



République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et  
de la Recherche Scientifique



Université Echahid Chikh Larbi Tébessi - Tébessa

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

Département des Mathématiques et Informatique

## Mémoire

Présenté en vue de l'obtention du diplôme de MASTER

Filière : (Mathématiques/Informatique)

Option : Systèmes et multimédia

Par

KOUACHI Raiane

# Classification audio basée sur les représentations spectrographiques

*Date de soutenance Juin 2024*

### ***Devant le jury***

<i>Dr. Abdelatif SAHRAOUI</i>	<i>MCA</i>	<i>Université Larbi Tébessi</i>	<i>Président</i>
<i>Dr. Walid BOUAMRA</i>	<i>MCB</i>	<i>Université Larbi Tébessi</i>	<i>Examineur</i>
<i>Mr. Abdellatif GAHMOUSSE</i>	<i>MAA</i>	<i>Université Larbi Tébessi</i>	<i>Encadreur</i>
<i>Dr. Mohamed GASMI</i>	<i>MCA</i>	<i>Université Larbi Tébessi</i>	<i>Co-Encadreur</i>

*Année universitaire : 2023/2024*

## **Résumé:**

Dans ce projet, nous nous concentrons sur les maladies cardiovasculaires en visant le développement d'un système de pointe reposant sur l'intelligence artificielle et l'apprentissage automatique, en mettant particulièrement l'accent sur les techniques d'apprentissage profond. Ce système a pour but d'améliorer le diagnostic et la détection des anomalies du rythme cardiaque en utilisant les ensembles de données CirCor DigiScope Phonocardiogram 2022 et la base de données des sons cardiaques fœtaux de l'Université de Shiraz.

Pour entraîner notre modèle à reconnaître les sons cardiaques, nous avons utilisé des architectures avancées telles que ResNet50 et VGG19. Les résultats ont été extrêmement prometteurs, confirmant l'efficacité de notre approche. En particulier, le modèle VGG19 a atteint une précision remarquable de 96,88 %, soulignant ainsi la robustesse et l'exactitude de notre système.

## المخلص

في هذا المشروع نتعامل مع المجال الطبي المتعلق بأمراض القلب، وهدفنا هو تطوير نظام يعتمد على الذكاء الاصطناعي والتعلم الآلي، ولا سيما التعلم العميق، ويساعد هذا النظام في التشخيص للكشف عن تشوهات ضربات القلب باستخدام مجموعة بيانات The CirCor DigiScope و Phonocardiogram Dataset 2022 Shiraz University Fetal Heart Sounds Database.

من أجل تدريب النموذج على التعرف على أصوات القلب استخدمنا نماذج متقدمة مثل: VGG19، ResNet50 وكانت النتائج التي تم الحصول عليها مرضية للغاية، حيث أثبتت التجارب من النهج المقترح فعاليته من خلال تحقيق نسبة مقدرة بـ 96.88 % باستخدام النموذج التحويلي VGG19 .

# **Abstract:**

In this project, we focus on the medical domain of cardiovascular diseases. Our objective is to develop an AI and machine learning-based system, particularly leveraging deep learning techniques, to assist in diagnosing and detecting heart rhythm abnormalities. We utilize the CirCor DigiScope Phonocardiogram Dataset 2022 and the Shiraz University Fetal Heart Sounds Database for this purpose.

To train the model for recognizing heart sounds, we employed advanced architectures such as ResNet50 and VGG19. The results were highly promising, with our experiments validating the effectiveness of the proposed approach. Specifically, the VGG19 model achieved an impressive accuracy rate of 96.88%.

# **REMERCIEMENTS**

Je remercie Dieu Tout-Puissant de m'avoir donné force, courage et patience tout au long de ces années.

Toute ma gratitude et remerciement vont à mes encadrants Mr GAHMOUSSE Abdelatif et Mr GASMI Mohammed, pour avoir accepté de diriger ce travail, leurs encouragements, leur disponibilité et pour leur rigueur scientifique, pour leurs nombreux conseils qui ont contribué à la réalisation de ce travail, pour la confiance qu'ils m'ont accordée, et leur soutien scientifique et moral, qui m'ont été d'une grande aide durant Période de travail, merci beaucoup.

Je remercie également les membres du jury pour l'honneur qu'ils m'ont fait, en acceptant d'examiner ce travail.

J'adresse mes remerciements à tous les professeurs qui ont participé à notre enseignement tout au long de notre parcours pédagogique et pour leur soutien scientifique et moral.

Enfin, je remercie ma famille pour sa présence continue à mes côtés et son soutien à mon égard tout au long de mon parcours éducatif .

**Raiane**

# Table des matières

Introduction générale .....	01
-----------------------------	----

## **Chapitre 01: Concepts et Outils pour la Reconnaissance du son numérique**

1- Présentation du son .....	04
1-1 Signal analogique/numérique .....	05
1-1-1 La différence entre signal analogique/numérique. ....	06
1-1-2 Conversion entre analogique et numérique. ....	06
1-2 Des formats de fichiers de données audio.....	07
1-2-1 Le format audio compressé (sans perte) .....	07
1-2-2 Le format audio compressé (avec perte).....	08
1-2-3 Le format audio non compressé (sans perte).....	08
2- Analyse audio .....	09
2-1 Le rôle d'analyse audio.....	09
2-2 Principes de la transformation de données audio .....	09
2-3 Les filtres et l'équaliseur .....	12
2-3-1 Les filtres .....	12
2-3-2 Un égaliseur (EQ).....	13
2-3-3 Les types d'égaliseurs .....	13
2-3-4 Effet d'égalisation .....	14
2-4 Le Spectrogramme .....	15
2-4-1 Les types de spectrogrammes.....	16
3- Modèles de reconnaissance du son.....	18
3-1 Types de modèles de reconnaissance du son.....	19
3-2 Applications de la reconnaissance du son .....	19

## **Chapitre 02 : Santé Cardiaque**

1- Le Cœur humain .....	23
1-1 L'importance du Cœur .....	24
1-2 Les maladies du cœur .....	24

1-3 Types des maladies cardiaques.....	25
2- Outils pour le diagnostic de cardiologie .....	26
2-1 Un stéthoscope .....	26
2-2 L'ECG (Électrocardiogramme) .....	27
2-3 Phonocardiography.....	28
2-4 Logiciel d'analyse du son cardiaque .....	28
2-5 Difficultés des outils de diagnostic cardiaque.....	28
3-L'IA dans les maladies cardiologie .....	29
3-1 Jeux de données .....	29
3-1-1 MIT-BIH .....	30
3-1-2 PASCAL .....	30
3-1-3 Shiraz.....	31
3-1-4 CirCor 2022.....	31
3-2 Travaux connexes.....	32

### **Chapitre 03 Approche proposée**

1-Outils et bibliothèques .....	38
1-1 Distribution des Classes du jeux de données .....	40
1-2 Distribution des Classes du jeu de données Shiraz.....	40
1-3 Distribution des Classes du jeu de données Circor22.....	41
2- Méthodologie .....	43
2-1 Prétraitement des données audios .....	44
2-2 Classification basées Transfert Learning .....	45
2-2-1 ResNet50 .....	46
2-2-2 VGG19 .....	46
2-3 Résultats .....	47
3- Analyse et Filtrage des jeux de données basée sur le balayage rythmique	50
3-1 Catégorisation des Segments .....	53
3-2 Statistiques des jeux de données après filtrage des bruits .....	55
3-3 Prétraitement des données audios après filtrage des bruits.....	56

3-4 Résultats .....	58
4- Discussion des Résultats .....	63
Conclusion général .....	66
Bibliographie.....	69

## Liste de figures

Figure.1.1: la normalisation de son analogique / numérique .....	05
Figure.1.2: Conversion entre analogique et numérique .....	07
Figure.1.3: La transformation rapide de Fourier (FFT) .....	10
Figure.1.4: MFCC spectre .....	11
Figure.1.5: WAV forme .....	11
Figure.1.6: STFT spectre.....	12
Figure.1.7: L'égaliseur paramétrique.....	13
Figure.1.8: Spectrogramme .....	15
Figure.1.9: Les types de spectrogrammes.....	17
Figure.1.10: Spectrogramme de Mel.....	18
Figure.2.1: fonctionnement du cœur.....	24
Figure.2.2: Le stéthoscope .....	26
Figure.2.3: L'ECG.....	27
Figure.2.4: Phonocardiography signal.....	28
Figure.3.1. Méthodologie proposée .....	44
Figure.3.2 Extraction de la zone d'intérêt. ....	45
Figure.3.3 Matrice de confusion de la phase test en utilisant ResNet50 .....	48
Figure.3.4. Courbes de l'Accuracy et Loss en utilisant ResNet50.....	48
Figure.3.5. Matrice de confusion de la phase test en utilisant VGG19.....	49
Figure.3.6. Courbes de l'Accuracy et Loss en utilisant VGG19 .....	50
Figure.3.7. Résultats du balayage rythmique de quelques échantillons du jeu de donnée Shiraz.....	52
Figure.3.8. Méthodologie proposée avec Filtrage de bruits et Filtre à basse fréquence .....	57
Figure.3.9. Extraction de la zone d'intérêt après filtre à basse fréquence.....	58
Figure.3.10 Matrice de confusion de la classification VGG19 et ResNet50.....	59
Figure.3.11. Courbes VGG19 avec Shiraz LFF 300 .....	60
Figure.3.12. Courbes ResNet50 avec Shiraz LFF 300 .....	61
Figures.3.13 Matrice de confusion & Courbes VGG19 avec Circor22 LFF 100 .....	62

