

UNIVERSITÉ DE BORDEAUX
Collège Sciences et Technologies
STAGE RECHERCHE MASTER 2

IDENTIFICATION AUTOMATIQUE DE CARACTÉRISTIQUES SONORES

APPLICATION AUX SONS PULMONAIRES

CHAMBRES GAËTAN

Sujet proposé par :

DESAINTE-CATHERINE Myriam
HANNA Pierre

Enseignant référant :

FALLERI Jean-Rémi

21 mars 2019

Remerciements

Tout d'abord, j'adresse toute ma gratitude à Monsieur Pierre HANNA, en qualité de maître de stage, ainsi qu'à Madame Myriam DESAINTE-CATHERINE, co-encadrante, pour leur disponibilité, leur soutien, leur compréhension et leur attention. Leur confiance et le partage régulier ont fait de ce stage un apport crucial de connaissances et d'expériences.

Je tiens également à remercier Monsieur Jean-Rémi FALLERI qui aura été un référent de stage disponible et réactif.

Je souhaite aussi remercier Madame Annick MERSIER pour sa disponibilité et sa réactivité qui ont permis d'organiser au mieux ce stage, ainsi que pour son aide et son soutien tout au long de l'année. Je la remercie également, elle, ainsi que toute l'équipe du SCRIME, pour leur accueil chaleureux.

J'adresse également mes remerciements à toutes les personnes qui ont pu m'aider lors de mes recherches durant ce stage, toutes celles qui auront lu et répondu à mes mails de questionnements, et en particulier à Monsieur Gyorgy KURTAG pour sa contribution dans le domaine de la recherche bibliographique sur les sons pulmonaires.

Je souhaite aussi remercier Monsieur Thibault KELLER pour sa disponibilité, son dynamisme et son support technique qui ont permis de disposer de tous les outils nécessaires à l'organisation de ce stage.

Pour finir, je tiens à remercier mon entourage et mes proches qui m'ont soutenu toute l'année, tant dans mes études que lors de ce stage, ainsi que ceux qui m'ont conseillé dans la rédaction de ce rapport ou l'ont relu.

Résumé

En médecine moderne, il est courant que les spécialistes travaillent avec des informations visuelles ou audio, calculées ou mesurées informatiquement. On parle dans ce cas d'images ou de sons numériques. Il existe plusieurs procédés informatiques qui permettent d'analyser, et, dans certain cas, de reconnaître ces informations numériques. À une époque où le nombre de spécialistes de la santé est insuffisant par rapport au travail que représente les diagnostics des patients toujours plus nombreux, il est nécessaire de proposer aux médecins des outils pour les aider.

Le sujet de ce stage a été proposé dans l'objectif d'étudier la faisabilité d'un tel outil : aider les pneumologues à identifier les pathologies chez leurs patients ou former les futurs médecins à reconnaître ces pathologies. Ce stage a consisté à étudier les méthodes possibles pour identifier les sons pulmonaires de patients afin de reconnaître : soit leur état de santé, soit la pathologie dont ils sont atteints, ou encore les symptômes respiratoires audibles.

Ce rapport résume le travail qui a été réalisé sur une période de quatre mois et demi : du 15 Mars 2018 au 31 Juillet 2018. Une étude de l'existant permet de lier informatique musicale, traitement du signal et sons pulmonaires, puis de présenter une base de données d'enregistrements pulmonaires. Les travaux détaillés permettent, par la suite, d'observer quatre approches différentes de la classification de sons pulmonaires. Ces approches font apparaître des résultats assez hétérogènes. Si dans certains cas, les pathologies des patients sont correctement prédites dans plus de 90% des cas, certaines pathologies demeurent difficiles à identifier. Néanmoins, l'ensemble des résultats est encourageant, et permettra de poursuivre ces travaux de recherche lors d'une thèse.

Exceptés les logos officiels présents sur la page de garde, toutes les images, tous les tableaux et tous les schémas présents dans ce rapport sont des créations originales de Gaëtan CHAMBRES. Elles ont toutes été réalisées dans le cadre du stage pour compléter ce rapport.

Table des matières

1	Introduction	1
1.1	Domaine du Stage	1
1.1.1	Le SCRIME	1
1.1.2	Les missions du SCRIME	1
1.1.3	L'équipe du SCRIME	2
1.2	Pourquoi ce projet	2
1.3	Contenu du rapport	2
2	Étude de l'existant	4
2.1	Identification sonore	4
2.1.1	Le son numérique	4
2.1.2	L'informatique et le traitement du signal sonore	5
2.2	Domaine médical	6
2.2.1	La médecine et le multimédia	6
2.2.2	Médecine et informatique	7
2.3	Les sons pulmonaires	8
2.3.1	Les bruits normaux	8
2.3.2	les bruits anormaux	8
3	Base de donnée existante	11
3.1	Le challenge ICBHI	11
3.1.1	Présentation	11
3.1.2	Organisation	11
3.1.3	Répartition	12
3.2	Les données	13
3.2.1	Les pathologies	13
3.2.2	Les annotations	17
3.2.3	Les résultats du challenge	18

4	Bases de travail, méthodes et outils	21
4.1	Approche par apprentissage machine	21
4.1.1	Problème de classification	21
4.1.2	Arbre de décision	24
4.2	Outils de développement	26
4.2.1	Veille scientifique	26
4.2.2	Programmation	26
4.3	Représentation du contenu audio	27
4.3.1	Segmentation de la respiration	27
4.3.2	Calcul des descripteurs audio	27
4.3.3	Exploitation des données	30
5	Les différentes approches étudiées	32
5.1	Classification en fonction des types de cycles respiratoires	32
5.1.1	Les modèles de classification	32
5.1.2	Mise en oeuvre	34
5.1.3	Résultats obtenus	38
5.2	Classification en fonction des différentes pathologies	39
5.2.1	Intérêt	39
5.2.2	Mise en oeuvre	40
5.2.3	Résultats obtenus selon différentes répartitions	41
5.3	Classification appliquée aux enregistrements	44
5.3.1	Intérêts	44
5.3.2	Mise en oeuvre	45
5.3.3	Résultats	45
5.4	Classification à l'échelle du patient	46
5.4.1	Intérêt	46
5.4.2	Mise en oeuvre	46
5.4.3	Résultats	47
6	Bilan	48
6.1	Discussion	48
6.2	Apport personnel	49
6.3	Perspectives	50
	Bibliographie	52
	Identification audio	52
	Informatique médicale	53
	Études de sons cardiaques	53
	Études de sons pulmonaires	53
	Challenge ICBHI	54

Essentia	54
Réseaux de neurones profonds	54
Divers	55
Annexes	57
Partie 1 : Fichier de descripteurs	57
Partie 2 : Matrices de confusion	59
Partie 3 : Publication scientifique	61

Chapitre 1

Introduction

1.1 Domaine du Stage

1.1.1 Le SCRIME

Ce stage a été effectué au sein du Studio de Création et de Recherche en Informatique et Musiques Expérimentales (ou Électro-acoustiques).[33]

Ce Groupement d'Intérêt Scientifique (GIS) et Artistique est rattaché à l'équipe Image et Son du Laboratoire Bordelais de Recherche en Informatique (LaBRI).

Les champs d'activités du SCRIME s'étendent de la recherche scientifique à la création artistique, en passant par la formation, la diffusion des musiques contemporaines et la pédagogie en milieu scolaire et universitaire.

Le SCRIME est soutenu par le Ministère de la Culture et de la Communication, la Ville de Bordeaux, la Région Aquitaine, la Direction Régionale des Affaires Culturelles (DRAC), le Centre National de la Recherche Scientifique (CNRS) et Bordeaux INP.

1.1.2 Les missions du SCRIME

Les différentes missions du SCRIME [33] sont :

- Favoriser la collaboration entre scientifiques et artistes à des fins de recherche et de création,
- Favoriser la diffusion internationale des résultats des recherches et des oeuvres artistiques,
- Promouvoir les collaborations scientifiques et artistiques avec d'autres centres de recherche et de création en France ou à l'étranger,
- Favoriser les actions pédagogiques dans les institutions partenaires ainsi que dans les autres établissements publics d'enseignement et de création.

1.1.3 L'équipe du SCRIME

L'équipe de ce groupement d'intérêt scientifique et artistique est composée de :

Myriam Desainte-Catherine : Directrice scientifique du SCRIME, chercheuse en informatique et professeur des Universités

Edgar Nicouleau : Directeur artistique du SCRIME, compositeur et professeur de musique électro-acoustique

Annick Mersier : Gestionnaire du SCRIME

Pierre Hanna : Chercheur en informatique et analyse musicale

Thibaud Keller : Réalisateur en informatique musicale et ingénieur

Raphaël Marczak : Chercheur en game-studies et ingénieur informatique

Gyorgy Kurtag : Artiste et compositeur

1.2 Pourquoi ce projet

Actuellement, l'identification automatique de caractéristiques sonores est un enjeu majeur dans le domaine du son numérique. Une identification automatique fiable est peu simple de part la complexité de la description audio ou musicale, la robustesse des méthodes existantes mais aussi la masse de données audio à traiter. Dans le domaine musical, il existe de nombreux travaux visant à classer des titres selon différents critères, le genre musical par exemple [1] [2].

En médecine moderne, les diagnostics sont souvent établis selon des informations visuelles ou audio qui peuvent être calculées ou mesurées par ordinateur. On est alors en présence d'images ou de sons numériques, qui peuvent être analysés automatiquement. Si le calcul automatique de descripteurs du signal audio permet la classification de genre sur des titres musicaux ou d'identifier certains types de sons (avions, trains), il est alors envisageable que l'approche soit intéressante pour classer d'autres types de sons comme des sons cardiaques ou pulmonaires.

S'il existe déjà plusieurs travaux concernant l'étude d'enregistrements cardiaques, l'étude de sons pulmonaires a été moins explorée, en particulier avec des méthodes récentes comme l'apprentissage machine. Réaliser ce stage impliquait donc d'étudier l'analyse de sons pulmonaires à l'aide de méthodes d'apprentissage machine et lier ainsi le domaine de l'informatique musicale avec le domaine médical.

1.3 Contenu du rapport

Durant ce stage, plusieurs objectifs étaient visés. Le premier travail (Chapitre 2) aura été un travail bibliographique visant à recenser les différents travaux déjà

réalisés sur l'analyse de sons pulmonaires. Cette étape est nécessaire pour se renseigner sur les caractéristiques du signal audio que l'on souhaite étudier dans un enregistrement de respiration, et également sur la manière d'obtenir ces informations.

Une fois l'état de l'art des méthodes effectué, il a été nécessaire de choisir une base de données qui permettait de tester les méthodes (Chapitre 3). Il fallait trouver une base de données avec un grand nombre d'éléments, uniformisée et annotée, pour pouvoir finalement poser les bases de travail (Chapitre 4).

Pour finir, il aura fallu tester plusieurs méthodes de classification de sons pulmonaires afin de comparer les différents résultats (Chapitre 5).

Cela a permis de tirer des conclusions (Chapitre 6) quant à l'apport d'une approche par classification grâce à l'apprentissage machine sur l'étude des sons de la respiration mais également les perspectives envisageables pour poursuivre les travaux de ce stage.

Chapitre 2

Étude de l'existant

2.1 Identification sonore

2.1.1 Le son numérique

Depuis plusieurs décennies, grâce au développement de l'informatique et d'internet, le son s'est numérisé. C'est particulièrement visible par la fulgurante démocratisation du streaming musical [34] et les changements de mode de consommation de la musique [35]. Chaque enregistrement sonore est un signal, et ce signal peut être informatiquement représenté grâce à la technique de l'échantillonnage.

Cette technique consiste à mesurer des valeurs à des intervalles réguliers sur le signal, afin d'obtenir une suite numérique qui permet de représenter ce signal.

La représentation informatique du signal sonore permet de nombreux traitements, comme la modification du signal en appliquant des transformations sur les valeurs numériques qui le représente, ou encore de l'identification de caractéristiques. Ces caractéristiques sonores peuvent être représentées par des descripteurs.

Il s'agit de valeurs numériques qui vont représenter des données particulières, contenues dans le signal audio. De même, on peut calculer la variation de ces données dans deux domaines différents : le domaine fréquentiel et le domaine temporel [3]. Le domaine fréquentiel permet de visualiser les variations de fréquences qui composent le signal audio. Le domaine temporel quant à lui permet de visualiser les variations d'amplitudes du signal en fonction du temps. Cela permet par exemple de repérer la présence d'événements audio comme des battements. Les figures 2.1 et 2.2 correspondent respectivement aux représentations temporelle et spectrale d'un cycle respiratoire d'une personne saine.

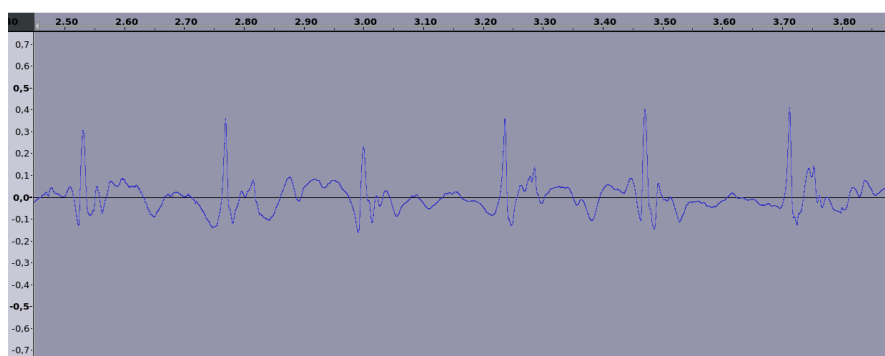


FIGURE 2.1 – Représentation temporelle d’un cycle respiratoire d’une personne saine. Abscisses : amplitude normalisée et ordonnées : temps en secondes.

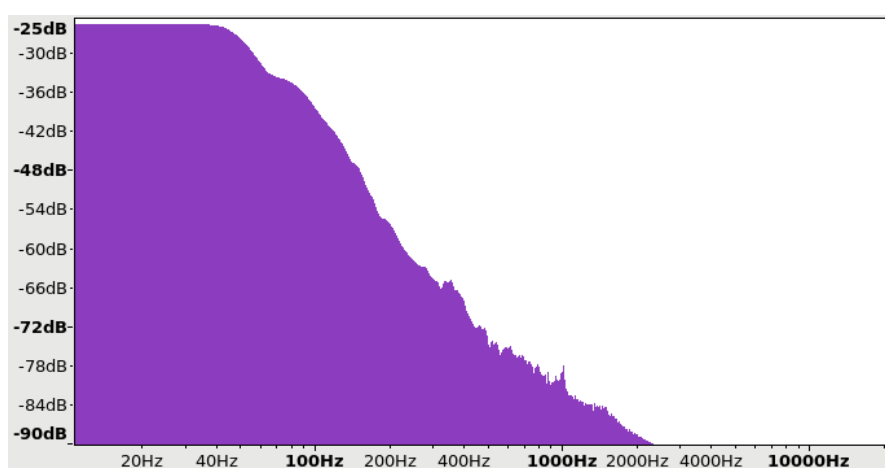


FIGURE 2.2 – Représentation fréquentielle d’un cycle respiratoire d’une personne saine. Abscisses : amplitude en décibels et ordonnées : fréquence en hertz.

2.1.2 L’informatique et le traitement du signal sonore

Les descripteurs audio peuvent permettre d’identifier des particularités, perceptibles ou non, dans le signal, et ainsi, d’effectuer une classification en fonction de ces informations. En effet, les variations des différentes valeurs d’un signal n’évoluent pas de la même manière entre une sonnerie téléphonique et une personne qui parle par exemple, c’est pourquoi une classification est possible en fonction de ces valeurs [4].

Pour obtenir une telle classification de manière automatique, une méthode couramment utilisée est celle de l’apprentissage machine [5]. Cette méthode consiste à établir un algorithme qui va être capable de reconnaître des éléments en fonction

des données qu'il a déjà rencontré précédemment. Plus concrètement, entraîner un modèle consiste à lui donner en entrée un certain nombre d'éléments avec leur description. Lorsque le modèle est entraîné, on peut présenter un nouvel élément. Le modèle calculera alors différentes distances afin d'établir quel est la description d'élément connu qui ressemble le plus à celui nouvellement présenté. Il va alors attribuer le label, c'est à dire l'identifiant, de l'élément qui est le plus ressemblant selon lui.

Le développement des méthodes d'apprentissage machine lié à l'amélioration des performances de calculs a permis d'établir des modèles de classification capable d'identifier des enregistrements audio [6] ou bien de reconnaître certains types de sons comme les bruits environnementaux [7].

L'amélioration des performances a permis également l'apparition de logiciels tels que *OSSR*¹ (Orelia Sound Source Recognition Software) qui permet d'identifier, dans un enregistrement audio, des sons tels que des avions, des trains ou du trafic routier [8]. Ce logiciel est essentiellement utilisé dans le domaine professionnel pour évaluer les impacts acoustiques de ces sources sonores dans un environnement. En urbanisme par exemple, des constructeurs immobiliers cherchent à savoir si une voie ferrée ou une ligne aérienne produisent des nuisances sonores sur le lieu de construction d'un nouveau bâtiment. Pour ce faire, un enregistrement est effectué sur le lieu de construction, et le logiciel essaie d'identifier s'il retrouve, dans cet enregistrement, des informations sonores qui pourrait être le bruit d'un avion ou d'un train.

Cette approche par reconnaissance sonore peut être appliquée au domaine médical, en particulier en cardiologie ou pneumologie, car les spécialistes travaillent régulièrement avec des enregistrement audio. Par exemple, on peut chercher à identifier la présence, non pas d'un train dans un environnement sonore de ville, mais d'une caractéristique sonore particulière dans la respiration d'un patient.

2.2 Domaine médical

2.2.1 La médecine et le multimédia

Comme il l'a été dit, en médecine, les informations importantes peuvent être acquises de différentes façons. Ces informations, quelque soit le moyen d'acquisition, sont le plus souvent représentées sous formes d'images ou de sons. Certains diagnostics sont même directement liés à des informations visuelles ou audio. Par exemple, un médecin peut utiliser un scanner ou un IRM (Imagerie par Résonance Magnétique) pour avoir une représentation du fonctionnement de certains

1. <http://www.orelia.fr/fr/environnement/ossr-le-codage-automatique-des-sources-de-bruit.html>

organes, tandis que la radiographie peut informer de l'état d'une zone osseuse d'un patient. D'autres spécialistes utilisent de l'information audio pour établir leurs diagnostics, en utilisant un stéthoscope ou une échographie par exemple. Être capable d'identifier correctement des problèmes de santé requiert du temps et de l'expérience, seulement, d'après les statistiques de l'organisation Mondiale de la Santé (OMS) [36]. Toutefois, seulement 45% des états membres respectent le ratio médecin/population recommandé (1 médecin pour 1000 habitants).

Avec de si mauvaises statistiques et en prenant en compte qu'un diagnostic implique d'écouter individuellement un patient pendant une durée non compressible et le fait que les spécialistes de la santé sont déjà débordés, les conditions de travail sont très mauvaises pour les spécialistes et des erreurs de diagnostic peuvent être commises.

C'est pourquoi, actuellement, trouver de nouvelles manières d'aider les médecins à gagner du temps est une priorité. Des outils automatiques et fiables peuvent aider à diagnostiquer plus de personnes ou bien aider les spécialistes à commettre moins d'erreur liées à leur charge de travail. En plus d'une aide au diagnostic, il est également possible de proposer aux patients des outils personnalisés pour la surveillance médicale.

