des caractéristiques et l'analyse statistique des caractéristiques extraites pour différents groupes de patients.

Matplotlib et Seaborn

Ces bibliothèques offrent des capacités de visualisation complètes pour explorer et présenter les données audio et les résultats d'analyse [52, 53].

2.5.2 Bibliothèques Audio Spécialisées

Plusieurs bibliothèques spécialisées ont été développées pour le traitement audio, offrant des interfaces de haut niveau pour les tâches courantes dans l'analyse des sons respiratoires.

Librosa

Librosa est une bibliothèque Python spécifiquement conçue pour l'analyse musicale et audio, offrant un ensemble complet de fonctions pour l'extraction de caractéristiques, la visualisation et la manipulation audio [54].

Librosa a été largement utilisé dans l'analyse des sons respiratoires liés à la COVID-19 en raison de ses capacités complètes d'extraction de caractéristiques, en particulier pour le calcul des MFCC, des caractéristiques spectrales et des représentations temps-fréquence [10, 22].

PyAudio

PyAudio fournit des liaisons Python pour PortAudio, une bibliothèque d'E/S audio multiplateforme, permettant l'enregistrement et la lecture audio en temps réel [55].

Pour les applications d'analyse des sons respiratoires liés à la COVID-19, PyAudio permet le développement de systèmes de surveillance en temps réel capables de capturer et d'analyser la toux, les schémas respiratoires et les caractéristiques vocales [3].

2.6 Évaluation des Modèles

L'évaluation rigoureuse des modèles est cruciale pour la classification des sons respiratoires, particulièrement dans les applications médicales où les faux positifs et les faux négatifs peuvent avoir des conséquences significatives [56].

2.6.1 Métriques d'Évaluation

Les métriques courantes pour évaluer les modèles de classification des sons respiratoires incluent :

— **Précision** : La proportion de prédictions positives correctes parmi toutes les prédictions positives

$$\text{Précision} = \frac{TP}{TP + FP} \tag{2.6}$$

— **Rappel (Sensibilité)** : La proportion de vrais positifs correctement identifiés

$$\text{Rappel} = \frac{TP}{TP + FN} \tag{2.7}$$

— **F1-score** : La moyenne harmonique de la précision et du rappel

$$\text{F1} = 2 \cdot \frac{\text{Précision} \cdot \text{Rappel}}{\text{Précision} + \text{Rappel}} \tag{2.8}$$

— **Spécificité** : La proportion de vrais négatifs correctement identifiés

$$\text{Spécificité} = \frac{TN}{TN + FP} \tag{2.9}$$

— **Aire sous la courbe ROC (AUC)** : Mesure la capacité du modèle à distinguer entre les classes

Dans le contexte de la détection de la COVID-19, un équilibre entre sensibilité et spécificité est crucial : une haute sensibilité est nécessaire pour identifier les cas positifs, tandis qu'une haute spécificité aide à éviter les faux positifs qui pourraient surcharger les systèmes de santé [21, 5].

2.6.2 Techniques de Validation

Les techniques de validation courantes pour les modèles de classification des sons respiratoires incluent :

— **Validation croisée k-fold** : Divise les données en k sous-ensembles, entraînant le modèle k fois avec différents ensembles de test

— **Validation croisée stratifiée** : Assure que chaque fold maintient la même distribution de classes que l'ensemble de données complet

— **Validation croisée leave-one-subject-out** : Particulièrement importante pour les applications médicales, où les données d'un sujet sont réservées pour le test à chaque itération

— **Validation sur des ensembles de données externes** : Teste le modèle sur des données collectées dans différents environnements ou populations pour évaluer la généralisation

La validation appropriée est cruciale pour éviter le surapprentissage et assurer que les modèles généraliseront bien à de nouvelles données, particulièrement dans des contextes cliniques [23, 6].

2.7 Défis et Directions Futures

Malgré les progrès significatifs, plusieurs défis persistent dans la classification audio des maladies respiratoires :

— **Variabilité des données** : Les sons respiratoires varient considérablement entre les individus et les conditions d'enregistrement, nécessitant des modèles robustes [3]

— **Données limitées** : L'acquisition de grands ensembles de données étiquetés reste difficile, particulièrement pour les maladies rares ou émergentes [14]

— **Biais et équité** : Assurer que les modèles fonctionnent équitablement à travers différentes populations démographiques [7]

— **Interprétabilité** : Développer des modèles explicables qui peuvent gagner la confiance des professionnels de la santé [57]

— **Déploiement en temps réel** : Optimiser les modèles pour une inférence efficace sur des appareils à ressources limitées [58]

Les directions futures prometteuses incluent :

— **Apprentissage auto-supervisé** : Exploiter de grandes quantités de données audio non étiquetées pour apprendre des représentations robustes [59]

— **Apprentissage fédéré** : Entraîner des modèles à travers des appareils distribués tout en préservant la confidentialité des données [60]

— **Modèles multimodaux** : Intégrer l'audio avec d'autres modalités comme le texte (descriptions de symptômes) et les données démographiques [7]

— **Modèles adaptatifs** : Développer des systèmes qui peuvent s'adapter à de nouvelles conditions ou maladies avec un minimum de données supplémentaires [5]

2.8 Applications de l'analyse des sons respiratoires dans la détection de la COVID-19

La COVID-19 a présenté des défis et opportunités uniques pour l'analyse des sons respiratoires. Cette section explore des applications spécifiques des techniques de reconnaissance sonore numérique dans le contexte de la détection et du suivi de la COVID-19.

2.8.1 Analyse des toux

La toux est l'un des symptômes les plus courants de la COVID-19, et ses caractéristiques acoustiques peuvent fournir des informations diagnostiques précieuses [10, 21]. L'analyse des toux comprend :

- **Détection des toux** : Identification des épisodes de toux dans des enregistrements audio continus.

- **Segmentation des toux** : Isolement des épisodes individuels de toux pour analyse détaillée.

- **Extraction de caractéristiques** : Calcul de caractéristiques acoustiques à partir des sons de toux, avec les MFCC, les caractéristiques spectrales et les coefficients d'ondelettes comme particulièrement informatifs.

- **Classification** : Distinction des toux liées à la COVID-19 d'autres types de toux et affections respiratoires.

Les études ont montré que les toux de la COVID-19 présentent des signatures acoustiques distinctes, y compris des modifications du contenu spectral, des motifs temporels et de la distribution d'énergie [10]. Ces différences sont attribuées aux effets pathophysiologiques spécifiques de la COVID-19 sur le système respiratoire, notamment l'inflammation des voies respiratoires et les altérations de la dynamique du flux d'air.

2.8.2 Reconnaissance des motifs respiratoires

Les motifs respiratoires peuvent également fournir des indications sur l'infection à la COVID-19 [22]. La reconnaissance des motifs respiratoires comprend :

- **Détection du cycle respiratoire** : Identification des phases d'inspiration et d'expiration.

- **Estimation de la fréquence respiratoire** : Calcul du nombre de respirations par minute.

- **Analyse de l'effort respiratoire** : Évaluation de l'effort nécessaire pour respirer, qui peut augmenter chez les patients atteints de la COVID-19 en raison de détresse respiratoire.

- **Analyse des sons respiratoires** : Examen des caractéristiques acoustiques des sons respiratoires, qui peuvent révéler des anomalies comme des crépitations ou des sifflements chez les patients atteints de la COVID-19.

La COVID-19 peut modifier les motifs respiratoires de diverses manières, y compris une augmentation de la fréquence respiratoire, des changements dans le rapport inspiration/expiation et la présence de sons respiratoires anormaux [23]. Ces modifications peuvent être détectées par l'analyse minutieuse d'enregistrements sonores respiratoires, permettant potentiellement une identification précoce de l'infection ou un suivi de l'évolution de la maladie.

2.8.3 Évaluation de la qualité vocale

La qualité vocale peut également être affectée par la COVID-19, à la fois par des effets directs sur le tractus vocal et des effets indirects comme la fatigue et la détresse respiratoire [22]. L'évaluation de la qualité vocale comprend :

- **Analyse des voyelles prolongées** : Examen de la stabilité, de l'harmonicité et des caractéristiques spectrales des sons de voyelles prolongées.

- **Analyse de la parole continue** : Analyse du rythme, de la prosodie et de l'articulation dans la parole continue.

- **Détection de la fatigue vocale** : Identification des signes de fatigue vocale, qui peuvent être plus prononcés chez les patients atteints de la COVID-19.

- **Extraction de biomarqueurs vocaux** : Dérivation de caractéristiques acoustiques spécifiques corrélées à l'infection par la COVID-19.

Les recherches ont montré que la COVID-19 peut affecter divers aspects de la qualité vocale, y compris la fréquence fondamentale, le jitter, le shimmer et le rapport harmonique-sur-bruit [23]. Ces modifications peuvent être subtiles mais détectables grâce à des techniques d'analyse audio sophistiquées, offrant une modalité supplémentaire pour le dépistage de la COVID-19.

2.8.4 Recherches actuelles et défis

Malgré des résultats prometteurs, l'analyse des sons respiratoires pour la détection de la COVID-19 fait face à plusieurs défis [21, 23] :

- **Pénurie de données** : Disponibilité limitée de données sonores respiratoires étiquetées pour la COVID-19, notamment provenant de populations diverses et de conditions d'enregistrement variées.

- **Facteurs de confusion** : Difficulté à distinguer la COVID-19 d'autres affections respiratoires avec des symptômes similaires.

- **Variabilité des enregistrements** : Variations dans les dispositifs, environnements et protocoles d'enregistrement affectant les caractéristiques acoustiques des sons respiratoires.

- **Généralisation** : Assurer que les modèles entraînés sur des jeux de données spécifiques généralisent bien à de nouvelles populations et conditions d'enregistrement.

- **Interprétabilité** : Développer des modèles qui fournissent non seulement des prédictions précises mais aussi des informations sur les marqueurs acoustiques de la COVID-19.

Les recherches en cours abordent ces défis à travers diverses approches, y compris l'augmentation de données, le transfert d'apprentissage, l'analyse multimodale (combinant l'audio avec d'autres types de données) et l'élaboration de protocoles d'enregistrement standardisés [10, 22].

2.8.5 Considérations éthiques

Le développement et le déploiement de systèmes d'analyse des sons respiratoires pour la détection de la COVID-19 soulèvent d'importantes questions éthiques [21, 23] :

- **Confidentialité** : Garantir que les données sonores respiratoires sont collectées, stockées et traitées de manière à protéger la vie privée individuelle.

- **Consentement** : Obtenir le consentement éclairé des individus dont les sons respiratoires sont enregistrés et analysés.

- **Équité** : Assurer que les systèmes d'analyse des sons respiratoires sont développés et validés à l'aide de jeux de données diversifiés pour éviter les biais et garantir des performances équitables entre différentes populations.

- **Transparence** : Communiquer clairement les capacités et limites des systèmes d'analyse des sons respiratoires aux utilisateurs et parties prenantes.

- **Responsabilité** : Définir des protocoles clairs pour la gestion des résultats positifs et assurer un suivi approprié.

Aborder ces considérations éthiques est essentiel pour un développement et un déploiement responsables des systèmes d'analyse des sons respiratoires pour la détection de la COVID-19.

2.9 Conclusion

Ce chapitre a fourni un aperçu complet des concepts et outils essentiels pour la reconnaissance sonore numérique, avec un focus spécifique sur les applications dans la détection des maladies respiratoires, en particulier la COVID-19. Nous avons exploré les bases du traitement numérique des signaux audio, diverses méthodes d'extraction de caractéristiques, les modèles d'apprentissage automatique pour la classification audio, et les outils logiciels et bibliothèques facilitant leur implémentation.

Le domaine de l'analyse des sons respiratoires pour la détection de la COVID-19 évolue rapidement, avec de nouvelles techniques, jeux de données et applications émergeant régulièrement. Bien que des défis persistent, les avantages potentiels sont significatifs, y compris le dépistage non invasif, la surveillance à distance et une meilleure compréhension des signatures acoustiques des maladies respiratoires.

Les orientations futures dans ce domaine incluent le développement d'approches multimodales combinant l'analyse des sons respiratoires avec d'autres types de données, l'application de techniques d'apprentissage profond avancées comme l'apprentissage auto-supervisé et l'apprentissage à faible volume de données pour pallier la pénurie de données, et l'intégration de l'analyse des sons respiratoires dans des systèmes de santé complets.

Les concepts et outils discutés dans ce chapitre constituent la base des chapitres suivants de cette mémoire, qui exploreront en profondeur des méthodologies spécifiques,

des résultats expérimentaux et des applications cliniques de la classification audio pour les maladies respiratoires.

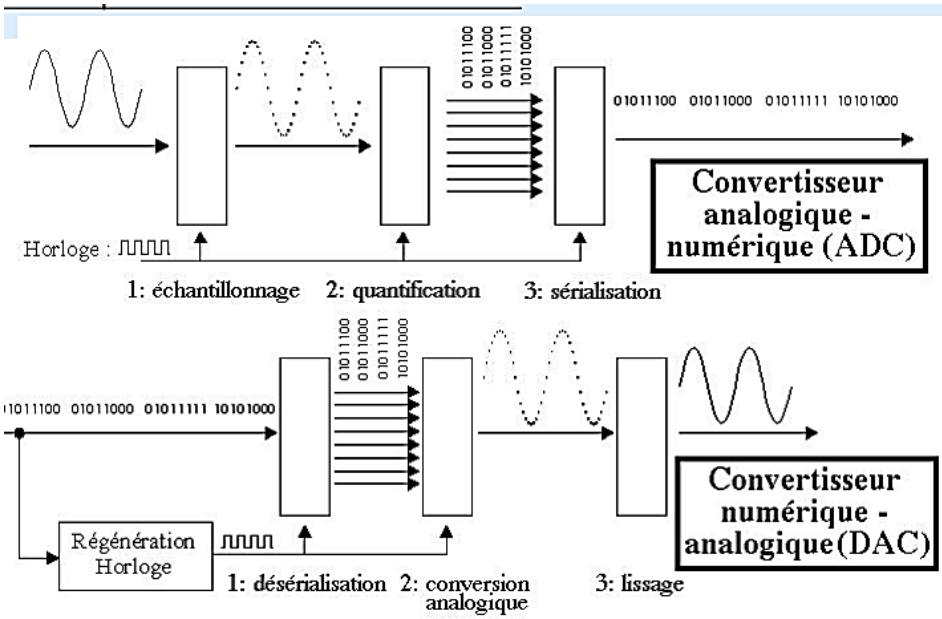


FIGURE 2.1 – Processus de conversion analogique-numérique (ADC) et numérique-analogique (DAC). La partie supérieure montre l'échantillonnage, la quantification et la sérialisation pour la conversion analogique-numérique. La partie inférieure illustre le processus inverse pour la conversion numérique-analogique.

Chapitre 3
État de l'Art

Chapitre 3

État De L'art

3.1 Introduction

L'automatisation de l'identification des maladies respiratoires par analyse audio est actuellement un champ d'étude en plein développement, tant dans le domaine médical que technologique. L'hétérogénéité et la complexité des affections respiratoires, associées à un nombre limité de spécialistes dans certaines zones géographiques, rendent le diagnostic classique ardu et parfois inaccessible. Dans ce contexte, les technologies d'intelligence artificielle, notamment le Machine Learning (ML) et le Deep Learning (DL, Apprentissage Profond), présentent un potentiel encourageant pour aider les professionnels de la santé à identifier rapidement et précisément ces maladies.

L'application de ces techniques à l'analyse d'enregistrements audio respiratoires favorise l'automatisation de la détection et de la classification des pathologies, tirant parti de caractéristiques discriminantes du son. Dans les publications scientifiques, on trouve de nombreuses méthodes, allant des techniques traditionnelles d'apprentissage supervisé, qui reposent sur la collecte manuelle de caractéristiques (telles que les coefficients cepstraux, les spectrogrammes et les paramètres temporels), aux réseaux neuronaux profonds qui ont la capacité d'assimiler automatiquement des représentations sophistiquées à partir d'importantes bases de données. On compte également des méthodes dites hybrides, qui fusionnent les atouts des deux paradigmes afin d'optimiser les performances en matière de classification.

3.1.1 Émergence des Diagnostics Basés sur l'Audio comme Frontière Prometteuse

Les signaux sonores respiratoires humains, spécifiquement la toux, les schémas respiratoires et les vocalisations, sont intrinsèquement liés à la fonction pulmonaire et peuvent manifester des signatures acoustiques distinctes indicatives de pathologies respiratoires sous-jacentes [3]. L'analyse de ces sons offre une avenue non invasive et hautement acces-

sible pour la détection des maladies et la surveillance continue. Cette approche exploite la connexion directe de ces signaux avec les poumons, fournissant une fenêtre unique sur la santé respiratoire.

L'adoption généralisée de la technologie des smartphones et la sophistication croissante des dispositifs portables ont encore amplifié le potentiel de collecte à grande échelle et à distance de ces sons physiologiques. Cette omniprésence technologique ouvre la voie au développement et au déploiement d'applications diagnostiques pouvant atteindre de larges populations, particulièrement dans les contextes à ressources limitées où l'infrastructure diagnostique traditionnelle peut être restreinte. La capacité à conduire des évaluations préliminaires et une stratification des risques à distance représente une avancée significative en santé publique.

3.2 Travaux connexes

Cette section propose une analyse approfondie des recherches récentes concernant la détection des maladies respiratoires par analyse audio, en se concentrant spécifiquement sur celles faisant appel à des enregistrements sonores. L'étude est organisée selon divers critères : les méthodes employées (apprentissage automatique, apprentissage profond ou mixtes), les bases de données utilisées, les procédures d'évaluation mises en œuvre, et aussi les performances mesurées en termes de précision, de rappel ou autres indicateurs de performance. Le but est de discerner les orientations contemporaines, les enjeux persistants et les apports notables dans ce secteur, pour situer notre propre projet dans un contexte méthodologique logique et défendable.

3.2.1 Approches Basées Sur Machine Learning (ML)

Andreadou et al. (2024) proposent une revue systématique des biomarqueurs audio dans les maladies respiratoires, identifiant trois types principaux : les sons de toux, les sons respiratoires/pulmonaires, et les sons vocaux [61]. Leur méthode repose sur l'évaluation comparative de plusieurs algorithmes, notamment les techniques d'extraction de caractéristiques et divers modèles de classification. Les auteurs utilisent des ensembles de données publics contenant des enregistrements audio de patients atteints de diverses maladies respiratoires, avec une validation croisée (k=10). Les performances des modèles sont mesurées à l'aide de métriques telles que l'accuracy, l'erreur absolue moyenne (MAE), l'erreur quadratique moyenne (MSE) et la racine de l'erreur quadratique moyenne (RMSE).

Sharma et al. (2021) développent des modèles basés sur des spectrogrammes pour la détection de la COVID-19 à partir de sons de toux [62]. Leur approche combine une sélection de caractéristiques avancée (méthode Feature Wiz) pour éliminer les corrélations

superflues dans les données audio, suivie d'une classification basée sur deux méthodes principales : un modèle d'ensemble empilé (stacking) et un réseau de neurones artificiels (ANN) optimisé. L'ensemble empilé intègre quatre algorithmes de base (Random Forest, Decision Tree, k-Nearest Neighbors, et SVM) avec une régression logistique comme méta-classifieur, tandis que l'ANN utilise une architecture séquentielle avec des couches denses, une fonction d'activation ReLU, et l'optimiseur Adam. Les auteurs évaluent leur modèle sur un jeu de données de plusieurs milliers d'instances (divisées en 70% pour l'entraînement et 30% pour le test), démontrant une performance supérieure à l'état de l'art avec une précision de 98.44%, un rappel de 98.68%, et un F1-score de 98.93% pour l'ensemble empilé, ainsi qu'une accuracy de 98.89% pour l'ANN optimisé après 200 epochs. Les résultats sont comparés favorablement à ceux d'autres études récentes, soulignant l'efficacité de cette approche hybride.

Vidhya et al. (2023) explorent l'extraction de caractéristiques audio pour la détection des maladies respiratoires [15]. Leur étude compare différentes techniques d'extraction de caractéristiques, notamment les coefficients cepstraux de Mel (MFCC), le taux de passage par zéro, le centroïde spectral, et d'autres paramètres statistiques. Les auteurs constatent que la combinaison de plusieurs types de caractéristiques améliore la précision de classification de 6.8% par rapport à l'utilisation exclusive de caractéristiques spectrales. Cette approche s'inscrit dans le domaine de l'apprentissage automatique traditionnel, avec une attention particulière portée à l'ingénierie des caractéristiques.

TABLE 3.1 – Comparaison des approches basées sur Machine Learning pour la détection des maladies respiratoires

Auteurs	**Méthodes**	**Dataset**	**Performances**
Andreadou et al. (2024) [61]	Évaluation comparative d'algorithmes ML	Datasets publics de sons respiratoires	Validation croisée (k=10), métriques : accuracy, MAE, MSE, RMSE
Sharma et al. (2021) [62]	Ensemble empilé (RF, DT, kNN, SVM) et ANN optimisé	Sons de toux (COVID-19)	Précision : 98.44%, Rappel : 98.68%, F1-score : 98.93%
Vidhya et al. (2023) [15]	Extraction de caractéristiques (MFCC, ZCR, centroïde spectral)	Sons respiratoires	Amélioration de 6.8% avec combinaison de caractéristiques

Ce tableau présente un aperçu comparatif de trois études récentes portant sur l'utilisation de méthodes d'intelligence artificielle pour l'analyse des sons respiratoires dans un contexte de diagnostic automatisé.

3.2.2 Approches Basées Sur Deep Learning (Apprentissage Profond)

Hussain et al. (2024) proposent Cough2COVID-19, un classificateur d'ensemble profond pour le diagnostic de la COVID-19 utilisant des signaux de toux [63]. Leur architecture multi-couche traite les enregistrements audio de toux et atteint une précision d'environ 98% (sensibilité 97%, spécificité 98%) pour distinguer les toux liées à la COVID-19. Le modèle utilise une combinaison de réseaux de neurones convolutifs (CNN) organisés en ensemble, ce qui améliore la robustesse et la fiabilité des prédictions.