## Liste de tableaux

Table.3.1. Distribution des Classes du jeu de données Shiraz .....	41
Table.3.2. Distribution des Classes du jeu de données Circor22 .....	43
Tableau.3.3 Métrique de la classification en utilisant ResNet50 .....	47
Tableau.3.4. Métrique de la classification en utilisant VGG19 .....	49
Tableau.3.5. Statistiques Histogramiques des durées d'enregistrements.....	51
Tableau.3.6. Résultats du balayage rythmique et Histogramique sur le jeu de donnée Shiraz .....	53
Tableau.3.7. Résultats du balayage rythmique et histogramique sur le jeu de donnée Circor22.....	54
Tableau.3.8. Distribution des classes après filtrage de bruits (Shiraz) .....	55
Tableau.3.9. Distribution des classes après filtrage de bruits (Circor22) .....	55
Tableau.3.10. Résultats de la classification TL sur le jeu de données Shiraz après filtrage de bruits .....	59
Tableau.3.11. Résultats de la Classification en utilisant le VGG19 et ResNet50 sur les trois lots après l'application du filtre de basse fréquence (LFF). .....	60
Tableau.3.12. Résultats de la Classification en utilisant le VGG19, ResNet50 et MobilNet après l'application du filtre de basse fréquence (LFF 100) sur les échantillons élus du Circor22. ....	61

# **I**ntroduction générale

Les soins de santé sont un ensemble complet de services et d'interventions visant à protéger la santé publique des personnes contre la maladie, ce qui assure une vie saine, améliorant ainsi la qualité de vie et augmentant l'espérance de vie.

Les maladies cardiovasculaires sont parmi les plus répandues et difficiles à traiter. La détection précoce des maladies cardiovasculaires est très importante pour réduire les risques et améliorer les résultats. Les maladies cardiaques sont devenues l'une des principales causes de décès dans le monde.

La santé et les soins cardiaques bénéficient grandement de l'IA dans de nombreux aspects qui travaillent pour assurer la santé du cœur et du corps, et l'IA aide à diagnostiquer et à traiter les problèmes cardiaques plus rapidement en analysant de nombreuses données médicales.

Dans ce contexte, l'un de ces systèmes a été choisi pour être étudié dans ce mémoire, Un système intelligent basé sur le Transfert Learning pour la classification des sons à partir de leurs spectrogrammes.

Il existe de nombreux défis liés à la classification des sons cardiaques comme naturels ou anormaux. La méthode traditionnelle d'attribution du son cardiaque par les médecins a des limites en matière de précision et de cohérence, en s'appuyant fortement sur l'expérience du médecin. Il est nécessaire de disposer de méthodes de mécanisme plus fiables pour classer les sons cardiaques permettant un diagnostic précoce et une intervention rapide des maladies cardiaques qui constituent un problème de santé mondial majeur.

L'objectif principal est de fournir une solution efficace et rapide alimentée par l'IA qui peut identifier avec précision les sons cardiaques naturels et anormaux, en particulier dans les endroits où l'accès aux soins

cardiaques spécialisés est limité. Ainsi, on peut améliorer la précision et la cohérence de la classification des sons cardiaques pour la détection précoce des maladies cardiaques, en utilisant l'apprentissage automatique et l'apprentissage profond plutôt que des méthodes traditionnelles.

Notre étude se fonde sur les trois chapitres suivants : le premier chapitre intitulé « Concepts et outils pour la reconnaissance du son numérique », dans lequel on présente une brève description de la reconnaissance de son, on rappelle leur historique et parlons en général à tout ce qui concerne le son numérique et les principes de la transformation de données audio. Le deuxième chapitre entame la thématique de la "Santé cardiaque", les soins cardiaques, les maladies associées, ainsi que les travaux basés sur l'IA dans ce domaine. On termine par un dernier chapitre illustrant l'approche proposée. Ce chapitre contient la modélisation de l'approche proposée ; une présentation des outils et des bibliothèques exploitées, suivi par l'implémentation et l'expérimentation réalisée et suivi par les résultats obtenus et leurs discussions. A la fin de ce mémoire nous concluons par une conclusion générale qui résume nos contributions et perspectives.

# **C**hapitre 01:

## **Concepts et Outils pour la Reconnaissance du son numérique**

Dans nos vies, les voix jouent un rôle essentiel et important, qu'elles soient amusantes ou ennuyeuses, basses ou hautes, calmes ou puissantes, affectant notre humeur et nos décisions. Ces informations sont constamment traitées par notre cerveau pour nous donner des preuves fondamentales et vitales sur notre environnement. Cependant, il est possible d'analyser avec précision les signaux vocaux pour nous fournir des informations plus explicites grâce aux avancées technologiques et scientifiques actuelles.

Ces développements sont passés par des étapes historiques fondamentales dans les années 1950, les premiers systèmes d'identification de nombres parlés ont évolué; Dans les années 1960, l'introduction de modèles cachés de Markov (HMM) pour la reconnaissance de la parole; Dans les années 1970, des systèmes étendus de reconnaissance de la parole ont été développés; Dans les années 1980, les réseaux de neurones synthétiques ont émergé pour une identification adéquate; Dans les années 1990, grâce aux progrès des algorithmes et des capacités informatiques, il y a eu des améliorations importantes dans les performances d'identification; Dans les années 2000, le développement de systèmes de reconnaissance du son en temps réel et d'applications grand public comme la reconnaissance vocale sur les smartphones .Enfin, de 2010 à aujourd'hui, l'apprentissage profond a émergé pour une bonne reconnaissance, conduisant à des progrès significatifs dans les derniers développements [1] [2].

## **1- Présentation du son**

Le son est l'une des formes d'énergie qui émet lorsque les particules d'air vibrent et se déplacent dans un modèle particulier appelé ondes sonores.

Ces ondes sont des ondes longitudinales qui nécessitent un milieu matériel pour se propager, tel que l'air ou l'eau. Elles sont caractérisées par leur amplitude, qui correspond au volume perçu du son, leur fréquence, qui

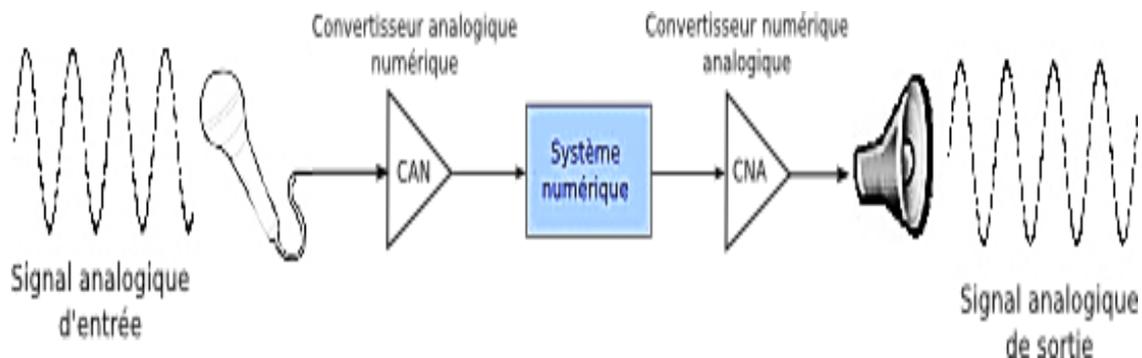
détermine la hauteur perçue (ou le pitch), et leur longueur d'onde, qui est liée à leur vitesse de propagation [3] [4] [5].

### 1-1 Signal analogique/numérique

Le son analogique est la transformation des vibrations sonores en signaux électriques variant exactement dans le même temps et dans les mêmes proportions. À l'inverse, un signal numérique est un signal discontinu qui ne fluctue que pendant des moments spécifiques.

La numérisation de signal est une méthode utilisée pour stocker, dupliquer et transmettre des signaux sans perte, effectuer des traitements numériques et fusionner différents signaux sur les mêmes supports, offrant précision, stabilité et modularité.

Le but de la numérisation est de transformer le signal analogique qui contient une quantité infinie d'amplitudes en un signal numérique contenant lui une quantité finie de valeurs.



**Figure.1.1: la normalisation de son analogique / numérique [6]**

La quantification du son peut différer avant et après la transmission en raison de processus tels que la compression audio, les perturbations pendant la transmission sur un réseau, et les conversions numériques-analogiques. Chaque étape du processus de capture, de transmission et de restitution du son peut influencer la qualité finale de l'audio quantifié.