L'informatique peut apporter ces nouveaux outils grâce à la constante amélioration des capacités d'analyse. En effet, il existe déjà des exemples d'outils médicaux fiables et rapides basés sur des analyses multimédia [9].

2.2.2 Médecine et informatique

Chaque bilan cardiaque ou pulmonaire inclut une auscultation durant laquelle le spécialiste médical écoute les sons provenant du corps du patient à l'aide de différents outils (stéthoscope, sonographe). Cela montre à quel point l'analyse du son est importante pour la détection des pathologies cardiaques et pulmonaires. Les maladies cardiaques étant la principale cause de mortalité dans le monde [37], de nombreux travaux ont été réalisés au sujet de l'identification des sons cardiaques [10, 11]. Une revue [12] publiée en 2012 traitant des techniques de traitement du signal appliquées à l'analyse des sons cardiaques conclut que *"l'habilité à correctement acquérir, analyser, et interpréter les signaux des sons du coeur pour améliorer les systèmes de diagnostics cliniques est devenu une priorité"*. Depuis, il y a eu beaucoup d'améliorations, comme le traitement des enregistrements effectués en environnement bruyant [13] par exemple. De plus, de nouvelles méthodes permettent d'améliorer drastiquement les contributions au domaine, telles que l'apprentissage machine et l'utilisation de réseaux de neurones profonds [26]. Ces approches permettent de meilleures performances en vision par ordinateur [27, 28] ou en analyse audio [29, 30].

Les sons respiratoires restent moins étudiés que les sons cardiaques, mais il existe tout de même certains travaux [14, 15]. Une revue [16] publiée en 2007 souligne que les précédentes recherches permettent d'identifier des marqueurs, des particularités dans la respiration, comme les crépitants ou les sibilants. Comme expliqué précédemment, avec l'apparition de l'apprentissage machine et les réseaux de neurones profonds, les performances en détection sonore et classification des sons ont été grandement améliorées. Le domaine médical [32], en particulier les études sur les sons pulmonaires, a profité de cette progression [17, 18]. En effet, grâce à l'identification des particularités sonores dans la respiration, l'analyse des sons pulmonaires peut alors être ramenée à un problème de classification [19] : le genre de problème que l'apprentissage machine peut résoudre de manière performante.

2.3 Les sons pulmonaires

L'auscultation pulmonaire fait partie des manipulations de l'examen clinique pulmonaire. Le principe de cette auscultation est d'écouter le son de la respiration pour identifier des particularités sonores. Malgré des difficultés à instaurer des nomenclatures normalisées [20], on distingue tout de même certaines catégories de particularités sonores. Selon la sémiologie pulmonaire [21], les deux principales catégories sont les sons (ou bruits) anormaux et normaux, respectivement qui contiennent des particularités sonores ou non.

2.3.1 Les bruits normaux

Les bruits normaux que l'on peut entendre dans la respiration d'un patient sont appelés *murmures vésiculaire*. Ces bruits ("*adventitious sound*" en nomenclature anglaise) sont essentiellement définis par l'absence des particularités sonores recensées dans la catégorie des bruits anormaux.

2.3.2 les bruits anormaux

Les sons respiratoires classés comme bruits anormaux sont ceux qui possèdent une ou des particularités audibles. On parle également de *sons adventices* ("*adventitious sound*" en nomenclature anglaise). Selon la nomenclature française, il existe de nombreux types de *sons adventices*, tels que les *crépitants* (secs ou humides), les *sibilants* (simples ou complexes), les *ronchis*, les *râles*, les *ronflements*, etc. Durant ce stage et dans le but d'uniformiser le travail avec les bases de données existantes, le choix a été fait de travailler avec la sémiologie anglo-saxonne. Ainsi, les bruits pulmonaires sur lesquels le travail s'est concentré sont catégorisés de la manière suivante :

Les "crackles" : Il s'agit de sons discontinus, explosifs, qui apparaissent généralement à l'inspiration. Par analogie, on peut imaginer les crépitements que pourrais produire de l'huile versée sur une poêle chaude. Leur apparition peut indiquer des pathologies du tissu pulmonaire. Ils sont caractérisés par leur forme d'onde, leur durée et leur position dans le cycle respiratoire. Les figures 2.3 et 2.4 représentent le signal audio d'un cycle respiratoire présentant des *crackles*. On peut remarquer des évènements sonores éparses sur la représentation temporelle (figure 2.3, des pics d'amplitude proches dans le temps).

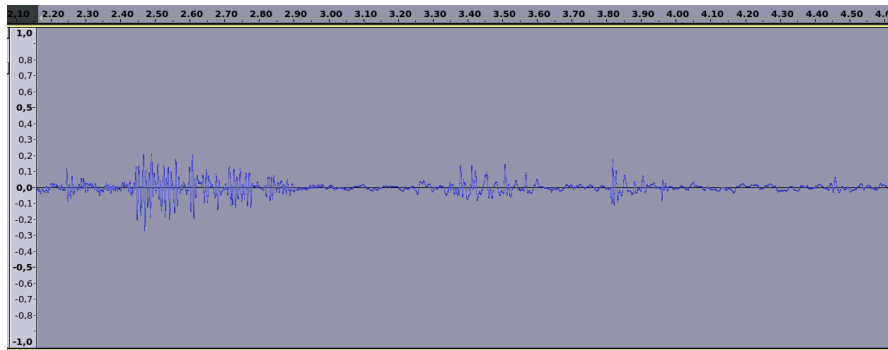


FIGURE 2.3 – Représentation temporelle d'un cycle respiratoire présentant des *crackles*. Abscisses : amplitude normalisée et ordonnées : temps en secondes.

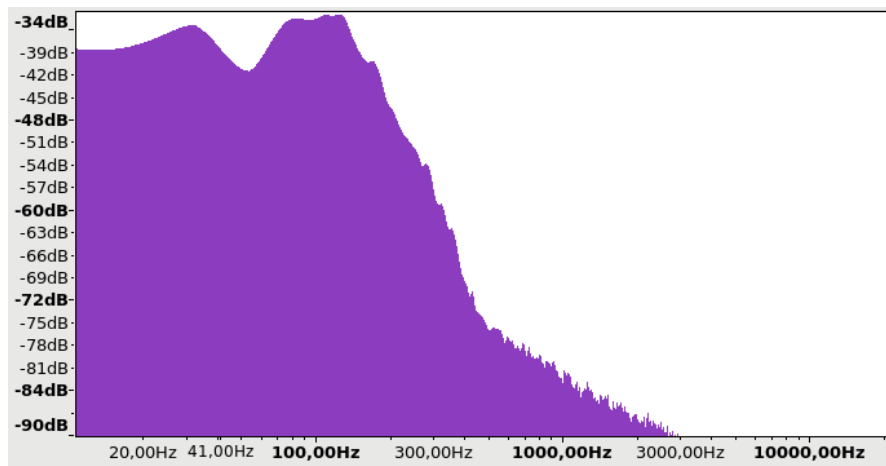


FIGURE 2.4 – Représentation fréquentielle d'un cycle respiratoire présentant des *crackles*. Abscisses : amplitude en décibels et ordonnées : fréquence en hertz.

Les "*wheezes*" : Il s'agit de sons continus qui présentent une hauteur perceptible. On peut entendre un sifflement pendant la respiration. Ils apparaissent par exemple dans l'asthme et la broncho-pneumopathie chronique obstructive. Acoustiquement, il est caractérisé par une forme d'onde avec une fréquence dominante généralement supérieure à 100 Hz et une durée supérieure à 100 ms. De plus, le son doit inclure au moins 10 vibrations successives. Les "*wheezes*" peuvent présenter une seule fréquence (mono-phonique) ou bien plusieurs fréquences (polyphonique). Les figures 2.5 et 2.6 représentent le signal audio d'un cycle respiratoire présentant des *wheezes*. On peut remarquer sur la représentation fréquentielle, figure 2.6, une densité de fréquences plus importante en particulier entre 1000Hz et 3000Hz.

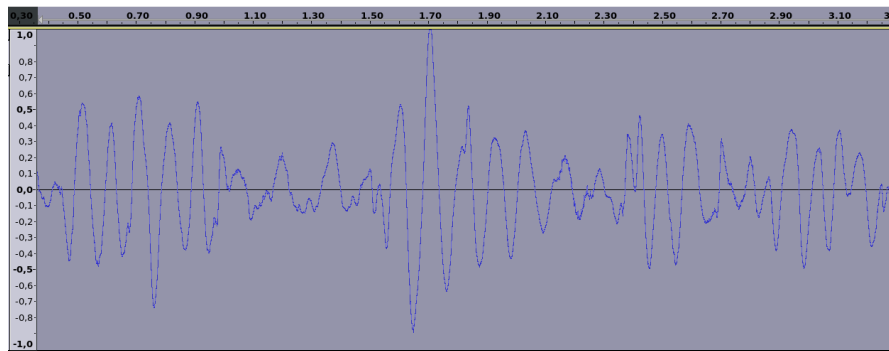


FIGURE 2.5 – Représentation temporelle d'un cycle respiratoire présentant des *wheezes*. Abscisses : amplitude normalisée et ordonnées : temps en secondes.

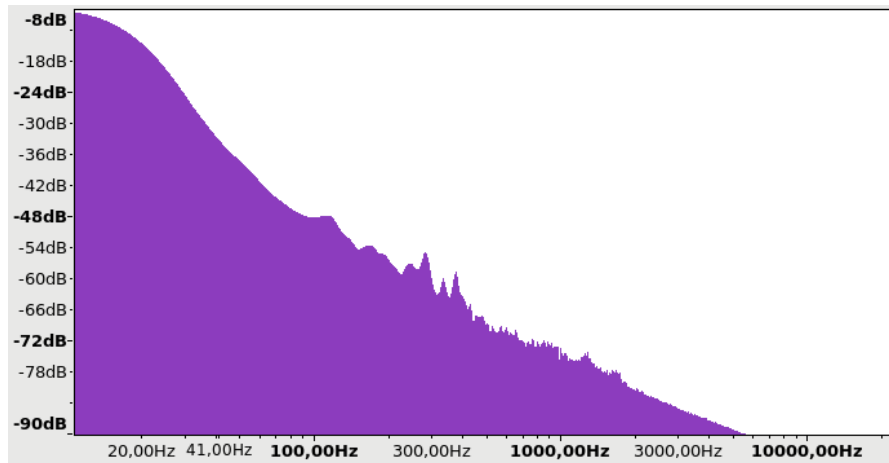


FIGURE 2.6 – Représentation fréquentielle d'un cycle respiratoire présentant des *wheezes*. Abscisses : amplitude en décibels et ordonnées : fréquence en hertz.

Chapitre 3

Base de donnée existante

3.1 Le challenge ICBHI

3.1.1 Présentation

En 2017, *The International Conference on Biomedical and Health Informatics* a organisé un challenge scientifique ayant comme objectif d'identifier la présence de *crackles* et de *wheezes* au sein des enregistrements pulmonaires de plusieurs patients. Pour ce challenge, une base de donnée spécifique [22] était fournie. Elle a été réalisée en collectant des enregistrements de sons respiratoires de deux différentes équipes de recherche, L'équipe ESSUA [38] (*Escola Superior De Saude Da Universidade de Aveiro* à Aveiro au Portugal, et l'équipe AUTH [39] (*Aristotle University of THessaloniki*, à Thessalonique en Grèce.

Les enregistrements ont été effectués sur 126 patients différents (128 initialement, mais 2 d'entre eux ont été retirés de la base de données finale par les organisateurs), comprenant enfants, adultes et personnes âgées, en bonne santé ou bien présentant différentes pathologies. La composition détaillée de la base de données se trouve en partie 3.1.3

3.1.2 Organisation

Le challenge a été organisé en deux parties, la partie de "*train*" et la partie de "*test*". Pour cela, la base de données de tous les enregistrements pulmonaires a été séparée en deux catégories, la partie publique et la partie privée. Lors de la phase de "*train*", les équipes de recherches participant au concours avaient seulement accès à la partie publique des enregistrements, avec leurs annotations. Durant cette phase, les équipes pouvaient tester leur méthodes de classification et utiliser les annotations pour effectuer de l'apprentissage machine. Durant la seconde phase, la phase de "*test*", les participants avaient accès à la partie privée, c'est à dire au

reste des enregistrements, mais sans leurs annotations. Cela permet de tester la précision des différentes méthodes selon un procédé uniformisé et équitable pour chaque participant.

3.1.3 Répartition

La base de données a été séparée en 2 parties, la partie publique appelée "*train*" et la partie privée appelée "*test*". Les différents enregistrements ont été séparés de manière à ce que tous les enregistrements d'un même patient soient dans une seule et même partie. De plus, le ratio suivant a été respecté : 2/3 des enregistrements se situent dans la partie "*train*" pour 1/3 dans la partie "*test*". Cette répartition a été respectée pour permettre de tester des méthodes d'apprentissage machine, qui ont besoin d'avoir plus de données pour s'entraîner qu'il n'y en a à classer. La table 3.1 permet de résumer la répartition des patients, des enregistrements et des différents types de cycles respiratoires selon les différentes parties de la base de donnée. Cette table permet déjà de faire apparaître certaines limites, comme le déséquilibre des deux parties de la base de données entre les cycles avec *crackles* et les cycles avec *wheezes*. Cela permet également de remarquer que les cycles respiratoires normaux sont en majorité dans chacune des deux parties.

TABLE 3.1 – Répartition des données dans la base de donnée du challenge ICBHI. [22]

Base de donnée	Partie " <i>train</i> "	Partie " <i>test</i> "	Total
Nb. patients	79	49	128
Nb. enregistrements	539	381	920
Nb. cycles avec " <i>crackles</i> "	1247	617	1864
Nb. cycles avec " <i>wheezes</i> "	533	353	886
Nb. cycles avec " <i>both</i> "	366	140	506
Nb. cycles normaux	2152	1490	3642
Nb. total de cycles	4298	2600	6898

3.2 Les données

3.2.1 Les pathologies

Dans cette base de données figurent des enregistrements audio de patients ayant des pathologies différentes. Le schéma en figure 3.1 présente les différents organes impliqués dans le processus de respiration et leur position. Grâce à ce schéma, il sera possible de visualiser quelles sont les zones atteintes pour chacune des pathologies présente dans cette base de données.

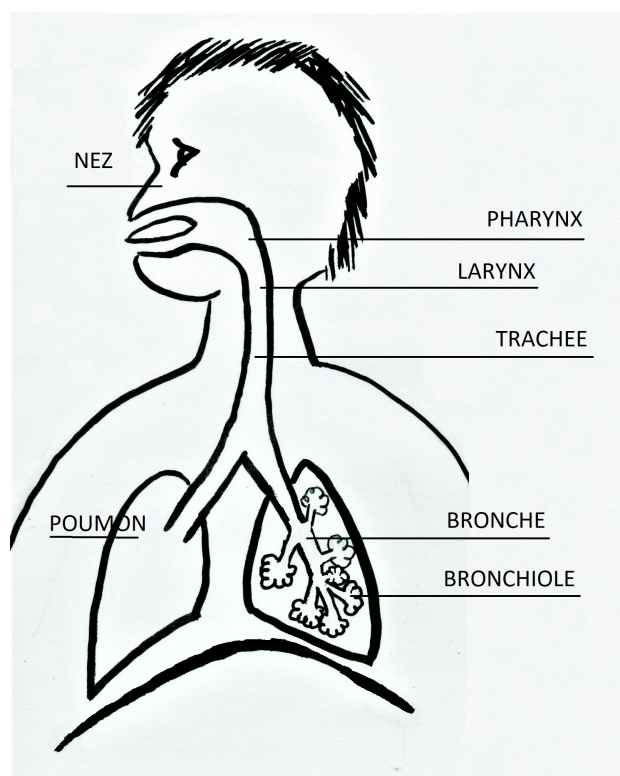


FIGURE 3.1 – Schéma du système respiratoire.

Les différentes pathologies, au nombre total de sept, présentées par les patients de cette base de données sont listées ci-après. Toujours dans le but d'uniformiser les recherches, la dénomination anglophone des pathologies est indiquée, et sera utilisée dans la suite des travaux.

Asthme : Gêne respiratoire souvent caractérisée par un sifflement audible à la respiration. L'ensemble du système respiratoire est impacté par cette pathologie, comme indiqué en figure 3.2. Selon la notation anglophone, on utilisera l'appellation "*Asthma*".

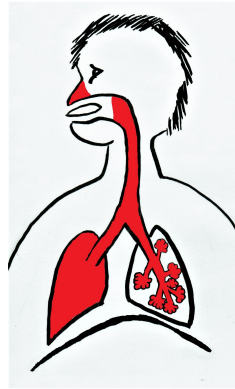


FIGURE 3.2 – Zones du système respiratoire atteintes pour un patient qui présente la pathologie "*Asthma*".

Infection des voies respiratoires inférieures : Il s'agit d'une infection affectant les voies aériennes inférieures incluant la trachée, les bronches et les poumons (voir figure 3.3). Lorsqu'elles ne sont pas traitées, elles peuvent évoluer vers une pneumonie ou une bronchiolite. Ces infections se manifestent principalement par la présence de la toux, des difficultés à respirer, des sifflements dans la respiration et de la fièvre. Selon la notation anglophone ("*Lower Respiratory Tract Infection*"), on utilisera l'appellation "*LRTI*".



FIGURE 3.3 – Zones du système respiratoire atteintes pour un patient qui présente la pathologie "*LRTI*".

Pneumonie Il s'agit d'une infection respiratoire dont les symptômes peuvent varier suivant le germe responsable. On retrouve tout de même régulièrement la présence d'un foyer de crépitants, et la zone la plus impactée concerne généralement les bronchioles (voir figure 3.4). Selon la notation anglophone, on utilisera l'appellation "*Pneumonia*".

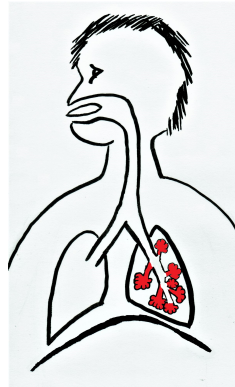


FIGURE 3.4 – Zones du système respiratoire atteintes pour un patient qui présente la pathologie "*Pneumonia*".

Bronchiectasie : Les patients atteints de cette maladie rare ont les voies respiratoires aériennes définitivement endommagées (voir schéma 3.5), et sont donc plus sensibles aux différentes infections respiratoires. Parmi les symptômes typiques qui peuvent apparaître, on retrouve la toux et la présence de sons anormaux et de sifflements dans la respiration. Selon la notation anglophone, on utilisera l'appellation "*Bronchiectasis*".



FIGURE 3.5 – Zones du système respiratoire atteintes pour un patient qui présente la pathologie "*Bronchiectasis*".

Bronchiolite Il s'agit d'une infection qui touche les bronchioles (petites bronches) comme indiqué sur le schéma 3.6. Elle apparaît particulièrement chez les nourrissons. Parmi les symptômes, on peut noter la présence de toux sèche, et une respiration rapide et sifflante. Selon la notation anglophone, on utilisera l'appellation "*Bronchiolitis*".



FIGURE 3.6 – Zones du système respiratoire atteintes pour un patient qui présente la pathologie "*Bronchiolitis*".

Infection des voies respiratoires supérieures Il s'agit d'infection affectant les voies respiratoires supérieures telles que le nez, le pharynx et le larynx. Le schéma 3.7 indique les zones concernées. Les principaux symptômes sont la toux, le renflement, des congestions nasales et le mal de gorge. Selon la notation anglophone ("*Upper Respiratory Tract Infection*"), on utilisera l'appellation "*URTI*".