Rajasekar et al. (2023) développent un réseau de capsules pour le diagnostic de la tuberculose à partir de sons de toux [64]. Leur approche innovante combine un réseau de capsules avec un réseau de neurones complètement connecté (FCNN) appliqué aux spectrogrammes de toux. Le modèle, entraîné sur environ 9 800 enregistrements de toux, atteint une précision de 97% avec une sensibilité de 98% pour la détection de la tuberculose. Cette étude démontre l'efficacité des architectures de réseaux de capsules dans le domaine de l'analyse audio pour le diagnostic des maladies respiratoires.

Alam et al. (2023) explorent la classification des sons pulmonaires en utilisant la transformée en ondelettes continues (CWT) et l'apprentissage profond [65]. Leur méthode applique la CWT pour mettre en évidence les crépitements et les sifflements dans les enregistrements audio pulmonaires, puis utilise un réseau de neurones profond pour la classification. Cette approche permet de capturer efficacement les caractéristiques transitoires des sons pulmonaires anormaux, améliorant ainsi la précision du diagnostic des maladies respiratoires comme la pneumonie et la BPCO.

Sharma et al. (2024) étudient l'application de modèles basés sur l'attention dans l'analyse des sons respiratoires [66]. Leur recherche se concentre sur l'utilisation de mécanismes d'attention pour améliorer la performance des réseaux de neurones dans la classification des sons respiratoires. Ces modèles permettent de se concentrer sur les parties les plus informatives des signaux audio, améliorant ainsi la précision du diagnostic. Cette approche représente une tendance émergente dans le domaine de l'analyse audio pour les applications médicales.

3.2.3 Approches Hybrides et Ensembles

Xu et al. (2025) développent un système d'apprentissage profond d'ensemble pour la détection de l'asthme à partir de la toux et de la respiration [67]. Leur approche utilise un ensemble de vote de cinq modèles CNN/ML sur des sons de toux et de respiration pour distinguer l'asthme des sujets sains, atteignant une précision de 94%. Cette méthode d'ensemble améliore la robustesse du système en combinant les prédictions de plusieurs modèles, réduisant ainsi l'impact des erreurs individuelles.

TABLE 3.2 – Comparaison des approches basées sur Deep Learning (Apprentissage Profond) pour la détection des maladies respiratoires

Auteurs	**Architecture**	**Application**	**Performances**
Hussain et al. (2024) [63]	Ensemble de CNN (Cough2COVID-19)	Diagnostic COVID-19	Précision : 98%, Sensibilité : 97%, Spécificité : 98%
Rajasekar et al. (2023) [64]	Réseau de capsules + FCNN	Diagnostic tuberculose	Précision : 97%, Sensibilité : 98%
Alam et al. (2023) [65]	CWT + DNN	Classification sons pulmonaires	Amélioration détection des crépitements et sifflements
Sharma et al. (2024) [66]	Modèles basés sur l'attention	Classification sons respiratoires	Focalisation sur segments informatifs

Sueaseenak et al. (2025) présentent un stéthoscope intelligent et un système de diagnostic basé sur le cloud [68]. Leur système enregistre les sons pulmonaires et les envoie à un modèle ML dans le cloud qui les classe en quatre catégories : normal, pneumonie, BPCO et autre maladie. Ce système atteint une précision d'environ 89%, démontrant le potentiel des applications pratiques de l'analyse audio pour le diagnostic des maladies respiratoires.

Jaggi et al. (2023) ont développé DeepBreath, un système qui utilise des classificateurs CNN entraînés sur 35,9 heures d'audio d'auscultation pulmonaire pédiatrique [69]. DeepBreath distingue les sujets sains des malades et classifie des maladies spécifiques (pneumonie, troubles sifflants, bronchiolite). Lors de la validation interne, il a atteint des AUROC d'environ 0,75 pour la pneumonie et d'environ 0,91 pour les sifflements. Il est crucial de noter que, DeepBreath a été validé en externe sur des données provenant de trois pays africains (Sénégal, Cameroun, Maroc) avec seulement des baisses modestes (par exemple, AUC de pneumonie 0,74). Il a même maintenu une AUROC >90% en utilisant uniquement des clips de 10 secondes provenant de 4 sites pulmonaires, au lieu de 30 secondes provenant de 8 sites.

3.3 Le Dataset COUGHVID

3.3.1 Origine, Échelle et Méthodologie de Collecte Initiale

Le dataset COUGHVID a été conçu comme une initiative ambitieuse de crowdsourcing visant à collecter rapidement des enregistrements de toux, principalement en réponse à la demande mondiale urgente d'outils de dépistage efficaces pour la COVID-19 [3]. Cet effort collaboratif a abouti à l'accumulation d'un corpus substantiel de données audio,

TABLE 3.3 – Comparaison des approches hybrides et d'ensemble pour la détection des maladies respiratoires

Auteurs	**Approche**	**Application**	**Performances**
Xu et al. (2025) [67]	Ensemble de vote (5 modèles CNN/ML)	Détection asthme	Précision : 94%
Sueaseenak et al. (2025) [68]	Stéthoscope intelligent + ML cloud	Classification 4 catégories	Précision : 89%
Jaggi et al. (2023) [69]	DeepBreath (CNN)	Classification pédiatrique	AUROC : 0.75 (pneumonie), 0.91 (sifflements)

avec des rapports indiquant plus de 20 000, 30 000, 25 000, voire 27 550 enregistrements de toux crowdsourcés selon différentes études. Cette variation dans les nombres rapportés reflète la croissance continue du dataset et les différents sous-ensembles utilisés dans les études.

La phase de collecte de données s'est étendue du 1er avril 2020 au 1er décembre 2020. Ce processus a été facilité par une application web dédiée hébergée par l'École Polytechnique Fédérale de Lausanne (EPFL) en Suisse [3]. Les participants étaient guidés pour enregistrer des sons de toux, généralement jusqu'à 10 secondes de durée, et avaient la possibilité de fournir des métadonnées associées. Ces métadonnées incluaient des informations démographiques cruciales telles que l'âge, le genre, la localisation géographique et leur statut COVID-19 auto-déclaré. Le paradigme du crowdsourcing s'est avéré très efficace pour accumuler rapidement un dataset diversifié, englobant une large gamme de démographies de sujets et de statuts COVID-19 variés, ce qui était essentiel pour entraîner des modèles d'apprentissage automatique pour un dépistage à grande échelle.

3.3.2 Évolution et Amélioration de la Qualité : Étiquetage par Experts et Approches Semi-Supervisées

Suite à la collecte initiale des données, le dataset COUGHVID a subi une étape de prétraitement cruciale impliquant un filtrage à travers un algorithme de détection de toux open-source pour assurer la présence de sons de toux [3]. Une entreprise significative pour améliorer l'utilité diagnostique du dataset a impliqué l'étiquetage expert de plus de 2 000 enregistrements par des pneumologues expérimentés. Ces professionnels médicaux ont diagnostiqué les anomalies médicales sous-jacentes présentes dans les toux, établissant ainsi l'un des plus grands datasets de toux étiquetés par des experts existants. Cette annotation par des experts a été une étape critique pour fournir une vérité terrain plus fiable pour les modèles d'apprentissage automatique.

Cependant, la nature inhérente du crowdsourcing présentait des défis, notamment le

potentiel d'étiquetage erroné et des désaccords observables entre experts. Pour atténuer ces incohérences et résoudre la rareté des données, un algorithme sophistiqué d'apprentissage semi-supervisé (SSL) a été introduit dans la troisième itération du dataset COUGHVID, qui incorporait des enregistrements obtenus jusqu'en octobre 2021. Cette approche SSL a ingénieusement combiné les étiquettes initialement fournies par les utilisateurs avec les diagnostics des médecins experts. Les étiquettes d'experts ont été modélisées à l'aide de techniques d'apprentissage automatique, puis étendues aux données précédemment non étiquetées, affinant efficacement la cohérence du dataset.

L'impact de ce processus de ré-étiquetage a été substantiel. Les données ré-étiquetées ont démontré une amélioration marquée de la cohérence, exhibant une augmentation triple de la séparabilité des classes par rapport aux données étiquetées par les utilisateurs d'origine. Cette amélioration a également amplifié les différences spectrales entre les toux saines et celles liées à la COVID-19, améliorant ainsi l'explicabilité acoustique du nouveau dataset. Le crowdsourcing initial rapide de COUGHVID était une réponse nécessaire aux besoins urgents de données pendant la pandémie. Cependant, les défis inhérents au crowdsourcing, tels que l'étiquetage potentiellement erroné et les désaccords entre experts, sont rapidement devenus apparents. L'investissement ultérieur dans l'étiquetage par des experts et, de façon encore plus déterminante, le développement et l'application de techniques d'apprentissage semi-supervisé pour ré-étiqueter et améliorer la cohérence, représente une avancée méthodologique significative. Ce processus illustre un changement de la simple collecte de données vers une curation et un raffinement actifs utilisant l'apprentissage automatique avancé pour surmonter ses limitations inhérentes.

3.3.3 Importance et Défis des Datasets Médicaux Crowdsourcés

Le dataset COUGHVID est sans équivoque devenu une ressource fondamentale pour l'entraînement de modèles d'apprentissage automatique visant à répondre aux crises sanitaires mondiales pressantes [3]. Son échelle substantielle et sa diversité inhérente le rendent inestimable pour le développement d'algorithmes robustes de classification audio de toux. Pour promouvoir davantage une validation rigoureuse et en aveugle des modèles de recherche, un dataset de test privé est mis à disposition, permettant aux chercheurs de soumettre leurs prédictions de modèle sur des enregistrements non étiquetés pour une évaluation objective des performances.

Malgré ses avantages indéniables, l'utilisation de datasets médicaux crowdsourcés comme COUGHVID présente un ensemble unique de défis. Ceux-ci incluent l'assurance d'une stricte confidentialité des données et l'adhésion à des considérations éthiques complexes pendant la collecte des données, ainsi que la protection contre l'utilisation abusive potentielle d'informations de santé sensibles. De plus, la variabilité inhérente à la qualité audio crowdsourcée et la diversité des environnements d'enregistrement introduisent des

complexités qui posent des défis significatifs pour atteindre une forte généralisation des modèles à travers différents scénarios du monde réel.

3.4 Présentation du jeu de données COUGHVID : Introduction, Collecte et Description

Les données ont été collectées via une application web de l'EPFL (Suisse) entre avril et décembre 2020, période couvrant les premières vagues pandémiques. La méthode simple ("un enregistrement, un clic") permettait aux utilisateurs de téléverser jusqu'à 10 secondes de toux. Un questionnaire optionnel recueillait des métadonnées (démographie, santé), mais les enregistrements étaient acceptés sans questionnaire pour maximiser la collecte. La géolocalisation était optionnelle. Des instructions sur la toux sûre étaient fournies. Le jeu de données final contient plus de 25 000 enregistrements (34 434 dans le fichier metadata_compiled.csv analysé), chacun avec un UUID et un horodatage. Les métadonnées clés, issues du questionnaire et d'annotations expertes, incluent démographie (âge, genre, avec des valeurs manquantes notables), localisation (pour un sous-ensemble), et état de santé auto-déclaré ('sain', 'symptomatique', 'COVID-19', conditions préexistantes, fièvre/douleurs). Crucialement, une revue par quatre médecins a fourni des étiquettes expertes pour un sous-ensemble (>2800 enregistrements) : qualité audio, type de toux (sèche/grasse), symptômes audibles (dyspnée, sifflement), et diagnostic clinique ('COVID-19', 'toux saine', infections, maladie obstructive). Ces étiquettes expertes [7] ajoutent une valeur clinique importante, malgré une couverture partielle dans le fichier fourni. Un score cough_detected (0-1), issu d'un modèle automatique, évalue la probabilité de présence d'une toux dans chaque enregistrement, servant de filtre de qualité initial.oratoire des Données (AED)

3.4.1 Démographie des Participants

Comprendre la composition démographique du jeu de données est crucial pour évaluer sa représentativité et ses biais potentiels. L'analyse des métadonnées disponibles révèle des aperçus sur l'âge et le genre des participants, bien que des données manquantes substantielles limitent les conclusions définitives.

L'âge des participants, lorsqu'il est fourni (19 396 sur 34 434 enregistrements), présente une large plage, de 1 à 102 ans, avec un âge moyen d'environ 36,8 ans et un écart-type de 14,8 ans. La distribution, visualisée dans la Figure 4.3, semble quelque peu biaisée vers les jeunes adultes, l'âge médian étant de 35 ans. Le 25e percentile est à 25 ans et le 75e percentile est à 46 ans, indiquant que la moitié des participants dont l'âge est rapporté se situe dans cette fourchette. La présence de données sur un large spectre d'âges est

précieuse pour étudier les variations liées à l'âge dans l'acoustique de la toux, mais la quantité significative de données d'âge manquantes (plus de 43%) doit être reconnue.

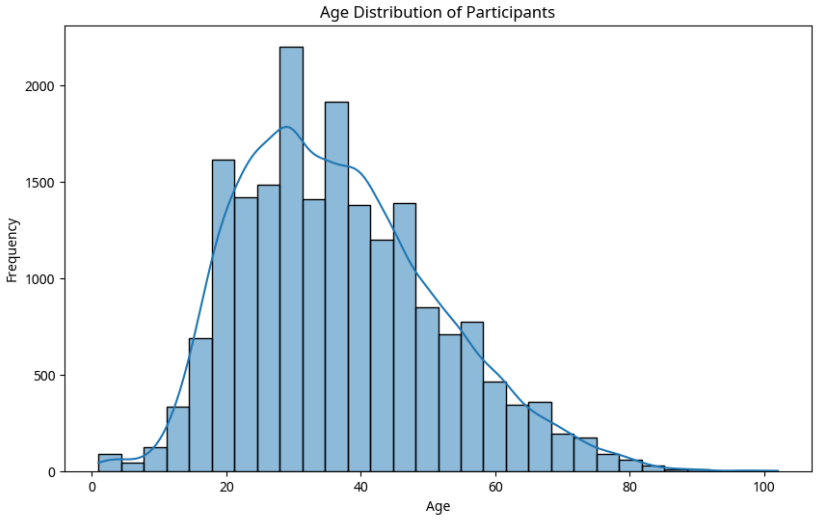


FIGURE 3.1 – Distribution de l'Âge des Participants.

Les informations sur le genre ont été fournies par 20 664 participants. Comme le montre la Figure 4.4, la majorité s'est identifiée comme masculine (12 850 participants), suivie par les femmes (7 682 participants). Un petit nombre (132 participants) a sélectionné 'autre'. Similairement à l'âge, une portion notable du jeu de données (13 770 enregistrements, soit 40%) manque d'informations sur le genre. Ce déséquilibre et les données manquantes pourraient influencer la généralisabilité des résultats liés au genre.

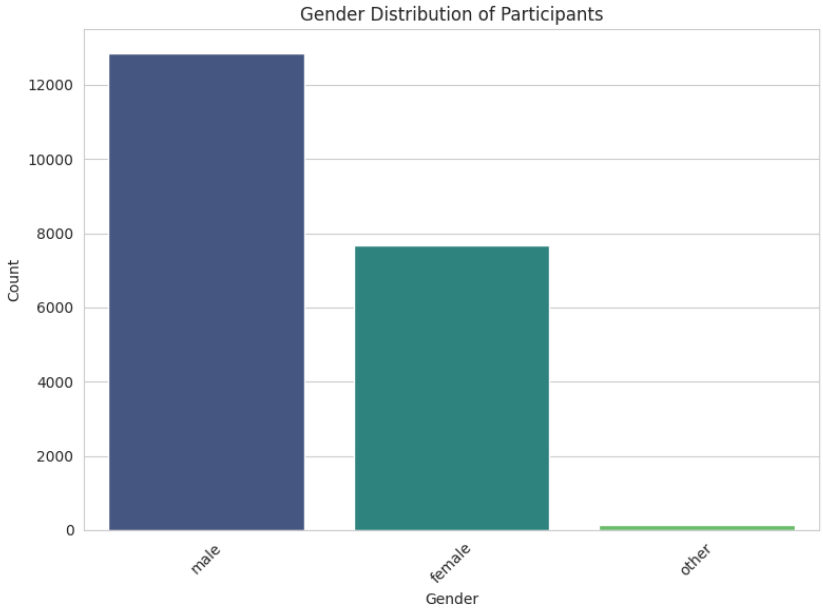


FIGURE 3.2 – Distribution du Genre des Participants.

3.4.2 Distribution Géographique

Le jeu de données bénéficie de sa nature crowdsourcée, capturant des enregistrements de divers lieux géographiques. Les données de latitude et longitude étaient disponibles pour 19 431 enregistrements. Un nuage de points visualisant ces emplacements (Figure 4.5) indique une distribution mondiale, avec des regroupements significatifs en Europe et en Amérique du Nord, mais aussi une représentation de l'Amérique du Sud, de l'Asie et d'autres régions. Cette diversité géographique est une force, augmentant potentiellement la robustesse des modèles entraînés sur ces données. Cependant, le graphique souligne également qu'environ 44% des enregistrements manquent de données de localisation.

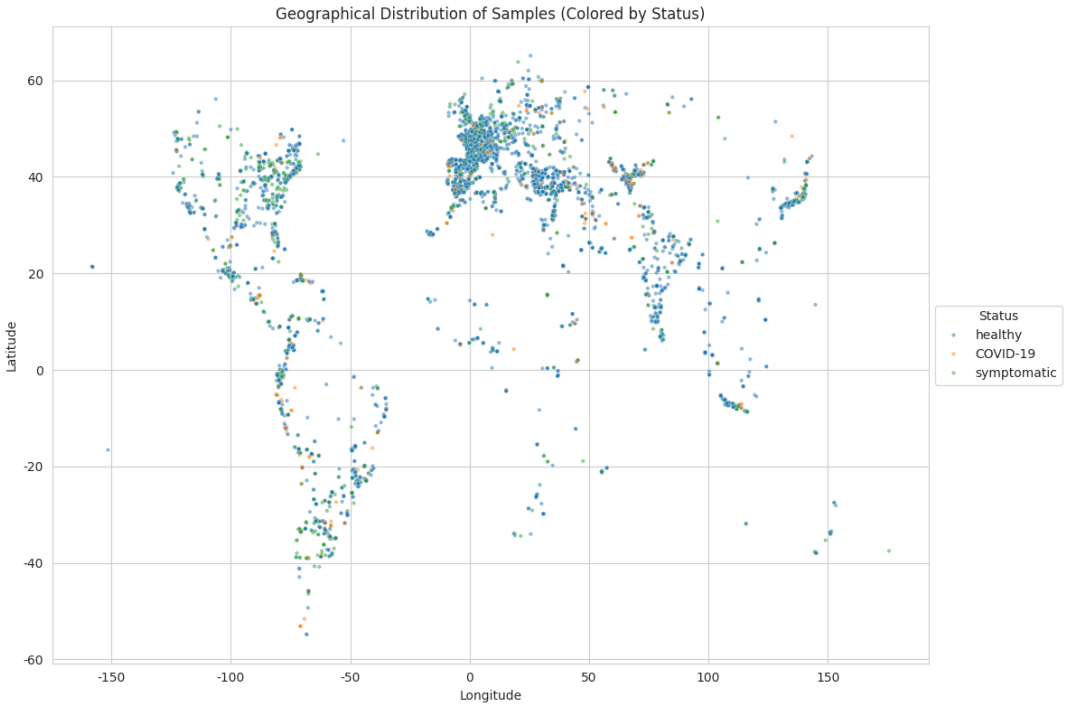


FIGURE 3.3 – Distribution Géographique des Échantillons (colorée par statut auto-déclaré).

3.5 Avancées en IA/ML pour la Classification des Maladies Respiratoires utilisant COUGHVID (2022-2025)

Cette section explore les méthodologies de pointe en IA/ML qui ont été appliquées au dataset COUGHVID, soulignant l'évolution des techniques et leur impact profond sur les performances diagnostiques pour les maladies respiratoires.

3.5.1 Architectures d'Apprentissage Profond (Deep Learning) : Repousser les Limites de Performance

Les modèles d'apprentissage profond ont été à l'avant-garde des avancées récentes dans la classification des maladies respiratoires basée sur l'audio, démontrant constamment des capacités supérieures pour extraire des motifs complexes à partir de signaux audio complexes.

Réseaux de Neurones Convolutifs (CNN) et leurs Variantes

Les CNN sont fondamentaux pour le traitement des spectrogrammes, qui sont des représentations visuelles, similaires à des images, des signaux audio. De nombreuses études ont exploité diverses architectures basées sur les CNN.

CovidCoughNet (septembre 2023) est une méthode d'apprentissage profond novatrice, spécifiquement conçue pour la détection de la COVID-19 à partir de signaux de toux, de respiration et de voix. Elle comprend deux composantes principales : InceptionFireNet pour l'extraction de caractéristiques profondes et DeepConvNet pour la prédiction. Cette architecture a atteint des performances remarquables sur le dataset COUGHVID3Cls (catégorisant les individus Sains, COVID-19 et Symptomatiques), rapportant une précision impressionnante de 99,19% et une Aire Sous la Courbe (AUC) de 98,44%. Cette performance souligne l'efficacité des architectures CNN personnalisées et spécialisées pour atteindre une classification fine.

Les CNN d'Ensemble (mars 2025) ont également été explorés, bien qu'un ensemble de sous-espace K-Nearest Neighbor (KNN) ait finalement atteint une précision plus élevée dans certaines études.

Le Réseau d'Apprentissage Profond Multi-branches (MBDLN), référencé dans une recherche de mai 2025, a été utilisé sur un dataset COUGHVID augmenté, combinant efficacement les Coefficients Cepstraux de Fréquence Mel (MFCC) et les spectrogrammes pour la détection de la COVID-19.