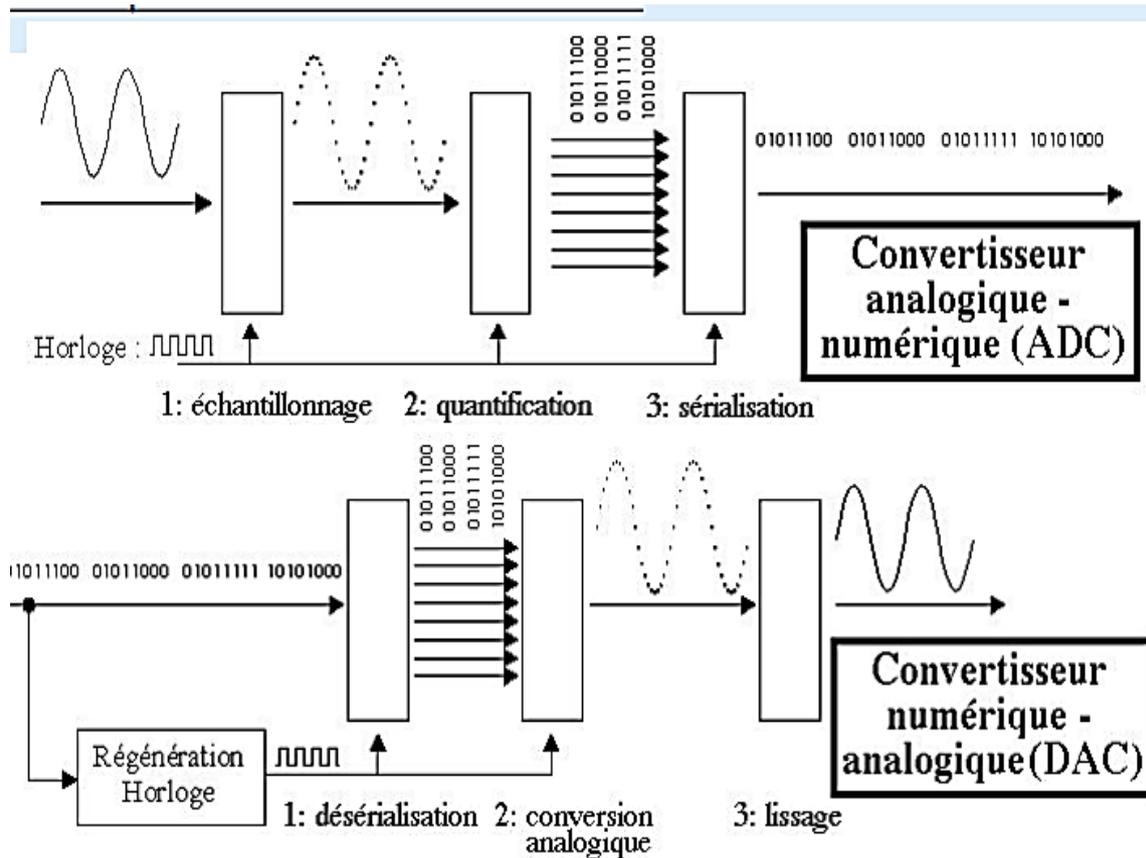
**1-1-1 La différence entre signal analogique/numérique.**

La différence entre les deux réside dans la façon dont les signaux (en l'occurrence sonores) sont traités et sauvegardés. Si, dans le cas de la technologie numérique, un signal sonore est converti en système binaire (à base de 1 et de 0), la technologie analogique conserve le signal sonore sous sa forme non codée. Par ailleurs, lorsqu'on considère qu'un signal analogique peut prendre un nombre infini de valeurs alors qu'un signal numérique ne peut en prendre que deux, il semble que l'analogique est le mieux placé des deux. Les signaux analogiques ont cependant des inconvénients qui font que la plupart des systèmes électroniques modernes sont numériques.

**1-1-2 Conversion entre analogique et numérique.**

La conversion d'un signal analogique en signal numérique est réalisée par un convertisseur analogique numérique (ADC en anglais pour: Analogue to Digital Converter). La réciproque est assurée par un DAC (Digital to Analogue Converter).

Lorsqu'un signal est converti de l'analogique en numérique, il faut choisir une résolution pour la conversion. La résolution détermine le nombre de valeurs numériques possibles en correspondance des valeurs analogiques [6] [7] [8] [9].



**Figure.1.2: Conversion entre analogique et numérique [7]**

## 1-2 Des formats de fichiers de données audio

Le fichier audio est la forme numérique de l'audio analogique qui existe depuis de très nombreuses années. Pour cela, on utilise un code informatique qui permet de convertir des langages audio en différents formats.

### 1-2-1 Le format audio compressé (sans perte)

La compression sans perte est une technique de compression qui permet de préserver une qualité CD du fichier source original tout en réduisant la taille du fichier à un niveau raisonnable. Les formats audio compressés « lossless » couramment utilisés incluent :

- FLAC (Free Lossless Audio Codec).
- ALAC (Apple Lossless Audio Codec).

Les formats audio compressés sans perte utilisent des algorithmes de compression qui préservent l'intégrité totale des informations contenues dans le flux audio. Le son est ainsi exactement le même que celui de la source originale.

### **1-2-2 Le format audio compressé (avec perte)**

Les formats audio compressés réduisent la taille des données en supprimant les informations détectées comme « superflues ». Cela permet de stocker plus de données audio dans moins d'espace disque.

Cette perte d'informations peut se traduire par une atténuation des cymbales, une réponse aux grave plus douce, un flou de l'attaque ou même une certaine dégradation de la vibration de la corde d'une basse. L'expérience musicale se retrouve dans une certaine mesure dénaturée.

Les formats audio compressés couramment utilisés incluent :

- MQA (de l'anglais, Master Quality Authenticated).
- MP3 (MPEG-1 Audio Layer III).
- AAC (Augmented and Alternative Communication).

### **1-2-3 Le format audio non compressé (sans perte)**

Un format audio non compressé est un format d'enregistrement qui ne réduit pas la taille des données en supprimant une part des informations. Il conserve donc toutes les données d'origine, garantissant une qualité sonore maximale. Les formats audio non compressés couramment utilisés incluent.

- DSD (de l'anglais « Direct Stream Digital »).
- AIFF (Audio Interchange File Format).
- WAV (Waveform Audio File Format) [10] [11].

## **2-Analyse audio**

L'analyse audio consiste à analyser et extraire des informations pertinentes à partir de signaux audios. Elle nécessite l'utilisation de différentes méthodes et algorithmes afin d'analyser les caractéristiques, les propriétés et le contenu des données audio. Il est possible de réaliser une analyse audio sur diverses sources audio, telles que la musique, la parole, les sons environnementaux et d'autres signaux acoustiques.

### **2-1 Le rôle d'analyse audio**

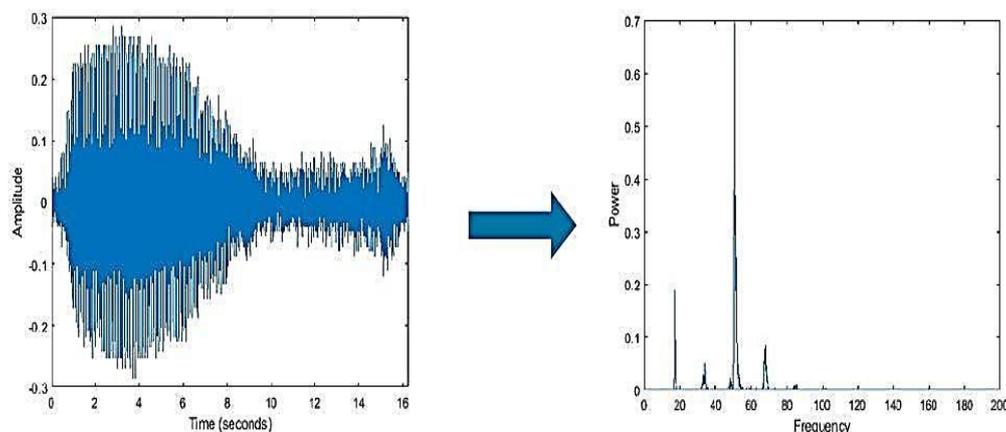
Le rôle de l'analyse audio est d'extraire d'un signal sonore l'information la plus pertinente possible afin de faire une séparation entre les sons. L'objectif de l'analyse audio peut varier selon l'application. Il peut s'agir de tâches telles que : le traitement du signal, l'extraction des fonctionnalités, la classification et la reconnaissance audio, l'analyse de la musique, l'analyse de la parole et l'analyse du contenu audio.

### **2-2 Principes de la transformation de données audio**

Nous devons expliquer certains concepts avant d'entrer dans les détails de la façon de gérer les fichiers audios. Ces phases apparaîtront dans presque chaque étape du processus, de l'agrégation de données robustes à l'obtention de prédictions d'apprentissage automatique. Il convient à noter que l'analyse audio implique un travail d'image plutôt qu'une écoute. Parce que l'utilisation d'images pour analyser et traiter les données audios pour différentes tâches d'apprentissage automatique est une méthode multivalente et efficace. L'analyse acoustique peut être effectuée sur des formes d'ondes sonores non traitées, des cartes spectrales ou d'autres représentations converties de signaux acoustiques [12] [13].