FIGURE 3.7 – Zones du système respiratoire atteintes pour un patient qui présente la pathologie "*URTI*".

Maladie pulmonaire obstructive chronique Il s'agit d'affections chroniques qui peuvent être de différentes causes, le tabagisme en tête. Cette pathologie implique l'inflammation des voies respiratoires, des bronches à la trachée, impliquant de fortes difficultés à respirer (voir schéma 3.8). Les personnes atteintes présentent généralement une toux persistante et une respiration sifflante. Selon la notation anglophone ("*Common Obstructive Pulmonary Disease*") on utilisera l'appellation "**COPD**".



FIGURE 3.8 – Zones du système respiratoire atteintes pour un patient qui présente la pathologie "**COPD**".

Il existe une huitième catégorie, celle des patients sains, qui contient uniquement les enregistrements de patients ne présentant aucune pathologie. On considère chacun de ces enregistrements comme **normaux**, c'est à dire ne présentant aucune particularité audible.

3.2.2 Les annotations

Comme expliqué précédemment, l'intérêt principal d'utiliser une base de données existante est de pouvoir bénéficier d'annotations déjà présentes et certifiées, bien souvent par des spécialistes. Pour cette base de données, chaque enregistrement audio comporte plusieurs cycles respiratoires successifs d'un patient donné. On parlera de cycles respiratoire pour l'enchaînement d'une inspiration suivie d'une expiration. Pour chacun de ces enregistrement audio, il existe un fichier texte du même nom, qui indique les informations concernant la position dans le temps de chaque cycle respiratoire, ainsi que s'il contient des particularités respiratoires audibles. La table 3.2 est un exemple de fichier d'annotation pour un enregistrement donné.

TABLE 3.2 – Contenu du fichier d’annotation pour l’enregistrement 107_2b4_Al_mc_AKGC417L [22].

Debut	Fin	<i>Crackles</i>	<i>Wheezes</i>
1.018	3.411	1	0
3.411	5.827	1	0
5.827	8.339	1	1
8.339	10.923	1	0
10.923	13.292	0	1
13.292	16.018	1	0
16.018	18.482	1	0
18.482	19.542	1	0

D’après cette table 3.2, chaque ligne correspond à un cycle respiratoire. Dans cet exemple précis, le premier cycle respiratoire débute à 1.018s (seconde), termine à 3.411s et ne contient que des *crackles*. Le cinquième cycle respiratoire commence à 10.923s, termine à 13.292s et ne contient que des *wheezes*. Enfin, d’après le score binaire indiqué, le troisième cycle contient à la fois des *crackles* et des *wheezes*, tandis qu’aucun cycle ne présente aucune particularité respiratoire (le score binaire aurait indiqué 0 pour les *crackles* et 0 pour les *wheezes*. Pour plus de détails sur la répartition des types de cycles respiratoires, se référer à la section 3.1.3.

3.2.3 Les résultats du challenge

Pour ce challenge, les consignes étaient de proposer des méthodes permettant de classer automatiquement les cycles respiratoires selon les bruits respiratoires présents. Afin de procéder à une évaluation équitable et uniformisée de chaque méthode, une métrique précise d’évaluation a été appliquée. Ainsi, les participants ont calculé la sensibilité Se et la spécificité Sp de leur résultats de prédiction, afin d’obtenir comme résultat final la précision Acc (pour *Accuracy* qui est la moyenne de la sensibilité et la spécificité ramenée en pourcentage). Cette précision Acc est également calculée lors de la phase *test*. Elle est alors nommée *Score* et elle est la valeur finale qui permettra de comparer la précision de chacune des méthodes

proposées. Voici les formules utilisées :

$$Se = \frac{(VP_C + VP_W + VP_B)}{(VP_C + FN_C) + (VP_W + FN_W) + (VP_B + FN_B)} \quad (3.1)$$

$$Sp = \frac{VP_N}{VP_N + FN_N} \quad (3.2)$$

$$Acc = \left(\frac{Se + Sp}{2} \right) * 100 \quad (3.3)$$

avec les termes suivants :

VPi signifie *Vrais Positifs* de la classe *i*. *Vrais Positifs* est le groupe d'éléments correctement classés dans la classe en question.

FNi signifie *Faux Négatifs* de la classe *i*. *Faux Négatifs* est le groupe d'éléments incorrectement classés. Ces éléments n'ont pas été classés dans la classe *i* alors qu'ils auraient dû l'être.

VPi+FNi correspond donc au nombre total d'éléments censés se trouver en classe *i*

i correspond aux différentes classes, *C* pour *crackles*, *W* pour *wheezes*, *B* pour *both* qui signifie "les deux" (comprendre à la fois *crackles* et *wheezes*), et enfin *N* pour *normal* (aucune particularité respiratoire audible).

Parmi les résultats soumis lors du challenge, deux méthodes se montrent particulièrement intéressantes car elles abordent la classification des sons pulmonaires tel qu'il était envisagé de le faire durant ce stage. Ainsi, les cycles respiratoires sont classifiés selon les quatre catégories mentionnées précédemment : *crackles*, *wheezes*, *both* et *normal*.

La première méthode : Nommée *HMM* pour *Hidden Markov Model*, elle est basée sur un modèle de Markov caché. Cette approche a été proposée par N. Jakovljević et T. Lončar-Turukalo [23]. Pour cette méthode, après un pré-traitement des enregistrements afin de réduire le bruit et les battements du coeur, des MFCC (*Mel-frequency Cepstral Coefficients*) sont utilisées en entrée d'un modèle de Markov caché. Ce modèle permet de calculer la probabilité qu'un cycle respiratoire a de contenir des *crackles* et des *wheezes*. Les meilleurs scores atteints par cette méthode lors de la phase *train* sont $Se = 0.4232$, $Sp = 0.5669$ et $Acc = 49.50\%$ tandis que le meilleur score atteint lors de la phase *test* est de 39.37% .

La seconde méthode : Nommée *NLSp* pour *Non-Linear Spectral features extraction*, elle correspond à une extraction non linéaire de descripteurs spectraux. Cette approche a été proposée par G. Serbes, S. Ulukaya et Y.P. Kahya [24]. Pour l'application de cette méthode, les enregistrements se voient d'abord appliquer un sous-échantillonnage de 44100Hz à 4000Hz et un filtre *Butterworth* qui permet d'altérer le moins possible la bande passante, afin de retirer des informations audio non pertinentes pour la classification. Après cela, une transformée en ondelettes non linéaire est appliquée. Cette décomposition permet d'obtenir trois canaux (aiguë, grave, résidu). Des descripteurs audio sont ensuite calculés sur chaque canal grâce à l'application d'une transformée de Fourier rapide. Au final, ces descripteurs sont utilisés en entrée d'un SVM (*Support Vector Machine*) qui est un modèle de prédiction. Pour cette méthode, le meilleur score obtenu durant la phase de *test* est de 49.86%. Pour permettre une comparaison uniforme est équitable, les résultats de spécificité et sensibilité présentés sont ceux qui présentent la précision *Acc* la plus proche du score final, à savoir : $Se = 0.4890$, $Sp = 0.7780$ et $Acc = 49.98\%$.

La table 3.3 résume les précisions des prédictions obtenues par ces deux méthodes.

TABLE 3.3 – Résumé des résultats obtenus grâce aux méthodes de l'état de l'art.

Méthoda	Se	Sp	Acc	Score
HMM	0.4232	0.5669	49.50%	39.37%
NLSp	0.4890	0.7780	49.98%	49.86%

Chapitre 4

Bases de travail, méthodes et outils

4.1 Approche par apprentissage machine

4.1.1 Problème de classification

L'objectif de ce stage était d'établir dans quelle mesure l'apprentissage machine est une bonne approche pour reconnaître automatiquement des sons pulmonaires. Pour ce faire, il est nécessaire d'approcher le problème comme un problème de classification. On doit choisir un modèle de prédiction et une forme de représentation du signal, comme les descripteurs audio par exemple. Pour une meilleure compréhension, les différentes étapes de la méthode de classification par apprentissage machine sont détaillées ci-dessous, avec un exemple simplifié comme illustration.

Première étape : Calcul des descripteurs pour la base de données *train*.

Dès le début, l'ensemble de la base de données est divisé en deux parties : une partie pour l'entraînement du modèle de prédiction et une partie pour le test. Lors de la première étape, on calcule une suite de descripteurs représentant le signal audio pour chacun des cycles respiratoires de la base de donnée *train*. Dans l'exemple donné en figure 4.1, cela correspond à prendre tous les objets présents dans la base de donnée d'entraînement (ici un cube et un cylindre) pour calculer les descripteurs de chacun de ces objets (respectivement représentés au dessus et en dessous des objets).

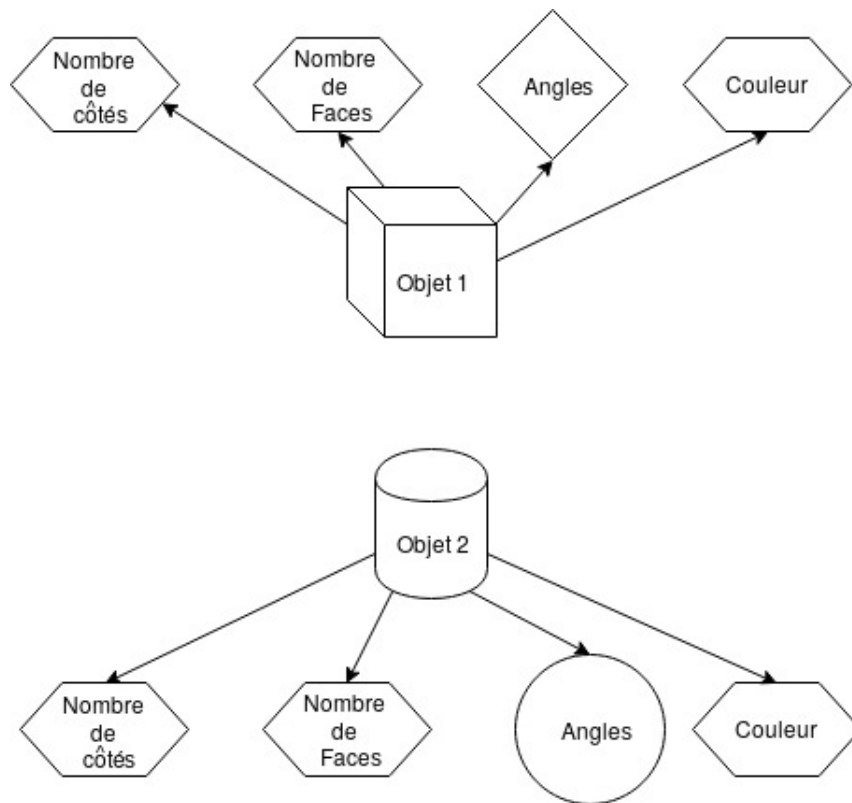


FIGURE 4.1 – Étape 1 : Exemple d'un entraînement de modèle, avec 2 objets distincts et chacun 4 valeurs de descripteurs.

Deuxième étape : Calcul des descripteurs de la base de données *test*.

Lors de la deuxième étape, on calcule une suite de descripteurs représentant le signal audio pour chacun des cycles respiratoires de la base de donnée *test*. Dans l'exemple donné figure 4.2, cela revient à présenter un nouvel objet au modèle, et calculer les mêmes descripteurs que ceux qui ont servis à l'entraînement du modèle.

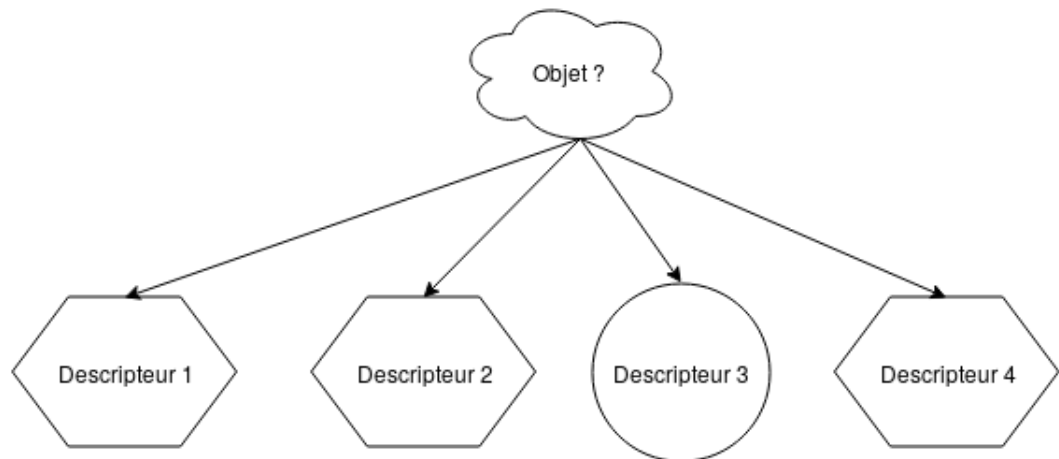


FIGURE 4.2 – Étape 2 : Exemple de présentation d'un nouvel élément avec ses descripteurs.

Troisième étape : Comparaison et classification de l'objet Une fois en possession des descripteurs du nouvel objet, le modèle de prédiction cherche les corrélations (ou les dé-corrélations) avec les suites de descripteurs de cycles respiratoires dont le type est connu. Cette comparaison permet d'attribuer une classe à un cycle respiratoire de type inconnu. Dans l'exemple illustratif figure 4.3, cela revient à comparer les suites de descripteurs. Ici, les descripteurs sont identiques excepté la troisième valeur. Cette troisième valeur est discriminante, car elle est différente du descripteur correspondant au cube, mais identique au descripteur correspondant au cylindre. De ce fait, on peut attribuer le label "*cylindre*" pour identifier le nouvel objet.

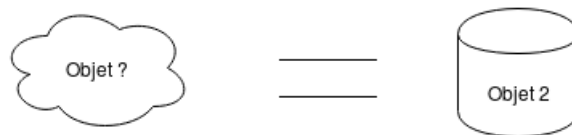


FIGURE 4.3 – Étape 3 : Exemple de comparaison et classification de l'objet inconnu.

En résumé, l'idée de la classification automatique est donc, pour un cycle respiratoire donné de type inconnu, selon la suite de descripteurs qui représentent son signal audio et au regard des descripteurs de cycles respiratoires que l'on connaît déjà, retrouver la classe à laquelle il semble appartenir.

La base de donnée ICBHI [22] sélectionnée pour ce projet contient les annotations nécessaires pour permettre cette approche. En effet, comme expliqué précédemment (voir partie 3.2.3) il est possible de classer les cycles respiratoires selon quatre classes, les cycles respiratoires contenant des *crackles*, ceux contenant des *wheezes*, ceux qui contiennent à la fois des *crackles* et des *wheezes* et les cycles respiratoires normaux.

4.1.2 Arbre de décision

Dans le domaine médical, il est nécessaire de comprendre pourquoi un diagnostic est établi ou bien quelles sont les raisons et les symptômes qui mènent à cette décision. Comprendre ce type de décisions "humaines" est nécessaire pour proposer une aide "informatique" au diagnostique.

Pour cette raison, le système d'arbre de décision a été choisi comme modèle d'apprentissage et de prédiction. En effet, les arbres de décision permettent de produire des procédures de classification mais permettent également l'utilisation d'outils afin de les rendre interprétables. De plus, les algorithmes d'apprentissage sont efficaces, rapides et permettent donc de tester plusieurs approches différentes en un temps restreint, même avec un jeu de données conséquent.

Le principe d'un arbre de décision est de schématiser une prise de décision en fonction d'une séquence de tests. Chaque élément est décrit par un ensemble de variables descriptives qui sont testées dans les différents noeuds de l'arbre. La figure 4.4 est un exemple d'arbre de décision à deux classes, *Malade* et *Sain*. Ce schéma a été inspiré des exemple du cours "*Apprentissage automatique : les arbres de décision*" de l'Université de Lille 3¹. L'arbre de décision comprend trois types de noeuds, le noeud racine (représenté sur le schéma par un cercle grisé) par lequel se fait l'accès à l'arbre, les noeuds internes (représentés sur le schéma par un cadre grisé) qui ont des descendants (ou enfants) qui sont eux-même des noeuds, et les noeuds terminaux, aussi appelés feuilles (représentés sur le schéma par du texte seul) qui n'ont pas de descendant. Les tests sont effectués dans les noeuds internes et les décisions sont prises dans les noeuds feuilles.

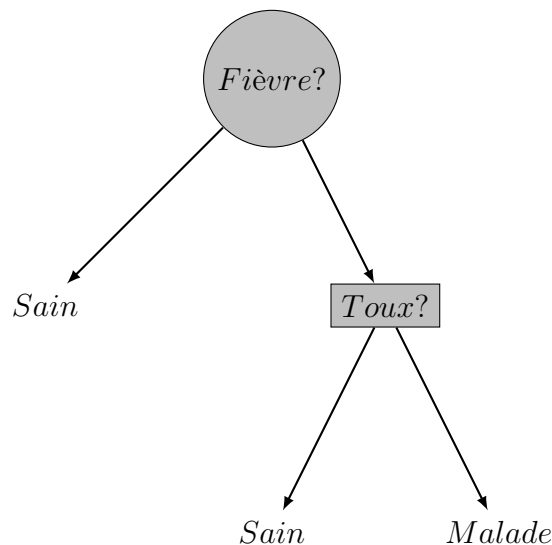


FIGURE 4.4 – Schéma d'un exemple d'arbre de décision qui permet d'établir si un patient est malade ou sain selon deux critères : fièvre et toux.

Cet arbre de décision est applicable à un ensemble d'éléments, les patients. Il

1. <http://www.grappa.univ-lille3.fr/polys/apprentissage/sortie004.html>

contient des noeuds internes qui correspondent aux tests suivants, en grisé sur le schéma : "Le patient a-t-il de la fièvre?" et "Le patient tousse-t-il?". Cette séquence de tests permet d'atteindre des feuilles, extrémités des segments en texte seul, et donc de classer le patient en deux catégories, *Malade* ou *Sain*.

L'arbre de décision est un modèle de classification. Comme indiqué en section 4.1.1, un tel modèle a besoin d'être entraîné. Cette phase d'entraînement équivaut à la construction de l'arbre. Pour cela, tous les éléments dont la classe est connue sont regroupés. Dans le cas de l'exemple indiqué par le schéma 4.4, il s'agit des patients. Ensuite, les éléments sont séparés en deux catégories en fonction des variables descriptives, de manière à être le plus discriminant possible. Cela permet d'obtenir deux catégories, ici par exemple, les patients malades avec de la toux et les autres patients. On répète alors cette étape en séparant à nouveau les données en fonction de leur descripteurs. Chacune de ces séparations (appelées "*split*" selon le terme anglais) équivalent à un test, c'est à dire un noeud interne de l'arbre de décision. Cette opération de séparation est à répéter jusqu'à ce que tous les patients soient catégorisés correctement, c'est à dire que l'on atteigne les feuilles de l'arbre, *malade* ou *sain*.

4.2 Outils de développement

Pour réaliser ce stage, plusieurs outils ont été nécessaires. Le sujet étant un sujet d'exploration expérimentale, aucun outils ou langages n'étaient imposés.