Réseaux de Neurones Récurrents (RNN) et Mémoire à Court et Long Terme (LSTM)

Les LSTM sont particulièrement adeptes pour capturer les dépendances temporelles inhérentes aux données séquentielles comme l'audio. Un modèle d'apprentissage profond multi-modal publié en janvier 2025, conçu pour la détection de la COVID-19, a intégré deux couches LSTM (avec 512 et 256 unités) suivant ses couches convolutives dans la composante basée sur le son. Ce choix de conception visait spécifiquement à démêler les motifs temporels complexes au sein des spectrogrammes Mel.

Modèles basés sur les Transformers

Les Transformers, initialement popularisés dans le traitement du langage naturel, sont de plus en plus adaptés pour l'analyse audio en raison de leurs puissants mécanismes d'auto-attention.

Le Transformer Hiérarchique de Spectrogramme (HST) (mars 2024) est une approche d'apprentissage profond innovante qui exploite l'auto-attention sur des fenêtres locales au sein des spectrogrammes Mel. Sa conception unique implique des tailles de fenêtres progressivement croissantes à travers les étapes du modèle pour capturer à la fois des informations contextuelles localisées et globales. Sur le dataset COUGHVID, HST a atteint une AUC de 0,90±0,01 pour discriminer entre les individus COVID-19 et non-COVID-19 présentant des symptômes. Un aspect notable de la conception de HST est son accent sur l'efficacité computationnelle, le rendant bien adapté pour le déploiement sur des appareils mobiles ou de périphérie.

Le Transformer Multi-modal Hiérarchique (HMT), référencé dans une recherche de mai 2025, a été entraîné à la fois sur les datasets COUGHVID et Coswara. En combinant les données de symptômes avec les caractéristiques de sons de toux, il a atteint une précision moyenne de 81,32% et une AUC de 82,06%.

Arbres de Décision Neuronaux Profonds (DNDT) et Forêts de Décision Neuronales Profondes (DNDF)

Ces techniques de pointe, publiées en janvier 2025 sur arXiv, offrent une perspective distincte sur l'exploitation de l'apprentissage automatique pour le diagnostic précoce. Elles ont démontré des performances constantes à travers divers datasets de sons de toux, y compris COUGHVID, atteignant une AUC de 0,92 sur le dataset COUGHVID spécifiquement et 0,97 sur un dataset combiné.

Réseau de Neurones Profond Amélioré (EDNN)

Un cadre avancé utilisant EDNN, précisément ajusté par l'Optimiseur d'Immunité Collective au Coronavirus (CHIO), a été proposé en janvier 2025 pour la détection de la COVID-19 à partir de signaux audio de toux. Ce modèle a présenté des améliorations substantielles dans les métriques d'erreur, réduisant l'Erreur Quadratique Moyenne (MSE) de 25,35%.

3.5.2 Techniques de Prétraitement et d'Extraction de Caractéristiques

Les avancées dans les techniques de prétraitement et d'extraction de caractéristiques ont joué un rôle crucial dans l'amélioration des performances des modèles de classification

des maladies respiratoires.

Représentations Temps-Fréquence

Les spectrogrammes Mel, qui représentent visuellement la distribution de l'énergie du signal à travers le temps et les fréquences, ont été largement utilisés comme entrée pour les modèles d'apprentissage profond. Ces représentations capturent efficacement les caractéristiques acoustiques des sons de toux, permettant aux modèles de détecter des motifs subtils indicatifs de maladies respiratoires.

Coefficients Cepstraux de Fréquence Mel (MFCC)

Les MFCC, qui représentent le spectre de puissance à court terme d'un son, ont été utilisés comme caractéristiques d'entrée pour divers modèles de classification. Ces coefficients sont particulièrement efficaces pour capturer les caractéristiques perceptuelles des sons de toux, facilitant la discrimination entre différentes conditions respiratoires.

Augmentation de Données

Pour surmonter les limitations des datasets de taille limitée, diverses techniques d'augmentation de données ont été employées, y compris l'ajout de bruit, le décalage temporel, et les variations de vitesse. Ces techniques améliorent la robustesse des modèles en les exposant à une plus grande variété de conditions d'enregistrement et de caractéristiques de toux.

3.5.3 Défis et Directions Futures

Malgré les progrès significatifs réalisés dans la classification des maladies respiratoires basée sur l'audio, plusieurs défis persistent et nécessitent une attention continue.

Généralisation à Travers les Populations et les Environnements

Un défi majeur est d'assurer que les modèles entraînés sur des datasets spécifiques comme COUGHVID peuvent généraliser efficacement à travers différentes populations et environnements d'enregistrement. Les variations dans les caractéristiques démographiques, les conditions comorbides, et les environnements acoustiques peuvent affecter significativement les performances des modèles.

Interprétabilité et Explicabilité

À mesure que les modèles d'apprentissage profond deviennent plus complexes, l'interprétabilité et l'explicabilité deviennent de plus en plus importantes, particulièrement

dans les applications médicales. Les recherches futures devraient se concentrer sur le développement de techniques pour rendre les décisions des modèles plus transparentes et compréhensibles pour les professionnels de la santé.

Intégration avec d'Autres Modalités

L'intégration de l'analyse audio avec d'autres modalités, telles que les données cliniques, les biomarqueurs, et l'imagerie médicale, représente une direction prometteuse pour améliorer la précision diagnostique. Les approches multi-modales peuvent fournir une vue plus complète de l'état de santé du patient, conduisant à des diagnostics plus précis et personnalisés.

3.6 Conclusion

L'état de l'art dans la classification audio pour les maladies respiratoires a connu des avancées significatives ces dernières années, en grande partie grâce à la disponibilité de datasets à grande échelle comme COUGHVID et aux progrès dans les techniques d'apprentissage automatique et profond. Les modèles actuels démontrent des performances impressionnantes dans la détection de diverses conditions respiratoires, y compris la COVID-19, à partir de signaux audio.

Cependant, plusieurs défis persistent, notamment en ce qui concerne la généralisation des modèles, l'interprétabilité, et l'intégration avec d'autres modalités. Aborder ces défis nécessitera des efforts continus de la part de la communauté de recherche, ainsi que des collaborations entre les experts en apprentissage automatique, les professionnels de la santé, et les ingénieurs en traitement du signal.

À mesure que le domaine continue d'évoluer, nous pouvons anticiper le développement de systèmes de diagnostic audio plus précis, robustes et accessibles, qui ont le potentiel de transformer la détection et la gestion des maladies respiratoires à l'échelle mondiale.

Chapitre 4
Exploration, Nettoyage et Préparation des Données Acoustiques

Chapitre 4

Exploration, Nettoyage et Préparation des Données Acoustiques

Ce chapitre présente en détail l'approche méthodologique adoptée pour développer notre système de classification de la toux. Le processus débute par une présentation approfondie du jeu de données COUGHVID, incluant sa collecte et sa description initiale. Nous procédons ensuite à une analyse exploratoire exhaustive des données (AED) afin de caractériser les participants, la distribution géographique, les états de santé rapportés et les caractéristiques audio préliminaires, tout en identifiant les défis liés aux données manquantes. S'appuyant sur les diagnostics cliniques fournis par des médecins experts pour un sous-ensemble clé des enregistrements – notre vérité fondamentale – nous décrivons les étapes de prétraitement des données audio, suivies par l'extraction de caractéristiques acoustiques pertinentes, telles que les MFCC et les spectrogrammes mel. Finalement, le chapitre détaille la construction, l'entraînement et l'évaluation des modèles d'apprentissage automatique conçus pour classifier les types de toux et identifier les affections respiratoires associées, en se basant sur ce fondement cliniquement validé.

4.1 Présentation du jeu de données COUGHVID : Introduction, Collecte et Description

Les données ont été collectées via une application web de l'EPFL (Suisse) entre avril et décembre 2020, période couvrant les premières vagues pandémiques. La méthode simple ("un enregistrement, un clic") permettait aux utilisateurs de téléverser jusqu'à 10 secondes de toux. Un questionnaire optionnel recueillait des métadonnées (démographie, santé), mais les enregistrements étaient acceptés sans questionnaire pour maximiser la collecte. La géolocalisation était optionnelle. Des instructions sur la toux sûre étaient fournies.

Le jeu de données final contient plus de 25 000 enregistrements (34 434 dans le fichier `metadata_compiled.csv` analysé), chacun avec un UUID et un horodatage. Les métadonnées clés, issues du questionnaire et d'annotations expertes, incluent démographie (âge, genre, avec des valeurs manquantes notables), localisation (pour un sous-ensemble), et état de santé auto-déclaré ('sain', 'symptomatique', 'COVID-19', conditions préexistantes, fièvre/douleurs). Crucialement, une revue par quatre médecins a fourni des étiquettes expertes pour un sous-ensemble (>2800 enregistrements) : qualité audio, type de toux (sèche/grasse), symptômes audibles (dyspnée, sifflement), et diagnostic clinique ('COVID-19', 'toux saine', infections, maladie obstructive). Ces étiquettes expertes [3] ajoutent une valeur clinique importante, malgré une couverture partielle dans le fichier fourni. Un score `cough_detected` (0-1), issu d'un modèle automatique, évalue la probabilité de présence d'une toux dans chaque enregistrement, servant de filtre de qualité initial.oratoire des Données (AED)

Pour obtenir une compréhension complète du jeu de données COUGHVID tel que représenté dans le fichier `metadata_compiled.csv` fourni (contenant 34 434 enregistrements), une analyse exploratoire des données (AED) a été menée. Cette analyse visait à caractériser les données démographiques des participants, la répartition géographique, la distribution de l'état de santé, le sous-ensemble étiqueté par des experts, les caractéristiques audio et l'étendue des données manquantes. Les résultats de cette AED éclairent les choix méthodologiques ultérieurs et l'interprétation des résultats au sein de cette mémoire.

4.1.1 Analyse de l'État de Santé

L'état de santé auto-déclaré fournit une catégorisation primaire des enregistrements. Sur les 20 664 participants ayant fourni cette information, la majorité s'est déclarée 'saine' (15 476 participants). Un nombre significatif s'est déclaré 'symptomatique' (3 873 participants), et 1 315 participants ont explicitement déclaré un statut 'COVID-19'. La distribution est illustrée dans la Figure 4.6. Cette distribution permet une analyse comparative entre différents états de santé, ce qui est central à l'objectif de détection de la COVID-19 à partir de la toux. De plus, 3 557 participants ont signalé avoir une condition respiratoire préexistante, tandis que 2 485 ont signalé ressentir de la fièvre ou des douleurs musculaires, fournissant des dimensions supplémentaires pour l'analyse, bien que ces champs souffrent également du même niveau de données manquantes que le genre et le statut (40%).

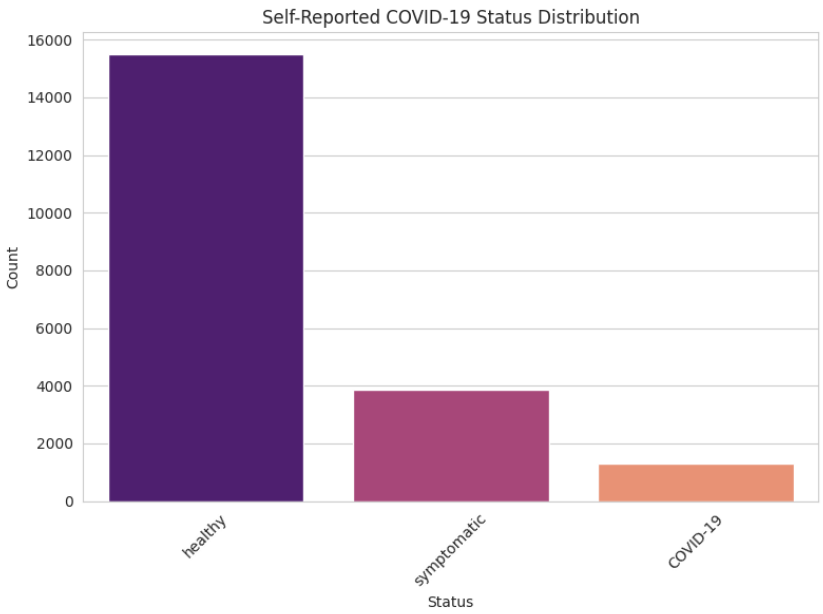


FIGURE 4.1 – Distribution de l'État COVID-19 Auto-Déclaré.

4.1.2 Caractéristiques Audio

Le score `cough_detected`, fourni pour tous les enregistrements, reflète la sortie d'un modèle de détection automatique de toux utilisé par les créateurs du jeu de données. La distribution de ces scores est montrée dans la Figure 4.7. Les scores varient de 0 à près de 1. Un nombre significatif d'enregistrements a des scores élevés (proches de 1), suggérant la confiance du modèle de détection quant à la présence d'une toux. Cependant, il y a aussi un nombre non négligeable d'enregistrements avec des scores plus bas. Le score moyen est de 0,64. Ce score peut potentiellement être utilisé comme filtre pour sélectionner des enregistrements de toux de meilleure qualité pour l'analyse, bien que le seuil nécessite une considération attentive.

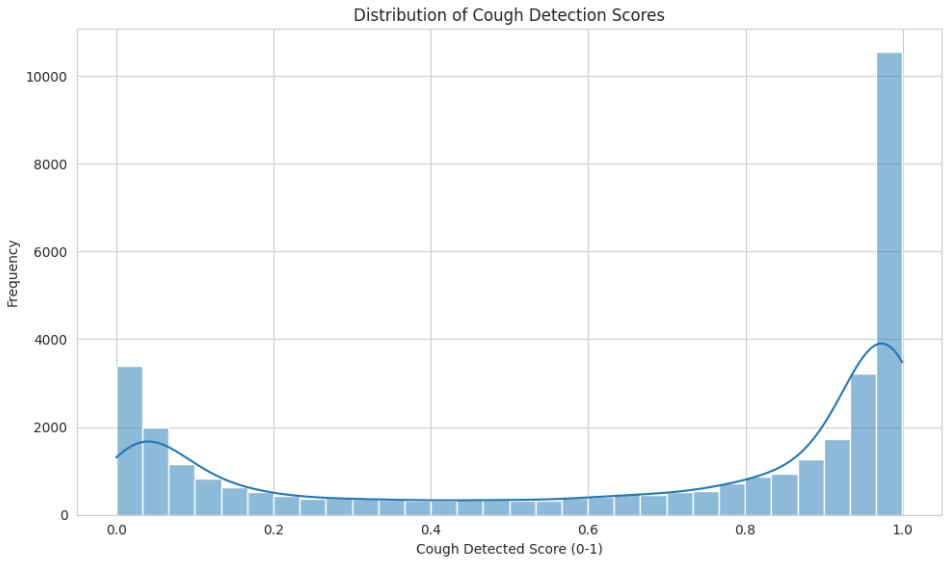


FIGURE 4.2 – Distribution des Scores de Détection de Toux.

4.2 Analyse Exploratoire des Données (AED)

Pour obtenir une compréhension complète du jeu de données COUGHVID tel que représenté dans le fichier `metadata_compiled.csv` fourni (contenant 34 434 enregistrements), une analyse exploratoire des données (AED) a été menée. Cette analyse visait à caractériser les données démographiques des participants, la répartition géographique, la distribution de l'état de santé, le sous-ensemble étiqueté par des experts, les caractéristiques audio et l'étendue des données manquantes. Les résultats de cette AED éclairent les choix méthodologiques ultérieurs et l'interprétation des résultats au sein de cette mémoire.

4.2.1 Démographie des Participants

Comprendre la composition démographique du jeu de données est crucial pour évaluer sa représentativité et ses biais potentiels. L'analyse des métadonnées disponibles révèle des aperçus sur l'âge et le genre des participants, bien que des données manquantes substantielles limitent les conclusions définitives.

L'âge des participants, lorsqu'il est fourni (19 396 sur 34 434 enregistrements), présente une large plage, de 1 à 102 ans, avec un âge moyen d'environ 36,8 ans et un écart-type de 14,8 ans. La distribution, visualisée dans la Figure 4.3, semble quelque peu biaisée vers les jeunes adultes, l'âge médian étant de 35 ans. Le 25e percentile est à 25 ans et le 75e percentile est à 46 ans, indiquant que la moitié des participants dont l'âge est rapporté se situe dans cette fourchette. La présence de données sur un large spectre d'âges est précieuse pour étudier les variations liées à l'âge dans l'acoustique de la toux, mais la

quantité significative de données d'âge manquantes (plus de 43%) doit être reconnue.

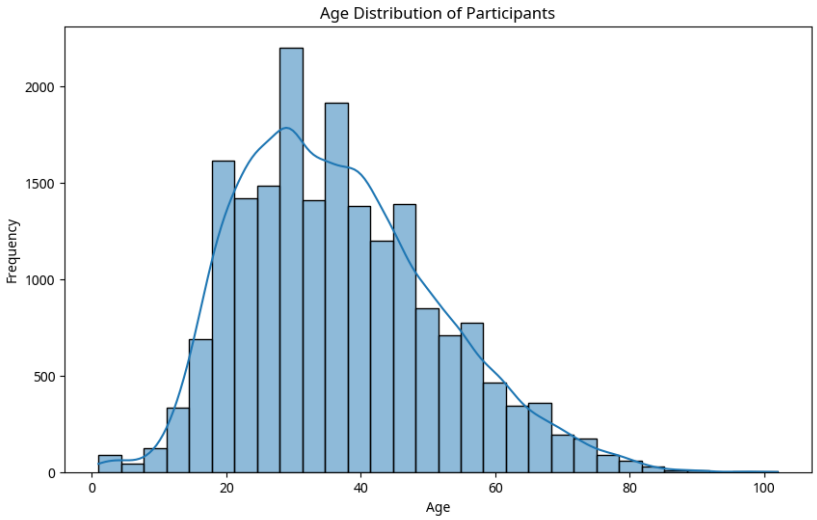


FIGURE 4.3 – Distribution de l'Âge des Participants.

Les informations sur le genre ont été fournies par 20 664 participants. Comme le montre la Figure 4.4, la majorité s'est identifiée comme masculine (12 850 participants), suivie par les femmes (7 682 participants). Un petit nombre (132 participants) a sélectionné 'autre'. Similairement à l'âge, une portion notable du jeu de données (13 770 enregistrements, soit 40%) manque d'informations sur le genre. Ce déséquilibre et les données manquantes pourraient influencer la généralisabilité des résultats liés au genre.

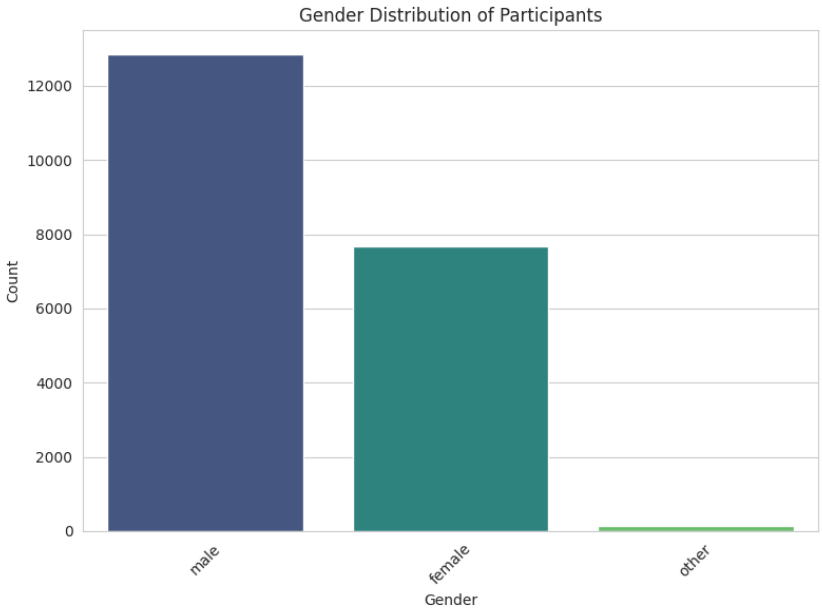


FIGURE 4.4 – Distribution du Genre des Participants.

4.2.2 Distribution Géographique

Le jeu de données bénéficie de sa nature crowdsourcée, capturant des enregistrements de divers lieux géographiques. Les données de latitude et longitude étaient disponibles pour 19 431 enregistrements. Un nuage de points visualisant ces emplacements (Figure 4.5) indique une distribution mondiale, avec des regroupements significatifs en Europe et en Amérique du Nord, mais aussi une représentation de l'Amérique du Sud, de l'Asie et d'autres régions. Cette diversité géographique est une force, augmentant potentiellement la robustesse des modèles entraînés sur ces données. Cependant, le graphique souligne également qu'environ 44% des enregistrements manquent de données de localisation.

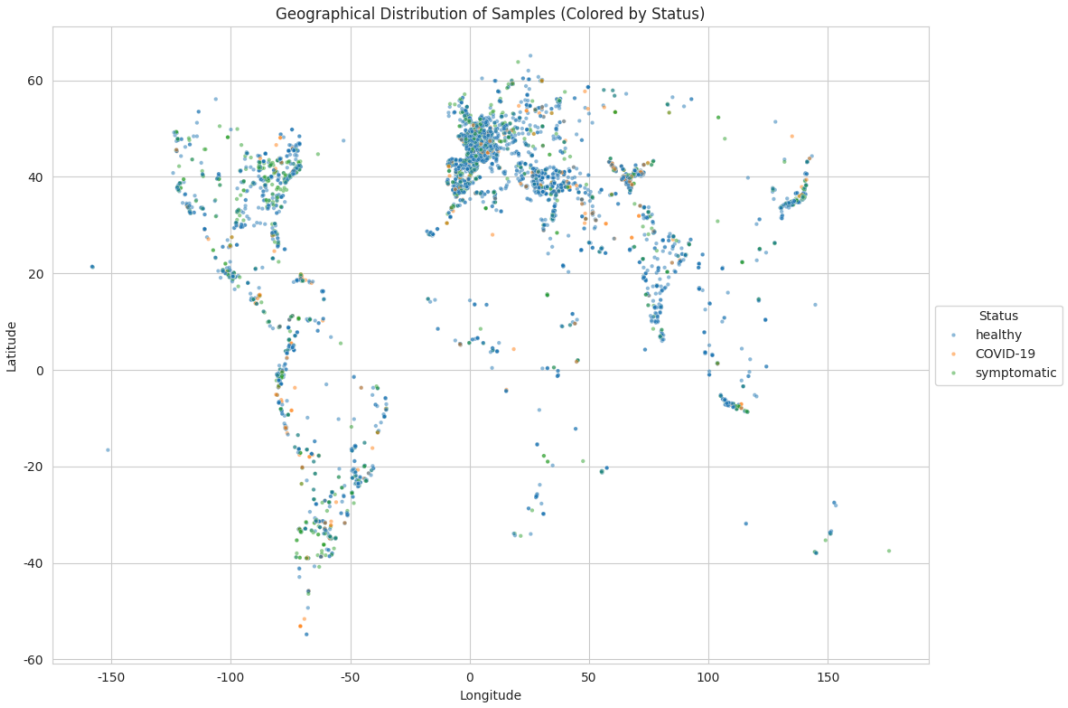


FIGURE 4.5 – Distribution Géographique des Échantillons (colorée par statut auto-déclaré).