- **La transformation de Fourier (FT):** est une fonction mathématique qui casse un signal en pointes de différentes amplitudes et fréquences. Nous l'utilisons pour convertir des formes d'onde en tracés de spectre correspondants pour regarder le même signal d'un angle différent et effectuer une analyse de fréquence. C'est un instrument puissant pour comprendre les signaux et les erreurs de dépannage.
- **FFT (Fast Fourier Transform) :** FFT est un algorithme qui calcule la transformation de Fourier discrète (DFT) d'une séquence ou son inverse. Il est couramment utilisé pour transformer un signal de son domaine d'origine (par exemple, le domaine temporel) au domaine de fréquence, ce qui est utile pour analyser le contenu en fréquence du signal.

La transformation rapide de Fourier (FFT) est l'algorithme qui calcule la transformation de Fourier [14] [15] .



**Figure.1.3: La transformation rapide de Fourier (FFT) [14]**

- **MFCCs**

Mel-Frequency Cepstral Coefficients (MFCCs): MFCCs sont une caractéristique largement utilisée dans le traitement audio et la reconnaissance vocale. Ils représentent le spectre de puissance à court terme d'un son sur une échelle de fréquence mel non linéaire, qui se rapproche de la

réponse du système auditif humain plus étroitement que les échelles de fréquence linéaires [16].

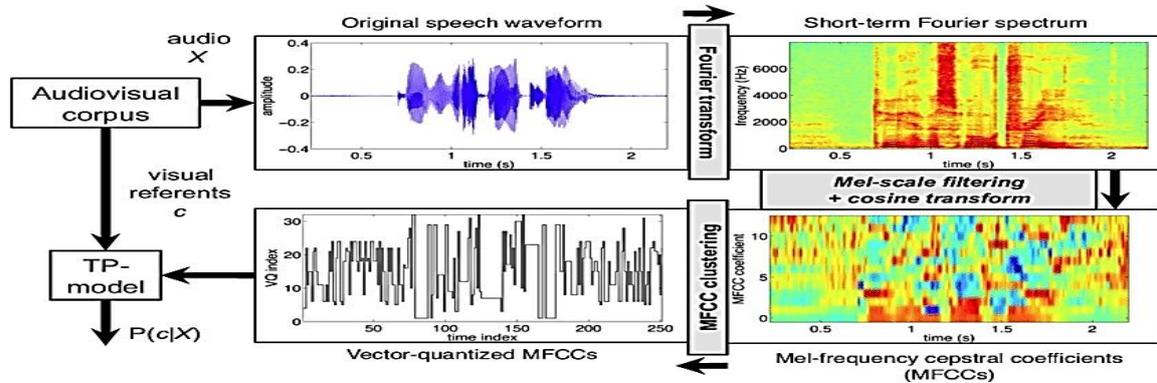


Figure.1.4: MFCC spectre [16]

- Une forme d'onde

Une forme d'onde (wavform) sont des représentations visuelles des signaux audio, reflétant les changements d'amplitude au fil du temps. Ils affichent le temps et l'amplitude, mais ne révèlent pas les fréquences [11].

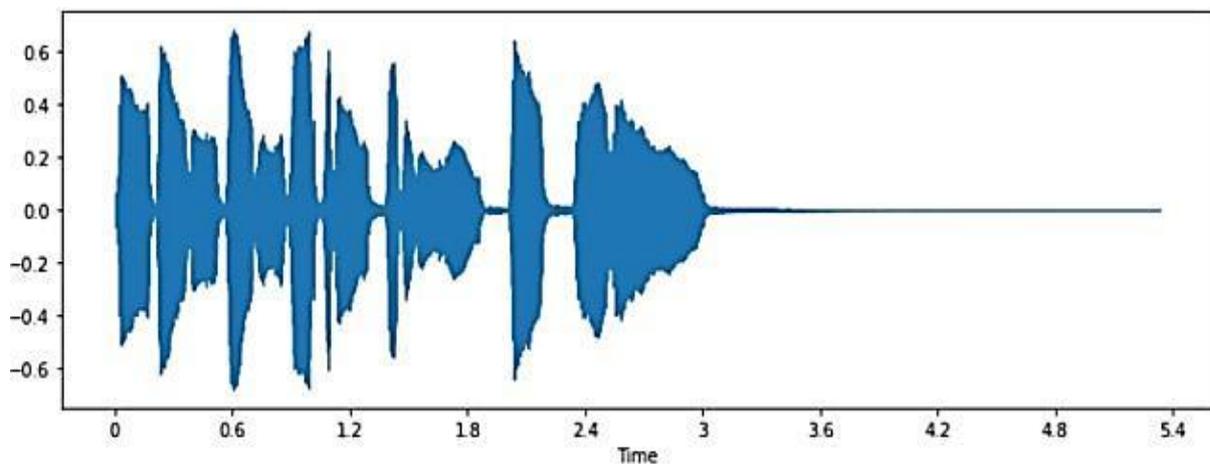


Figure.1.5: WAV forme

- STFT

La transformation de Fourier à court terme (STFT) est une séquence de Fourier qui transforme une forme d'onde en un spectrogramme. Utilisez la fonction stft pour calculer la transformée de Fourier à court terme (STFT) d'un signal dont les caractéristiques de spectre varient dans le temps [16][17]

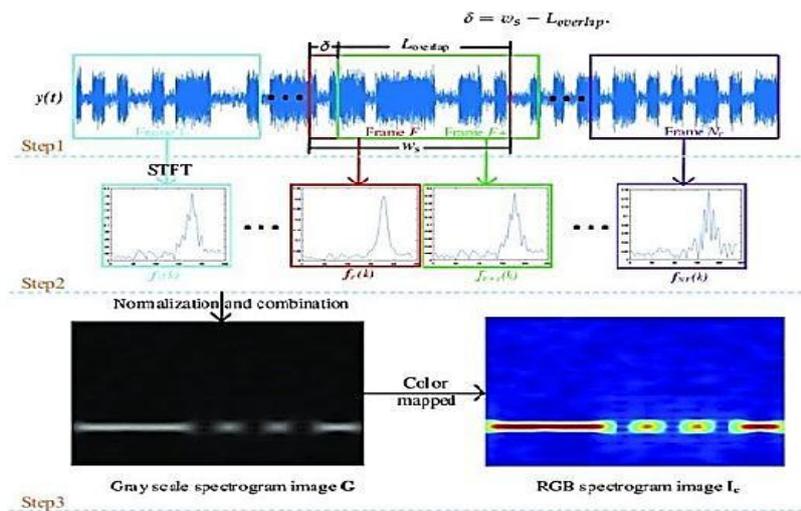


Figure.1.6: transformation STFT spectre [17]

## 2-3 Les filtres et l'équaliseur

La principale différence entre les filtres et les égaliseurs est que les filtres atténuent seulement (c.-à-d. réduisent) certaines fréquences du spectre, tandis que les égaliseurs peuvent soit augmenter, soit atténuer la force de bandes de fréquences particulières du spectre.

### 2-3-1 Les filtres

Les filtres audio sont utilisés pour ajuster le son et améliorer la qualité sonore d'un enregistrement ou d'une performance en direct. Ces filtres peuvent être des filtres analogiques ou des filtres audio numériques. Les filtres audio ont deux types :

- **Le filtre analogique:** est un type de filtre qui utilise des résistances, des condensateurs pour éliminer ou réduire certaines fréquences sonores. Il est moins précis que le filtre numérique, mais produit des sons riches, naturels et audacieux.
- **Le filtre audio numérique:** convertit un signal analogique en une version numérique applique des algorithmes pour modifier le son de façon plus précise et subtile.

### 2-3-2 Un égaliseur (EQ)

Un égaliseur (EQ) est un type de filtre qui corrige les pertes dans la transmission des signaux audio. c'est un outil utilisé pour modifier le son. Il comporte des filtres analogiques et des filtres audio numériques qui permettent de changer la fréquence du signal audio. Ces filtres audio modifient et affinent le son selon les paramètres choisis. La principale fonction d'un égaliseur est d'ajuster le ton d'un signal audio.

### 2-3-3 Les types d'égaliseurs

les types d'égaliseurs qui utilisent les filtres audio numériques:

- **L'égaliseur graphique (ou égaliseur à bande)** : est un outil utilisé pour atténuer les fréquences indésirables dans un signal audio. Il est composé de potentiomètres permettant d'augmenter ou de réduire des bandes de fréquence précises.
- **L'égaliseur paramétrique**: permet d'ajuster le niveau de chaque fréquence avec précision. Il dispose de trois paramètres principaux : la fréquence centrale, la largeur de bande (Q factor) et le gain. En ajustant ces paramètres, il est possible d'atténuer ou de renforcer des fréquences spécifiques dans le signal audio.