4.2.1 Veille scientifique

La globalité du travail de veille scientifique a été réalisée sur internet en utilisant plusieurs moteurs de recherches. Les articles scientifiques proviennent de sites de référencement scientifique tel que *ieeexplore.ieee.org*, *researchgate.net* ou encore *semanticscholar.org*.

Un véritable travail d'échange a également eu lieu avec d'autres membres de l'équipe du SCRIME (voir partie 1.1.1), comme avec G. Kurtag, artiste et compositeur, qui avait déjà travaillé sur le recensement de différents travaux sur la perception auditive des particularités de la respiration.

4.2.2 Programmation

Le langage de programmation *Python* (Version 2.7.13) a été choisi pour le travail réalisé lors de ce stage. Ce langage présente plusieurs avantages, en particulier d'être multi-plateformes et de proposer de nombreuses bibliothèques pour faciliter

le traitement de données, la gestion de fichiers de données et de leurs différents formats, le traitement du signal, mais également pour l'apprentissage automatique et les calculs de précisions.

Parmi les principales bibliothèques utilisées, le *Scipy kit* est un groupement de plusieurs bibliothèques qui a servi principalement au traitement des données, en particulier *Numpy* et *Pandas* pour les opérations et la gestion des tableaux et des matrices, et *Matplotlib* pour l'affichage de courbes statistiques ou de résultats.

La partie traitement du signal audio et calcul des descripteurs a été réalisée grâce à la bibliothèque *Essentia* [25], tandis que la partie apprentissage automatique et modèle de prédiction a été réalisée selon la bibliothèque *XGBoost* qui correspond à une implémentation d'arbres de décision boostés par gradient.

4.3 Représentation du contenu audio

4.3.1 Segmentation de la respiration

Comme expliqué partie 4.1, pour classifier les différents cycles respiratoires, il est nécessaire de représenter chacun des éléments à classifier par une suite de variables descriptives. En considérant le cycle respiratoire comme l'élément à traiter, il est nécessaire de repérer la position des cycles respiratoires dans les enregistrements audio, afin de pouvoir calculer un ensemble de descripteurs audio pour chacun d'entre eux.

Pour cette segmentation des enregistrements, il a été décidé de respecter les informations indiquées par les fichiers d'annotations de la base de données. En effet, pour chaque enregistrement de cette base, un fichier d'annotation indique la position de chacun des cycles respiratoires dans le temps grâce à des dates de début et de fin en secondes. Le schéma 3.2 disponible en section 3.2.2 permet de voir comment sont présentées ces informations. Cette décision implique de respecter les informations liées à la base de données choisie, mais permet un gain de temps considérable en s'abstrayant d'étudier une méthode de segmentation automatique de la respiration.

4.3.2 Calcul des descripteurs audio

Les particularités respiratoires que l'on cherche à identifier, les *crackles* et les *wheezes* étant définies par leur propriétés audibles (son continu ou discontinu, hauteur, intensité), il a été choisi de représenter chaque cycle respiratoire par des descripteurs audio (aussi appelés "*features*" selon le terme anglophone). Il existe plusieurs bibliothèques permettant de calculer un certain nombre de descripteurs

audio et parmi elles, *Essentia* a été choisi. La bibliothèque *Essentia* [25] présente plusieurs avantages, en particulier le calcul automatique de nombreux types de descripteurs audio, et la possibilité de calculer ces descripteurs de manière automatique sur des séquences sonores précises, dans notre cas, les cycles respiratoires. L'extracteur automatique de descripteurs *Freesound features extractor* a été utilisé car il a été pensé pour le calcul en masse de descripteurs sur de larges jeux de données. Cet extracteur est utilisé au sein du projet *Freesound*², une base de donnée collaborative comprenant tous types d'extraits audio. Les descripteurs obtenus permettent d'identifier le contenu des extraits proposés par les utilisateurs pour enrichir la base de données. En effet, il est possible de coder un profil de paramétrage de l'extracteur, et ainsi garantir les mêmes conditions de calcul des séquences de descripteurs pour tous les extraits audio souhaités.

La bibliothèque *Essentia* permet de calculer un très grand nombre de descripteurs différents, dont voici les principales catégories :

Low level : Informations spectrales telles que *Barkbands*, *Energybands*, *Melbands*, *MFCC*, etc. Ces informations sont pertinentes pour tous types de sons car elles sont basées sur le contenu spectral du signal, c'est à dire la composition sonore de l'extrait.

Rythme : Informations rythmiques telles que le nombre de battements, les intervalles de rythme, la position des battements, etc. Ces informations sont principalement utilisées dans le contexte musical pour identifier le rythme et ses battements sur un extrait d'un morceau de musique.

SFX : Informations basées sur l'harmonicité et l'inharmonicité de l'extrait. Ces informations sont pertinentes dans le domaine musical, mais peuvent également être utilisées dans d'autres domaines, car l'inharmonicité peut contenir des informations intéressantes en fonction des décalages de fondamentale.

Tonalité : Informations basées sur la tonalité de l'extrait. Ces informations sont très liées au domaine musical, car elles permettent d'identifier les tonalités et les suites de notes dans un extrait audio, en particulier pour un titre musical.

Parmi ces descripteurs, la catégorie qui semble la plus intéressante dans le cadre de ces travaux est la catégorie "*Low level*" qui signifie *bas niveau*. En effet, comme les sons que l'on cherche à représenter ne sont pas des extraits à caractères musicaux, il est plus pertinent de les représenter par des descripteurs bas niveaux

2. <https://freesound.org/>

calculés sur le signal audio, dans le domaine spectral par exemple. Afin de maximiser la représentation du signal audio, des valeurs statistiques ont été calculées pour chaque catégories de descripteurs, a savoir :

min : la valeur minimum.

max : la valeur maximale.

median : la médiane des valeurs.

mean : la moyenne des valeurs.

var : la variance.

stdev : la déviation standard.

dmean : la moyenne des dérivées premières.

dvar : la variance des dérivées premières.

dmean2 : la moyenne des dérivées secondes.

dvar2 : la variance des dérivées secondes.

En page suivante se trouve le profil de paramétrage de l'extracteur qui a permis de calculer les différents descripteurs audio :

```

profile = {
    stats = ['min', 'max', 'median', 'mean',
            'var', 'stdev', 'dmean', 'dvar',
            'dmean2', 'dvar2']

    'startTime': float(start_time),
    'endTime': float(end_time),
    'lowlevelFrameSize': 2048,
    'lowlevelHopSize': 1024,
    'lowlevelWindowType': 'blackmanharris62',
    'lowlevelSilentFrames': 'noise',
    'lowlevelStats': stats,

    'rhythmMethod': 'degara',
    'rhythmMinTempo': 40,
    'rhythmMaxTempo': 208,
    'rhythmStats': stats,

    'tonalFrameSize': 4096,
    'tonalHopSize': 2048,
    'tonalWindowType': 'blackmanharris62',
    'tonalSilentFrames': 'noise',
    'tonalStats': stats
}

features, features_frames =
    es.FreesoundExtractor(**profile)(audio_in)

```

Comme il est possible de le voir, il faut fournir à l'extracteur les dates de début *startTime* et fin *endTime* du cycle respiratoire à traiter, puis de paramétrer les différentes fenêtres d'échantillonnage à respecter pour les calculs. La dernière ligne correspond à l'appel à l'extracteur qui effectue le calcul et stocke les résultats dans des tableaux.

4.3.3 Exploitation des données

Afin d'en faciliter l'exploitation, toutes les données sont stockées dans un fichier *CSV* (*Comma-Separated Values*). Ce type de fichier correspond à un tableau de données qui contient les informations en ligne, séparées par un caractère de

séparation, souvent la virgule. Parmi les avantages à travailler avec des fichiers *CSV*, il y a la possibilité de consulter manuellement le fichier grâce à un éditeur de texte présentant la fonction tableur, mais également de pouvoir extraire ce fichier automatiquement dans les tableaux ou des matrices grâce aux nombreuses bibliothèques *Python*. Ainsi l'utilisation de ces données, en entrée du modèle de prédiction par exemple, est facilitée. La figure 6.3 présentée en annexe permet de visualiser un extrait de fichier *CSV* contenant des descripteurs.

Chapitre 5

Les différentes approches étudiées

S'agissant d'un stage de recherche expérimentale, la recherche de méthodes existantes et d'outils nécessaires correspond à une partie conséquente du travail. Toutefois, quatre approches différentes de classification ont pu être implémentées et testées. Ces quatre approches sont détaillées dans les sections qui suivent.

5.1 Classification en fonction des types de cycles respiratoires

La première approche consiste à classer les cycles respiratoires en fonction des différents particularités respiratoires, à savoir *crackles*, *wheezes*, *both* pour la catégorie comprenant à la fois *crackles* et *wheezes* et enfin la catégorie *normal* pour tous les cycles ne présentant pas de particularité respiratoire.

5.1.1 Les modèles de classification

Pour cette approche, deux manières de classer les cycles respiratoires ont été testées, une approche multi-classe et une approche mono-classe, afin de savoir laquelle des deux offre de meilleurs résultats.

L'approche multi-classe

Cette approche consiste à analyser la suite de descripteurs d'un cycle respiratoire, et à évaluer si cette suite semble plus proche de celle d'un cycle respiratoire contenant des *crackles*, d'un cycle respiratoire contenant des *wheezes*, d'un cycle respiratoire contenant à la fois des *crackles* et des *wheezes* ou d'un cycle respiratoire normal. Cette approche, la plus simple, permet d'obtenir un pourcentage

d'appartenance par classe, et de les comparer pour conserver le meilleur pourcentage. L'algorithme de décision est schématisé en figure 5.1. Sur ce schéma, les flèches illustrent la comparaison des séquences de descripteurs. Dans cet exemple, le cycle X, au vu des pourcentages obtenus par les prédictions, se voit attribuer le label *crackles*. Autrement dit, on estime, grâce au modèle de prédiction, que ce cycle respiratoire contient des *crackles*.

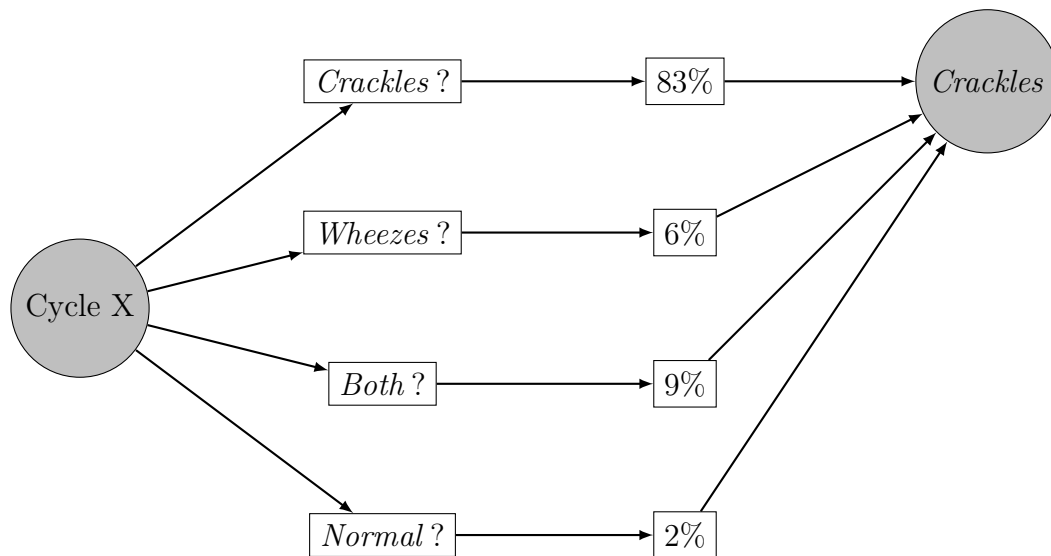


FIGURE 5.1 – Schématisation de l'algorithme de classification pour le modèle multi-classe, avec exemple de prédiction.

L'approche Mono-classe

Cette approche diffère de l'approche multi-classe dans le sens où, au lieu d'essayer de trouver un pourcentage d'appartenance pour chacune des classes possibles, on compare le cycle respiratoire à chacune des classes de manière successive. Plus précisément, cela consiste à étudier la séquence de descripteurs d'un cycle respiratoire, et à la comparer afin de déduire si sa séquence de descripteurs semble indiquer qu'il contient des *crackles* ou non. Une fois cette première discrimination effectuée, le modèle évalue si la suite de descripteur semble indiquer la présence de *wheezes* ou non. De ces deux analyse, on peut déduire à quelle classe appartient le cycle respiratoire en suivant le schéma décisionnel représenté en figure 5.2.

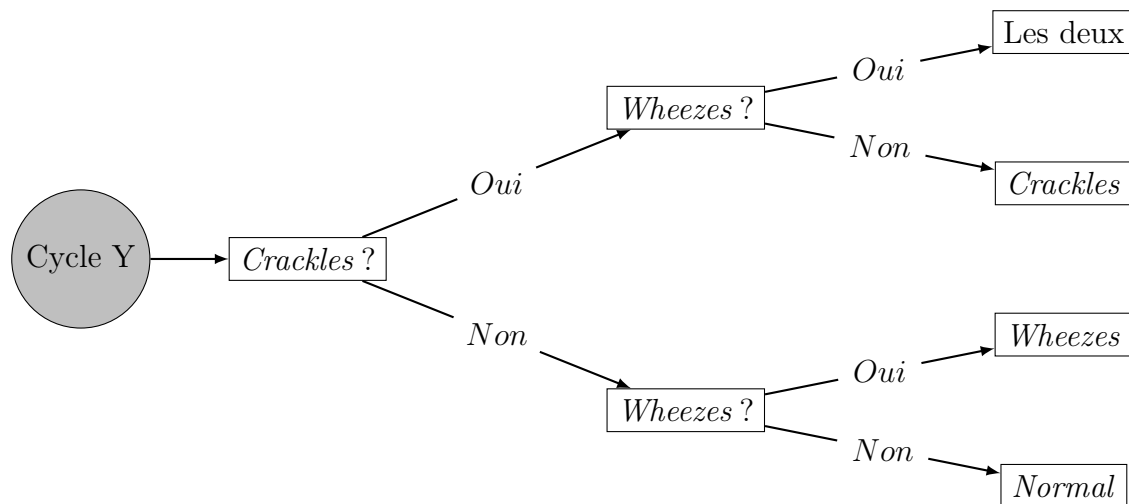


FIGURE 5.2 – Schématisation de l’algorithme de classification pour le modèle mono-classe.

5.1.2 Mise en oeuvre

Tandis que la mise en place du modèle de prédiction multi-classe est assez immédiate, la mise en place du modèle mono-classe a nécessité plus de réflexion, en particulier pour effectuer plusieurs classifications consécutives avec le même modèle d’arbre de décision. C’est pourquoi le code présenté est celui utilisé pour la classification mono-classe.

Afin de pouvoir les comparer équitablement, tous les résultats présentés dans ce rapport ont été obtenus avec le même paramétrage du modèle. Cet ensemble de paramètres a été établi empiriquement durant la progression de ce stage.

On notera que les variables et commentaires présents dans ce code sont en anglais. S’agissant d’un travail de recherche, ces travaux sont susceptibles d’être étudiés ou poursuivis par d’autres scientifiques. L’utilisation de l’anglais comme standard de codage permet une plus grande possibilité de diffusion. Cela entre dans le cadre de l’uniformité des travaux discutée en section 2.3.2.

L’essentiel du travail a consisté à sélectionner correctement les informations relatives aux enregistrements qui seront utilisées. Certaines informations, comme le type de micro utilisé ou la zone du corps sur laquelle l’enregistrement a été effectué, ne sont pas exploitées. Pour effectuer ce tri, les différentes informations fournies par la base de données sont récoltées dans des fichiers *CSV* séparés, qui

sont ensuite chargés et convertis en tableau de valeurs dans le code. Il est ainsi plus aisé de récupérer les informations jugées utiles.

De par la conception du modèle mono-classe, il est nécessaire de pouvoir disposer séparément des informations concernant les *crackles* et les *wheezes* pour pouvoir effectuer deux classifications distinctes successives, une par type de symptôme. Dans le code suivant, pour les cycles respiratoires du set *train*, les fichiers *CSV* sont chargés dans les tableaux de flottants nommés *train_dataset* et *train_info*. Ainsi, il est possible de récupérer les informations concernant la présence de *crackles* ou de *wheezes* dans des tableaux séparés, respectivement *classification_train_crackle* et *classification_train_wheeze*.

La même procédure est effectuée pour les cycles respiratoires du set *test* dont les informations seront stockées de la même manière dans des tableaux nommés *classification_test_crackle* et *classification_test_wheeze*. Les descripteurs des cycles respiratoires sont sauvegardés dans les tableaux *features_train* pour les cycles respiratoires de *train* et dans *features_test* pour les cycles du set *test*.

```
# loading first CSV
train_dataset = np.loadtxt(input_train, delimiter="," ,
                           skiprows = 1, usecols=range(1,nbcols_train))

# loading second CSV
train_info = np.loadtxt(info_train,delimiter = ',' ,
                        skiprows = 0,usecols=range(7,9))

# saving info about crackles
classification_train_crackle = train_info[:,0:1]
# saving info about wheezes
classification_train_wheeze = train_info[:,1:2]
# saving features for train cycles
features_train = train_dataset[:,1:]
```

Il est à noter que tous les fichiers sont triés de manière à contenir les informations dans le même ordre. Par exemple, la ligne 422 dans le premier fichier correspond à la séquence de descripteurs du 421 cycle respiratoire du set de données. Dans le second fichier, cette 422ème ligne correspond aux informations d'annotations de ce même 421ème cycle. Ce décalage d'une ligne est dû à la présence d'en-têtes informatives en première ligne des fichiers. De cette manière, les données sont conservées dans des tableaux différents, mais un cycle respiratoire donné correspond au même indice dans chacun de ces tableaux.