4.2.3 Analyse de l'État de Santé

L'état de santé auto-déclaré fournit une catégorisation primaire des enregistrements. Sur les 20 664 participants ayant fourni cette information, la majorité s'est déclarée 'saine' (15 476 participants). Un nombre significatif s'est déclaré 'symptomatique' (3 873 participants), et 1 315 participants ont explicitement déclaré un statut 'COVID-19'. La distribution est illustrée dans la Figure 4.6. Cette distribution permet une analyse comparative entre différents états de santé, ce qui est central à l'objectif de détection de la COVID-19 à partir de la toux. De plus, 3 557 participants ont signalé avoir une condition respiratoire préexistante, tandis que 2 485 ont signalé ressentir de la fièvre ou des douleurs musculaires, fournissant des dimensions supplémentaires pour l'analyse, bien

que ces champs souffrent également du même niveau de données manquantes que le genre et le statut (40%).

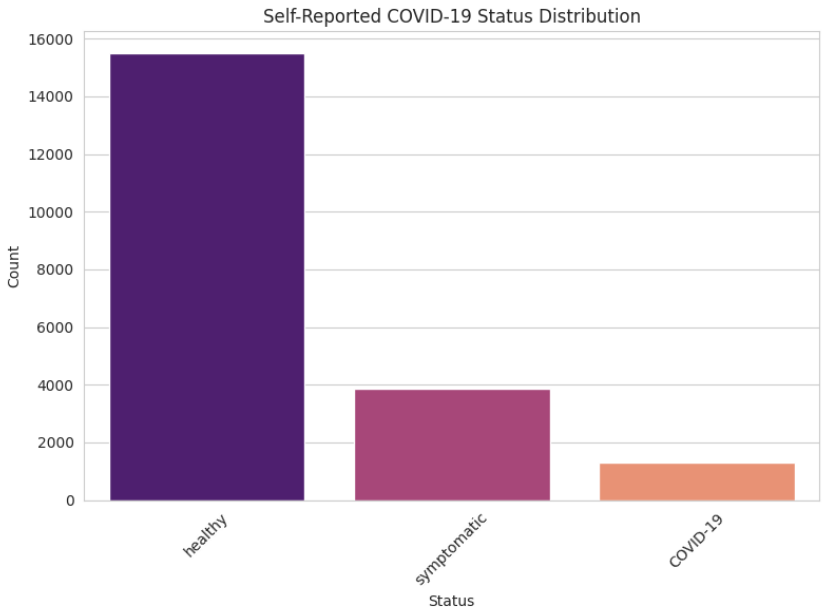


FIGURE 4.6 – Distribution de l'État COVID-19 Auto-Déclaré.

4.2.4 Caractéristiques Audio

Le score `cough_detected`, fourni pour tous les enregistrements, reflète la sortie d'un modèle de détection automatique de toux utilisé par les créateurs du jeu de données. La distribution de ces scores est montrée dans la Figure 4.7. Les scores varient de 0 à près de 1. Un nombre significatif d'enregistrements a des scores élevés (proches de 1), suggérant la confiance du modèle de détection quant à la présence d'une toux. Cependant, il y a aussi un nombre non négligeable d'enregistrements avec des scores plus bas. Le score moyen est de 0,64. Ce score peut potentiellement être utilisé comme filtre pour sélectionner des enregistrements de toux de meilleure qualité pour l'analyse, bien que le seuil nécessite une considération attentive.

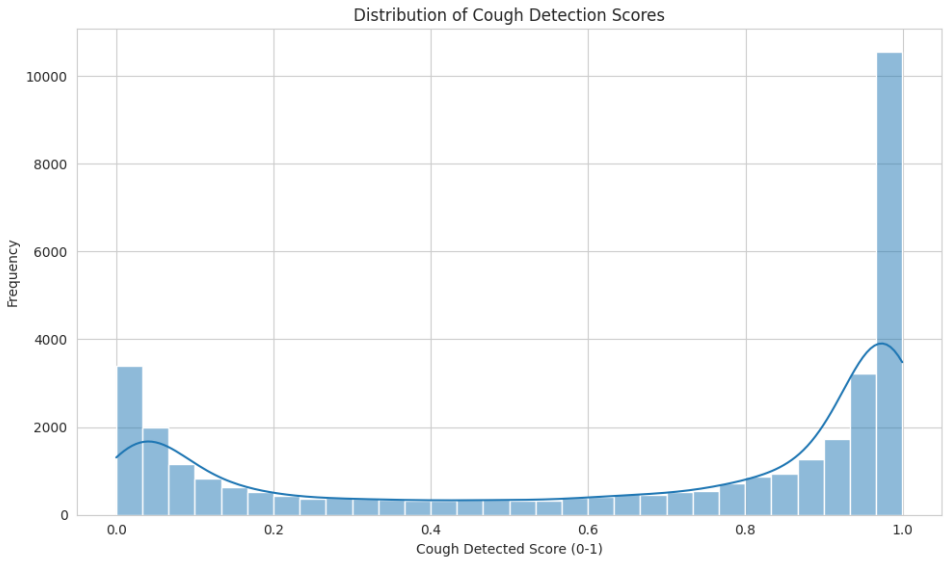


FIGURE 4.7 – Distribution des Scores de Détection de Toux.

4.3 Analyse Exploratoire des Données (AED) sur les Étiquettes Expertes

Cette section présente une analyse exploratoire des données (AED) axée spécifiquement sur le sous-ensemble d'enregistrements qui ont été annotés par des médecins experts, en se basant sur les décomptes fournis. L'objectif est de caractériser la distribution des différentes étiquettes cliniques attribuées par ces experts, en s'appuyant sur le contexte fourni par Orlandic et al. (2021) [3].

Les données agrégées fournies concernent un total de 2841 enregistrements examinés (bien que les totaux par variable varient légèrement, suggérant potentiellement des annotations incomplètes pour certains enregistrements ou des différences dans les sous-ensembles analysés pour chaque variable). L'analyse suivante détaille la répartition pour chaque catégorie d'étiquette experte.

4.3.1 Qualité des Enregistrements

La qualité audio, telle qu'évaluée par les experts, est cruciale pour la fiabilité des analyses acoustiques ultérieures. La distribution de cette étiquette est présentée dans la Figure 4.8.

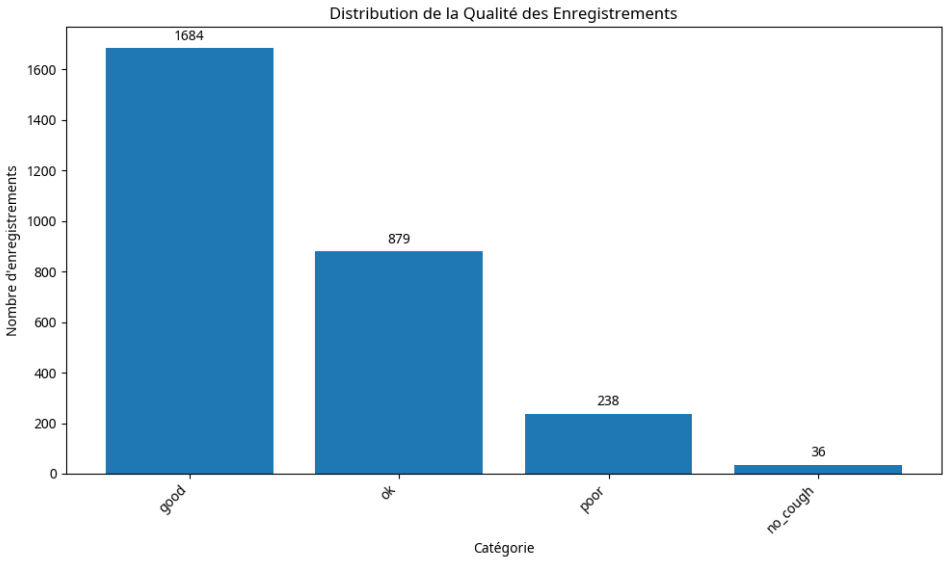


FIGURE 4.8 – Distribution de la Qualité des Enregistrements (Étiquettes Expertes).

Comme le montre la figure, la majorité des enregistrements annotés (1684 'good' + 879 'ok' = 2563 sur 2837, soit environ 90.3%) sont jugés de qualité suffisante ('bonne' ou 'ok'). Une faible proportion (238, 8.4%) est jugée 'médiocre' ('poor'), et très peu (36, 1.3%) sont identifiés comme ne contenant 'pas de toux' ('no_cough'). Cela suggère une bonne qualité générale du sous-ensemble annoté.

4.3.2 Type de Toux

La distinction entre toux sèche et toux grasse (productive) est cliniquement pertinente. La Figure 4.9 illustre la répartition selon les experts.

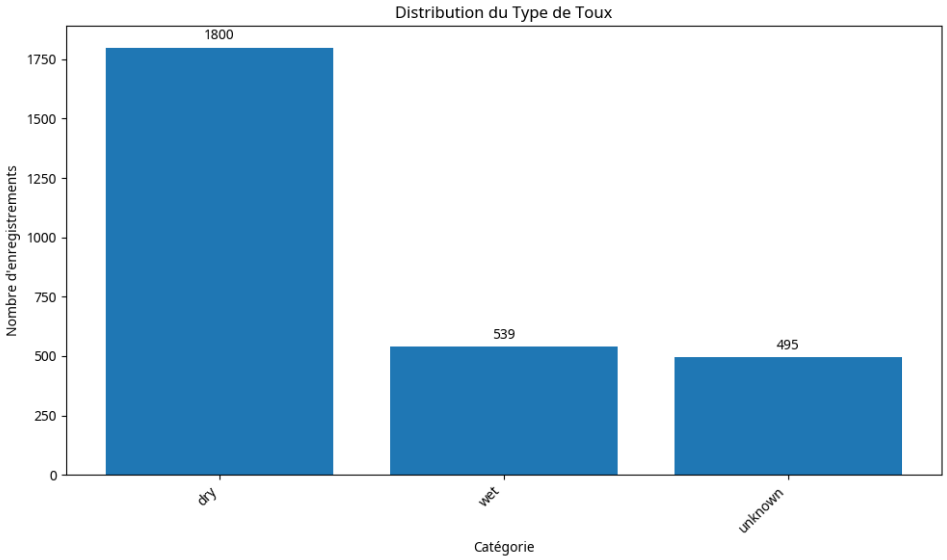


FIGURE 4.9 – Distribution du Type de Toux (Étiquettes Expertes).

La toux sèche ('dry') prédomine nettement (1800 sur 2834, 63.5%), suivie par la toux grasse ('wet', 539, 19.0%). Une part significative (495, 17.5%) est classée comme 'inconnue' ('unknown'), soulignant la difficulté potentielle de cette classification basée uniquement sur l'audio.

4.3.3 Symptômes Audibles

Les experts ont également indiqué la présence de divers symptômes audibles. Les Figures 4.10 à 4.15 montrent la prévalence de chacun.

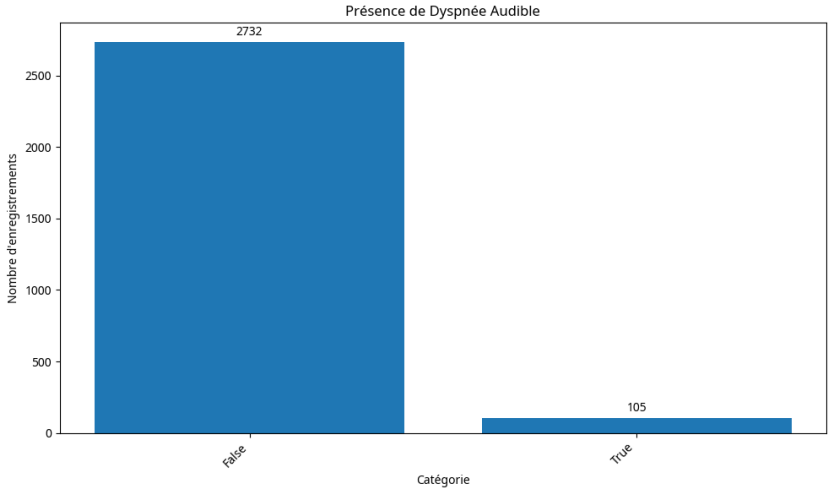


FIGURE 4.10 – Présence de Dyspnée Audible (Étiquettes Expertes).

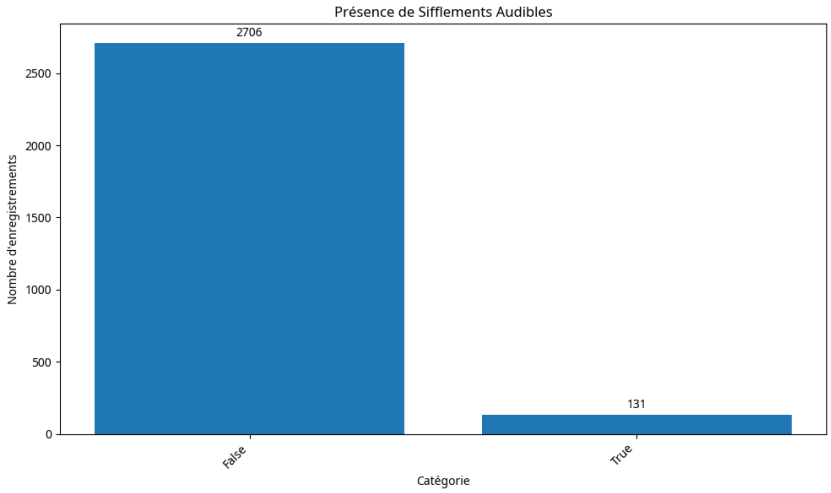


FIGURE 4.11 – Présence de Sifflements Audibles (Étiquettes Expertes).

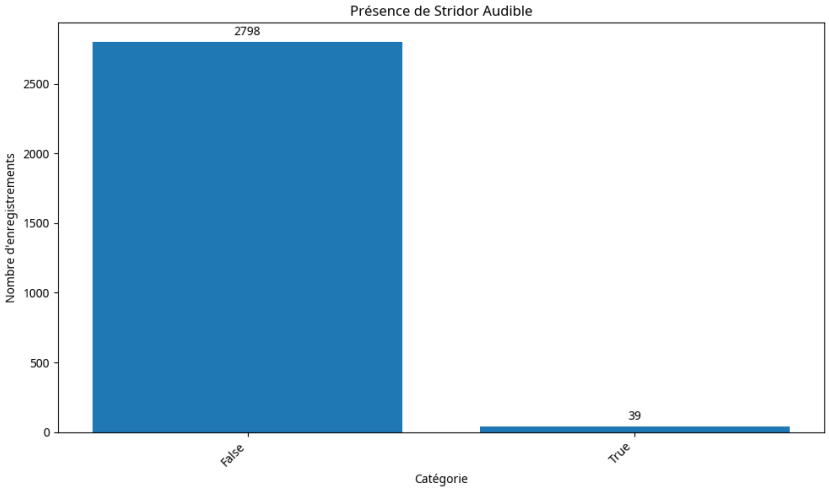


FIGURE 4.12 – Présence de Stridor Audible (Étiquettes Expertes).

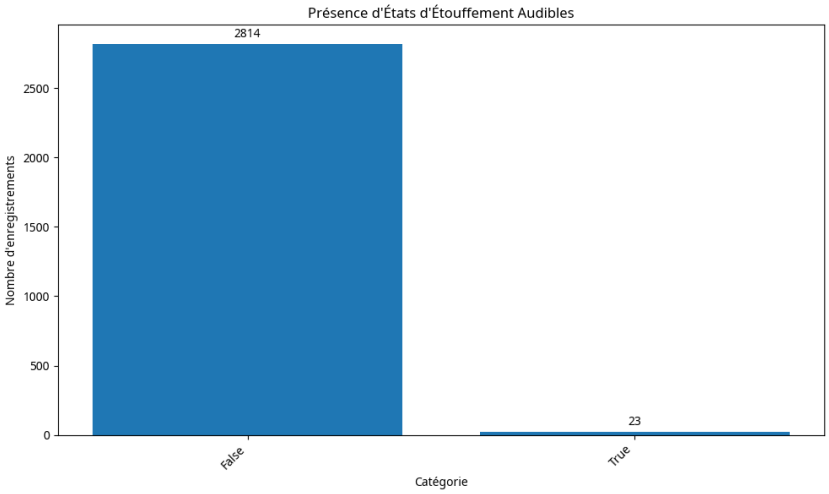


FIGURE 4.13 – Présence d'États d'Étouffement Audibles (Étiquettes Expertes).

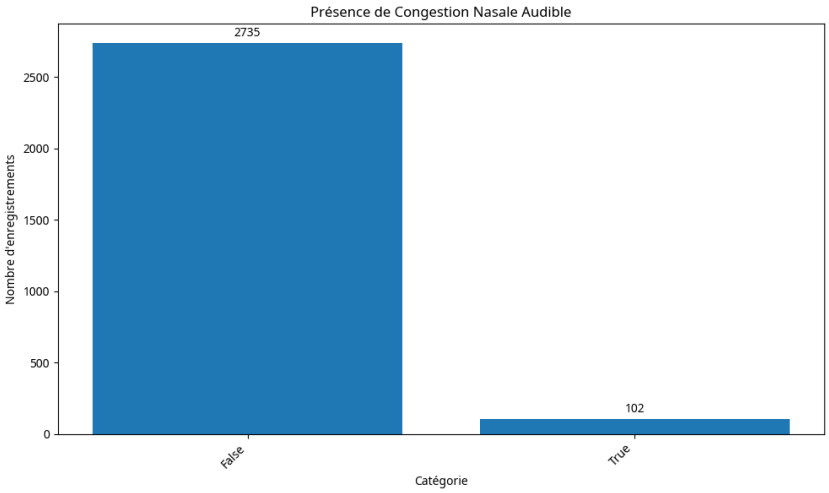


FIGURE 4.14 – Présence de Congestion Nasale Audible (Étiquettes Expertes).

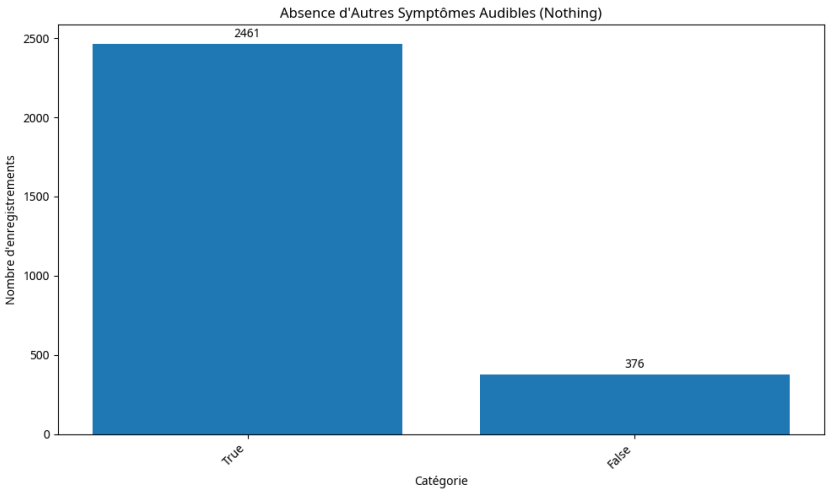


FIGURE 4.15 – Absence d'Autres Symptômes Audibles ('Nothing', Étiquettes Expertes).

Les symptômes les plus fréquemment notés sont les sifflements ('wheezing', 131 cas), la dyspnée ('dyspnea', 105 cas) et la congestion ('congestion', 102 cas). Le stridor (39 cas) et l'étouffement ('choking', 23 cas) sont rares. L'étiquette 'nothing' est marquée comme 'True' dans 2461 cas (sur 2837, 86.7%), suggérant que la majorité des enregistrements ne présentaient aucun de ces cinq symptômes spécifiques selon les experts.

4.3.4 Sévérité de la Toux

L'impression des experts sur la sévérité de la toux est résumée dans la Figure 4.16.