Figure.1.7: L'égaliseur paramétrique[18]

Les égaliseurs paramétriques disposent de plusieurs types de filtres ayant divers effets sur le signal audio :

- ✓ Les filtres passe-haut et passe-bas sont des filtres qui laissent passer les fréquences hautes et réduisent les fréquences basses en fonction de la fréquence de coupure (diagramme de Bode).
- ✓ Les filtres en plateau permettent d'augmenter ou de diminuer le volume au-dessus ou en dessous d'une certaine fréquence.
- ✓ Les filtres en cloche ou passe-bande augmentent ou réduisent le volume d'une fréquence ou d'une bande de fréquences spécifiques [18] [19] .

#### **2-3-4 Effet d'égalisation**

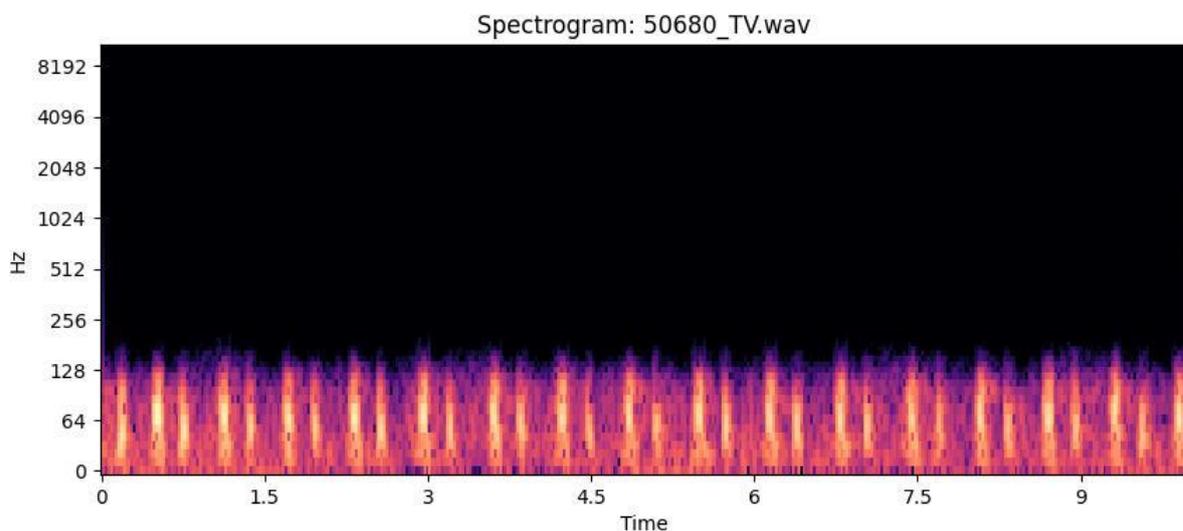
- ✓ L'effet Graphic Equalizer augmente ou coupe des bandes de fréquences spécifiques et fournit une représentation visuelle de la courbe d'égalisation résultante. Contrairement à l'égaliseur paramétrique, l'égaliseur graphique utilise des bandes de fréquence prédéfinies pour une égalisation rapide et facile.
- ✓ Les égaliseurs graphiques avec moins de bandes permettent un ajustement plus rapide; plus de bandes offrent une plus grande précision.
- ✓ Définit le niveau de précision pour l'égalisation. Des niveaux de précision plus élevés donnent une meilleure réponse en fréquence dans les plages inférieures.
- ✓ L'effet d'égalisation paramétrique fournit un contrôle maximal sur l'égalisation tonale. Contrairement à l'égaliseur graphique, qui fournit un nombre fixe de fréquences et de bandes passantes  $Q$ , l'égaliseur paramétrique vous donne un contrôle total sur les paramètres de fréquence,  $Q$  et de gain.

✓ L'égaliseur paramétrique utilise des filtres IIR (Infinite Impulse Response) de deuxième ordre, qui sont très rapides et fournissent une résolution de fréquence très précise. Par exemple, vous pouvez augmenter précisément une plage de 40 à 45 Hz. Les filtres FIR (Finite Impulse Response) comme l'égaliseur graphique fournissent cependant une précision de phase légèrement améliorée [20] [21] .

## 2-4 Le Spectrogramme

Un spectrogramme est un tracé qui montre la force du signal (ou l'intensité sonore) au fil du temps pour diverses fréquences présentes dans l'onde sonore.

Le temps, étant la variable indépendante, est tracé sur l'axe X. Différentes fréquences présentes dans la forme d'onde audio sont tracées sur l'axe Y. Une troisième dimension, représentée par des couleurs, est utilisée pour représenter la force (ou l'intensité) du signal pour une fréquence donnée à un moment donné.



**Figure.1.8: Spectrogramme**

le spectrogramme offrent une représentation puissante et multivalente des signaux audio, facilitant ainsi leur analyse, leur traitement et leur utilisation dans une variété de tâches d'apprentissage machine et de traitement du signal. On peut traiter les spectrogrammes comme des images, ce qui les

rend compatibles avec les architectures de réseaux neuronaux convolutionnels (CNNs), qui sont couramment utilisées pour analyser des images et des séquences temporelles. Tandis que, le traitement direct du signal audio permet d'obtenir une vision brute du signal dans le domaine temporel [22] [23] .

### 2-4-1 Les types de spectrogrammes

Il existe plusieurs types de spectrogrammes à tracer, parmi eux on cite:

- **Spectrogramme de puissance à fréquence linéaire**

Un spectrogramme de puissance à fréquence linéaire représente le temps sur l'axe des x, la fréquence en Hz sur une échelle linéaire sur l'axe des y et la puissance en dB.

- **Spectrogramme de puissance logarithmique**

Ce spectrogramme présente les mêmes informations à l'exception d'une échelle logarithmique sur l'axe y pour les fréquences.

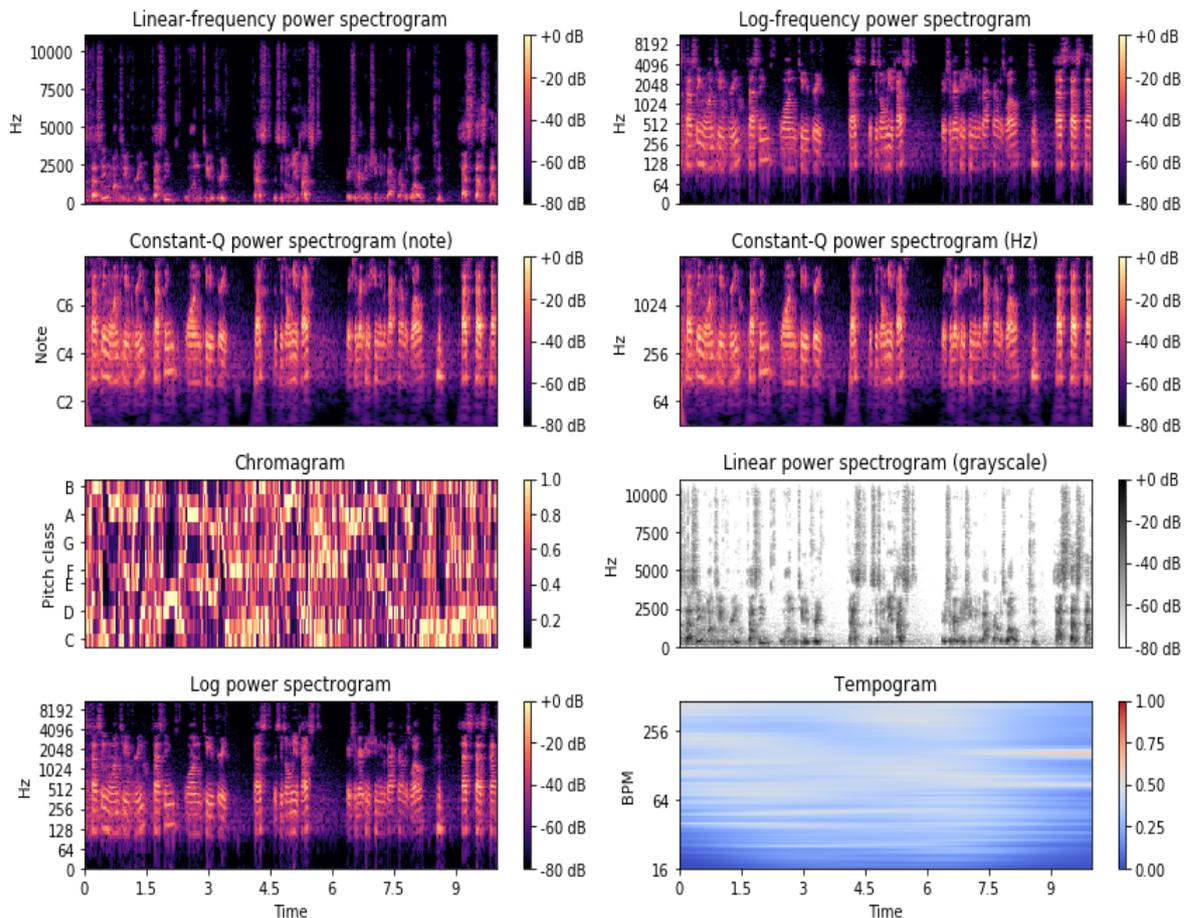
- **Spectrogramme de puissance Constant-Q**

Contrairement à la transformation de Fourier, mais similaire à l'échelle mel, la transformation constante-Q utilise un axe de fréquence logarithmique.