Une fois les informations initiales correctement stockées, vient le tour du modèle de classification. Parmi les lignes de code suivantes, la première permet l'entraînement du modèle, en prenant comme entrée le tableau qui contient les descripteurs des cycles du set *train* et celui contenant les informations de ces cycles à propos des *crackles*. Une fois le modèle entraîné, la seconde ligne permet le calcul de prédictions avec, en entrée, le tableau des descripteurs des cycles du set *test*. Finalement, une fois arrondies, les prédictions sont stockées dans le tableau nommé *crackle_predictions*.

```
# training
model.fit(features_train, classification_train_crackle)
# predicting
crackle_pred = model.predict(features_test)
# saving predictions
crackle_predictions = [round(value) for value in crackle_pred]
```

Ensuite, les prédictions sont parcourues pour établir la première étape de classification des cycles. Pour cela, suivant la prédictions obtenue pour chacun des cycles respiratoires (1 s'il contient des *crackles*, 0 sinon) on attribue la valeur 2, qui correspondra donc au numéro de classe des cycles avec *crackles*, sinon on lui attribue la valeur 0. Ces valeurs sont stockées dans un tableau final de classification nommé *saved_prediction*.

```
# creating the final array
saved_prediction = len(crackle_predictions)*[0]
# for every predicted cycles
for cpt in range(0, len(crackle_predictions)):
    if(crackle_predictions[cpt] == 1): # predicted as "crackles"
        saved_prediction[cpt] = 2 # classified as "crackles"
    else:
        saved_prediction[cpt] = 0 # classified as "no crackles"
```

Une fois l'étape terminée pour la reconnaissance des *crackles*, on reproduit le procédé pour les *wheezes*. On effectue un nouvel entraînement du modèle avec, cette fois, les informations relatives au *wheezes*, puis on effectue des prédictions sur les mêmes cycles que précédemment (ceux du set *test*). Les prédictions obtenues sont stockées dans le tableau *wheeze_predictions*.

```

# training
model.fit(features_train, classification_train_wheeze)
# predicting
wheeze_pred = model.predict(features_test)
#saving predictions
wheeze_predictions = [round(value) for value in wheeze_pred]

```

Une fois ce deuxième tour de prédiction terminé, il reste à comparer les résultats avec ceux obtenus au tour précédent pour déterminer la classification finale. Pour cela, il faut parcourir simultanément le tableau final de classification et le tableau de prédiction des *wheezes*. Ainsi, si le cycle a précédemment été classé parmi les *crackles* (valeur 2 dans *saved_predictions*) et qu'il a été prédit également parmi les *wheezes* (valeur 1 dans *wheeze_predictions*) alors on lui attribue la valeur 4 dans *saved_predictions*, valeur qui correspond à une classification finale dans la catégorie *both*. On effectue le même genre de comparaison pour les autres classes, *crackles* uniquement avec la valeur 2, *wheezes* uniquement avec la valeur 3 et *normal* avec la valeur 1.

```

# for every predicted cycles
for cpt in range(0, len(wheeze_predictions)):
    if(saved_prediction[cpt]==2): #if already crackle
        if(wheeze_predictions[cpt]==1): # and predicted as wheeze
            saved_prediction[cpt] = 4 # then classified in both
    elif(saved_prediction[cpt] != 2): # else if not crackle
        if(wheeze_predictions[cpt]==1): # and predicted as wheeze
            saved_prediction[cpt] = 3 # classified in wheeze
        else: # not wheezes and not crackle
            saved_prediction[cpt] = 1 # then classified as normal

```

5.1.3 Résultats obtenus

Grâce à cette approche, on obtient un tableau de prédictions qui contient les classes des cycles respiratoires du set *test*. La comparaison du tableau de prédictions avec le tableau des annotations permet d'établir précisément la qualité des résultats obtenus. Afin de pouvoir comparer avec les méthodes déjà proposées lors du challenge ICBHI, les mêmes métriques d'évaluation ont été respectées, à savoir la spécificité *Sp*, la sensibilité *Se* et la précision *Acc*, comme expliqué en section 3.1.

L'utilisation de l'extracteur automatique de la bibliothèque *Essentia* nous permettant d'obtenir plus de 3000 descripteurs, un premier essai de classification a été réalisé avec 13 coefficients de *MFCC* (*Mel-Frequency Cepstral Coefficient*) comme unique type de descripteur. De cette manière, la complexité et le temps de calcul de la classification sont drastiquement réduits. Cela permet de tester le code en un temps plus rapide avec une précision comparable avec celle obtenue grâce aux descripteurs *low level* (plus de 2500 coefficients).

Comme pour tout exercice de programmation, le code créé lors de ce stage a nécessité un nombre important de tests. La plupart d'entre eux consistent à vérifier la longueur des tableaux produits. Par exemple, il est nécessaire de s'assurer que le tableau de prédictions a la même taille que le nombre de cycles respiratoires que l'on cherche à classer.

Les résultats obtenus sont résumés dans le tableau 5.1. La sensibilité et la spécificité sont des valeurs comprises entre 0 et 1. Le meilleur résultat possible est 1, et la précision est exprimée en pourcentage.

TABLE 5.1 – Précision de la classification par arbre de décision selon la métrique du challenge ICBHI.

Modèle	Se	Sp	Acc
Mono/MFCC	0.2216	0.7604	49.10%
Mono/Lowlevel	0.3505	0.5819	46.62%
Multi/MFCC	0.2081	0.7805	49.43%
Multi/Lowlevel	0.3550	0.6255	49.02%
HMM	0.4232	0.5669	49.50%
NLSp	0.4890	0.7780	49.98%

En observant les résultats obtenus, on remarque que, malgré une précision *Acc* comparable à l'état de l'existant (méthodes *HMM* et *NLSp* détaillées en section 3.2.3), la spécificité et la sensibilité sont moins équilibrées. En effet, pour les résultats obtenus avec le modèle multi-classes et les coefficients *MFCC* en entrée, ligne

Multi/MFCC du tableau 5.1, on peut déduire que les cycles respiratoires normaux sont correctement classés 78% du temps, tandis que les cycles avec particularités respiratoires sont correctement classés seulement 20% du temps.

Pour mieux comprendre la disparité de ces résultats, des matrices de confusion ont été calculées. L'intégralité de ces matrices est disponible en deuxième partie des annexes de ce rapport. Dans cette partie, les explications seront basées sur l'exemple du modèle *Multi/MFCC*. La table 5.2 indique les prédictions obtenues, en nombre de cycles respiratoires, en utilisant ce modèle.

On y remarque une surévaluation des cycles respiratoires normaux. Cela s'explique par le fait qu'en utilisant seulement les 13 coefficients de *MFCC*, la séquence de descripteurs n'est pas suffisamment représentative du signal audio pour discriminer les cycles respiratoires. En effet, on remarque sur le tableau 5.1 des résultats plus équilibrés avec l'utilisation des descripteurs de bas niveaux.

TABLE 5.2 – Matrice de confusion des prédictions, en nombre de cycles, obtenues avec le modèle *Multi-classe* et les descripteurs *MFCC*.

	Cn	Cc	Cw	Cb	Total
Tn	1163	229	85	13	1490
Tc	374	211	31	1	617
Tw	265	63	20	5	353
Tb	103	29	8	0	140
Total	1905	532	144	19	2600

5.2 Classification en fonction des différentes pathologies

5.2.1 Intérêt

La partie 2.3.2 explique combien les bruits pulmonaires identifiables différents sont nombreux. Ce nombre peut augmenter en fonction des sémiologies variables selon les pays. Aussi, pour certains scientifiques, une classification de sons pulmonaires basée simplement sur une partie des symptômes (comme l'approche envisagée en section 5.1) peut être vue comme restrictive.

Afin de remédier à ce problème, une nouvelle approche a été envisagée. Cette approche permet d'exploiter une autre information des annotations de la base de données : les pathologies. L'idée est de classer les cycles respiratoires directement selon les pathologies plutôt qu'en essayant de distinguer les symptômes.

En effet, la séquence de descripteurs audio qui permet de représenter le signal correspondant à un cycle respiratoire peut contenir des informations différentes et complémentaires à la présence de *crackles* et/ou de *wheezes*. Le modèle de classification cherche alors les corrélations directement entre les séquences de descripteurs et les pathologies des différents patients.

5.2.2 Mise en oeuvre

Pour cette approche, plusieurs problèmes ont du être résolus. Le problème majeur est celui du déséquilibre de la base de données. La tableau 5.3 permet d'observer que les cycles respiratoires ne sont pas également distribués selon les pathologies.

TABLE 5.3 – Répartition du nombre de patients et de leurs cycles respiratoires en fonction des pathologies. L'abréviation "*Pat.*" utilisée signifie "*Patients*".

	Pat. <i>train</i>	Cycles <i>train</i>	Pat. <i>test</i>	Cycles <i>test</i>	Total Pat.	Total cycle
Asthme	1	6	0	0	1	6
LRTI	2	32	0	0	2	32
Pneum.	6	285	0	0	6	285
Bron.	6	90	1	14	7	104
Bronchio.	3	74	3	86	6	160
URTI	10	171	4	71	14	242
COPD	38	3493	26	2254	64	5747
Sain	13	147	13	175	26	322
Total	79	4298	47	2600	126	6898

Il est immédiatement remarquable que les patients atteints de la pathologie *COPD* sont sur-représentés par rapport aux autres pathologies. Par extension, le nombre de cycles respiratoires est également beaucoup plus important. Pour remédier à ce problème, il est nécessaire de pondérer l'entraînement du modèle. En effet, le modèle de prédiction cherche à optimiser le taux d'erreur globale, ce qui implique que la distribution des classes en prédictions dépend de la classe la plus représentée. Donner un poids aux classes permet donc de corriger cette erreur de distribution. A chaque classe, le poids attribué est défini suivant la règle suivante :

$$\text{Pondération} = \frac{\text{Nombre d'occurrences négatives de la classe } C}{\text{Nombre d'occurrences positives de la classe } C}$$

De cette manière, les éléments minoritaires gagnent en importance tandis que ceux majoritairement représentés en perdent. Une erreur sur une classe sous-représentée est considérée plus coûteuse qu’une erreur sur une classe sur-représentée.

La répartition des données ne respectant plus celle indiquée par le challenge ICBHI, il est moins pertinent de conserver la même métrique d’évaluation. Par conséquent, pour plus de clarté, une nouvelle formule définit la précision :

$$\text{Précision} = \frac{\text{Nombres de cycles correctement prédits}}{\text{Nombres de cycles totaux de cette pathologie}}$$

5.2.3 Résultats obtenus selon différentes répartitions

Pour corriger le problème de distribution des patients en fonction de leur pathologie, plusieurs résultats ont été calculés, avec des répartitions différentes entre les données d’entraînement et de test. Les premiers résultats ont été obtenus avec la répartition des cycles respiratoires correspondant à celle du challenge ICBHI. D’autres expériences ont été menées avec des séparations entraînement/test aléatoires sur l’ensemble de la base de données.

Répartition du challenge ICBHI

Le tableau 5.4 permet de visualiser les nombres de cycles respiratoires correctement classés et ceux incorrectement classés. Comme le précise la formule indiquée section 5.2.2, la colonne *Précision* de ce tableau est obtenue en divisant la colonne *Correct* par la colonne *Test*.

D’après le tableau 5.4, 99% des cycles respiratoires de la pathologie *COPD* sont correctement détectés. Ce résultat contraste avec les autres pourcentages obtenus, qui varient entre 0% (*Bronchiolitis*) et 50% (*Bronchiectasis*). Cette disparité des résultats peut s’expliquer par le déséquilibre trop important de la base de données, que la pondération ne suffit pas à corriger. On peut également expliquer cette différence par la possibilité que certains cas se confondent, avec des cycles respiratoires similaires en caractéristiques mais qui appartiennent à des pathologies différentes. Afin de tester la robustesse de ces résultats, d’autres classifications ont été réalisées avec des répartitions *train/test* différentes de celle du challenge ICBHI.

TABLE 5.4 – Classification en pathologies en fonction des descripteurs, des cycles respiratoires du set test du ICBHI Challenge. Les résultats sont donnés en nombre de cycles, et la précision en pourcentage.

	<i>Train</i>	<i>Test</i>	Total	Prédits	Correct	Précision (%)
Asthme	6	0	6	0	0	0.0
LRTI	32	0	32	21	0	0.0
Pneum.	285	0	285	31	0	0.0
Bron.	90	14	104	7	7	50.0
Bronchio.	74	86	160	11	0	0.0
URTI	171	71	242	97	15	21.1
COPD	3493	2254	5746	2329	2251	99.8
Sain	147	175	322	104	65	37.1
Total	4298	2600	6897	2600	2338	89.9

Répartition aléatoire

L’objectif de cette classification est de pouvoir mieux tester la robustesse de l’approche par pathologie. Par exemple, selon la répartition des données indiquée pour le challenge ICBHI, trois pathologies (*Asthma*, *LRTI* et *Pneumonia*) ne sont pas représentées dans le set *test*. Autrement dit, il n’est pas possible d’évaluer la reconnaissance des cycles respiratoires de ces pathologies avec une telle répartition.

Plusieurs ensembles de set *train* et *test* ont donc été créés avec une répartition aléatoire de toutes les données. Le nombre total de cycles respiratoires a été divisé en trois, deux tiers des données servant à l’entraînement du modèle tandis que le tiers restant est utilisé pour tester l’approche, c’est à dire calculer des prédictions. Les éléments répartis entre ces deux catégories étant choisis aléatoirement, l’entraînement et les prédictions sont effectués sur des cycles respiratoires différents à chaque fois. Afin d’être plus compréhensif, les deux ensembles de set, ainsi que les résultats respectivement obtenus, sont présentés dans les tableaux 5.5 et 5.6.

TABLE 5.5 – Classification des cycles respiratoires en pathologies en fonction des descripteurs. Séparation entraînement (63%) / test (33%) avec sélection aléatoire des données. Première classification, résultats en nombre de cycles et précision en pourcentage.

	<i>Train</i>	<i>Test</i>	Total	Prédits	Correct	Précision (%)
Asthme	5	1	6	1	1	100.0
LRTI	26	6	32	4	4	66.6
Pneum.	189	96	285	94	84	87.5
Bron.	69	35	104	32	32	91.4
Bronchio.	111	49	160	44	41	83.6
URTI	162	80	242	71	63	78.7
COPD	3840	1907	5746	1919	1900	99.6
Sain	218	104	322	113	96	92.3
Total	4620	2278	6898	2278	2221	97.5

TABLE 5.6 – Classification des cycles respiratoires en pathologies en fonction des descripteurs. Séparation entraînement (63%) / test (33%) avec sélection aléatoire des données. Seconde classification, résultats en nombre de cycles et précision en pourcentage.

	<i>Train</i>	<i>Test</i>	Total	Prédits	Correct	Précision (%)
Asthme	4	2	6	0	0	0.0
LRTI	20	12	32	4	4	33.3
Pneum.	185	100	285	98	91	91.0
Bron.	69	35	104	35	35	100.0
Bronchio.	115	45	160	36	34	75.5
URTI	170	72	242	79	62	86.1
COPD	3847	1900	5746	1917	1896	99.7
Sain	210	112	322	109	99	88.3
Total	4620	2278	6898	2278	2221	97.5

D’après les tableaux 5.5 et 5.6, si l’on se concentre sur la moyenne totale des prédictions, la précision de cycles correctement prédits atteint 97.5%. Ce résultat proche de la perfection nécessite d’être attentivement étudié. On peut mettre de côté la pathologie *COPD* qui était déjà correctement reconnue lors des résultats

précédents. Avec 99.8% selon la répartition du challenge contre 99.6% et 99.7% selon la répartition aléatoire, cela fait peu de différence. En revanche lorsque l'on étudie les autres pathologies, on remarque des améliorations significatives de la précision. Par exemple, pour les cycles respiratoires correspondants à la pathologie *Bronchiolitis*, la précision était de 0.0% (aucun cycle respiratoire correctement prédit) avec la répartition du challenge. Avec la répartition aléatoire, la précision atteint 83.6% selon le tableau 5.5 et 75.5% selon la seconde classification (tableau 5.6). Cette différence significative est sans explication mathématique ou informatique au premier abord. Cependant, l'écoute d'un lot représentatif d'enregistrements pulmonaires a permis d'identifier le phénomène à l'origine de résultats si différents.

Il s'agit d'un effet "*horse*", terme qui désigne, dans le domaine de l'apprentissage machine, "un système qui ne règle finalement pas le problème qu'il donne l'impression de résoudre" [31]. On parle de "*horse*" (ou bien d'effet "Hans le Malin" ou "*Clever Hans*" en anglais) lorsque que l'on qualifie un système de classification qui prédit correctement l'appartenance d'éléments à leurs classes, mais pour de mauvaises raisons. C'est le cas si, par exemple, un modèle de classification est censé identifier des sons en fonction de leurs enveloppes spectrales, mais qu'il parvient à le faire grâce à des méta-données comme la durée des enregistrements.

Le problème de la répartition aléatoire des données est que les enregistrements des patients peuvent être séparés entre le set *train* et le set *test*. Dans la répartition du challenge, tous les enregistrements d'un même patient se trouvent soit dans *train* soit dans *test*. Cette différence fait apparaître le problème de "*horse*".

Dans le cas présent, d'après les écoutes effectuées, les disparités de qualité des enregistrements peuvent expliquer le phénomène de "*horse*" rencontré. De nombreux enregistrement souffrent de bruits parasites, qui sont liés aux conditions d'enregistrements ou bien au matériel utilisé. Or, ces conditions d'enregistrements et le matériel utilisé restent identiques pour un même patient. De ce fait, les informations audibles relatives à ces conditions, se retrouvent dans les différents enregistrements d'un même patient, et causent ce cas de mauvais apprentissage.

5.3 Classification appliquée aux enregistrements

5.3.1 Intérêts

Une nouvelle approche a été expérimentée dans le but de diminuer la dimension des données classifiées afin de permettre des analyses sur la précision des classifications. L'idée est de reproduire le système de classification des cycles respiratoires en pathologies en fonction des séquences de descripteurs, mais en calculant cette séquence non pas sur la durée du cycle respiratoire, mais sur l'ensemble de l'enre-

gistement. Ainsi la décision d’attribution d’une pathologie se fera sur l’ensemble des cycles respiratoires d’un même enregistrement. De cette manière, on passe de 6898 cycles, à 919 enregistrements à traiter.

5.3.2 Mise en oeuvre

La mise en oeuvre reste la même que la classification effectuée lors de l’approche précédente, avec le même calcul de précision que celui détaillé section 5.2.2. Seul le calcul des descripteurs diffère. Il n’est plus effectué sur la durée d’un cycle respiratoire mais sur celle de la totalité de l’enregistrement d’un patient.

5.3.3 Résultats

Le tableau 5.7 regroupe les résultats obtenus en classifiant les enregistrements en pathologies.

TABLE 5.7 – Classification des enregistrements en pathologies. Résultats donnés en nombre d’enregistrements, précision en pourcentage.

	<i>Train</i>	<i>Test</i>	Total	Prédits	Correct	Précision (%)
Asthme	1	0	1	0	0	0.0
LRTI	2	0	2	0	0	0.0
Pneum.	37	0	37	9	0	0.0
Bron.	14	2	16	0	0	0.0
Bronchio.	7	6	13	0	0	0.0
URTI	46	6	52	9	1	16.6
COPD	463	330	793	332	325	98.4
Sain	18	17	35	11	7	41.1
Total	558	361	919	361	333	92.3

Il apparaît une fois de plus que les résultats sont corrects pour la catégorie correspondant à la pathologie *COPD*. Les autres pathologies sont très mal identifiées, en particulier à cause du fait qu’il n’y a pas un nombre assez grand d’enregistrements pour entraîner le modèle de manière suffisante. Les résultats obtenus grâce à cette approche ne sont donc pas assez fiables avec cette base de données.

5.4 Classification à l'échelle du patient

5.4.1 Intérêt

Une dernière approche a été expérimentée pour améliorer les résultats obtenus durant ce stage. Il s'agit de penser différemment le problème de la classification automatique. De manière générale, en informatique musicale et traitement du signal, la classification est particulièrement basée sur le contenu audio des éléments. C'est le cas des méthodes testées précédemment. La classification des cycles respiratoires est effectuée uniquement en fonction du contenu audio, représenté par les séquences de descripteurs.

Or, l'idée principale d'étudier les maladies pulmonaires avec des outils automatiques est de proposer un outils d'aide au diagnostic. Dans une telle optique, la question de la pertinence d'une classification uniquement basée sur le contenu se pose. Pour établir un diagnostic, un médecin va ausculter son patient, c'est à dire écouter et analyser sa respiration. Pour cela, le médecin va écouter le patient respirer durant un certain temps, et va donc prendre en compte plusieurs cycles respiratoires, écoutés à différents endroits du corps, pour prendre sa décision. De cette manière, il est possible de prendre une décision plus robuste car elle ne sera pas basée sur un seul cycle respiratoire, mais sur une succession de cycles. Il semble pertinent de reproduire ce système de fonctionnement avec la classification automatique.