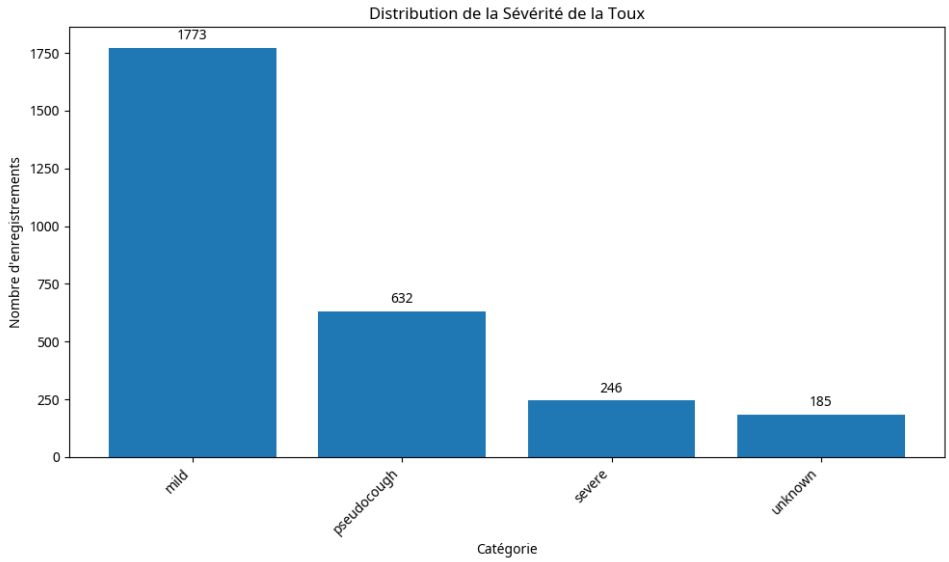


FIGURE 4.16 – Distribution de la Sévérité de la Toux (Étiquettes Expertes).

La plupart des toux sont jugées 'légères' ('mild', 1773 sur 2836, 62.5%). La catégorie 'pseudocough' (toux saine/non pathologique) est également fréquente (632, 22.3%). Les toux 'sévères' ('severe') sont moins courantes (246, 8.7%), et une part d'incertitude demeure ('unknown', 185, 6.5%).

4.3.5 Diagnostic Expert

Le diagnostic probable assigné par les experts est une information clé. Sa distribution est présentée en Figure 4.17.

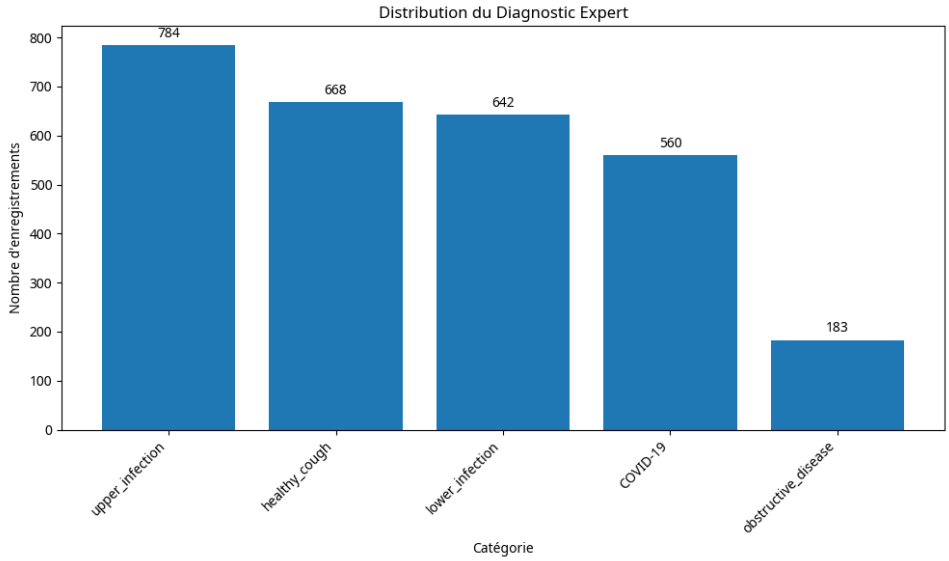


FIGURE 4.17 – Distribution du Diagnostic Expert (Étiquettes Expertes).

Les diagnostics les plus fréquents sont l'infection des voies respiratoires supérieures ('upper_infection', 784 sur 2837, 27.6%) et la toux saine ('healthy_cough', 668, 23.5%). L'infection des voies respiratoires inférieures ('lower_infection', 642, 22.6%) et la COVID-19 (560, 19.7%) sont également bien représentées. Les maladies obstructives ('obstructive_disease', 183, 6.5%) sont la catégorie la moins fréquente. La présence notable d'enregistrements étiquetés COVID-19 par des experts (560) est un atout majeur pour la recherche spécifique à cette maladie.

4.3.6 Interprétation et Implications des Distributions

L'analyse des distributions (Section 4.3) révèle plusieurs points importants :

— **Qualité et Fiabilité :** La prédominance des qualités 'good' et 'ok' (90%) renforce la confiance dans la capacité de ce sous-ensemble annoté à être utilisé pour des analyses acoustiques significatives. Les cas 'poor' et 'no_cough' sont suffisamment minoritaires pour être potentiellement exclus sans perte majeure d'information, ou traités avec prudence.

— **Prédominance de la Toux Sèche :** La forte proportion de toux sèches (63.5%) est cohérente avec les symptômes typiques de nombreuses infections virales des voies respiratoires supérieures et de la COVID-19, qui étaient des cibles d'intérêt lors de la collecte.
— **Incertitude de l'Annotation Audio :** Les catégories 'unknown' pour le type de toux (17.5%) et la sévérité (6.5%) mettent en évidence les limites de l'évaluation purement auditive. Certains aspects cliniques sont difficiles à discerner sans informations contextuelles supplémentaires. Cela souligne l'importance de considérer l'accord inter-annotateurs (non détaillé ici mais mentionné comme un aspect à évaluer dans de telles études) lors de l'utilisation de ces étiquettes.
— **Profil des Symptômes Audibles :** La rareté de certains symptômes comme le stridor et l'étouffement dans ce jeu de données peut rendre difficile l'entraînement de modèles spécifiques pour leur détection. L'étiquette 'nothing' (True pour 87%) suggère que la plupart des toux, même si elles sont associées à un diagnostic pathologique, ne présentent pas nécessairement ces symptômes audibles spécifiques de manière évidente pour les experts.
— **Diversité des Diagnostics :** Le jeu de données annoté couvre un éventail pertinent de conditions, incluant des infections hautes et basses, la COVID-19, des maladies obstructives et des toux saines. Cet équilibre relatif (à l'exception des maladies obstructives, moins fréquentes) est précieux pour développer des modèles de diagnostic différentiel. La confirmation experte de 560 cas de COVID-19 est particulièrement notable.
— **Relation Sévérité-Diagnostic :** Bien que non explorée ici faute de données croisées, l'intuition clinique suggère des liens entre la sévérité et le diagnostic (par ex., les infections basses ou la COVID-19 pourraient être plus souvent associées à une sévérité 'sévère'). Une analyse future sur les données individuelles pourrait explorer ces corrélations.

4.4 Préparation et Filtrage des Données Expertes

Suite à l'analyse exploratoire initiale (Section 4.2) , une étape cruciale de préparation et de filtrage des données a été entreprise pour constituer un jeu de données plus ciblé et de meilleure qualité pour l'entraînement et l'évaluation des modèles. Cette étape s'est concentrée sur l'exploitation des annotations fournies par les experts médicaux, qui, bien que couvrant un sous-ensemble plus restreint des données (comme discuté dans la Section 4.3), offrent une vérité terrain clinique plus fiable que les statuts auto-déclarés.

Le processus, a débuté par le chargement du fichier de métadonnées complet (`metadata_compiled.csv`). La première étape de filtrage a consisté à ne retenir que les enregistrements pour lesquels au moins un diagnostic expert était disponible. Les colonnes

de diagnostic expert considérées étaient `diagnosis_1`, `diagnosis_2`, `diagnosis_3`, et `diagnosis_4`. Les enregistrements sans aucune annotation de diagnostic par les experts ont donc été exclus de l'analyse ultérieure.

Ensuite, un filtrage basé sur la qualité audio, également évaluée par les experts, a été appliqué. Les colonnes correspondantes (`quality_1` à `quality_4`) ont été examinées. Tous les enregistrements pour lesquels au moins un expert avait indiqué une qualité *ńo_cough´* (pas de toux) ou *ṕoor´* (médiocre) ont été écartés. Ce filtrage vise à garantir que les signaux audio utilisés contiennent bien des événements de toux exploitables et que leur qualité est suffisante pour une analyse acoustique pertinente.

Une étape importante a ensuite été la consolidation des diagnostics multiples. Pour obtenir une étiquette de diagnostic unique par enregistrement, une stratégie spécifique a été mise en œuvre, basée sur le nombre d'experts ayant fourni un diagnostic (colonne `diagnosis_count`) :

— **Diagnostic unique :** Si un seul expert avait fourni un diagnostic (`diagnosis_count` = 1), ce diagnostic était directement retenu comme étiquette finale.

— **Diagnostics multiples :** Le code fourni applique une fonction `process_row` qui gère les cas où `diagnosis_count` est compris entre 1 et 4. . Pour `diagnosis_count` supérieurs à 1, elle retient le diagnostic le plus fréquent (vote majoritaire).

Cette procédure a permis d'attribuer une étiquette de diagnostic unique (colonne `diagnosis`) aux enregistrements conservés. La distribution finale des classes de diagnostic dans ce jeu de données filtré (avant segmentation et augmentation) était la suivante : *upper_infection* (695), *lower_infection* (553), *healthy_cough* (529), *COVID-19* (498), et *obstructive_disease* (170). Ce jeu de données affiné a servi de base aux étapes suivantes.

4.5 Segmentation des Événements de Toux

Une fois les enregistrements audio filtrés selon les critères de qualité et de diagnostic expert, l´tape suivante a consisté à isoler précisément les segments contenant les sons de toux au sein de chaque enregistrement. Une segmentation précise est essentielle pour garantir que les caractéristiques acoustiques extraites correspondent bien à l'événement d'intérêt.

Pour cela, un algorithme de segmentation basé sur l'énergie du signal a été employé, inspiré des approches courantes en détection d'activité vocale ou sonore [70]. L'idée fondamentale est de détecter les portions du signal dont l'énergie dépasse certains seuils adaptatifs, comme illustré schématiquement dans la Figure 4.18. L'algorithme fonctionne comme suit :

1. **Calcul de l'énergie RMS :** L'énergie moyenne quadratique (Root Mean Square - RMS) de l'ensemble du signal audio est d'abord calculée.

2. **Définition des Seuils :** Deux seuils adaptatifs sont définis par rapport à l'énergie

RMS : un seuil bas (`seg_th_l` = 0.1 * RMS) et un seuil haut (`seg_th_h` = 2.0 * RMS).
3. **Détection des Segments :** Le signal est parcouru. Le dépassement du seuil haut marque le début potentiel d'une toux. Un *padding* (marge de 0.1 ou 0.2s selon l'implémentation) est ajouté avant ce point.
4. **Détection de la Fin :** Lorsque l'énergie retombe sous le seuil bas pendant une durée minimale (tolérance de 0.01s), la fin est détectée. Un *padding* est ajouté après ce point.
5. **Filtrage par Durée :** Seuls les segments dont la durée (hors padding) dépasse une longueur minimale (`min_cough_len` = 0.2s) sont conservés.

Le résultat est une liste de segments audio correspondant aux événements de toux présumés, prêts pour l'extraction de caractéristiques.

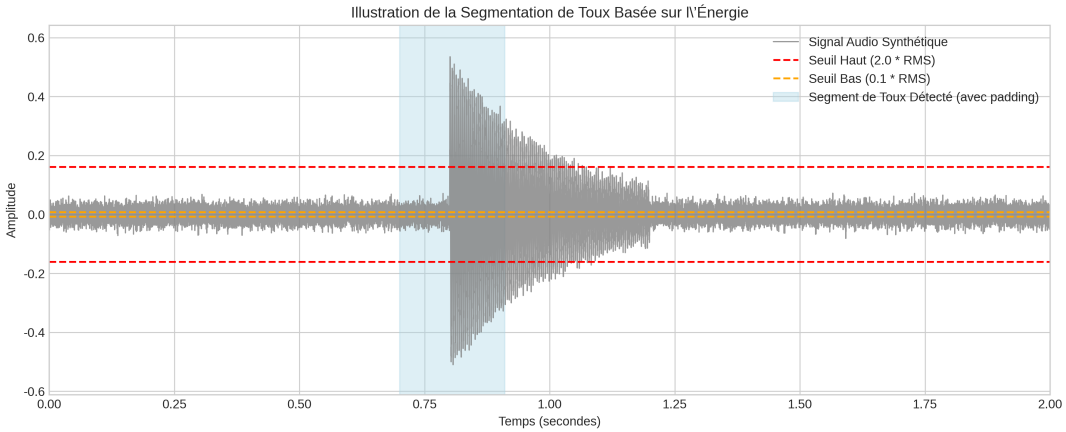


FIGURE 4.18 – Illustration du processus de segmentation basé sur l'énergie avec seuils adaptatifs.

4.6 Chargement et normalisation des segments audio

Une fois les enregistrements pertinents identifiés et les toux segmentées, l'étape suivante consistait à intégrer les données audio brutes à notre pipeline de traitement. Nous avons utilisé la bibliothèque `librosa` [71], un outil standard en analyse audio, pour charger chaque toux segmentée. Lors du chargement, l'audio a été automatiquement converti en un seul canal (mono) et rééchantillonné à une fréquence constante de 16 kHz. Cette normalisation garantit que toutes les analyses ultérieures fonctionnent sur des données aux caractéristiques uniformes.

Une décision clé à ce stade concernait la gestion des différentes longueurs des toux segmentées. Afin de préparer les données à une extraction de caractéristiques cohérente et à une entrée dans le modèle, nous avons défini une durée cible fixe pour chaque segment. Nous avons expérimenté plusieurs durées, testant des valeurs allant de 0,3 seconde à 1 seconde. Nos tests empiriques ont révélé qu'une durée cible de 0,5 seconde offrait un équilibre optimal, capturant suffisamment d'informations acoustiques pour la classifica-

tion sans introduire de silence excessif ni tronquer des parties cruciales des toux courtes. Par conséquent, chaque toux segmentée a été soit complétée par du silence (si elle était inférieure à 0,5 s), soit tronquée (si elle était supérieure) pour correspondre précisément à cette durée cible de 0,5 seconde, soit 8 000 échantillons à notre fréquence d'échantillonnage de 16 kHz. Cela a permis de créer des entrées audio de taille uniforme, prêtes à l'extraction de caractéristiques.

4.7 Augmentation des Données pour la Réduction du Surapprentissage

Afin däccroître efficacement la taille du jeu de données expert et, de manière cruciale, pour atténuer le risque de surapprentissage (overfitting) du modèle, nous avons employé des techniques däugmentation de données sur les segments de toux extraits, comme illustré à la Figure 4.19. Le surapprentissage se produit lorsquün modèle mémorise trop spécifiquement les données dëntraînement, y compris leur bruit et leurs particularités, ce qui entraîne de faibles performances sur de nouvelles données invisibles. Läugmentation de données aide à contrer ce phénomène en élargissant artificiellement la diversité de lënsemble dëntraînement, obligeant le modèle à apprendre des motifs plus généralisables. Notre approche, säppuyant sur la bibliothèque `audiomentations` [72], a généré trois versions augmentées pour chaque segment original valide (durée >= 0.5s). Cette stratégie est cohérente avec les conclusions de la littérature, telles que celles mises en évidence par Blankenmeier et al. (2023) [73], qui démontrent lëfficacité de certaines augmentations pour améliorer la généralisabilité des modèles acoustiques de santé. Löbjectif n¨tait pas seulement döbtenir plus de données, mais de construire des modèles plus robustes et moins sujets au surapprentissage.

Les transformations suivantes ont été appliquées aléatoirement (probabilité de 0.5 chacune) :

- **Ajout de Bruit Gaussien (AddGaussianNoise) :** Amplitude entre 0.001 et 0.015.
- **Étirement Temporel (TimeStretch) :** Facteur entre 0.8 et 1.25.
- **Décalage de Hauteur (PitchShift) :** Décalage entre -2 et +2 demi-tons.
- **Décalage Temporel (Shift) :** Décalage entre -10

Ces transformations simulent des variations acoustiques réalistes (bruit, vitesse, pitch, timing). Les segments augmentés ont été traités comme les originaux (ajustement de durée, extraction de caractéristiques) et ajoutés au jeu de données, multipliant par quatre le nombre total d¨chantillons.

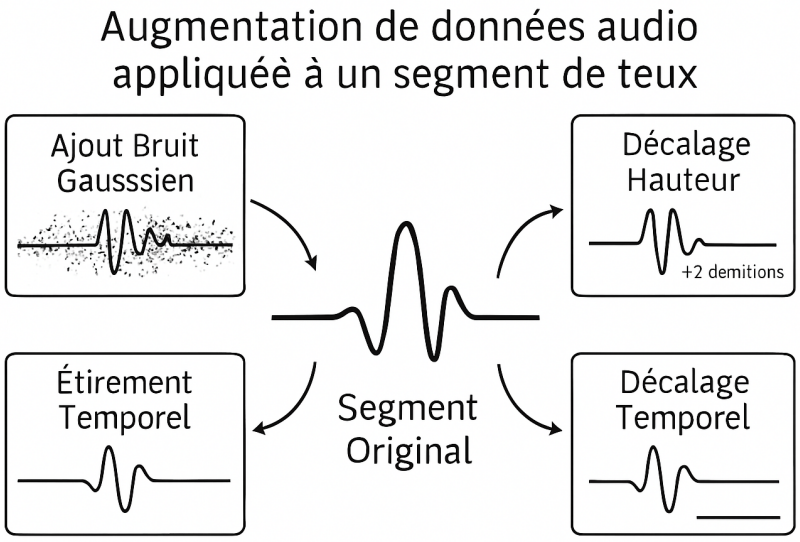


FIGURE 4.19 – Illustration conceptuelle des techniques däugmentation de données audio appliquées.

Stratégies dÄugmentation Avancées Au-delà des transformations standards appliquées ici, la recherche continue dëxplorer des stratégies däugmentation plus sophistiquées pour améliorer la robustesse et la performance des modèles sur les signaux acoustiques liés à la santé. Des études systématiques ont évalué lëfficacité de diverses augmentations et de leurs combinaisons, notamment dans le cadre de läpprentissage contrastif. Par exemple, Blankenmeier et al. (2023) [73] ont démontré les bénéfices synergiques de la combinaison de certaines augmentations temporelles, comme le décalage temporel circulaire et l¨tirement temporel, pour améliorer la généralisabilité des représentations acoustiques de santé, y compris sur des données de toux comme celles de COUGHVID. Däutres approches visent à simuler des conditions réelles ou à pallier le manque de données. Mishra et al. (2023) [74] ont montré lëfficacité de läugmentation par simulation de codec (en introduisant des artefacts de compression à bas débit) pour améliorer significativement la classification des sons cardiaques dans le domaine temporel, suggérant une piste pour simuler les dégradations potentielles lors de la transmission ou du stockage. Plus récemment, des techniques génératives avancées, telles que les modèles de diffusion audio conditionnels, sont explorées pour créer des échantillons synthétiques de haute fidélité. Kim et al. (2023) [75] proposent dütiliser de tels modèles génératifs combinés à un affinage adversarial pour aligner les caractéristiques des données synthétiques et réelles, améliorant ainsi la performance, notamment pour les classes minoritaires dans les tâches de classification des sons respiratoires. Lëxploration et la combinaison judicieuse de ces diverses stratégies däugmentation, des transformations classiques aux méthodes de simulation et génératives, apparaissent donc comme une voie prometteuse pour renforcer les

modèles dänalyse acoustique en santé.

4.8 Extraction de Caractéristiques Acoustiques

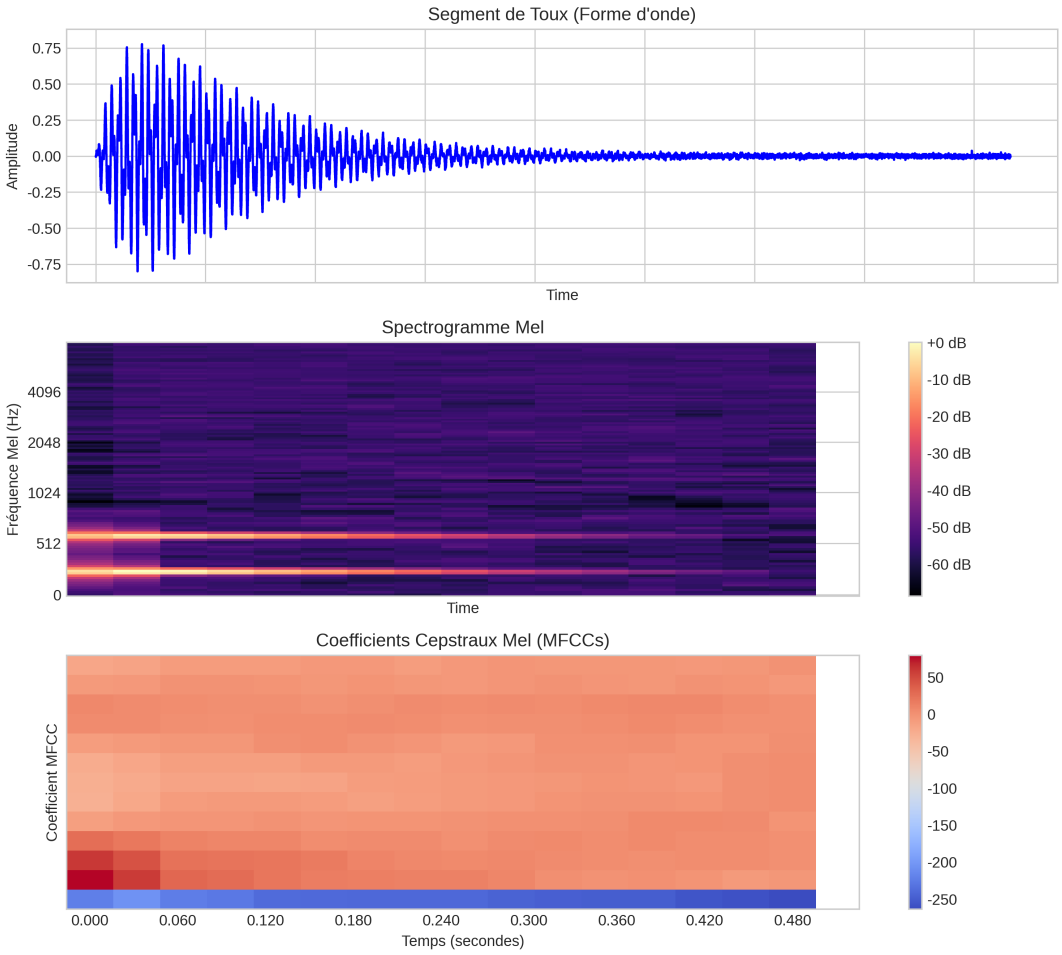


FIGURE 4.20 – Illustration de l'extraction de caractéristiques : Forme d'onde, Spectrogramme Mel et MFCCs.