- **Chromagramme**

Chromagramme affiche l'intensité de chaque hauteur (pitch)

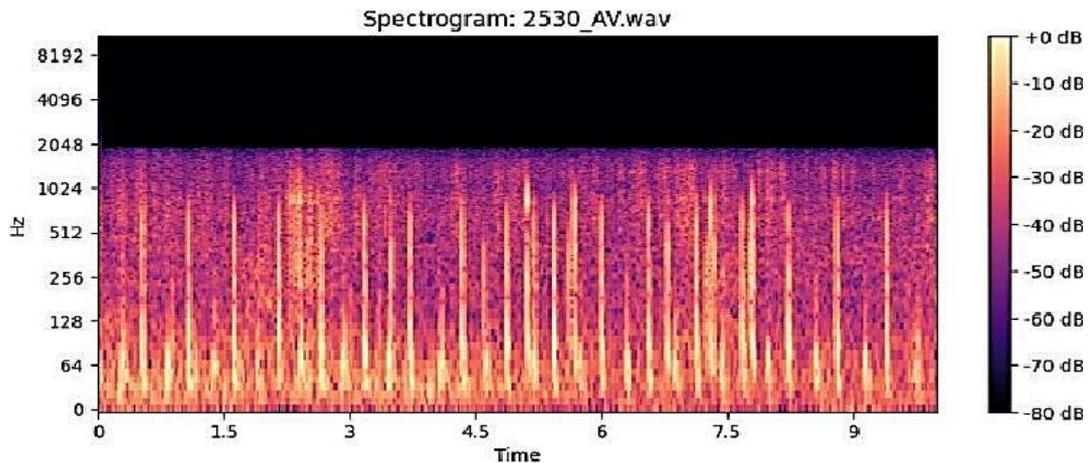
C,C#,D,D#,E,F,F#,G,G#,A,A#,B pour chaque intervalle de temps. Une propriété principale des caractéristiques chroma est qu'elles capturent les caractéristiques harmoniques et mélodiques de la musique, tout en étant robustes aux changements de timbre et d'instrumentation.



**Figure.1.9: Les types de spectrogrammes [23]**

•Un spectrogramme mel

Un spectrogramme mel est une variante du spectrogramme couramment utilisée dans les tâches de traitement de la parole et d'apprentissage automatique. Il est similaire à un spectrogramme en ce sens qu'il montre le contenu en fréquence d'un signal audio au fil du temps, mais sur un axe de fréquence différent. Dans un spectrogramme standard, l'axe de fréquence est linéaire et est mesuré en hertz (Hz). Pour créer un spectrogramme mel, le STFT est utilisé comme auparavant, divisant l'audio en segments courts pour obtenir une séquence de spectres de fréquence [23] [24].



**Figure.1.10: Spectrogramme de Mel**

### 3- Modèles de reconnaissance du son

Dans le domaine de la reconnaissance du son, deux grandes approches sont généralement utilisées :

- **Modèles basés sur des règles (modèles Classiques)**

Les modèles basés sur des règles sont conçus à partir d'une compréhension préalable des propriétés acoustiques des sons et des règles linguistiques. Ces modèles sont généralement élaborés par des experts du domaine qui définissent des règles spécifiques pour interpréter les sons et les signaux audio, comme l'utilisation de incluent les coefficients cepstraux en fréquence mél (MFCC), qui capturent les propriétés spectrales du son, et les modèles de Markov cachés (HMM), qui modélisent les séquences temporelles des caractéristiques acoustiques.

- **Modèles basés sur l'apprentissage automatique (modèle avancée) :**

Les modèles basés sur l'apprentissage automatique (ML) utilisent des algorithmes pour apprendre à partir des données, permettant ainsi la reconnaissance autonome des sons. Contrairement aux modèles basés sur des règles, ces modèles ne nécessitent pas une connaissance préalable exhaustive des propriétés acoustiques ou linguistiques. Au lieu de cela, ils apprennent à

partir d'exemples annotés pour développer des capacités de reconnaissance et d'interprétation des sons [25] [26].

### 3-1 Types de modèles de reconnaissance du son

Les modèles de reconnaissance du son sont des outils puissants qui utilisent diverses techniques pour interpréter et comprendre les sons. Voici un aperçu des différents types de modèles :

- **Modèles de reconnaissance de la parole** : Ces modèles sont spécialisés dans la compréhension de la parole humaine, transformant l'audio en texte transcrit. Ils sont essentiels pour les assistants vocaux et les systèmes de dictée.
- **Modèles de reconnaissance de musique** : Ils identifient les morceaux de musique, les genres et les artistes en analysant les caractéristiques audio. Ces modèles sont souvent utilisés dans les applications de reconnaissance musicale comme Shazam.
- **Modèles de reconnaissance des bruits** : Ces modèles détectent et classifient les sons environnementaux, tels que le trafic ou les alarmes incendie. Ils sont utiles pour la surveillance de la sécurité et l'analyse environnementale.
- **Modèles de reconnaissance d'événements acoustiques** : Ils identifient des événements spécifiques dans un enregistrement audio, comme des applaudissements ou des coups de feu, ce qui peut être crucial pour la surveillance et l'analyse d'événements [27].

### 3-2 Applications de la reconnaissance du son

La reconnaissance du son est un domaine technologique en pleine expansion avec de nombreuses applications pratiques. Voici une liste des applications dans divers domaines :

- **Interaction Homme-Machine** : comme systèmes de commande vocale utilisés dans les voitures, les maisons intelligentes, et les appareils électroménagers pour faciliter l'interaction sans utilisation des mains.
- **Sécurité et Surveillance** : détection des Sons d'Alarme utilisée pour détecter des alarmes incendie, des alertes de sécurité,
- **Divertissement** : utilisées des applications comme Shazam pour identifier des morceaux de musique en écoutant un extrait sonore.
- **Industrie** : cette technologie est employée pour améliorer la maintenance prédictive .
- **Santé** : la reconnaissance du son a un potentiel considérable, notamment pour l'analyse du sommeil par la reconnaissance des ronflements et du suivi de la respiration, ainsi que pour la détection des maladies respiratoires et le suivi de la condition physique à travers des applications mobiles de santé fournissant un suivi en temps réel et des conseils personnalisés.Elle a également trouvé des applications innovantes et cruciales dans le domaine de la santé cardiaque, améliorant ainsi les diagnostics et les soins pour les patients atteints de maladies cardiovasculaires en offrant des outils avancés pour le diagnostic, la surveillance continue, en utilisations dans: Analyse des Bruits Cardiaques ,Diagnostic des Maladies Cardiaques et Surveillance Cardiaque Continue [26][27] [28].

Dans ce chapitre, nous avons offert une brève introduction à la reconnaissance sonore, couvrant son évolution historique, les divers modèles et types de son, les approches globales et analytiques du traitement audio, ainsi que les méthodes de transformation des données audio. En outre, nous avons mentionné plusieurs travaux de reconnaissance sonore dans divers domaines, dont celui de la santé. Dans le prochain chapitre, nous approfondirons notre discussion en présentant en détail la santé cardiaque.

# **C**hapitre 02

## **Santé Cardiaque**

Les soins de santé constituent un ensemble complet de services et d'interventions visant à protéger la santé publique des personnes contre les maladies. Dans ce contexte, les soins de santé cardiaque sont très importants, car les maladies cardiaques sont l'une des principales causes de mortalité dans le monde.

La santé et les soins cardiaques bénéficient grandement de l'intelligence artificielle dans de nombreux aspects qui assurent la santé du cœur et du corps

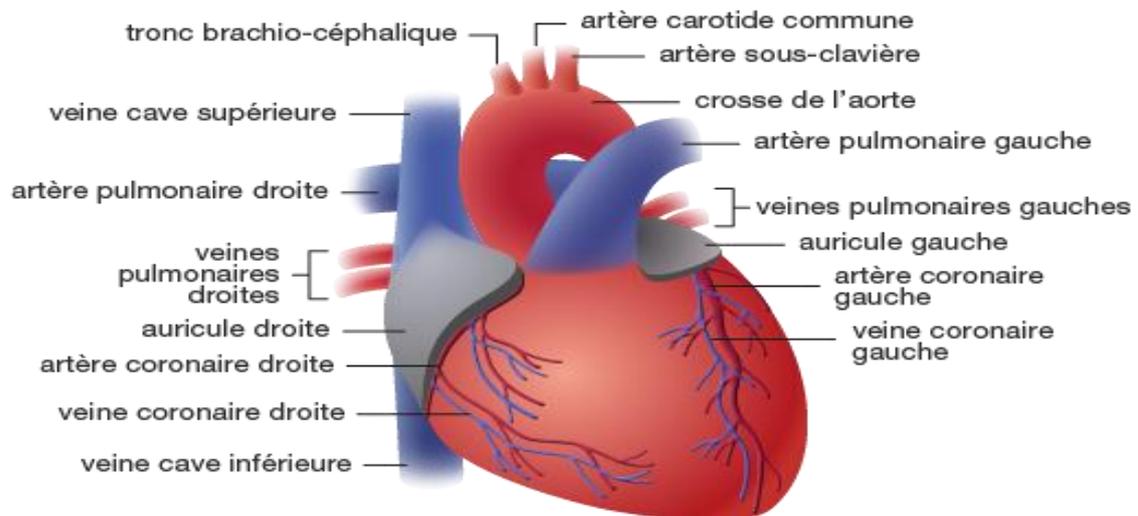
### **1-Le Cœur humain**

Le cœur est une masse musculaire creuse, essentiellement constituée de tissu cardiaque appelé myocarde, il est un organe musculaire situé dans la cage thoracique, derrière le sternum, et décalé légèrement sur la gauche chez la plupart des individus. Sa taille est environ 1,5 fois la taille du poing de la personne. Une paroi épaisse divise le cœur en deux parties, gauche et droite. Chaque partie comporte deux cavités, une oreillette et un ventricule, reliées par une valve.