5.4.2 Mise en oeuvre

Pour cela, la classification est effectuée en deux temps. La première partie sera appelée la partie "micro" tandis que la seconde partie sera appelée la partie "macro".

Ces deux noms font référence au fait que l'on cherche à analyser les sons pulmonaires "un étage au dessus", c'est à dire avec une vue plus globale, qui se situe au niveau du patient et plus seulement des cycles respiratoires à l'unité. La partie "micro" revient exactement à prédire les cycles un à un en fonction des particularités respiratoires comme détaillé en section 5.1. La partie "macro", quant à elle, consiste à examiner l'ensemble des prédictions précédemment obtenues pour tous les cycles d'un même patient. Selon ces prédictions, une décision est prise sur l'état du patient, à savoir s'il est plutôt susceptible d'être sain, ou malade.

Avec cette approche, il est possible d'identifier les patient dits "à risque", c'est à dire susceptibles d'être atteint d'une maladie pulmonaire, tandis que l'on parvient à écarter les patients sains. En effet, la décision est prise de la manière suivante : *« si tous les cycles respiratoires d'un même patient sont prédits comme normaux, alors le patient doit être sain, mais si l'un au moins des cycles est prédit comme*

ayant une particularité respiratoire, alors le patient doit être malade ».

5.4.3 Résultats

La matrice représentée dans le tableau 5.8 permet de visualiser les résultats obtenus grâce à cette approche. Il s'agit d'une matrice de confusion. Les lignes correspondent aux prédictions tandis que les colonnes correspondent à la réalité de l'état des patients. Ainsi, la première case de la première ligne indique le nombre de patients qui ont été prédits malades alors qu'ils étaient réellement malades, tandis que la seconde case de la première ligne indique le nombre de patients prédits comme malade alors qu'ils étaient en réalité sains. De la même manière, pour la seconde ligne, la première case indique le nombre de patients prédits comme sains alors qu'ils étaient malades, et la seconde case indique le nombre de patients prédits comme sain à juste titre.

TABLE 5.8 – Matrice de confusion pour la prédiction de l'état de santé des patients.

	Malade	Sain	Total
Malade	29	5	34
Sain	2	11	13
Total	31	16	47

De cette matrice de confusion 5.8, on peut obtenir des pourcentages de précision en calculant des ratios. On obtient alors le pourcentage de précision de bonnes réponses en faisant la somme des patients correctement prédits (cases en gris clair) et en divisant cette somme par le nombre total de patients (case en gris foncé). On obtient alors des détections correctes 85% du temps, c'est à dire que plus de 8 patients sur 10 sont correctement classés.

Bien que le nombre de patients soit assez limité par la base de données, les résultats sont très encourageants. Cette façon de penser la classification automatique en respectant le niveau du patient et non pas celui du contenu audio a fait l'objet d'une publication scientifique pour la conférence CBMI 2018 (*Content Based Multimedia Indexing*). Le document rédigé durant ce stage est disponible en troisième partie des annexes de ce rapport.

Chapitre 6

Bilan

6.1 Discussion

Les résultats obtenus durant ce stage permettent de faire ressortir que l'identification automatique des sons pulmonaires par apprentissage machine semble intéressant et pertinent. En effet, les précisions de prédictions obtenues pour les pathologies les plus représentées sont très encourageantes, comme les cycles respiratoires de patients atteints par la pathologie *COPD*, que l'on parvient à prédire correctement plus de 90% du temps. De plus, l'approche de classification au niveau du patient détaillée section 5.4.1 permet également de bons résultats, en estimant correctement 8 à 9 fois sur 10 si un patient semble atteint d'une pathologie respiratoire ou non. Ces résultats permettent d'envisager de poursuivre ces travaux pour pouvoir proposer un outils d'aide au diagnostic à des spécialistes de la santé et des pneumologues.

Cependant, des limites sont apparues durant le stage, comme l'impossibilité de détecter correctement certaines pathologies. En effet, la mauvaise représentation de ces pathologies dans la base de données, comme l'unique patient atteint d'asthme par exemple, empêche de travailler efficacement sur ces données et d'avoir un système de prédiction efficace et fiable.

Toutes les expériences réalisées et les résultats obtenus tendent à faire apparaître que, pour l'approche par apprentissage machine, les possibilités de travail sont directement liés aux bases de données utilisées. Or, dans le domaine médical, il est difficile d'obtenir des bases de données conséquentes et correctement annotées. Le secret médical, le coût financier des appareils nécessaires (pour l'enregistrement pulmonaire par exemple) et le manque de consensus entre les différents laboratoires médicaux et hôpitaux (par disciplines médicales ou par nationalités) entravent bien souvent la diffusion publique de données.

6.2 Apport personnel

Bien qu'ayant déjà acquis des notions d'apprentissage machine durant le cursus de Master 2 Informatique, ce stage aura nécessité d'approfondir mes connaissances dans la discipline, et de découvrir de nouvelles façons de réfléchir. Par exemple, la classification des cycles respiratoires en pathologies avec une répartition aléatoire de la base de données en entraînement et en test, aura nécessité beaucoup de réflexion. La découverte du phénomène de *horse* détaillé partie 5.2.3 m'aura appris de toujours se méfier des résultats obtenus, et qu'un résultat avec des chiffres intéressants n'est pas toujours un bon résultat.

D'autres limites sont apparues durant l'exercice de ce stage. Il aura parfois été compliqué de travailler avec des jeux de données de taille importante. C'était en effet un exercice inhabituel et nouveau. Cela aura nécessité également beaucoup de réflexion et de recherches afin d'optimiser la manière de stocker et traiter ces données. Cette réflexion s'ajoute à la réflexion expérimentale nécessaire dans un domaine de recherche comme ici, qui consiste à identifier les données à exploiter parmi le lot disponible.

Ce stage aura été également l'occasion de gagner en autonomie de travail, et en approche expérimentale. Le sujet était peu directif, avec une absence de cahier de charge. De ce fait, une large manoeuvre était possible dans le travail à réaliser, et une organisation rigoureuse était donc nécessaire. Pour progresser tout au long du stage, il aura fallu beaucoup échanger avec les encadrants, et parfois d'autres chercheurs dans le domaine, tous très disponibles et bienveillants. Cela a pu permettre de réfléchir à des approches ou des solutions pour les proposer aux encadrants lors de rendez-vous réguliers. Ce fonctionnement aura été très enrichissant, tant d'un point de vue organisationnel grâce à l'autonomie accordée couplée à des résultats attendus régulièrement, que d'un point de vue connaissances grâce aux différentes recherches dans le domaine de la santé, de la pneumologie, de l'informatique médicale ou encore de l'apprentissage machine, mais également et surtout grâce aux nombreux échanges avec les différents chercheurs et professionnels.

Ce stage m'aura également apporté l'opportunité de présenter les travaux réalisés lors d'une publication scientifique. Cette expérience m'a beaucoup appris sur le discours scientifique, la rédaction d'une publication et le comportement à avoir en recherche.

Les nombreux et différents apports de ce stage, en particulier la découverte concrète de la recherche informatique dans le domaine du traitement du son, m'ont conforté dans l'idée de poursuivre dans cette voie. J'ai donc décidé d'effectuer une thèse sur un sujet très proche de celui de ce stage.

6.3 Perspectives

Ces travaux seront donc poursuivis lors d'une thèse. Plusieurs nouvelles expériences et approches ont déjà été réfléchies. Tout d'abord, comme expliqué partie 6.1, les résultats obtenus dépendent fortement de la qualité de la base de données. Dans le cas présent, la base de données qui aura servi pour la totalité de ses travaux souffre de quelques inconvénients : la qualité variable des enregistrements, la présence de bruits et de sons parasites, le déséquilibre des pathologies représentées ou le manque d'uniformisation dans la répartition des données. Pour parer à ces problèmes, il est envisageable de travailler sur l'élaboration d'une nouvelle base de données, qui serait réalisée grâce au partenariat de chercheurs et de médecins de plusieurs laboratoires. L'élaboration d'une telle base de données nécessite d'identifier précisément quelles sont les informations souhaitées, quels sont les éléments qui doivent être présents (parmi les annotations par exemple) et quelle est la qualité minimum nécessaire d'enregistrement.

Il est également envisagé de travailler sur l'extraction des particularités respiratoires. Ces travaux consisteraient à identifier très précisément les informations audio d'un cycle respiratoire qui correspondent à un symptôme. Une fois correctement identifiées, on pourrait reproduire ces informations audio, et les ajouter au signal sonore d'un cycle respiratoire qui serait sans particularité audible. De cette manière, il serait possible de faire de l'enrichissement sonore, en modifiant un cycle respiratoire sain pour qu'il devienne un nouveau cycle contenant une particularité respiratoire. La principale application d'un tel travail sera l'enrichissement de base de données.

Travailler sur une base de données différente de celle utilisée durant ce stage implique de réfléchir à une méthode de segmentation automatique de la respiration en cycles respiratoires. En effet, la base de données du challenge fournie les indications de position des cycles respiratoires dans les enregistrements audio. Cela ne sera pas toujours le cas avec d'autres enregistrements audio, et il sera donc nécessaire de réfléchir aux critères audio qui permettent d'identifier une inspiration ou une expiration.

D'autres choses peuvent être faites pour améliorer le système de classification, comme le pré-traitement des enregistrements audio pour réduire le bruit ou les sons parasites. Cette étape est présente dans la plupart des travaux de l'existant, mais a été volontairement écartée durant ce stage, à la fois par gain de temps mais également pour éviter de perdre de l'information entre le signal d'origine et le signal d'entrée du modèle de classification. Une approche par réseau de neurones profond est également envisagée, en utilisant directement les enregistrements audio en entrée. De cette manière, le réseau de neurones fera ressortir les informations qui semblent pertinentes pour la classification. Ces informations peuvent être différentes des suites de descripteurs qui étaient précédemment utilisées pour

l'apprentissage machine.

Avec d'un côté des résultats prometteurs, comme la reconnaissance des cycles respiratoires de patients atteints de *COPD*, et de l'autre des résultats moins exploitables comme la reconnaissance des cycles respiratoires de patients atteints d'asthme, il y a tout intérêt à poursuivre ces recherches. La réalisation d'une thèse sur l'étude automatique des maladies pulmonaires permettra certainement d'aider encore plus les médecins et spécialistes de la santé lors de l'établissement de leurs diagnostics.

Bibliographie

Identification audio

- [1] A. J. H. Goulart, R. C. Guido, C. D. Maciel. "Exploring different approaches for music genre classification". Egyptian Informatics Journal, vol. 13, no. 2, pp. 59-63, 2012.
- [2] N. Scaringella, G. Zoia and D. Mlynek, "Automatic genre classification of music content : a survey," in IEEE Signal Processing Magazine, vol. 23, no. 2, pp. 133-141, 2006.
- [3] Z. Liu, J. Huang, Y. Wang, T. Chen, "Audio feature extraction and analysis for scene classification". Proceedings of First Signal Processing Society Workshop on Multimedia Signal Processing, Princeton, NJ, pp. 343-348, 1997.
- [4] E. Wold, T. Blum, D. Keislar, J. Wheaten, "Content-based classification, search, and retrieval of audio". IEEE MultiMedia, vol. 3, no. 3, pp. 27-36, 1996.
- [5] V. Venkatachalam, L. Cazzanti, N. Dhillon and M. Wells, "Automatic identification of sound recordings," in IEEE Signal Processing Magazine, vol. 21, no. 2, pp. 92-99, 2004.
- [6] N. Orio. "Automatic identification of audio recordings based on statistical modeling". Signal Processing, vol. 90, no. 4, pp. 1064-1076, 2010.
- [7] A. Rabaoui, M. Davy, S. Rossignol, Z. Lachiri, N. Ellouze, "Sélection de descripteurs audio pour la classification des sons environnementaux avec des SVMs mono-classe". GRETSI, Groupe d'Etudes du Traitement du Signal et des Images, 2007.
- [8] B. Defreville, S. Bloquet, G. Filippi, C. Aujard. "La reconnaissance automatique de sources sonores dans l'environnement". 10ème Congrès Français d'Acoustique, Lyon, France. Société Française d'Acoustique - SFA, 2010.

Informatique médicale

- [9] K. Schoeffmann, B. Münzer, M. Riegler, P. Halvorsen. "Medical Multimedia Information Systems (MMIS)". In Proceedings of the 2017 ACM on Multimedia Conference (MM '17). ACM, New York, NY, USA, pp. 1957-1958, 2017.

Études de sons cardiaques

- [10] X. Yang, F. Yang, L. Gobeawan, S.Y. Yeo, S. Leng, L. Zhong and Y. Su. "A multi-modal classifier for heart sound recordings". In Computing in Cardiology, vol. 43, 2017.
- [11] G.D. Clifford, C. Liu, B. Moody, J. Millet, S. Schmidt and Q. Li, I. Silva and R.G. Mark. "Recent advances in heart sound analysis". In Physiological Measurement, vol.38, no. 8, pp. E10-E25, 2017.
- [12] B. Emmanuel. "A review of signal processing techniques for heart sound analysis in clinical diagnosis". Journal of medical engineering and technology, vol. 36, pp. 303-307, 2012.
- [13] M. Abdollahpur, A. Ghaffari, S. Ghiasi and M. Mollakazemi. "Detection of pathological heart sounds", In Physiological Measurement, vol.38, no. 8, pp. 1616-1630, 2017.

Études de sons pulmonaires

- [14] H. Polat and I. Güler. "A simple computer-based measurement and analysis system of pulmonary auscultation sounds". In Journal of medical systems, vol. 28, no. 6, pp. 665-672, 2004.
- [15] R. X. A. Pramono, S. A. Bowyer, E. Rodríguez-Villegas. "Automatic adventitious respiratory sound analysis : A systematic review". PloS one, 2017.
- [16] S. Reichert and R. Gass and E. Andrès. "Analyse des sons auscultatoires pulmonaires (Analysis of respiratory sounds)". In ITBM-RBM, vol. 28, pp. 169-180, 2007.
- [17] G. Chang, and Y. Lai. "Performance evaluation and enhancement of lung sound recognition system in two real noisy environments". In Computer Methods and Programs in Biomedicine, vol. 97, no. 2, pp. 141-150, 2010.
- [18] A. Abbas and A. Fahim. "An Automated Computerized Auscultation and Diagnostic System for Pulmonary Diseases". In Journal of Medical Systems, vol. 34, no. 6, pp. 1149-1155, 2010.

- [19] C. Chen, W. Huang, T. Tan, C. Chang and Y. Chang. "Using K-Nearest Neighbor Classification to Diagnose Abnormal Lung Sounds". In *Sensors*, vol. 15, no. 6, pp. 13132-13158, 2015.
- [20] G. Postiaux. "Nomenclature stéthacoustique pulmonaire : pourquoi pas un consensus mondial ?". *Revue Mal Respir © SPLF*, vol. 16, pp. 1075-1090, 1999.
- [21] E. Andrès, S. Reichert, R. Gass, C. Brandt, Y. Kehayoff. "La séméiologie pulmonaire à l'ère de la médecine factuelle". *La Revue de Respir ©*, vol. 6, pp. 47-50, 2007.

Challenge ICBHI

- [22] B. M. Rocha, D. Filos, L. Mendes, I. Vogiatzis, E. Perantoni, E. Kaimakamis, P. Natsiavas, A. Oliveira, C. Jácome, A. Marques, R. P. Paiva, I. Chouvarda, P. Carvalho and N. Maglaveras. "A Respiratory Sound Database for the Development of Automated Classification". In *Precision Medicine Powered by pHealth and Connected Health*, pp. 33-37, 2017.
- [23] N. Jakovljević, T. Lončar-Turukalo. "Hidden Markov Model Based Respiratory Sound Classification". In Maglaveras N., Chouvarda I., de Carvalho P. (eds) *Precision Medicine Powered by pHealth and Connected Health. IFMBE Proceedings*, vol 66, 2018.
- [24] G. Serbes, S. Ulukaya, Y.P. Kahya. "An Automated Lung Sound Preprocessing and Classification System Based On Spectral Analysis Methods". In Maglaveras N., Chouvarda I., de Carvalho P. (eds) *Precision Medicine Powered by pHealth and Connected Health. IFMBE Proceedings*, vol 66, 2018.

Essentia

- [25] D. Bogdanov, N. Wack, E. Gómez, S. Gulati, P. Herrera, O. Mayor. "ESSENTIA : an Audio Analysis Library for Music Information Retrieval". *International Society for Music Information Retrieval Conference (ISMIR'13)*, pp. 493-498, 2013.

Réseaux de neurones profonds

- [26] X. Du, Y. Cai, S. Wang and L. Zhang. "Overview of deep learning". *31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, Wuhan, pp. 159-164, 2016.

- [27] Q. Wu, Y. Liu, Q. Li, S. Jin and F. Li. "The application of deep learning in computer vision". Chinese Automation Congress (CAC), Jinan, pp. 6522-6527, 2017.
- [28] E. Nishani and B. Çiço. "Computer vision approaches based on deep learning and neural networks : Deep neural networks for video analysis of human pose estimation". 6th Mediterranean Conference on Embedded Computing (MECO), Bar, pp. 1-4, 2017.
- [29] Y. Petetin, C. Laroche and A. Mayoue. "Deep neural networks for audio scene recognition". 23rd European Signal Processing Conference (EUSIPCO), Nice, pp. 125-129, 2015.
- [30] J. Li, W. Dai, F. Metze, S. Qu and S. Das. "A comparison of Deep Learning methods for environmental sound detection". IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, pp. 126-130, 2017.
- [31] B. L. Sturm. "A Simple Method to Determine if a Music Information Retrieval System is a "Horse"". In IEEE Transactions on Multimedia, vol. 16, no. 6, pp. 1636-1644, 2014.
- [32] V. Maknickas and A. Maknickas. "Recognition of normalabnormal phonocardiographic signals using deep convolutional neural networks and mel-frequency spectral coefficients". In Physiological Measurement, vol. 38, no. 8, pp. 1671-1684, 2017.

Divers

- [33] SCRIME. "Studio de Création et de Recherche en Informatique et Musiques Expérimentales". Online, 2018. url : <https://scrime.u-bordeaux.fr/> [consulté le 24-05-18]
- [34] E. Lannuzzi, "Statistiques significatives sur la consommation de musique dans le monde", Elise News, 20/09/2017. url : elise.news/2017/09/statistique-consommation-musique-monde/ [consulté le 18/07/2018]
- [35] C. Dubois, "L'avenir de la musique est dans le streaming", L'Expansion L'Express, 13/08/2017. url : https://lexpansion.lexpress.fr/high-tech/l-avenir-de-la-musique-est-dans-le-streaming_1934440.html [consulté le 18/07/18]
- [36] World Health Organization. "Density of physicians". Online, 2017. url : http://www.who.int/gho/health_workforce/physicians_density/en/ [consulté le 18-05-2018].
- [37] World Health Organization. "Cardiovascular Diseases (cvds)". Online, 2017. url : <http://www.who.int/mediacentre/factsheets/fs317/en/> [consulté le 16-05-2018].

- [38] School of Health Sciences of Aveiro. Aveiro, Portugal.
<https://www.university-directory.eu/Portugal/School-of-Health-Sciences-University-of-Aveiro-ESSUA.html> [Online, accessed 08-06-18]
- [39] Aristotle University of Thessaloniki. Thessaloniki, Grece.
<https://www.auth.gr/en> [Online, accessed 08-06-18]

Annexes

Partie 1 : Fichier de descripteurs

Cette annexe citée en 4.3.3, présente la figure 6.3 qui permet de visualiser un extrait de fichier *CSV* contenant des descripteurs. La première ligne indique les informations contenues dans les colonnes du fichier. La première colonne indique le nom de l'enregistrement, la seconde colonne indique le numéro de classe attribué en fonction de la pathologie, et les colonnes restantes correspondent aux valeurs des descripteurs audio.