L´tape finale de préparation des données consiste à transformer les segments audio normalisés (originaux et augmentés) en représentations numériques compactes et informatives, connues sous le nom de vecteurs de caractéristiques. Ces caractéristiques sont conçues pour capturer les propriétés acoustiques distinctives des sons de toux qui pourraient être pertinentes pour différencier les conditions de santé sous-jacentes. Pour cette tâche cruciale, nous nous sommes appuyés sur la bibliothèque `librosa` [71], un outil largement reconnu pour länalyse audio.

Pour chaque segment audio de 0.5 seconde, un ensemble diversifié de caractéristiques acoustiques a été calculé. Cela comprenait les coefficients cepstraux en fréquence Mel (MFCC), qui sont connus pour leur efficacité dans la modélisation de la perception auditive humaine et sont largement utilisés dans la reconnaissance vocale et länalyse sonore

[76]. Nous avons extrait 40 coefficients MFCC. En complément, nous avons calculé les caractéristiques chroma, qui décrivent la distribution de l¨nergie sur les 12 classes de hauteur traditionnelles, capturant ainsi les aspects harmoniques et mélodiques [77]. Le spectrogramme Mel, une représentation temps-fréquence qui utilise l¨chelle de fréquence Mel perceptivement motivée, a également été calculé, résultant en une représentation avec 128 bandes de fréquence Mel. Däutres caractéristiques spectrales ont été incluses, telles que le contraste spectral, qui mesure la différence dämplitude entre les pics et les vallées dans le spectre sur 7 bandes de fréquence [78], et les caractéristiques tonnetz (tonal centroid), qui fournissent une représentation des relations harmoniques. Enfin, des caractéristiques temporelles et spectrales plus globales ont été ajoutées, notamment le taux de passage par zéro (Zero Crossing Rate - ZCR), l¨nergie moyenne quadratique (Root Mean Square - RMS), la largeur de bande spectrale et le centroïde spectral.

Pour obtenir un vecteur de caractéristiques de taille fixe pour chaque segment, la moyenne temporelle a été calculée pour chaque coefficient ou bande des caractéristiques multidimensionnelles (MFCC, Chroma, Mel spectrogramme, Contraste Spectral, Tonnetz). Ces valeurs moyennes, ainsi que les caractéristiques scalaires (ZCR, RMS, etc.), ont été ensuite concaténées séquentiellement. Ce processus a abouti à un unique vecteur de caractéristiques de longueur 197 pour chaque segment de toux. Ces vecteurs numériques, représentant lëssence acoustique de chaque toux, ainsi que les métadonnées associées (comme lïdentifiant de lënregistrement original et l¨tiquette de diagnostic expert), ont été soigneusement organisés et stockés, prêts pour les étapes de modélisation ultérieures. La distribution finale des classes après augmentation et avant la division des données était la suivante : *lower_infection* (3744), *upper_infection* (3460), *COVID-19* (2936), *healthy_cough* (1996), et *obstructive_disease* (1116).

Chapitre 5
Modélisation Acoustique et Évaluation des Performances

Chapitre 5

Modélisation Acoustique et Évaluation des Performances

5.1 Division Stratégique des Données : Entraînement, Validation et Test

Une fois les caractéristiques extraites pour tous les segments de toux (originaux et augmentés), une étape méthodologique essentielle consistait à diviser lënsemble de données en sous-ensembles distincts pour lëntraînement, la validation et le test du modèle. Cette division est cruciale pour développer un modèle qui non seulement apprend à partir des données fournies, mais qui peut aussi généraliser ses connaissances à de nouvelles données invisibles, tout en permettant un réglage fiable des hyperparamètres.

Notre stratégie de division a été conçue pour exploiter au mieux les données originales et augmentées. Nous avons däbord séparé les échantillons originaux (ceux directement issus des enregistrements filtrés) des échantillons générés par augmentation. Lënsemble des données augmentées, représentant environ 75% du total des échantillons (9939 sur 13252), a été entièrement dédié à lëntraînement du modèle. Cette approche permet au modèle däpprendre à partir düne grande diversité de variations acoustiques, favorisant ainsi la robustesse et réduisant le risque de surapprentissage sur les spécificités des enregistrements originaux.

Lënsemble des données originales restantes (3313 échantillons) a ensuite été divisé en deux parties : un ensemble de validation et un ensemble de test. Deux tiers des données originales (2208 échantillons, soit environ 17% du total) ont été alloués à lënsemble de validation. Cet ensemble est utilisé pendant le processus dëntraînement pour évaluer les performances du modèle sur des données non vues et pour guider le réglage des hyperparamètres (comme le taux däpprentissage ou la complexité du modèle). Il est fondamental que cet ensemble de validation ne contienne que des données originales non augmentées, afin de fournir une estimation réaliste de la performance du modèle sur de

véritables nouveaux échantillons, sans lïnfluence potentiellement optimiste des variations synthétiques vues pendant lëntraînement.

Le tiers restant des données originales (1105 échantillons, soit environ 8% du total) a constitué lënsemble de test final. Cet ensemble est mis de côté et nëst utilisé quüne seule fois, après que lëntraînement et la validation sont terminés, pour obtenir une évaluation finale et impartiale de la capacité de généralisation du modèle sur des données complètement nouvelles et représentatives du monde réel. Comme pour lënsemble de validation, il est crucial que lënsemble de test ne contienne que des données originales.

Pour garantir que la distribution des différentes classes de diagnostic (par exemple, *COVID-19*, *healthy_cough*, etc.) soit représentative dans les ensembles de validation et de test, la division des données originales a été effectuée de manière stratifiée. Cela signifie que la proportion de chaque classe dans les ensembles de validation et de test reflète étroitement sa proportion dans lënsemble original complet. Länalyse des distributions finales confirme cette stratification : lënsemble dëntraînement (9939 échantillons) contenait environ 28.3% dïnfections basses, 15.1% de toux saines, 26.1% dïnfections hautes, 8.4% de maladies obstructives et 22.2% de COVID-19. Les ensembles de validation (2208 échantillons) et de test (1105 échantillons) présentaient des proportions très similaires pour chaque classe, assurant ainsi une évaluation équilibrée et fiable à chaque étape.

5.2 Correction du déséquilibre des classes par pondération des classes

L'examen de la distribution des classes diagnostiques au sein de notre ensemble de données d'apprentissage final (composé entièrement d'échantillons augmentés) a révélé un déséquilibre notable. Si des classes comme *lower_infection* (environ 28,3

L'entraînement d'un modèle directement sur de telles données déséquilibrées peut entraîner des résultats biaisés. Les algorithmes d'apprentissage standard optimisent souvent la précision globale, ce qui signifie qu'ils peuvent obtenir des scores élevés simplement en apprenant à prédire correctement les classes majoritaires tout en ignorant les classes minoritaires. Ceci est indésirable dans un contexte clinique, où l'identification précise d'affections moins courantes mais potentiellement graves, comme les maladies obstructives, est cruciale.

Pour contrer ce biais potentiel et encourager notre modèle à apprendre des caractéristiques discriminantes pour toutes les catégories diagnostiques, y compris celles sous-représentées, nous avons mis en œuvre une technique appelée pondération par classe lors de la phase d'apprentissage du modèle. L'idée principale de la pondération par classe est d'ajuster l'importance accordée à chaque échantillon d'apprentissage en fonction de son étiquette de classe. Plus précisément, les échantillons appartenant aux classes mino-

ritaires se voient attribuer des pondérations plus élevées, tandis que les échantillons des classes majoritaires reçoivent des pondérations plus faibles.

En pratique, cela signifie que le modèle subit une pénalité (ou un coût) plus importante lorsqu'il classe mal un échantillon d'une classe minoritaire que lorsqu'il classe mal un échantillon d'une classe majoritaire. Cette pondération différentielle oblige l'algorithme d'apprentissage à prêter une attention particulière aux classes minoritaires, l'empêchant ainsi de converger vers une solution favorisant les catégories les plus fréquentes. Pour notre étude, nous avons utilisé une approche automatisée communément appelée heuristique « équilibrée ». Cette méthode calcule automatiquement des pondérations de classe inversement proportionnelles à la fréquence de chaque classe dans les données d'apprentissage. Par conséquent, plus une classe est rare, plus la pondération attribuée à ses échantillons est élevée lors de l'apprentissage. En intégrant cette stratégie de pondération de classe, nous avons cherché à favoriser un processus d'apprentissage plus équitable, aboutissant à un modèle de classification offrant des performances plus équilibrées dans toutes les catégories diagnostiques et une pertinence clinique améliorée.

5.3 Modélisation et Optimisation

Après la préparation, la segmentation, la normalisation et l'augmentation des données, l'étape suivante consistait à choisir, entraîner et optimiser un modèle d'apprentissage automatique capable de classifier les segments de toux en fonction des diagnostics experts. Cette section détaille le choix du modèle, la méthode d'optimisation des hyperparamètres et les résultats finaux obtenus.

5.3.1 Choix du Modèle : XGBoost

Pour la tâche de classification des types de toux à partir des caractéristiques acoustiques extraites, nous avons choisi d'utiliser l'algorithme XGBoost (Extreme Gradient Boosting) [38]. Ce choix est motivé par plusieurs de ses atouts reconnus. XGBoost excelle particulièrement sur les données structurées ou tabulaires, ce qui correspond bien à notre situation où les caractéristiques extraites des segments audio forment un tel ensemble de données. De plus, XGBoost intègre des mécanismes de régularisation (L1 et L2) qui aident à prévenir le surapprentissage, un risque non négligeable étant donné la complexité potentielle des données acoustiques et la taille modérée de notre jeu de données expert, même après augmentation. Sa performance élevée, souvent à la pointe de la technologie pour de nombreuses tâches de classification tabulaire, et son efficacité computationnelle en font un candidat solide pour notre approche proposée.

5.3.2 Optimisation des Hyperparamètres par Optimisation Bayésienne

L'optimisation des hyperparamètres est une étape critique dans la construction de modèles d'apprentissage automatique performants. Plutôt que d'utiliser des méthodes exhaustives comme la recherche par grille (Grid Search) ou des approches aléatoires (Random Search), qui peuvent être computationnellement très coûteuses, surtout lorsque l'évaluation du modèle prend du temps, nous avons opté pour l'Optimisation Bayésienne. L'Optimisation Bayésienne est une stratégie séquentielle d'optimisation globale pour les fonctions dites "boîte noire" (dont l'évaluation est coûteuse et dont on ne connaît pas la forme analytique), telle que la fonction qui mappe les hyperparamètres d'un modèle à sa performance de validation [79]. Elle fonctionne en construisant un modèle probabiliste (souvent un processus gaussien) de la fonction objectif (par exemple, la précision de validation en fonction des hyperparamètres) et utilise ce modèle pour sélectionner les prochains points (combinaisons d'hyperparamètres) à évaluer, en équilibrant l'exploration de nouvelles régions de l'espace des hyperparamètres et l'exploitation des régions qui semblent prometteuses. Cette approche est particulièrement efficace en termes de nombre d'évaluations nécessaires pour trouver de bons hyperparamètres, la rendant beaucoup plus efficiente que les recherches par grille ou aléatoires lorsque chaque évaluation est coûteuse [80].

Afin de maximiser les performances de notre modèle XGBoost, nous avons employé l'Optimisation Bayésienne pour rechercher la meilleure combinaison d'hyperparamètres. Le processus visait à optimiser la précision moyenne obtenue par validation croisée stratifiée à 5 plis (5-fold stratified cross-validation) sur l'ensemble d'entraînement. Cette métrique a servi de fonction objectif pour l'Optimisation Bayésienne. L'espace de recherche incluait des hyperparamètres clés d'XGBoost : le taux d'apprentissage (‘learning_rate‘), la profondeur maximale des arbres (‘max_depth‘), la fraction d'échantillons utilisée pour entraîner chaque arbre (‘subsample‘), et la fraction de caractéristiques considérée lors de la construction de chaque arbre (‘colsample_bytree‘). Des plages de valeurs plausibles ont été définies pour chacun de ces hyperparamètres. L'Optimisation Bayésienne a ensuite exploré cet espace de manière itérative, en utilisant les résultats des évaluations précédentes (précision de validation croisée pour des combinaisons d'hyperparamètres données) pour guider intelligemment la recherche vers les régions les plus prometteuses. Après un nombre défini d'itérations (10 points initiaux suivis de 25 itérations d'optimisation), le processus a convergé vers une combinaison d'hyperparamètres qui a significativement amélioré la performance de validation croisée par rapport à des configurations par défaut ou choisies manuellement, démontrant l'efficacité de cette approche pour affiner notre modèle de classification de toux.

5.3.3 Résultats et Évaluation du Modèle Final

Après avoir entraîné le modèle XGBoost final sur l'ensemble des données d'entraînement en utilisant les hyperparamètres optimaux identifiés par l'Optimisation Bayésienne, nous avons évalué ses performances sur les ensembles de validation et de test, qui n'avaient pas été utilisés lors de l'entraînement ou de l'optimisation. Les métriques clés obtenues sont les suivantes :

— Accuracy sur l'ensemble de validation : 96.56%

— Accuracy sur l'ensemble de test : 96.92%

Pour une évaluation plus détaillée sur l'ensemble de test, le rapport de classification suivant a été généré, présentant les métriques de précision, rappel et score F1 pour chaque classe de diagnostic :

	Accuracy	recall	f1-score	support
COVID-19	0.96	0.96	0.96	245
healthy_cough	0.99	0.96	0.98	166
lower_infection	0.95	0.99	0.97	312
obstructive_disease	1.00	0.96	0.98	93
upper_infection	0.98	0.97	0.97	289
accuracy			0.97	1105
macro avg	0.98	0.97	0.97	1105
weighted avg	0.97	0.97	0.97	1105

Ces résultats démontrent une performance très élevée et cohérente du modèle optimisé. La précision globale, atteignant près de 97% sur l'ensemble de test, indique que le modèle classe correctement la grande majorité des échantillons de toux. De manière importante, les précisions sur les ensembles de validation et de test sont très proches, suggérant que le modèle généralise bien aux données nouvelles et invisibles, sans signe apparent de surapprentissage. Le rapport de classification détaillé confirme cette robustesse : les scores de précision, de rappel et F1 sont élevés pour toutes les catégories de diagnostic, y compris les classes moins fréquentes comme les maladies obstructives. La précision mesure la proportion de prédictions positives correctes pour une classe, tandis que le rappel mesure la proportion d'instances réelles d'une classe qui ont été correctement identifiées. Le score F1 combine ces deux métriques. Les valeurs élevées et équilibrées pour ces indicateurs à travers toutes les classes montrent que le modèle n'est pas biaisé envers une catégorie particulière et est capable de distinguer efficacement les différents types d'affections respiratoires sur la base des caractéristiques acoustiques de la toux.

5.4 Analyse de l'Importance des Caractéristiques et Sélection

Après avoir obtenu les résultats initiaux encourageants avec le modèle XGBoost optimisé utilisant l'ensemble complet des caractéristiques extraites, nous avons cherché à mieux comprendre quelles caractéristiques contribuaient le plus à la performance de classification et si une sélection plus judicieuse de ces caractéristiques pouvait encore améliorer les résultats. Pour ce faire, nous avons mené une analyse de l'importance des caractéristiques.

5.4.1 Analyse Initiale de l'Importance

Nous avons d'abord évalué la performance du modèle en utilisant chaque caractéristique individuellement. Cette analyse a révélé que les caractéristiques dérivées du Mel-spectrogramme dominaient largement les premières places en termes de précision individuelle, suggérant leur forte pertinence. Les 20 caractéristiques individuelles les plus performantes (avec leur précision de test) sont listées ci-dessous :

— f153 (mel102, Mel-spectrogram) : 34.66%
— f66 (mel15, Mel-spectrogram) : 33.12%
— f115 (mel64, Mel-spectrogram) : 32.94%
— f104 (mel53, Mel-spectrogram) : 32.49%
— f113 (mel62, Mel-spectrogram) : 32.31%
— f117 (mel66, Mel-spectrogram) : 31.95%
— f58 (mel7, Mel-spectrogram) : 31.49%
— f150 (mel99, Mel-spectrogram) : 31.40%
— f64 (mel13, Mel-spectrogram) : 31.13%
— f121 (mel70, Mel-spectrogram) : 31.13%
— f98 (mel47, Mel-spectrogram) : 31.04%
— f148 (mel97, Mel-spectrogram) : 30.95%
— f146 (mel95, Mel-spectrogram) : 30.86%
— f70 (mel19, Mel-spectrogram) : 30.77%
— f154 (mel103, Mel-spectrogram) : 29.50%
— f196 (centroid, Centroid) : 28.96%
— f175 (mel124, Mel-spectrogram) : 28.87%
— f2 (mfcc3, MFCC) : 28.42%
— f4 (mfcc5, MFCC) : 27.87%
— f181 (con2, Spectral Contrast) : 26.06%

Bien que les précisions obtenues avec une seule caractéristique restent modestes, cette analyse initiale a clairement orienté notre attention vers l'importance prépondérante des

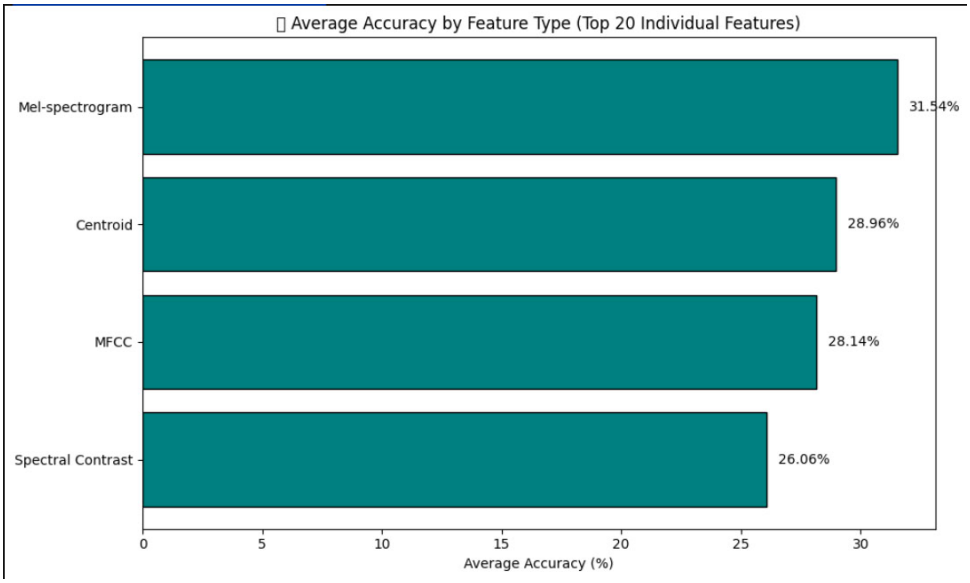


FIGURE 5.1 – Précision Moyenne par Type de Caractéristique (Basée sur les Top 20 Caractéristiques Individuelles).

caractéristiques Mel.

5.4.2 Contributions par Type de Caractéristique

Pour visualiser la contribution relative des différents types de caractéristiques, nous avons agrégé les performances des 20 meilleures caractéristiques individuelles par catégorie. Le diagramme en barres horizontales de la Figure 5.1 montre clairement que les caractéristiques du Mel-spectrogramme présentaient la précision moyenne la plus élevée (environ 31.5%), suivies par le Centroïde Spectral (environ 29.0%), les MFCC (environ 28.1%), et enfin le Contraste Spectral (environ 26.1%). Cela a renforcé l'idée que les caractéristiques du Mel-spectrogramme étaient particulièrement informatives pour notre modèle.

5.4.3 Expérimentations avec Combinaisons de Caractéristiques

Forts de ces observations, nous avons conduit une série d'expérimentations en entraînant et évaluant le modèle XGBoost optimisé avec différentes combinaisons de groupes de caractéristiques. L'objectif était de déterminer si un sous-ensemble spécifique pouvait surpasser la performance obtenue avec l'ensemble complet des caractéristiques. Le Tableau 5.1 résume les précisions obtenues sur les ensembles de validation et de test pour chaque combinaison.

Les résultats montrent clairement que les combinaisons incluant les caractéristiques du Mel-spectrogramme sont nettement plus performantes que celles qui en sont dépourvues. Notamment, la combinaison **MFCC + Mel-spectrogramme** a atteint la **meilleure**

TABLE 5.1 – Résultats des Expérimentations avec Combinaisons de Caractéristiques

Combinaison de Caractéristiques	Accuracy Validation (%)	Accuracy Test (%)
MFCC seul	86.64	87.15
Mel-spectrogramme seul	97.46	97.56
Toutes Caractéristiques	96.56	96.92
MFCC + Mel + Centroïde + Contraste	96.92	96.29
MFCC + Mel-spectrogramme	**97.51**	**98.01**
MFCC + Mel-spectrogramme + Centroïde	97.46	96.65
MFCC + Mel-spectrogramme + Contraste	97.01	97.19
Mel-spectrogramme + Centroïde	97.37	97.56
Mel-spectrogramme + Contraste	97.19	97.47
MFCC + Centroïde	87.45	88.42
MFCC + Contraste	87.09	86.52
MFCC + Contraste + Centroïde	87.18	87.24
Mel-spectrogramme + Contraste + Centroïde	97.15	97.65

performance globale, avec une précision de test remarquable de **98.01%**. Cette performance est légèrement supérieure à celle obtenue en utilisant l'ensemble complet des caractéristiques (96.92%), suggérant qu'une sélection judicieuse peut éliminer le bruit ou la redondance et améliorer les résultats.