Le cœur fonctionne comme une pompe qui, grâce à ses contractions régulières, propulse le sang dans tout l'organisme et assure ainsi l'alimentation en oxygène du corps entier. Chaque jour, le cœur pompe environ 8'000 litres de sang.

Quatre valves cardiaques, situées entre les oreillettes et les ventricules d'une part, et à la sortie des ventricules d'autre part, empêchent, lorsqu'elles sont fermées, le reflux du sang dans le mauvais sens. La fermeture des valves produit le son familier du battement du cœur.

Cet organe vital, joue un rôle crucial dans le corps humain en pompant le sang à travers tout l'organisme. Cette fonction permet de fournir de l'oxygène et des nutriments essentiels aux cellules, assurant ainsi le bon fonctionnement de tous les systèmes corporels [29].



**Figure.2.1: fonctionnement du cœur [29]**

## 1-1 L'importance du Cœur

L'importance du cœur dans le maintien de la vie est indéniable, car toute défaillance peut avoir des conséquences graves. Les maladies cardiaques, qui représentent une des principales causes de mortalité dans le monde, peuvent être causées par divers facteurs tels que l'hypertension, le diabète, le tabagisme, une alimentation déséquilibrée et le manque d'exercice physique. Parmi les types de maladies cardiaques, on retrouve les cardiopathies ischémiques, les insuffisances cardiaques, les arythmies et les maladies valvulaires. La prévention, le diagnostic précoce et les traitements adaptés sont essentiels pour gérer ces pathologies et améliorer la qualité de vie des patients [30].

## 1-2 Les maladies du cœur

Les maladies cardiovasculaire comprennent un large éventail de conditions qui affectant le cœur et les vaisseaux sanguins. Ceux-ci finissent par gêner, voire empêcher la circulation du sang qui alimente le cœur, le cerveau ou les jambes, provoquant des angines de poitrine, des infarctus, des AVC, et des artérites, parmi eux :

### 1-3 Types des maladies cardiaques

- **Insuffisance cardiaque** : L'insuffisance cardiaque se produit lorsque le cœur ne peut pas pomper suffisamment de sang pour répondre aux besoins du corps. Cela peut entraîner des symptômes tels que l'essoufflement, la fatigue et l'accumulation de liquide dans les tissus [31].
- **La malformations cardiaques**: Ce sont des malformations congénitales qui affectent la structure naturelle et fonctionnelle du cœur. Ces anomalies dans la valve, le contact entre les ventricules, les défauts de barrière peuvent causer ces anomalies différents symptômes de difficulté respiratoire et de fatigue. Ils sont diagnostiqués à l'examen clinique et par échographie cardiaque [32].
- **Infarctus du myocarde** : connu sous le nom de crise cardiaque, c'est une condition où une partie du muscle cardiaque est endommagée en raison d'un apport sanguin insuffisant. Cela se produit généralement en raison d'une obstruction des artères coronaires par un caillot sanguin [33].
- **Valvulopathies cardiaques** : Les valvulopathies cardiaques se réfèrent à des anomalies des valves cardiaques, qui régulent le flux sanguin à travers le cœur par exemple l'insuffisance mitrale, la valve mitrale ne se ferme pas complètement, entraînant un retour du sang vers l'oreillette gauche [34].
- **Endocardite** : l'endocardite est une infection de la muqueuse interne du cœur, appelée l'endocarde. Cette condition peut affecter les valves cardiaques, les parois des cavités cardiaques ou d'autres structures cardiaques. L'endocardite peut survenir lorsque des bactéries, des virus ou d'autres agents pathogènes pénètrent dans la circulation sanguine et colonisent la surface interne du cœur, généralement en raison d'une blessure ou d'une atteinte préexistante des valves cardiaques [35].

## 2- Outils pour le diagnostic de cardiologie

Dans le domaine médical l'audio contient des informations importantes pour les médecins, ces informations nous aident à diagnostiquer la maladie. L'enregistrement et l'analyse des sons du cœur, appelés auscultation cardiaque, sont des éléments fondamentaux de l'évaluation médicale de la santé cardiaque. Pour ce faire, les professionnels de la santé ont recours à une gamme d'outils sophistiqués conçus pour capturer et analyser les divers aspects de l'activité cardiaque tels que : Stéthoscope, électrocardiographe (ECG), échocardiographe, écouteurs médicaux, phonocardiographe, Holter cardiaque, appareils d'enregistrement audio et de lecture. Ces outils sont une partie essentielle de la pratique de la médecine moderne et aident les médecins et les professionnels de la santé à analyser les sons médicaux avec précision et efficacité, pour évaluer la santé cardiaque et diagnostiquer les problèmes éventuels, parmi eux :

### 2-1 Un stéthoscope

Le stéthoscope est un dispositif essentiel dans le domaine médical, permettant la captation, l'amplification et la transmission de sons physiologiques. Une fois plaquée sur le corps du patient, la membrane vibre, transmettant ainsi les ondes acoustiques au pavillon.



**Figure.2.2: Le stéthoscope [36]**

Le pavillon est la pièce maîtresse du stéthoscope, assure l'amplification et la transmission des ondes acoustiques vers la tubulure (le tube en plastique souple) puis vers les lyres, pour finir aux oreilles du professionnel de santé, en outre part, le stéthoscope se divise en trois grandes familles, chacune possédant des

caractéristiques bien précises et adaptées aux besoins des professionnels de santé : les stéthoscopes « infirmiers » ; les stéthoscopes « généralistes »; les stéthoscopes « cardiologiques ». Parmi eux les stéthoscopes « cardiologiques » s'adressent aux professionnels les plus exigeants en matière d'acoustique médical, car ils proposent une sensibilité acoustique exceptionnelle pour un diagnostic sûr et précis [36].

## 2-2 L'ECG (Électrocardiogramme)

L'ECG (Électrocardiogramme) est un examen médical non invasif qui permet d'enregistrer l'activité électrique du cœur. Il s'agit d'un outil de diagnostic essentiel en cardiologie. Il permet de détecter les anomalies du rythme cardiaque (arythmies), et il aide à identifier les problèmes dans la conduction électrique du cœur. Il y a deux type de ECG .

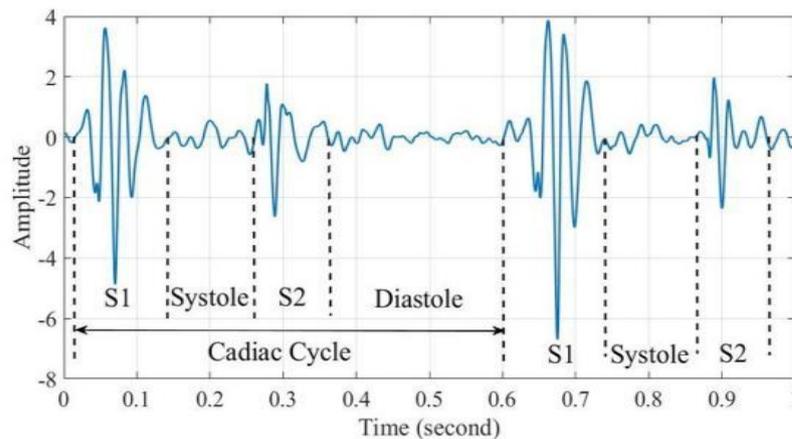


Figure.2.3: L'ECG [37]

- **Échocardiographie transthoracique (TTE)** : Utilise des ondes échographiques pour créer des images du cœur. Il peut capturer les sons cardiaques dans le cadre de la composante échographique Doppler, qui mesure le flux sanguin à travers le cœur et les valves.
- **Échocardiographie transoesophagienne (ETO)** : Consiste à insérer une sonde avec un transducteur échographique dans l'œsophage pour obtenir des images plus proches du cœur [38].

### 2-3 Phonocardiography

La phonocardiographie est l'enregistrement graphique des bruits et des souffles cardiaques. Elle joue un rôle important pour comprendre et interpréter les sons cardiaques, la phonocardiographie permet de surmonter les limitations de l'audition humaine et d'obtenir une représentation graphique fidèle de l'ensemble des sons cardiaques, bruits et souffles, pour mieux les analyser et les interpréter [39].