Cette figure est l'extrait d'un fichier comportant plus de 500 lignes (une par enregistrement dans le set *train*) et plus de 4000 colonnes (Nombre de valeurs de descripteurs calculées).

filename	pathologie	lowlevel.average_loudness	lowlevel.barkbands.dmean_0	lowlevel.barkbands.dmean_1	lowlevel.barkbands.dmean_2	lowlevel.barkbands.dmean_3	lowlevel.barkbands.dmean_4
105_1b1_Tc_sc_LitC2SE		5.0.000387538195355	0.00189192	0.000535471	2.5436e-05	2.02723e-05	3.26313e-05
106_2b1_Pl_mc_LitC2SE		6.0.00315306056291	0.000707509	0.00132077	0.000891378	0.000411917	0.000156926
106_2b1_Pr_mc_LitC2SE		6.0.00522426096722	0.12931	0.00891294	0.00252656	0.000197669	2.96318e-05
107_2b3_Al_mc_AKGC417L		6.0.0774864703417	0.522625	0.10576	0.00596835	0.00175916	0.000760166
107_2b3_Ar_mc_AKGC417L		6.0.106787465513	0.50024	0.149786	0.00921415	0.00476486	0.00305959
107_2b3_Ll_mc_AKGC417L		6.0.0175734143704	0.0649929	0.0512965	0.00668122	0.00149565	0.00105132
107_2b3_Lr_mc_AKGC417L		6.0.0542866289616	0.248186	0.142323	0.00507636	0.00111498	0.000587902
107_2b3_Pl_mc_AKGC417L		6.0.0764852911234	0.468909	0.0772555	0.023087	0.0132898	0.016343
107_2b3_Pr_mc_AKGC417L		6.0.165863156319	0.55486	0.13626	0.0134844	0.00542368	0.00738883
107_2b3_Tc_mc_AKGC417L		6.0.02398920618	0.158018	0.046195	0.00114091	0.000618999	0.000539414
107_2b4_Al_mc_AKGC417L		6.0.0452436171472	0.445747	0.0925054	0.0045385	0.001596	0.00065925
107_2b4_Ar_mc_AKGC417L		6.0.214166209102	0.564289	0.180309	0.0110031	0.00465732	0.00311968
107_2b4_Ll_mc_AKGC417L		6.0.00512166181579	0.0693431	0.0435212	0.00479028	0.0010342	0.000646344
107_2b4_Lr_mc_AKGC417L		6.0.066005140543	0.249957	0.149337	0.00491825	0.000919825	0.000514975
107_2b4_Pl_mc_AKGC417L		6.0.0681183114648	0.419663	0.0464279	0.0109474	0.00557408	0.00408898
107_2b4_Pr_mc_AKGC417L		6.0.160482883453	0.641362	0.142204	0.0148109	0.0066102	0.00627051
107_2b4_Tc_mc_AKGC417L		6.0.00205140979961	0.144338	0.035293	0.00109468	0.000572908	0.000484129
107_2b5_Al_mc_AKGC417L		6.0.0533164590597	0.464357	0.0937951	0.00615466	0.00163187	0.000684813
107_2b5_Ar_mc_AKGC417L		6.0.00984249077737	0.310688	0.104033	0.00645416	0.00229109	0.00127747
107_2b5_Ll_mc_AKGC417L		6.0.00695236586034	0.0628383	0.0434784	0.00496904	0.00123347	0.000815896
107_2b5_Lr_mc_AKGC417L		6.0.155275672674	0.380334	0.208901	0.0122779	0.0026602	0.00128598
107_2b5_Pl_mc_AKGC417L		6.0.0370704457164	0.384903	0.0543625	0.0167013	0.00961948	0.00929626
107_2b5_Pr_mc_AKGC417L		6.0.236976689041	0.636286	0.132798	0.0194445	0.00756589	0.0105822
107_2b5_Tc_mc_AKGC417L		6.0.0160297490656	0.21217	0.0521727	0.00186063	0.00106518	0.00115444
107_3p2_Al_mc_AKGC417L		6.0.00106129096821	0.127924	0.0462746	0.00278168	0.00114533	0.000264007
107_3p2_Ar_mc_AKGC417L		6.0.0174879804254	0.227382	0.109008	0.0122457	0.0113324	0.00375505
107_3p2_Ll_mc_AKGC417L		6.0.0305279903114	0.092887	0.00845717	0.000448289	0.00025059	0.000119626
107_3p2_Lr_mc_AKGC417L		6.0.037161719054	0.0668668	0.00836537	0.00030786	0.000186623	6.28917e-05
107_3p2_Pl_mc_AKGC417L		6.0.00967487879097	0.0570609	0.011676	0.00165754	0.000676471	0.000373211
107_3p2_Pr_mc_AKGC417L		6.0.000507963821292	0.0927325	0.0886542	0.00627054	0.00343073	0.00333244
107_3p2_Tc_mc_AKGC417L		6.0.0755652934313	0.387499	0.100583	0.00895404	0.00469922	0.0022443
108_1b1_Al_sc_LitC2SE		1.0.0975051447749	0.024034	0.00523275	4.47143e-05	8.81831e-06	8.24054e-06
110_1b1_Pr_sc_Mediton		6.0.00786138419062	0.341441	0.00993222	0.00314212	0.000978407	0.000960943
110_1p1_Al_sc_Mediton		6.0.0179404839873	0.415202	0.0241057	0.00163796	0.000190581	8.77765e-05

Partie 2 : Matrices de confusion pour le classement des cycles respiratoires en fonction des symptômes *crackles* et *wheezes*

Dans cette annexe sont présentés les différents tableaux comprenant les résultats détaillés des prédictions obtenues lors de la classification des cycles respiratoires en fonction des symptômes audibles. La table 6.1 présente la matrice des prédictions obtenues avec le modèle *Mono-classe* en utilisant les coefficients de *MFCC* en entrée. La table 6.2 regroupe les résultats obtenus avec le modèle *Mono-classe* avec les descripteurs *bas-niveau* en entrée. Enfin, les tables 6.3 et 6.4 permettent de visualiser les matrices de prédictions obtenues avec le modèle *Multi-classe*, respectivement avec les coefficients *MFCC* en entrée, puis les descripteurs *bas-niveau*.

Pour chacune des matrices, C_i correspond au nombre de cycles respiratoires correctement prédits pour la classe i . T_i correspond au nombre de cycles qui auraient dû être prédits dans la classe i . Enfin, i correspond aux différentes classes : n pour *normal*, c pour "*crackles*", w pour "*wheezes*" et b pour "*both*".

TABLE 6.1 – Matrice de confusion des prédictions obtenues avec le modèle *Mono-classe* et les coefficients de *MFCC* en tant que descripteurs.

	Cn	Cc	Cw	Cb	Total
Tn	1133	224	111	22	1490
Tc	357	221	23	16	617
Tw	246	71	25	11	353
Tb	105	25	10	0	140
Total	1841	541	169	49	2600

TABLE 6.2 – Matrice de confusion des prédictions obtenues avec le modèle *Mono-classe* et les descripteurs *bas-niveau* en entrée.

	Cn	Cc	Cw	Cb	Total
Tn	867	250	228	145	1490
Tc	298	274	22	23	617
Tw	156	43	97	57	353
Tb	65	29	28	18	140
Total	1386	596	375	243	2600

TABLE 6.3 – Matrice de confusion des prédictions obtenues avec le modèle *Multi-classe* et les coefficients de *MFCC* en tant que descripteurs.

	Cn	Cc	Cw	Cb	Total
Tn	1163	229	85	13	1490
Tc	374	211	31	1	617
Tw	265	63	20	5	353
Tb	103	29	8	0	140
Total	1905	532	144	19	2600

TABLE 6.4 – Matrice de confusion des prédictions obtenues avec le modèle *Multi-classe* et les descripteurs *bas-niveau* en entrée.

	Cn	Cc	Cw	Cb	Total
Tn	932	200	307	51	1490
Tc	283	256	33	45	617
Tw	165	38	121	29	353
Tb	68	35	20	17	140
Total	1448	529	481	142	2600

Partie 3 : Publication scientifique

Cette publication a été rédigée dans le cadre de l'appel à contribution lancé par la conférence *Content-Based Multimedia Indexing* (CBMI) qui se déroule les 4, 5 et 6 septembre 2018 à La Rochelle en France.

Cette publication met en valeur l'approche détaillée partie 5.4.1, qui propose des prédictions de classification à l'échelle du patient. La contribution principale de cet article est la mise en valeur d'une dichotomie en classification automatique. Dans certains cas, en particulier dans le domaine du traitement du son et du signal, la bonne vision n'est pas systématiquement une approche basée sur le contenu. Il est parfois bénéfique d'avoir un point de vue plus général du problème d'identification. Dans le cas présent, analyser l'ensemble des prédictions pour un patient permet d'avoir d'excellents résultats, tandis que les prédictions simplement basées sur le contenu audio des cycles respiratoires donnent des résultats plutôt moyens.

La publication a été rédigée en anglais selon les consignes imposées par la conférence, la mise en page respecte les normes IEEE, et les encadrants de ce stage ont co-signé le document.

Automatic detection of patient with respiratory diseases using lung sound analysis

Gaëtan CHAMBRES

Univ. Bordeaux, LaBRI,
Bordeaux INP, CNRS, UMR 5800
Talence, FRANCE

Pierre HANNA

Univ. Bordeaux, LaBRI
Bordeaux INP, CNRS, UMR 5800
Talence, FRANCE

Myriam DESAINTE-CATHERINE

Univ. Bordeaux, LaBRI
Bordeaux INP, CNRS, UMR 5800
Talence, FRANCE

Abstract—In modern medicine, every cardiac assessment or respiratory check-up includes an audio auscultation during which one the medical specialist listens to sounds from the patient body with different tools (stethoscope, sonography). This shows how important sound analysis is for heart and lungs disease detection. During the ICBHI 2017 challenge, a database of 920 records acquired from 126 subject, was used to find a method that predicted if a respiratory cycle contains, or not, adventitious sounds like crackles, wheezes or both of them. The team which submits the best results reached around 50% of correct detection. Using a machine learning approach with a boosted decisional tree model and more audio features leads to the same results. A new approach consists in creating a new model at the patient level, which is able to decide if a patient sounds sick or not, by taking as input the predicted results of the first classification model. This new model permits to reach 85% of good predictions and could be used as a tool for helping doctors to make better diagnosis.

Index Terms—Lung, pulmonary disease, respiratory sound, audio features, classification, machine learning, patient level, diagnosis, pathology

I. INTRODUCTION

In modern medicine, diagnostics are more often related to visual or audio information. Medical information can be acquired from different methods, and it is mainly represented to the medical specialist as images or sounds. Being able to correctly identify health problem regarding this information requires experience and time, but according to the World Health Organization (WHO) statistics [1], 45% of the WHO Member States report to have less than 1 physician per 1000 population, the WHO ratio recommendation. With such bad statistics, taking into account that a diagnosis implies to study individually every patient during a non compressible time and that health specialists are already overbooked, conditions are not optimum for them, and mistakes can be made. This is why finding new ways to help doctors to save time is a priority. Automatic and reliable tools can help to diagnosis more people, or it can help specialists to make less mistakes due to the work overload. Computer science can bring this new tools.

Since several decades, computer science improvements constantly ameliorate the ability to analyze automatically media data in real time. Indeed the diagnosis-help tools must be able

to process image and/or audio information. Computer science could help nursing staff or doctors by proposing faster and reliable tools for diagnosis [2]. It could give to the patient customizable tools for medical monitoring.

Every cardiac assessment or respiratory check-up includes an audio auscultation in which one the medical specialist listens to sounds from the patient body with different tools (stethoscope, sonography). This shows how important sound analysis is for heart and lungs disease detection. As heart disease remains the leading cause of death globally [3], there are many studies about heart sound identification [4]. A review [5] published in 2012 about signal processing techniques for heart sound analysis conclude that "the ability to correctly acquire, analyze and interpret heart sound signals for improved clinical diagnostic processes has become a priority". Since then, there are lots of improvements, for processing records taken in noisy environments [6] for example. Furthermore, new kinds of methods drastically improve the domain, as machine learning and deep learning [7]. These approaches contribute a lot to computer vision [8], [9] or audio analysis [10], [11]. By extension, there are several applications on medical studies and heart sound analysis [12], [13].

Analysis of respiratory sounds is less studied, but there are some existing works [14]. A review [15] published in 2007, highlights that previous researches permit to identify markers, such as crackles or wheezes. As said before, with the apparition of machine and deep learning, the performances of sound detection and classification are drastically improved [7] and studies about lung sound analysis take advantage of this progression [16] [17]. With identified markers, lung sound analysis can be transformed into a classification problem [18], a kind of problem that machine learning performs well to solve. This seems to be a relevant approach, but this kind of analysis more often focuses on the properties of the sound record, and not always on the patient level. Indeed, a doctor will listen several expirations and several inspirations, at several places on the patient body before taking a decision. This article raises the problem of using classification results based on sound content (micro-level) to take a decision at the patient level (macro-level), suggesting it the patient sounds sick or healthy.

The main contribution of this work is to suggest a new approach in audio classification. In some cases, here for lung pathologies, machine learning for audio classification based on sound content is not the best solution, or at least not alone. In this study, a machine learning approach is presented and performs as well as the previous state of the art. Using this classification model and extrapolating the results to take a decision on the patient level leads to better results.

II. LUNG SOUND ANALYSIS

A. Adventitious respiratory sounds

According to Reichert and al [15], an adventitious sound relates to *"additional respiratory sounds superimposed on normal breath sounds. It can be continuous (like wheezes) or discontinuous (such as crackles). Some of them (like squawks) have both characteristics. The presence of such sounds usually indicates pulmonary disorders."* Wheeze and crackle are the main categories of adventitious sounds. Crackles, which can reflect a pathology on lung tissue or airways, are discontinuous and explosive sounds. They are characterized by their specific waveform, their duration, and their location in the respiratory cycle. In opposition, a wheeze is a continuous sound with a musical character. It is characterized by periodic waveforms with a dominant frequency usually over 100 Hz and with a duration over 100 ms. When wheezes are observed, they usually indicate airways obstruction. A diagnosis-help tool could be seen as a classification problem, with a prediction model based on algorithm that detects this kind of markers. To be able to create a classification model, an annotated database with sufficient subjects and information is needed.

B. ICBHI 2017 Challenge - The database

The 2017 International Conference on Biomedical Health Informatics (ICBHI) organized a scientific challenge about automatic identification of crackles and wheezes in respiratory sounds. For this challenge, a specific database [19] was created. This database was made by collecting records of respiratory sounds from different research teams with different recording tools.

The work discussed in this paper is based on this database because it is diversified and fully annotated. For a given patient, one or several records are provided with one annotation file by record. Table I is an example of the content of an annotation file for a given record.

According to this table I, for each cycle contained in the given record, their position in time is given, and information about containing adventitious sound or not is given as a binary score. In this example, *cycle 1* starts at 1.018 second, ends at 3.411 second and only contains crackles. *Cycle 5* starts at 10.923 second, ends at 13.292, and only contains wheezes. According to the binary score, *cycle 3* contains both crackles and wheezes, and no respiratory cycle of this record sounds normal, with no adventitious sound (binary score would be 0 for crackles and 0 for wheezes).

Table II indicates the repartition of patients, records and the different type of cycles for the train and test set. This table

TABLE I
CONTENT OF THE ANNOTATION FILE FOR THE RECORD
107_2B4_AL_MC_AKGC417L [19]

Cycles	Start	End	Crackles	Wheezes
cycle 1	1.018	3.411	1	0
cycle 2	3.411	5.827	1	0
cycle 3	5.827	8.339	1	1
cycle 4	8.339	10.923	1	0
cycle 5	10.923	13.292	0	1
cycle 6	13.292	16.018	1	0
cycle 7	16.018	18.482	1	0
cycle 8	18.482	19.542	1	0

shows one limitation of the database. Crackles and wheezes repartition is not well balanced between the train and test sets. Another important information is the fact that there is a majority of normal cycles in both sets. Despite the unbalance, it was decided to keep this train-test separation because it was the split used during the ICBHI Challenge [19]. Having the same split permits to compare the obtained results to the state of the art.

TABLE II
REPARTITION OF THE DATA IN THE ICBHI CHALLENGE DATABASE [19]

Database	Train set	Test set	Total
Nb of patients	79	49	128
Nb of records	539	381	920
Nb of cycles with crackles	1247	617	1864
Nb of cycles with wheezes	533	353	886
Nb of cycles with both	366	140	506
Nb of normal cycles	2152	1490	3642
Nb total of cycles	4298	2600	6898

C. ICBHI 2017 Challenge - Submitted results

The challenge was divided in two parts, train and test. During the training phase, research teams accessed the public part of the database (the train set) to test their methods. During the test phase, research teams accessed the private part of the database (the test set) to be able to compute the accuracy Acc of their method. This last score was the final one in the evaluation of the method. It is computed by making the mean of sensitivity Se and specificity Sp for the predictions. As a reminder, these are the used formulas :

$$Se = \frac{(TP_C + TP_W + TP_B)}{(TP_C + FN_C) + (TP_W + FN_W) + (TP_B + FN_B)} \quad (1)$$

$$Sp = \frac{TP_N}{TP_N + FN_N} \quad (2)$$

$$Acc = \left(\frac{Se + Sp}{2} \right) * 100 \quad (3)$$

where TP_i stands for *TruePositive* in class i and FN_i for *FalseNegative* in class i . TP_i corresponds to the number of correctly classed elements in class i and $TP_i + FN_i$ the number total of elements that should be in class i . The different classes i are C for crackles, W for wheezes, B for both and N for normal.

It was decided to choose two of the results proposed during the ICBHI Challenge as the main comparison for this work, because these research teams proposed two different methods to classify respiratory cycles into the four categories of adventitious sound described before. As the idea is to propose a new way of thinking using machine learning, a focus is made on similar method of classification. The table III summarizes the results of both methods.

TABLE III
SUMMARY OF THE STATE OF THE ART RESULTS

Method	Se	Sp	Acc	Score
HMM	0.4232	0.5669	49.50%	39.37%
NLSp	0.4890	0.7780	49.98%	49.86%

For the first method, *HMM* stands for Hidden Markov Model classification. After preprocessing records to reduce noise and the heart beats, Mel-Frequency Cepstral Coefficients (MFCC) are used as inputs in a hidden Markov model. The model permits to compute the probability for a respiratory cycle to contain crackles or wheezes. The best score during train phase is $Se = 0.4232$, $Sp = 0.5669$ and $Acc = 49.50\%$, and the best score during test phase is $Score = 39.37\%$.

For the second method, *NLSp* stands for Non-Linear Spectral features extraction method. After preprocessing records to reduce the sample rate from 44100Hz to 4000Hz, and minimize noise effect with a Butterworth band pass filter, a non-linear resonance based wavelet decomposition is applied. This decomposition gives three channels (high,low,residue) and features are computed in each channel with a short time Fourier transform. Finally, the classification prediction is computed by a Support Vector Machine (SVM). The best score during the test phase is $Score = 49.86\%$ while the method gave better results on the train phase. To be able to compare, these are the closest results obtained during the train phase : $Se = 0.4890$, $Sp = 0.7780$ and $Acc = 49.98\%$.