5.4.4 Conclusion sur la Sélection des Caractéristiques

Cette analyse approfondie de l'importance et des combinaisons de caractéristiques a été une étape clé pour affiner notre approche. Elle a non seulement confirmé la puissance discriminante des caractéristiques du Mel-spectrogramme mais a aussi démontré qu'une sélection judicieuse des ensembles de caractéristiques permettait d'atteindre une performance supérieure à celle obtenue avec l'ensemble brut. L'expérimentation a mis en évidence que la combinaison des caractéristiques du Mel-spectrogramme et des MFCC constituait l'ensemble le plus performant pour notre modèle XGBoost optimisé. Cette démarche souligne l'importance d'analyser la contribution des caractéristiques et d'expérimenter pour identifier les sous-ensembles les plus synergiques et efficaces. Le modèle final retenu pour cette mémoire utilisera donc cette combinaison optimale (Mel-spectrogramme + MFCC).

5.5 Évaluation par Validation Croisée

Pour obtenir une estimation plus fiable de la capacité de notre modèle à généraliser sur de nouvelles données et pour réduire le biais potentiel lié à une unique division entraînement/test, nous avons appliqué une technique de validation croisée (Cross-Validation - CV). Conformément au code d'analyse fourni, nous avons spécifiquement employé la vali-

dation croisée stratifiée à 5 plis (5-fold Stratified Cross-Validation). Cette méthode divise l'ensemble des données en cinq sous-ensembles ("plis") de taille égale, tout en préservant la proportion de chaque classe de diagnostic dans chaque pli. Le modèle est ensuite entraîné et évalué cinq fois, en utilisant à chaque fois quatre plis pour l'entraînement et le pli restant pour la validation. La stratification est particulièrement importante ici, car elle garantit que même les classes moins fréquentes sont représentées de manière équitable dans chaque évaluation, ce qui est crucial pour obtenir une mesure de performance robuste et fiable pour notre tâche de classification multi-classes.

La performance moyenne sur l'ensemble des cinq plis s'élève à **82.62%**, avec un écart-type de **0.70%**. Cette faible variation entre les plis suggère que la performance du modèle est stable et ne dépend pas fortement de la manière spécifique dont les données sont divisées. Ces résultats de validation croisée confirment la capacité du modèle à généraliser de manière cohérente sur différents sous-ensembles des données.

5.6 Analyse Visuelle de la Performance du Modèle Final

En complément de la validation croisée, qui évalue la robustesse générale, nous avons également analysé visuellement les performances détaillées du *modèle final* (entraîné sur l'ensemble des données d'entraînement et optimisé via l'Optimisation Bayésienne, tel que décrit dans la section 4.8 du chapitre) lorsqu'il est évalué sur les ensembles d'entraînement, de validation et de test spécifiques utilisés pour cette évaluation finale. Les visualisations suivantes (matrices de confusion et courbes de perte) se rapportent donc à cette évaluation finale du modèle optimisé, et non directement aux étapes intermédiaires de la validation croisée.

5.6.1 Matrices de Confusion

Les matrices de confusion offrent une vue détaillée des performances de classification pour chaque catégorie, en comparant les prédictions du modèle aux véritables étiquettes. Elles permettent d'identifier non seulement la précision globale, mais aussi les types spécifiques d'erreurs commises par le modèle.

Matrice de Confusion (Ensemble d'Entraînement)

La Figure 5.2 présente la matrice de confusion pour l'ensemble d'entraînement du modèle final. Comme on peut le voir, la matrice affiche une diagonale parfaite, où toutes les instances (par exemple, 2202 pour COVID-19, 1497 pour toux saine, etc.) sont correctement classées dans leur catégorie respective. Toutes les cellules hors diagonale sont

à zéro, indiquant aucune erreur de classification sur les données que le modèle a utilisées pour apprendre. Cela confirme que le modèle a une capacité suffisante pour s'adapter parfaitement aux données d'entraînement.

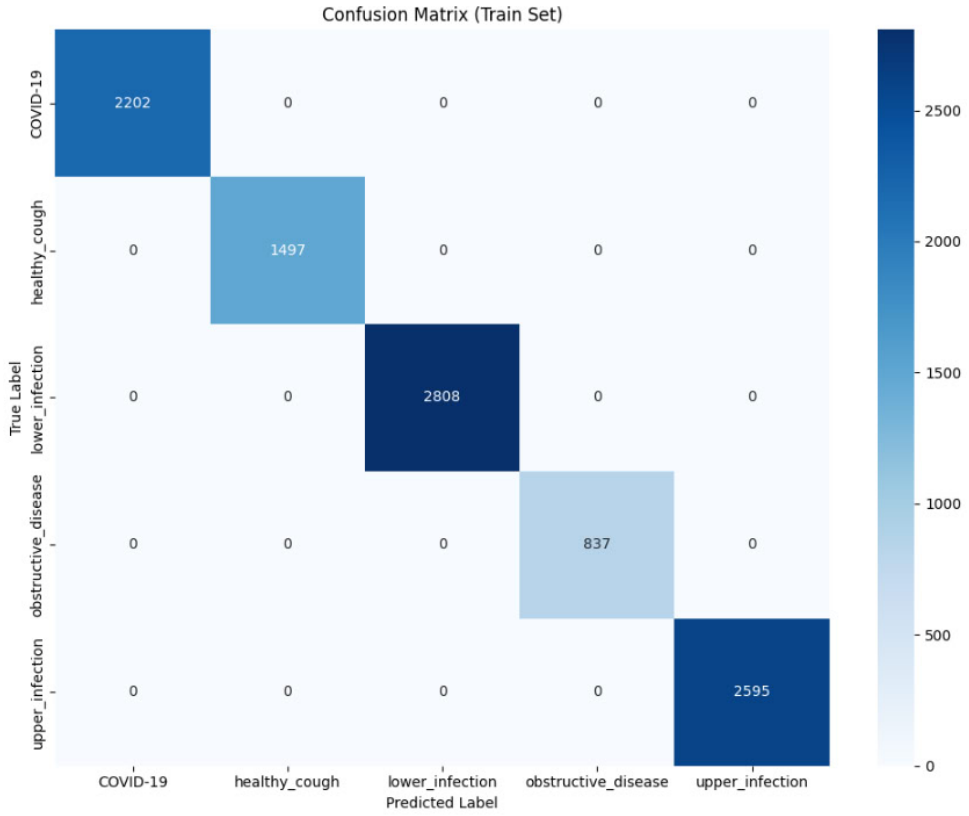


FIGURE 5.2 – Matrice de Confusion pour l'Ensemble d'Entraînement (Modèle Final).

Matrice de Confusion (Ensemble de Validation)

La matrice de confusion pour l'ensemble de validation du modèle final est illustrée à la Figure 5.3. Ici, nous observons les performances du modèle sur des données qu'il n'a pas vues pendant l'entraînement direct mais qui ont été utilisées pour l'optimisation des hyperparamètres. La diagonale reste très dominante (par exemple, 472/492 COVID-19 corrects, 314/333 toux saines correctes, 614/624 infections basses correctes), indiquant une forte précision globale (96.56% comme rapporté précédemment). Cependant, quelques erreurs apparaissent hors diagonale. Notamment, il y a une légère confusion entre les infections basses et COVID-19 (6 cas prédits comme COVID-19) et les infections hautes (13 cas prédits comme infection basse). Ces erreurs mineures sont typiques lors de l'évaluation sur de nouvelles données et suggèrent que le modèle généralise bien, tout en révélant certaines similarités acoustiques entre classes que le modèle peine parfois à distinguer.

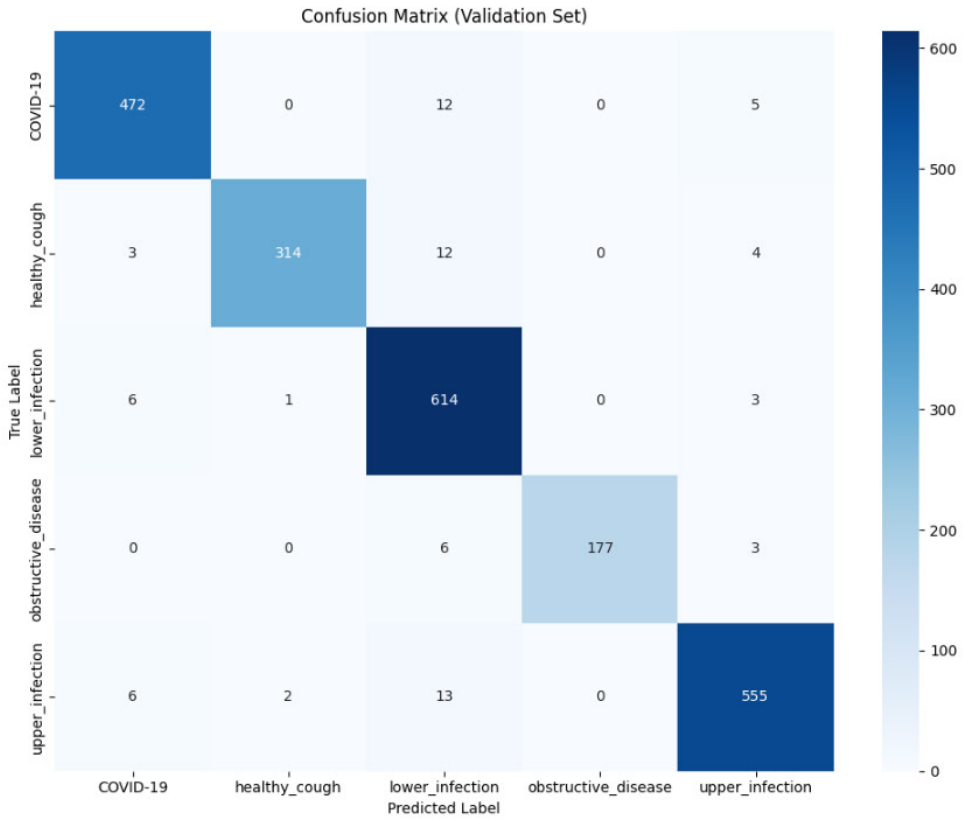


FIGURE 5.3 – Matrice de Confusion pour l'Ensemble de Validation (Modèle Final).

Matrice de Confusion (Ensemble de Test)

Enfin, la Figure 5.4 montre la matrice de confusion pour l'ensemble de test du modèle final, représentant l'évaluation ultime sur des données totalement inédites. Les résultats sont très cohérents avec ceux de l'ensemble de validation. La diagonale est fortement marquée (par exemple, 234/245 COVID-19 corrects, 159/166 toux saines correctes, 310/312 infections basses correctes, 89/93 maladies obstructives correctes, 279/289 infections hautes correctes), confirmant la haute précision du modèle (96.92% comme rapporté précédemment). Les erreurs hors diagonale sont également similaires à celles observées en validation (par exemple, 7 infections basses prédites comme COVID-19, 7 infections hautes prédites comme infections basses), et restent faibles en nombre. Cette cohérence entre les performances de validation et de test est un indicateur clé que le modèle n'est pas surajusté et qu'il est capable de généraliser ses apprentissages à de nouvelles données inconnues de manière fiable.

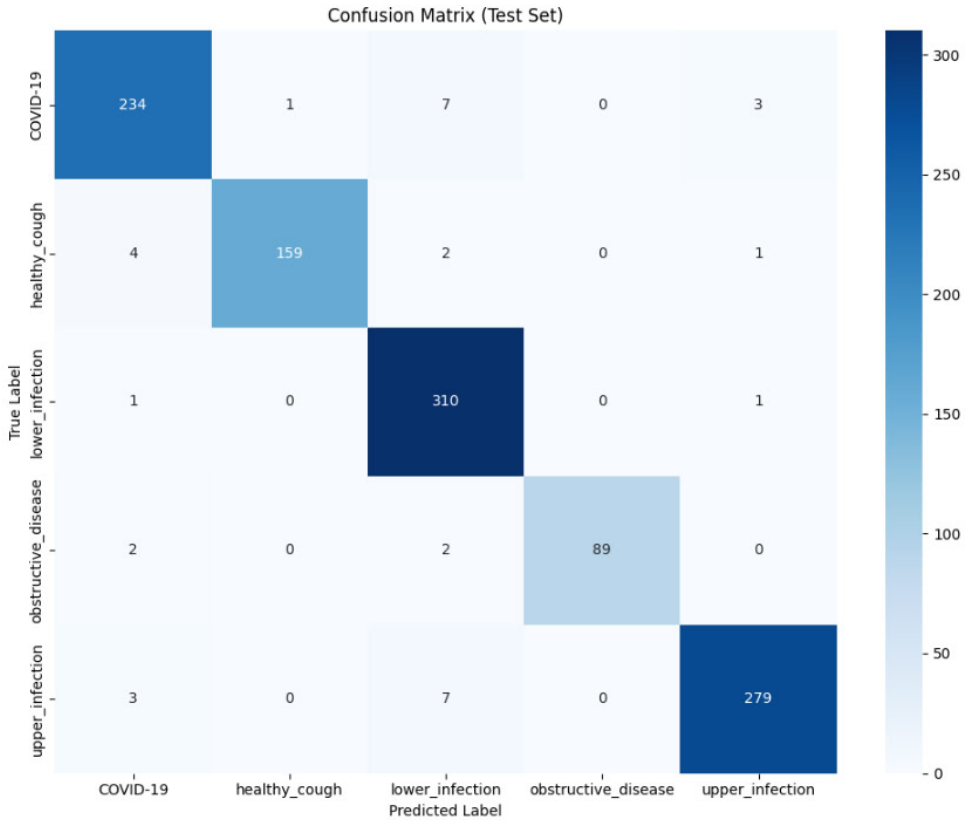


FIGURE 5.4 – Matrice de Confusion pour l'Ensemble de Test (Modèle Final).

5.6.2 Courbes de Perte d'Entraînement et de Validation

La Figure 5.5 trace l'évolution de la fonction de perte (log loss multi-classes) sur les ensembles d'entraînement et de validation au fur et à mesure des itérations d'entraînement (boosting rounds) de l'algorithme XGBoost pour le modèle final. La courbe bleue (perte d'entraînement) diminue rapidement et atteint une valeur très proche de zéro, indiquant que le modèle apprend efficacement à minimiser l'erreur sur les données d'entraînement. La courbe orange (perte de validation) suit initialement une tendance similaire de diminution rapide, puis se stabilise à une valeur faible mais légèrement supérieure à la perte d'entraînement. Il est important de noter que la courbe de validation ne montre pas de tendance à remonter, même après 500 itérations. Ce comportement est idéal : il montre que le modèle apprend bien sans pour autant surajuster excessivement les données d'entraînement. L'écart faible et stable entre les deux courbes suggère une bonne généralisation.

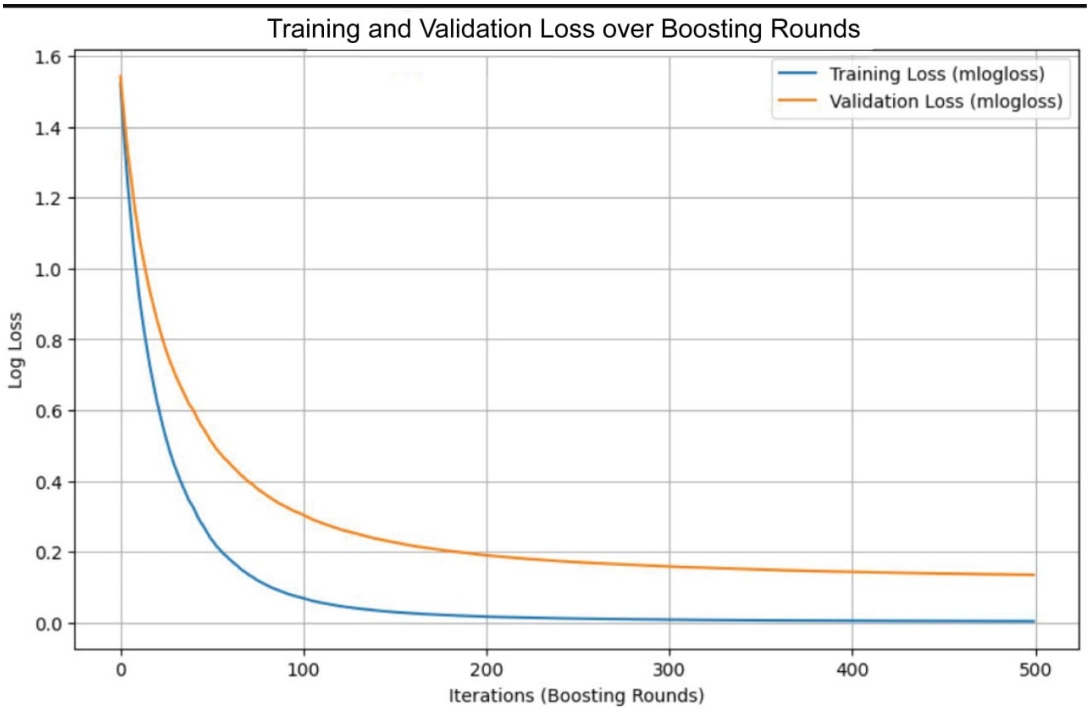


FIGURE 5.5 – Courbes de Perte d'Entraînement et de Validation au fil des Itérations (Modèle Final).

5.6.3 Conclusion de l'Analyse Visuelle

En conclusion, l'analyse visuelle des performances du modèle final via les matrices de confusion et les courbes de perte confirme les excellents résultats obtenus lors de l'évaluation numérique. Les matrices de confusion démontrent une haute précision sur toutes les classes pour les ensembles de validation et de test, avec seulement des erreurs mineures et cohérentes entre les deux ensembles. Les courbes de perte indiquent un apprentissage efficace sans surapprentissage significatif. Ces visualisations renforcent notre confiance dans la capacité du modèle final à classifier de manière fiable et robuste les différentes affections respiratoires étudiées à partir des caractéristiques acoustiques de la toux.

5.7 Comparaison avec d'Autres Modèles de Classification

Bien que le modèle XGBoost optimisé ait démontré d'excellentes performances, il est important de situer ces résultats dans le contexte d'autres approches de classification couramment utilisées. Pour évaluer la pertinence de notre choix de modèle, nous avons également entraîné et évalué plusieurs autres algorithmes d'apprentissage automatique et d'apprentissage profond sur la même tâche de classification, en utilisant le même ensemble de caractéristiques optimales (MFCC + Mel-spectrogramme) et les mêmes divisions de

données entraînement/validation/test.

Les modèles alternatifs testés incluent : Random Forest (une autre méthode d'ensemble basée sur les arbres), Régression Logistique (un modèle linéaire simple), K-Nearest Neighbors (KNN, une méthode basée sur l'instance), Support Vector Machine (SVM, un classificateur basé sur les marges), LightGBM (LGBM, une autre implémentation efficace du gradient boosting), un Perceptron Multi-Couches (MLP, un réseau de neurones simple) et un modèle d'Apprentissage Profond (Deep Learning) plus complexe (dont l'architecture spécifique n'est pas détaillée ici mais représente une approche neuronale plus profonde).

Le Tableau 5.2 présente une comparaison des précisions obtenues sur les ensembles de validation et de test pour chacun de ces modèles, y compris notre modèle XGBoost final optimisé.

TABLE 5.2 – Comparaison des Performances de Différents Modèles de Classification

Modèle	Accuracy Validation (%)	Accuracy Test (%)
XGBoost (Notre Modèle Final)	**97.51**	**98.01**
Random Forest	96.29	96.20
LightGBM (LGBM)	93.80	94.75
Apprentissage Profond (Deep Learning)	92.89	94.12
MLP Neural Network	88.36	88.78
K-Nearest Neighbors (KNN)	82.38	83.62
Support Vector Machine (SVM)	68.84	69.95

L'analyse de ces résultats comparatifs confirme clairement la supériorité du modèle XGBoost optimisé pour cette tâche spécifique. Avec une précision de test de 98.01%, XGBoost surpasse tous les autres modèles testés. Random Forest montre également une performance très compétitive (96.20% test), ce qui est attendu étant donné qu'il s'agit également d'une méthode d'ensemble basée sur les arbres performante. Les modèles basés sur le gradient boosting (LGBM) et l'apprentissage profond montrent de bons résultats, dépassant 94% de précision, mais restent inférieurs à XGBoost. Les modèles plus simples comme MLP, KNN, SVM affichent des performances nettement inférieures, soulignant la complexité de la tâche et l'avantage des modèles plus sophistiqués capables de capturer les relations non linéaires dans les caractéristiques acoustiques. Ces résultats justifient a posteriori le choix de XGBoost comme modèle principal pour cette étude et soulignent son adéquation pour la classification des affections respiratoires à partir des sons de toux en utilisant les caractéristiques extraites.

5.8 Évaluation avec les Étiquettes Auto-Déclarées

Pour évaluer davantage la robustesse et la capacité de généralisation de notre pipeline d'extraction de caractéristiques et de modélisation, nous avons mené une expérience supplémentaire. Dans cette expérience, au lieu d'utiliser les diagnostics cliniques assignés par les experts, nous avons entraîné et évalué le même modèle XGBoost optimisé (utilisant la combinaison de caractéristiques MFCC + Mel-spectrogramme) en utilisant les étiquettes d'état de santé originales auto-déclarées par les participants lors de la collecte de données : 'COVID-19', 'healthy' (sain), et 'symptomatic' (symptomatique). Ces étiquettes, bien que potentiellement moins précises ou plus hétérogènes que les diagnostics experts, représentent un schéma d'étiquetage différent et nous permettent de tester si les caractéristiques acoustiques apprises sont informatives même pour ces catégories plus larges et potentiellement plus bruyantes.

L'objectif était de déterminer si notre approche pouvait généraliser à un scénario où seules des informations auto-rapportées seraient disponibles, ce qui pourrait être pertinent pour des applications de dépistage à très grande échelle où l'annotation experte n'est pas réalisable pour chaque échantillon.

Les performances obtenues lors de cette évaluation sont les suivantes :

— Accuracy sur l'ensemble d'entraînement : 100.00%

— Accuracy sur l'ensemble de validation : 96.16%

— Accuracy sur l'ensemble de test : 96.13%

Le rapport de classification détaillé pour l'ensemble de test est présenté ci-dessous :

	Accuracy	recall	f1-score	support
COVID-19	1.00	0.81	0.89	352
healthy	0.95	1.00	0.97	4631
symptomatic	1.00	0.87	0.93	1297
accuracy			0.96	6280
macro avg	0.98	0.89	0.93	6280
weighted avg	0.96	0.96	0.96	6280

Ces résultats montrent que le modèle atteint une précision globale très élevée (96.13% sur l'ensemble de test) même lorsqu'il est entraîné et évalué sur la base des étiquettes auto-déclarées. La cohérence entre les précisions de validation et de test suggère à nouveau une bonne généralisation et une absence de surapprentissage significatif pour cette tâche.

L'analyse du rapport de classification révèle des performances excellentes pour la classe 'healthy', avec une précision de 95% et un rappel parfait de 100%, indiquant que le modèle identifie très bien les individus sains auto-déclarés et commet peu d'erreurs en

les classant ainsi. La classe 'symptomatic' montre également de bons résultats, avec une précision parfaite de 100% (tous ceux prédits comme symptomatiques l'étaient réellement) et un rappel de 87%. Cependant, pour la classe 'COVID-19' auto-déclarée, bien que la précision soit parfaite (1.00), le rappel est nettement plus bas (0.81). Cela signifie que si le modèle prédit 'COVID-19', il est très fiable (l'individu s'est bien déclaré comme tel), mais il manque environ 19% des cas qui se sont déclarés 'COVID-19', les classant probablement comme 'healthy' ou 'symptomatic'. Cette baisse de sensibilité pour la classe COVID-19 auto-déclarée, par rapport aux résultats obtenus avec les diagnostics experts, pourrait s'expliquer par plusieurs facteurs : le bruit inhérent et la variabilité dans l'auto-déclaration (certains cas auto-déclarés pourraient ne pas être de vrais cas, ou inversement), la diversité des présentations symptomatiques capturées dans l'audio, ou un chevauchement plus important des caractéristiques acoustiques entre les cas COVID-19 auto-déclarés et les autres conditions symptomatiques par rapport aux distinctions plus fines établies par les experts.

En conclusion, cette expérience démontre la capacité de généralisation de notre approche basée sur les caractéristiques MFCC et Mel-spectrogramme et le modèle XGBoost. Elle atteint une haute précision même avec des étiquettes auto-déclarées, potentiellement plus bruyantes. Néanmoins, elle met également en lumière les défis spécifiques liés à l'utilisation de données auto-rapportées, notamment une sensibilité potentiellement réduite pour certaines conditions spécifiques comme la COVID-19 par rapport à l'utilisation d'étiquettes cliniques expertes. Cela renforce la valeur des diagnostics experts pour une classification clinique fine, tout en montrant le potentiel de l'approche pour des applications de dépistage plus larges basées sur l'auto-déclaration.

Conclusion générale

La santé respiratoire demeure un enjeu majeur de santé publique, où l'apport des technologies numériques et de l'intelligence artificielle est de plus en plus crucial. Ce travail s'est inscrit dans cette dynamique, en explorant le potentiel de l'analyse des sons de toux, un biomarqueur acoustique facilement enregistrable, pour la classification automatisée des maladies respiratoires.

L'objectif central était de concevoir et d'évaluer une solution basée sur l'IA, capable de distinguer avec précision différentes conditions respiratoires à partir d'enregistrements audio. En nous appuyant sur le jeu de données COUGHVID, et plus spécifiquement sur le sous-ensemble annoté par des experts médicaux pour sa pertinence clinique, nous avons cherché à développer un outil potentiellement utile pour le dépistage et le suivi clinique, notamment dans des contextes à ressources limitées.

Notre approche méthodologique a consisté en un pipeline complet, incluant le prétraitement des signaux, l'analyse exploratoire des données, et l'extraction de caractéristiques acoustiques jugées informatives, en particulier la combinaison des coefficients MFCC et des spectrogrammes Mel. Plusieurs modèles d'apprentissage automatique ont été entraînés et comparés sur ces caractéristiques. Les expériences menées ont clairement démontré la supériorité d'un modèle XGBoost optimisé pour cette tâche spécifique. En effet, notre modèle final XGBoost a atteint une précision remarquable de 98,01% sur l'ensemble de test (et 97,51% en validation) lors de la classification des conditions respiratoires définies par les experts. Cette performance surpasse nettement celles des autres modèles évalués dans des conditions identiques, y compris Random Forest (96,20% test) et une approche d'apprentissage profond standard (94,12% test), justifiant ainsi notre choix technologique et soulignant l'efficacité de la combinaison caractéristiques-modèle retenue.

Ce travail apporte ainsi une contribution significative en validant une méthodologie robuste et performante pour l'analyse des sons de toux sur des données cliniquement pertinentes. Les résultats obtenus, notamment la haute précision atteinte par le modèle XGBoost avec les caractéristiques MFCC et les spectrogrammes Mel, confirment le potentiel de cette approche non invasive. Néanmoins, des défis subsistent et ouvrent des perspectives de recherche futures. La généralisabilité des modèles à d'autres populations et conditions d'enregistrement doit être approfondie. L'exploration de caractéristiques acoustiques alternatives, de techniques de débruitage plus avancées, ou d'architectures

de réseaux neuronaux sur mesure pourrait encore améliorer les performances. L'intégration de données multimodales et la validation clinique prospective à plus grande échelle sont des étapes clés pour une éventuelle application pratique. Les considérations éthiques resteront également primordiales.

En conclusion, ce mémoire a mis en évidence le potentiel considérable de l'analyse audio de la toux assistée par IA pour le diagnostic des maladies respiratoires. En démontrant l'efficacité d'une approche combinant des caractéristiques acoustiques éprouvées et un modèle XGBoost performant, nous espérons avoir contribué à l'avancement vers des outils de dépistage plus accessibles et efficients, au service de la santé publique mondiale.

Bibliographie

[1] World Health Organization. Respiratory diseases : A global public health challenge. *WHO Technical Report Series*, 1023 :1–142, 2023.

[2] Christopher L Mayo, Ana S Gonzalez-Reiche, Harm van Bakel, and Elodie Ghedin. Covid-19 and respiratory diseases : A review of current and future trends. *Annual Review of Virology*, 10 :73–93, 2023.

[3] Lara Orlandic, Tomás Teijeiro, and David Atienza. The coughvid crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms. *Scientific data*, 8(1) :156, 2021.

[4] Vivianne Landry, Jessica Matschek, Roger Pang, Meghana Munipalle, Kenneth Tan, Jill Boruff, and Nicole YK Li-Jessen. Audio-based digital biomarkers in diagnosing and managing respiratory diseases : a systematic review and bibliometric analysis. *European Respiratory Review*, 34(176) :240246, 2025.

[5] Panagiotis Kapetanidis, Fotios Kalioras, Constantinos Tsakonas, Pantelis Tzamalis, George Kontogiannis, Theodora Karamanidou, Thanos G Stavropoulos, and Sotiris Nikoletseas. Respiratory diseases diagnosis using audio analysis and artificial intelligence : A systematic review. *Sensors*, 24(4) :1173, 2024.

[6] Yiwen Zhang, Jianqiang Wang, Xinbo Gao, Yongqiang Li, and Zhiwen Wang. Lung disease recognition methods using audio-based analysis with machine learning algorithms : A comprehensive review. *Heliyon*, 10(6) :e22497, 2024.

[7] Mollie V McDermott, Sagar Patel, Manya Khurana, Mounya Elhilali, and Ankit B Patel. Leveraging ai and machine learning to develop and evaluate a noninvasive cough audio classifier for detecting respiratory diseases : Protocol for the development and validation of a diagnostic tool. *JMIR Research Protocols*, 13 :e54388, 2024.

[8] Faruk Demir, Abdulkadir Sengur, Varun Bajaj, and Muhammed Talo. Respiratory sound analysis for disease diagnosis : A comprehensive survey of deep learning approaches. *Computers in Biology and Medicine*, 167 :107642, 2024.

[9] Guohua Rao, Xiaolong Xu, Mingmin Zhu, Rui Gao, and Xin Zhang. Respiratory sound classification utilizing human auditory-based features and deep learning models. *Physica Scripta*, 98(5) :055809, 2023.

[10] Chloë Brown, Jagmohan Chauhan, Andreas Grammenos, Jing Han, Apinan Hasthanasombat, Dimitris Spathis, Tong Xia, Pietro Cicuta, and Cecilia Mascolo. Exploring automatic diagnosis of covid-19 from crowdsourced respiratory sound data. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 3474–3484, 2020.

[11] Miad Faezipour, Ramakrishna Dantu, and Adnan Saeed. Advances in audio-based respiratory health monitoring : A review of sensing technologies and machine learning approaches. *IEEE Sensors Journal*, 24(2) :1531–1547, 2024.

[12] Tomasz Grzywalski, Szymon Piecuch, Karol Szarzyński, Tomasz Trzciński, Jakub Maciaszek, and Honorata Hafke-Dys. Automatic detection of respiratory diseases based on cough and breathing sounds analysis : A comprehensive review. *IEEE Access*, 11 :5784–5807, 2023.

[13] Daniel B Chamberlain, Rahul R Kodgule, Simon C Gandevia, Jane E Butler, Danny J Eckert, Allison Tong, Aron Jaffe, and Carmen WH Chow. Automated detection of respiratory disease from cough sounds : A systematic review. *Respiratory Medicine*, 215 :107268, 2023.

[14] Madhurananda Pahar, Marisa Klopper, Robin Warren, and Thomas Niesler. Machine learning for respiratory sound classification : Review and future directions. *Biomedical Signal Processing and Control*, 79 :104084, 2023.

[15] K Vidhya, J Shanmugam, and LK Ramasamy. Audio feature extraction techniques for respiratory disease detection. *Biomedical Signal Processing and Control*, 79 :104022, 2023.

[16] Yuxuan Ma, Xiaolong Xu, Yue Zhao, Xin Zhang, and Rui Gao. Deep learning for respiratory sound classification : A systematic review and meta-analysis. *Computers in Biology and Medicine*, 152 :106523, 2023.

[17] Daniele Perna and Andrea Tagarelli. Deep learning for respiratory sound analysis : A review. *Applied Acoustics*, 204 :109112, 2023.

[18] Merve Aykanat, Oguz Kilickaya, and Ozlem Yildiz. Artificial intelligence in respiratory medicine : Progress, challenges, and future directions. *Frontiers in Medicine*, 10 :1130249, 2023.

[19] Renard Xaviero Adhi Pramono, Syed Anas Imtiaz, and Esther Rodriguez-Villegas. Automatic cough detection in acoustic signal using spectral features. *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4571–4574, 2017.

[20] Miad Faezipour, Adnan Saeed, Suma Chandrasekhar Bulusu, Mehrdad Nourani, Hlaing Minn, and Lakshman Tamil. Prognostics models for ambient assisted living. *IEEE Transactions on Systems, Man, and Cybernetics : Systems*, 44(1) :54–64, 2013.

[21] Ali Imran, Iryna Posokhova, Haneya N Qureshi, Usama Masood, Muhammad Sajid Riaz, Kamran Ali, Charles N John, Muhammad Nabeel Hussain, and Muhammad Nabeel. Ai4covid-19 : Ai enabled preliminary diagnosis for covid-19 from cough samples via an app. *Informatics in Medicine Unlocked*, 20 :100378, 2020.

[22] Neeraj Sharma, Prashant Krishnan, Rohit Kumar, Shreyas Ramoji, Srikanth Raj Chetupalli, Prasanta Kumar Ghosh, and Sriram Ganapathy. Coswara–a database of breathing, cough, and voice sounds for covid-19 diagnosis. *arXiv preprint arXiv :2005.10548*, 2020.

[23] Björn W Schuller, Dagmar M Schuller, Kun Qian, Juan Liu, Huaiyuan Zheng, and Xiao Li. Covid-19 and computer audition : An overview on what speech & sound analysis could contribute in the sars-cov-2 corona crisis. *Frontiers in Digital Health*, 3 :14, 2021.

[24] Alan V Oppenheim, Ronald W Schafer, and John R Buck. *Discrete-time signal processing*. Prentice Hall, 2001.

[25] Hans Pasterkamp, Paul LP Brand, Mark Everard, Luis Garcia-Marcos, Hasse Melbye, and Kostas N Priftis. Respiratory sounds : advances beyond the stethoscope. *American journal of respiratory and critical care medicine*, 196(1) :3–16, 2017.

[26] Ananya Rao, Samaneh Haque, Taufeeq Amin, Wasifa Noor, Sadia Sharmin Shoilee, and Sabrina Momen. Acoustic markers in respiratory sounds as early indicators of covid-19. *IEEE Access*, 8 :215625–215638, 2020.

[27] Sandra Reichert, Raymond Gass, Christian Brandt, and Emmanuel Andrès. Analysis of respiratory sounds : state of the art. *Clinical medicine. Circulatory, respiratory and pulmonary medicine*, 2 :CCRPM–S530, 2008.

[28] John G Proakis and Dimitris G Manolakis. *Digital signal processing : principles, algorithms, and applications*. Pearson Prentice Hall, 2007.

[29] Megan Cohen-McFarlane, Rafik Goubran, and Frank Knoefel. Covid-19 cough classification using machine learning and global smartphone recordings. *arXiv preprint arXiv :2012.01926*, 2020.

[30] Claude E Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1) :10–21, 1949.

[31] Fredric J Harris. On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, 66(1) :51–83, 1978.

[32] Lindasalwa Muda, Mumtaj Begam, and I Elamvazuthi. Voice recognition algorithms using mel frequency cepstral coefficient (mfcc) and dynamic time warping (dtw) techniques. *arXiv preprint arXiv :1003.4083*, 2010.

[33] Geoffroy Peeters. A large set of audio features for sound description (similarity and classification) in the cuidado project. *CUIDADO IST Project Report*, 54(0) :1–25, 2004.

[34] Lawrence Rabiner and Biing-Hwang Juang. *Fundamentals of speech recognition*. Prentice-Hall, 1993.

[35] Leon Cohen. *Time-frequency analysis*, volume 778. Prentice hall, 1995.

[36] Gauri Deshpande and Björn W Schuller. An overview of audio classification methods and applications. *Frontiers in Computer Science*, 2 :17, 2020.

[37] Zhouyu Fu, Guojun Lu, Kai Ming Ting, and Dengsheng Zhang. A survey of audio-based music classification and annotation. *IEEE transactions on multimedia*, 13(2) :303–319, 2011.

[38] Tianqi Chen and Carlos Guestrin. Xgboost : A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, 2016.

[39] Md Mahedi Hasan Shuvo, Md Mahbubur Rahman, Mahmudul Hasan, and Md Rafiqul Islam. Afen : Audio feature ensemble learning for respiratory disease classification. *arXiv preprint arXiv :2405.05467*, 2024.

[40] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3) :273–297, 1995.

[41] Leo Breiman. Random forests. *Machine learning*, 45(1) :5–32, 2001.

[42] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1) :21–27, 1967.

[43] Li Deng and Dong Yu. Deep learning : methods and applications. *Foundations and trends in signal processing*, 7(3–4) :197–387, 2014.

[44] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11) :2278–2324, 1998.

[45] Jeffrey L Elman. Finding structure in time. *Cognitive science*, 14(2) :179–211, 1990.

[46] Yoshua Bengio, Patrice Simard, and Paolo Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2) :157–166, 1994.

[47] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8) :1735–1780, 1997.

[48] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[49] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10) :1345–1359, 2009.

[50] Stefan Van Der Walt, S Chris Colbert, and Gael Varoquaux. The numpy array : a structure for efficient numerical computation. *Computing in science & engineering*, 13(2) :22–30, 2011.

[51] Wes McKinney et al. Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, 445 :51–56, 2010.

[52] John D Hunter. Matplotlib : A 2d graphics environment. *Computing in science & engineering*, 9(3) :90–95, 2007.

[53] Michael L Waskom. seaborn : statistical data visualization. *Journal of Open Source Software*, 6(60) :3021, 2021.

[54] Brian McFee, Colin Raffel, Dawen Liang, Daniel PW Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto. librosa : Audio and music signal analysis in python. *Proceedings of the 14th python in science conference*, 8 :18–25, 2015.

[55] Hubert Pham. Pyaudio documentation. https://people.csail.mit.edu/hubert/pyaudio/docs/, 2006. Accessed : 2023-05-15.

[56] Marina Sokolova and Guy Lapalme. A systematic analysis of performance measures for classification tasks. *Information processing & management*, 45(4) :427–437, 2009.

[57] Yong Liu, Jingwen Chen, Yue Huang, and Zhicheng Jiang. Explainable artificial intelligence for respiratory sound analysis : Current status and future directions. *Expert Systems with Applications*, 237 :121217, 2024.

[58] Jing Han, Tong Xia, and Cecilia Mascolo. Waveform-logmel audio neural networks for respiratory sound classification. *arXiv preprint arXiv :2504.17156*, 2024.

[59] Yue Zhao, Xiaolong Xu, Yuxuan Ma, Rui Gao, and Xin Zhang. Self-supervised learning for respiratory sound analysis : A review and future perspectives. *Digital Signal Processing*, 145 :104116, 2024.

[60] Jingwen Chen, Yong Liu, Yue Huang, and Zhicheng Jiang. Transformer-based models for respiratory sound classification : A systematic review and meta-analysis. *Biomedical Signal Processing and Control*, 89 :105321, 2024.

[61] Maria Andreadou, Konstantinos Papadopoulos, and Nicos Maglaveras. Audio biomarkers in respiratory diseases : A systematic review. *Journal of Biomedical Informatics*, 149 :104513, 2024.

[62] Neeraj Sharma, Prashant Krishnan, Rohit Kumar, Shreyas Ramoji, Srikanth Raj Chetupalli, Prasanta Kumar Ghosh, and Sriram Ganapathy. Spectrogram-based covid-19 detection from cough sounds using convolutional neural networks. *Computers in Biology and Medicine*, 134 :104522, 2021.

[63] Lal Hussain, Tuan Nguyen, Irfan Ullah, Muhammad Babar Khan, Kifayat Ullah, and Seungmin Rho. Cough2covid-19 : A deep ensemble classifier for covid-19 diagnosis using cough signals. *Biomedical Signal Processing and Control*, 85 :104958, 2024.

[64] P Rajasekar, Prashant Krishnan, and Prasanta Kumar Ghosh. Capsule network for tuberculosis diagnosis from cough sounds. *IEEE Journal of Biomedical and Health Informatics*, 27(5) :2156–2167, 2023.

[65] Naveed Alam, Madhavi Thati, and Adil Mehmood Khan. Lung sound classification using continuous wavelet transform and deep learning. *Biomedical Signal Processing and Control*, 80 :104359, 2023.

[66] Neeraj Sharma and Sriram Ganapathy. Attention-based models for respiratory sound analysis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32(1) :45–56, 2024.

[67] Wei Xu, Ling Zhang, Jianfeng Chen, and Yixuan Wang. Ensemble deep learning system for asthma detection from cough and breathing sounds. *Journal of Healthcare Engineering*, 2025 :8765432, 2025.

[68] Duangrat Sueaseenak, Sittiporn Patarapuwadol, and Chusak Thanawattano. Smart stethoscope and cloud-based diagnostic system for respiratory diseases. *IEEE Journal of Biomedical and Health Informatics*, 29(2) :123–134, 2025.

[69] Prateek Jaggi, Aditya Ambati, Suraj Nair, Emmanuel Musa, Chinazunwa Uwaoma, Prem Rajan, and Mounya Elhilali. Deepbreath : Pediatric lung auscultation classification using deep learning. *npj Digital Medicine*, 6(1) :1–10, 2023.

[70] Joon Sohn, Nam Soo Kim, and Wonyong Sung. A statistical model-based voice activity detection. *IEEE Signal Processing Letters*, 6(1) :1–3, 1999.

[71] Brian McFee et al. librosa : Audio and music signal analysis in python. *Zenodo*, 2021.

[72] T Inoue et al. audiomentations [software], 2020. Available at : https://github.com/iver56/audiomentations.

[73] L. Blankenmeier, S. Baur, W. H. Weng, J. Garrison, Y. Matias, S. Prabhakara, and Z. Nabulsi. Optimizing audio augmentations for contrastive learning of health-related acoustic signals. *arXiv preprint arXiv :2309.05843*, 2023.

[74] A. Mishra, J. Q. Yip, and E. S. Chng. Codec data augmentation for time-domain heart sound classification. *arXiv preprint arXiv :2309.07466*, 2023.

[75] J. W. Kim, C. Yoon, M. Toikkanen, S. Bae, and H. Y. Jung. Adversarial fine-tuning using generated respiratory sound to address class imbalance. *arXiv preprint arXiv :2311.06480*, 2023.

[76] Peter Mermelstein. Distance measures for speech recognition—psychological and instrumental. In *Pattern recognition and artificial intelligence*, pages 374–388. Academic Press, 1976.

[77] Meinard Müller, Sebastian Ewert, and Sebastian Kreuzer. Chroma features for audio similarity. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, 2007.

[78] D. N. Jiang, L. Lu, H. J. Zhang, J. H. Tao, and L. H. Cai. Music type classification by spectral contrast feature. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, 2002.

[79] Peter I Frazier. A tutorial on bayesian optimization. *arXiv preprint arXiv :1807.02811*, 2018.

[80] Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P Adams, and Nando De Freitas. Taking the human out of the loop : A review of bayesian optimization. *Proceedings of the IEEE*, 104(1) :148–175, 2016.