III. PROPOSED METHOD

As the goal is to demonstrate the importance of adopting a patient-level point of view for the classification, the proposed method is divided in two parts. The first part is about the classification of the respiratory cycles depending of the adventitious sounds and the second part is about extrapolate the classification results to consider the patient classification. These parts are respectively named the micro-level part and the macro-level part.

A. Micro-level part

The main goal of this part is to classify individually every respiratory cycles depending if adventitious sounds are detected or not. For that, all records are taken one by one, and for each records, features are extracted on the signal window containing every cycles. The classification of each cycle is computed with a boosted decisional tree, which gives, according to the features, the probability to belong to every class. The higher probability is kept as the decision.

1) *Features computation*: Features were computed thanks to the *Essentia* library [22]. The automatic *Freesound* features extractor is used, which is designed for batch computations on large sound collections and is used by *Freesound*¹. It is configured from a profile specifying the parameters to use for the features computation : in this case, a frame size of 2048, a hop size of 1024, a Blackmann-Harris window and finally, the time window corresponding to the respiratory cycle. A complete description of computed features can be found on the *Essentia* website². The different categories are the *lowlevel features* (spectral information like barkbands, energybands, melbands, mfcc, etc), the *rhythm features* (beats loudness, bpm, etc), the *SFX features* (harmonicity and inharmonicity information) and the *tonal features* (chords strength, tuning frequency, etc).

To maximize the chance to correctly represent the audio signal, statistical values are computed from the features: standard deviation, variance, minimum and maximum, median, mean, means of first and second derivatives and variances of first and second derivatives.

After several classification tests based on the features, it appears that not all of them were relevant. A focus was done on *lowlevel features* because they provide spectral information, which is useful to identify wheezes, and information about punctual events, which is useful to identify crackles.

2) *The classification model*: The classification system is based on a boosted decisional tree algorithm. Each leaf of the tree will be a class, and the branches will be the features conjunction to reach the different leaves. Two approach with different kind of classification model were tested, a monaclass one and a multiclass one.

The monaclass approach is a succession of two monaclass model, where each tree have two leaves: *crackles* and *no crackles* for the first one, then *wheezes* and *no wheezes* for the second one. They respectively gives the probability for every respiratory cycles to contain crackles and then, the probability for every cycles to contain wheezes. The final decision is taken according to the diagram on figure 1 where the *YES* and *NO* ways are decided according to the highest probability.

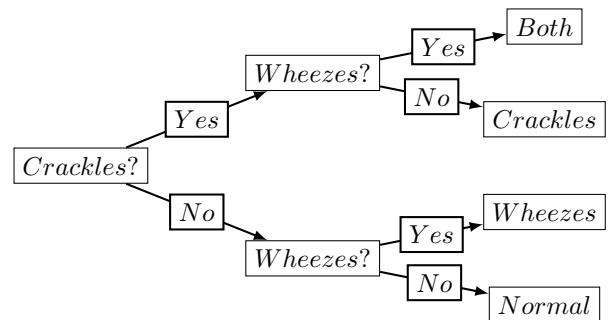


Fig. 1. Monaclass model : Decision diagram for final prediction

¹<https://freesound.org/>

²http://essentia.upf.edu/documentation/freesound_extractor.html

The second kind of model is a multiclass model, where a unique tree has 4 leaves for the 4 different classes: *normal*, *crackles*, *wheezes* or *both*. The final decision is taken by keeping the highest prediction of the four, one for each possible class.

3) *Results*: The system takes all the features as input and makes features conjunction, represented as branches of the tree, to reach the leaves. For a given respiratory cycle, by making the conjunction of the features, probability to belong to each class is given. The system is a boosted tree model. That means when the model is trained, several iterations are done to re-train the cycles which were not correctly classified. This permits to be more robust and precise. Parameters of the boosted decisional tree model were empirically set : the max depth of trees is 3, the maximum of boosting iterations is 100 and the learning rate is 0.1.

In the state of the art, the main features used are the mel-frequency cepstral coefficient (MFCC). To have a reliable comparison, a first bench of tests was done only using the 13 first mel-frequency cepstral coefficients as input features. In a search of improved precision, new results are computed using the *lowlevel* features. This two bench of tests were done for both monaclass and multiclass model. All the results are summarized in table IV, where *Mono/mfcc* and *Mono/lowlevel* indicate predictions made with the monaclass model, respectively with MFCC and lowlevel features as input. In the same way, *Multi/mfcc* and *Multi/lowlevel* indicate predictions made with the multiclass model, respectively with MFCC and lowlevel features as input.

TABLE IV
SUMMARY OF THE PREDICTION RESULTS

Model	Se	Sp	Acc
Mono/mfcc	0.2216	0.7604	49.10%
Mono/lowlevel	0.3505	0.5819	46.62%
Multi/mfcc	0.2081	0.7805	49.43%
Multi/lowlevel	0.3550	0.6255	49.02%

Although the accuracy is comparable to the state of the art because of the computation formula, sensitivity and specificity are not balanced. For the monaclass model with only MFCC features for example, normal cycles are correctly classed 76% of the time, but cycles with adventitious sounds are correctly classed only 22% of the time. Taking a look at the different confusion matrices permits to explain visually this problem. Table V refers to the confusion matrix of *Mono/mfcc*, table VI to *Mono/lowlevel*, table VII to *Multi/mfcc* and table VIII to *Multi/lowlevel*. In each table, C_i corresponds to the number of correctly predicted cycles for the class i , T_i corresponds to the number of cycle that should be predicted to be in class i and i corresponds to the different classes : n for normal, c for crackles, w for wheezes and b for both crackles and wheezes.

TABLE V
CONFUSION MATRIX OF THE PREDICTION OBTAINED WITH THE MONOCLASS MODEL AND MFCC FEATURES AS INPUT.

	Cn	Cc	Cw	Cb	Total
Tn	1133	224	111	22	1490
Tc	357	221	23	16	617
Tw	246	71	25	11	353
Tb	105	25	10	0	140
Total	1841	541	169	49	2600

TABLE VI
CONFUSION MATRIX OF THE PREDICTION OBTAINED WITH THE MONOCLASS MODEL AND LOWLEVEL FEATURES AS INPUT.

	Cn	Cc	Cw	Cb	Total
Tn	867	250	228	145	1490
Tc	298	274	22	23	617
Tw	156	43	97	57	353
Tb	65	29	28	18	140
Total	1386	596	375	243	2600

TABLE VII
CONFUSION MATRIX OF THE PREDICTION OBTAINED WITH THE MULTICLASS MODEL AND MFCC FEATURES AS INPUT.

	Cn	Cc	Cw	Cb	Total
Tn	1163	229	85	13	1490
Tc	374	211	31	1	617
Tw	265	63	20	5	353
Tb	103	29	8	0	140
Total	1905	532	144	19	2600

TABLE VIII
CONFUSION MATRIX OF THE PREDICTION OBTAINED WITH THE MULTICLASS MODEL AND LOWLEVEL FEATURES AS INPUT.

	Cn	Cc	Cw	Cb	Total
Tn	932	200	307	51	1490
Tc	283	256	33	45	617
Tw	165	38	121	29	353
Tb	68	35	20	17	140
Total	1448	529	481	142	2600

These confusion matrices permit to show several information. The models using MFCCs are better at classifying normal cycles. For the monaclass model, the precision of normal cycles correctly predicted is increased from 58% with lowlevel features (table VI) to 77% with MFCCs (table V). It is the same for the multiclass model, with a precision increased from 62% (table VIII) to 78% (table VII). On the other hand, crackles and wheezes are more correctly predicted when using the lowlevel features. It seems normal as lowlevel features catch more information about different audio events, for the crackles in particular which are less linked to spectral features. Finally, these matrices permit to see that the multiclass model is more efficient for wheezes detection and less precise for crackles detection. The precision for wheezes detection is increased from 27% (table VI) to 34% (table VIII) but the precision for crackles detection is decreased from 44% (table VIII) to 41% (table VI).

To have more readable results, table IX summarizes the results discussed from the confusion matrices, where each line of the table IX corresponds to a confusion matrix. The following values are provided in percent: the specificity Sp and the sensitivity Se_i for the three classes with c for crackle, w for wheeze and b for both. Acc stands for accuracy, with the same definition to table IV.

TABLE IX
SUMMARY OF THE CORRECTLY PREDICTED CYCLES FOR EACH CLASS IN PERCENT AND TOTAL ACCURACY

Model	Sp	Se_c	Se_w	Se_b	Acc
Mono/mfcc	76	36	7	0	49.10
Mono/lowlevel	58	44	27	13	46.62
Multi/mfcc	78	34	6	0	49.43
Multi/Lowlevel	63	41	34	12	49.02

B. Macro-level part

As the main idea of studying lung diseases with an automatic model is to propose a model which could help the lung specialist by saving time, the goal of the macro-level part is to suggest a "diagnostic" taking into account the totality of the predictions previously computed. Here, "diagnostic" means is the patient sounds sick or healthy. As a doctor not only listens one time the lung of his patient, but several times at different area of the body, we compute the different kind of cycle ratio, predicted for all the cycles of one patient. Depending on what kind of cycles appears the most, a decision is taken.

Has seen before, the classification models in the micro-level part give the probability for each respiratory cycles to be in each class depending the adventitious sounds. These results are quite good, but the hope was that when we look all the decisions for all the cycles of a patient, a trend appears. For example, for a patient with lot of crackles in his respiration, not all the cycles have to contain crackles. It could appear that one or more of them are normal. This is what looking at a trend is interesting. If this trend appears to be symptomatic, it is possible to suggest that the patient is sick. To identify the trend, the results are packed for all the cycles of the same patient. This leads to have a list of class for each patient, corresponding to the classes of the different cycles. After generating this list for every patient we want to classify, it remains to count how many cycles are normal, how many contains crackles, etc. The ratios are naturally obtained by:

$$Ratio_i = \frac{Cycles_i}{TC} \quad (4)$$

where $Cycles_i$ is the number of cycles predicted to be in class i and TC is the total of cycles analyzed for the current patient. Taking as goal to identify healthy patient, the decision should be taken depending the following question : *Is there a majority of normal cycles ?* and the two answers will be *if yes, the patient should be healthy* and *if no, the patient should be sick*. But, as explained before, and as confirmed by taking a deep look at the database, symptoms are non-systematical. It seems that crackles and wheezes do not appear on every

cycles of a sick patient and a lot of them have a majority of normal respiratory cycles, even the sick ones. This is why the decisive question is a bit transformed as follows: *Does the patient only have normal cycles ?* This is possible thanks to the problematic: it is more important to not miss a sick patient, thinking he is healthy, than to miss an healthy patient, thinking he is sick.

Keeping in mind that the classification results were not perfect, a threshold is used on the ratio. Two different thresholds have been empirically tested : $threshold = 0$ means that a patient is considered healthy if there is no cycle predicted with adventitious sounds at all (every cycles are predicted as normal), and $threshold = 0.1$ means that a patient is considered sick if less than 10% of the total of cycles are predicted as having adventitious sound. These 10% were empirically set to test the contribution of the threshold, but best results were obtained with $threshold = 0$. This proves that the respiratory cycles classification on the micro-level part is sufficiently strong to be used as input in the macro-level classification. The final results of the number of classified patient are visible in table X.

According to this table, 85% of the patients in the test dataset were correctly classified. It is important to keep in mind that the total number of patient is limited. For a more robust conclusion, the method should be extended to a bigger database.

TABLE X
CONFUSION MATRIX OF THE PREDICTION OF THE PATIENT HEALTH STATE

	Sick	Healty	Total
Sick	29	5	34
Healty	2	11	13
Total	31	16	47

IV. CONCLUSION

A. Discussion of the results

Looking at the sound content on respiratory sound seems to be difficult for a classification. Our classification cannot do better than the state of the art, and it seems difficult to have more than 50% of accuracy on the test set of this database. This is certainly due to the variable quality of the records, with sometimes a lot of noise and interferences. The unbalanced data could also be a cause of bad predictions, because it affects the model training. However, the new approach proposed permits to use these results to take a decision at the patient level. As a specialist will listen several cycles at different places on the patient body, we proposed to not only focus on the audio content for each respiratory cycle, but to predict a trend for each patient. This macro-level approach gives good results with an accuracy up to 85 %. It is possible to consider this model could be extended as automatic guide for diagnosis, and could help professionals by saving time or for educational purpose.

B. Perspectives

Like it was done in previous studies, the records could be preprocessed to avoid noise and interference. This could give better result for the respiratory cycles classification, but it is necessary to prove that the symptom information is not modified by the filtering step.

In macro-level part, some problems appear with the database. For example, some of the patient were annotated as having URTI for Upper Respiratory Tract Infection, but after taking a look at their records annotations it seems that no cycle present an adventitious sound : all recorded cycles are normal, with no crackle nor wheeze. With our system, we are not able to consider this kind of patient as sick. If we do not consider this type of subjects, the score could be increased up to $accuracy = 90\%$.

A new approach for decision on patient level was discussed about establishing a correlation between the presence of crackles, wheezes or both and a particular pathology. Unfortunately, for 3 of the 7 pathologies contained in the dataset, Asthma, Lower Respiratory Tract Infections (URTI) and Pneumonia, there is no subject in the test set. Trying to find a correlation for these pathologies implies to consider a new repartition of the database, and lose the reliability of comparing the results to the state of the art.

The utility of features is to represents the content of the audio signal. Another idea is to compute features on the respiratory record and train a classification model to identify directly the pathology. Crackles and wheezes are ones of the main adventitious sounds, but they are not the only ones [15], so it could be interesting to not restrict the information to them. These could help to make the difference between patient with URTI and healthy patient. For the 3 other pathologies (Common Obstructive Pulmonary Disease, Bronchiolitis and Bronchiectasis), finding a profile by training the model must be a very complex problem, because of the unbalanced and limited database. Correctly identify the features reliable to detect adventitious sounds will help in the process of establishing a new and normalized database.

ACKNOWLEDGMENT

This work was done during a Master internship at LaBRI, supported by Univ. Bordeaux. We are thankful to our colleague Gyorgy Kurtag who provided expertise that greatly assisted the research. Finally, a very special thanks goes to Annick Mersier and the whole SCRIME³ team for providing all that was necessary for the smooth running of the internship.

REFERENCES

- [1] World Health Organization. Density of physicians. http://www.who.int/gho/health_workforce/physicians_density/en/, 2017. [Online; accessed 18-May-2018].
- [2] Klaus Schoeffmann, Bernd Mnzer, Michael Riegler, and Pl Halvorsen. 2017. Medical Multimedia Information Systems (MMIS). In Proceedings of the 2017 ACM on Multimedia Conference (MM '17). ACM, New York, NY, USA, 1957-1958.
- [3] World Health Organization. Cardiovascular diseases (cvds). <http://who.int/mediacentre/factsheets/fs317/en/>, 2017. [Online; accessed 16-May-2018].
- [4] X. Yang, F. Yang, L. Gobeawan, S.Y. Yeo, S. Leng, L. Zhong and Y. Su, A multi-modal classifier for heart sound recordings. in Computing in Cardiology, vol. 43, 2017.
- [5] B. Emmanuel. A review of signal processing techniques for heart sound analysis in clinical diagnosis. Journal of medical engineering and technology. vol 36, pages 303-7. 2012.
- [6] M. Abdollahpur, A. Ghaffari, S. Ghiasi and M. Mollakazemi, Detection of pathological heart sounds, in Physiological Measurement, vol.38, num. 8, pp 1616-1630, 2017.
- [7] X. Du, Y. Cai, S. Wang and L. Zhang, "Overview of deep learning," 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), Wuhan, 2016, pp. 159-164.
- [8] Q. Wu, Y. Liu, Q. Li, S. Jin and F. Li, "The application of deep learning in computer vision," 2017 Chinese Automation Congress (CAC), Jinan, 2017, pp. 6522-6527.
- [9] E. Nishani and B. io, "Computer vision approaches based on deep learning and neural networks: Deep neural networks for video analysis of human pose estimation," 2017 6th Mediterranean Conference on Embedded Computing (MECO), Bar, 2017, pp. 1-4.
- [10] Y. Petetin, C. Laroche and A. Mayoue, "Deep neural networks for audio scene recognition," 2015 23rd European Signal Processing Conference (EUSIPCO), Nice, 2015, pp. 125-129.
- [11] J. Li, W. Dai, F. Metze, S. Qu and S. Das, "A comparison of Deep Learning methods for environmental sound detection," 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, 2017, pp. 126-130.
- [12] G.D. Clifford, C. Liu, B. Moody, J. Millet, S. Schmidt and Q. Li, I. Silva and R.G. Mark, Recent advances in heart sound analysis, in Physiological Measurement, vol.38, num. 8, pp E10-E25, 2017.
- [13] V. Maknickas and A. Maknickas, Recognition of normalabnormal phonocardiographic signals using deep convolutional neural networks and mel-frequency spectral coefficients, in Physiological Measurement, vol. 38, num. 8, pp 1671-1684, 2017.
- [14] H. Polat and I. Güler, A simple computer-based measurement and analysis system of pulmonary auscultation sounds, in Journal of medical systems, vol. 28, num. 6, pp.665-672, 2004.
- [15] S. Reichert and R. Gass and E. Andrès, Analyse des sons auscultatoires pulmonaires (Analysis of respiratory sounds), in ITBM-RBM, vol. 28, pp 169-180, 2007.
- [16] G. Chang, and Y. Lai, Performance evaluation and enhancement of lung sound recognition system in two real noisy environments, in Computer Methods and Programs in Biomedicine, vol. 97, num. 2, pp 141-150, 2010.
- [17] A. Abbas and A. Fahim, An Automated Computerized Auscultation and Diagnostic System for Pulmonary Diseases; in Journal of Medical Systems, vol. 34, num. 6, pp 1149-1155, 2010.
- [18] C. Chen, W. Huang, T. Tan, C. Chang and Y. Chang, Using K-Nearest Neighbor Classification to Diagnose Abnormal Lung Sounds, in Sensors, vol. 15, num. 6, pp 13132-13158, 2015.
- [19] B. M. Rocha, D. Filos, L. Mendes, I. Vogiatzis, E. Perantoni, E. Kaimakamis, P. Natsiavas, A. Oliveira, C. Jcome, A. Marques, R. P. Paiva, I. Chouvarda, P. Carvalho and N. Maglaveras, Respiratory Sound Database for the Development of Automated Classification, In Precision Medicine Powered by pHealth and Connected Health, 2017, pp 33-37
- [20] Jakovljevi N., Lonar-Turukalo T. (2018) Hidden Markov Model Based Respiratory Sound Classification. In: Maglaveras N., Chouvarda I., de Carvalho P. (eds) Precision Medicine Powered by pHealth and Connected Health. IFMBE Proceedings, vol 66. Springer, Singapore
- [21] Serbes G., Ulukaya S., Kahya Y.P. (2018) An Automated Lung Sound Preprocessing and Classification System Based OnSpectral Analysis Methods. In: Maglaveras N., Chouvarda I., de Carvalho P. (eds) Precision Medicine Powered by pHealth and Connected Health. IFMBE Proceedings, vol 66. Springer, Singapore
- [22] Bogdanov, D., Wack N., Gmez E., Gulati S., Herrera P., Mayor O., et al. (2013). ESSENTIA: an Audio Analysis Library for Music Information Retrieval. International Society for Music Information Retrieval Conference (ISMIR'13). 493-498.

³<https://scrime.u-bordeaux.fr/>