**Figure.2.4: Phonocardiography signal [40].**

### 2-4 Logiciel d'analyse du son cardiaque

Analyse du son cardiaque informatisée : Logiciel qui analyse les sons cardiaques enregistrés pour détecter les anomalies. Ce logiciel peut aider à diagnostiquer des conditions telles que des troubles valvulaires, des anomalies septales et d'autres problèmes cardiaques structurels [41].

### 2-5 Difficultés des outils de diagnostic cardiaque

Que le médecin utilise un cardiogramme ou des "électrodes" pour enregistrer l'activité du cœur électrique et la convertir en un graphique (électrocardiogramme) à travers lequel les médecins peuvent évaluer la santé du cœur. Mais il ne révèle pas toutes les informations, parce que ces outils ont des négatifs et montrent des contraintes physiques et techniques spécifiques, et au fil du temps, ils deviennent obsolètes, affectant ainsi la qualité sonore et le médecin ne peut pas diagnostiquer avec précision la maladie. De plus, le stéthoscope est

l'une des méthodes les plus faciles et les plus conviviales qui dépend entièrement du médecin et peut affecter son audition, et ces outils peuvent également être coûteux pour certaines personnes et organisations de santé. Tous ces facteurs influencent les décisions d'utiliser ces outils et les interventions médicales appropriées.

### **3-L'IA dans les maladies cardiologie**

L'intelligence artificielle (IA) révolutionne le domaine de la cardiologie en offrant des outils avancés pour le diagnostic, le traitement et la gestion des maladies cardiovasculaires. Grâce à l'analyse de grandes quantités de données médicales, l'IA peut identifier des modèles et des tendances qui échappent souvent aux méthodes traditionnelles, permettant ainsi des diagnostics plus précis et précoces. Les algorithmes d'apprentissage automatique sont utilisés pour interpréter les images médicales, comme les échocardiogrammes et les IRM cardiaques, aidant les cardiologues à détecter des anomalies subtiles. En outre, les systèmes d'IA peuvent surveiller en temps réel les signes vitaux des patients, prédire les événements cardiaques imminents et personnaliser les plans de traitement en fonction des besoins individuels. Ces avancées technologiques promettent d'améliorer considérablement les résultats pour les patients tout en optimisant l'efficacité des soins en cardiologie. [30] .

#### **3-1 Jeux de données**

Les jeux de données en cardiologie font référence à des ensembles de données cliniques et/ou de signaux cardiaques collectés à partir de patients atteints de différentes pathologies cardiaques.

Ces bases de données ont été créées dans le but de faciliter la recherche en cardiologie, en fournissant aux chercheurs des données précieuses pour la compréhension des maladies cardiaques et la création de nouvelles approches

pour le diagnostic et le traitement. Plusieurs jeux de données en cardiologie sont disponibles publiquement, nous citons :

### 3-1-1 MIT-BIH

La base de données d'arythmie MIT-BIH comprend différentes arythmies les signaux qui sont annotés indépendamment par deux ou plusieurs cardiologues selon leur type d'arythmie. Chaque enregistrement comprend des signaux ECG à deux canaux qui sont la dérivation modifiée du membre II (MLII) et l'une des dérivations modifiées V1, V2 ou V5. En raison de la déformation du deuxième canal. La durée de chaque signal est de 30 minutes avec une fréquence d'échantillonnage de 360 Hz et est filtrée par un filtre passe-bande de 0,1 à 100 Hz. La base de données MITBIH est bien connue pour être déséquilibrée par le nombre non égal de battements ECG pour chaque arythmie, ce qui détériore la précision des modèles DNN et CNN [42].

### 3-1-2 PASCAL

L'ensemble de données du concours PASCAL challenge se compose de deux parties (A & B) qui ont été générés à partir de deux sources différentes. Dataset-A a été obtenu auprès du grand public via l'application iSethoscope Pro- iPhone, et Dataset-B a été généré à partir d'un examen clinique dans les hôpitaux utilisant DigiScope. L'ensemble de données-A a été divisé en cinq catégories, c.-à-d. normal, murmure, extrasystole, artefacts et non étiqueté, et ayant un total de 176 échantillons de toutes les catégories de sons cardiaques en format wav. La catégorie non étiquetée a été exclue dans cette étude. Dans l'ensemble de données A, chaque échantillon comprend un taux de fréquence d'échantillonnage de 4000 i-e. Le dossier normal contient 44100 (4,97 s) et le dossier murmure comprend un taux de fréquence d'échantillonnage total de 110250 (4,96 s) [43].

### 3-1-3 Shiraz

La base de données sur les sons du cœur fœtal de l'Université de Shiraz (SU) (SUFHSDB) contient des enregistrements de phonocardiogramme fœtal et maternel (PCG) de 109 femmes enceintes en grossesses simple et jumelle. Les enregistrements ont été réalisés à l'hôpital Hafez de l'Université des sciences médicales de Shiraz, à Shiraz, en Iran, auprès de mères âgées de 16 à 47 ans avec un stéthoscope numérique placé sur le bas-ventre maternel. Au total, 99 sujets ont enregistré un signal, trois ont enregistré deux cas de jumeaux et sept ont été enregistrés individuellement, ce qui donne 119 enregistrements. La durée moyenne de chaque enregistrement est d'environ 90 secondes. Le taux d'échantillonnage était généralement de 16000 Hz avec une quantification 16 bits et quelques enregistrements à 44100 Hz. Les données ont été enregistrées en mode large bande du stéthoscope numérique, avec une réponse en fréquence de 20 Hz à 1 kHz [44].

### 3-1-4 CirCor 2022

La base de données CirCor DigiScope Phonocardiogram contient un ensemble de données disponibles sur [physionet.org](http://physionet.org), comprenant 5272 enregistrements de battements de cœur provenant de 1568 patients. Les enregistrements sont divisés en deux catégories : normaux/anormaux et présence/absence de murmures. La durée des enregistrements varie de 4,8 à 80,4 secondes. Les données comprennent des informations détaillées sur les patients telles que l'identification, la taille, le poids, l'âge, le sexe, ainsi que des caractéristiques importantes. Les enregistrements sont issus de quatre régions d'auscultation centrales : la valve pulmonaire (PV), la valve tricuspide (TV), la valve aortique (AV) et la valve mitrale (MV), annotées respectivement comme 1, 2, 3 et 4 pour les périodes S1, systolique, S2 et diastolique [56].

### 3-2 Travaux connexes

Plusieurs travaux de la reconnaissance du son dans le domaine de la santé cardiaque, spécifiquement les rythmes cardiaques, vont être abordés dans ce qui suit.

D'abord, les auteurs de ce travail présente ResHNet, un cadre d'apprentissage profond conçu pour la classification des sons cardiaques normaux et anormaux. Il utilise des spectrogrammes générés à partir de signaux sonores cardiaques 1D et forme un classificateur de réseau résiduel sur ces spectrogrammes. À l'aide de l'ensemble de données PhysioNet/CinC 2016 Challenge, ResHNet, une architecture réseau résiduelle profonde, est formée pour apprendre les fonctionnalités des images de spectrogramme. L'approche atteint des résultats impressionnants, avec une précision d'environ 97 % sur un ensemble de tests indépendants [45].

Par ailleurs, l'article a introduit une approche innovante pour la classification binaire des sons cardiaques, en distinguant les états 'normaux' et 'anormaux'. Il répond au besoin pressant de détection précoce des maladies cardiovasculaires. Le modèle proposé intègre des caractéristiques de bispectrum pour capturer les attributs non-linéaires complexes des sons cardiaques, associés à l'architecture de pointe Vision Transformer (ViT) pour les tâches de classification. Les résultats montrent la performance robuste du modèle, présentant une grande précision même dans les sons cardiaques discernant des patientes enceintes, dépassant les prouesses de classification des cardiologues expérimentés [28].

De plus, dans l'étude intitulée "Arythmic Heartbeat Classification Using 2D Convolutional Neural Networks", une méthode pour détecter les arythmies dans les signaux ECG est introduite. Cette approche utilise des images de battements ECG 2D comme entrée pour l'entraînement d'une nouvelle architecture de réseau de neurones convolutifs (CNN) spécialement conçue pour la classification de l'arythmie. La recherche démontre une efficacité

remarquable, atteignant une précision globale exceptionnelle de 99,7%, avec des taux de sensibilité et de spécificité impressionnants de 99,7% et 99,22%, respectivement. La méthodologie consistait à utiliser les données ECG de la base de données MIT-BIH, englobant cinq classes d'arythmie distinctes. Ces résultats soulignent le potentiel des modèles d'apprentissage profond formés sur les représentations ECG 2D comme un outil fiable et robuste pour la détection automatisée de l'arythmie cardiaque [46].

En outre, cette étude est proposée pour diagnostiquer efficacement les maladies cardiaques. Le modèle présenté utilise un réseau de neurones récurrent (RNN) avec des unités de mémoire à long terme (LSTM) pour classer huit types différents de signaux audios de battement de cœur. Il a été développé à l'aide du défi PASCAL et des ensembles de données du défi PhysioNet 2017. Le modèle RNN-LSTM proposé a atteint des précisions de classification de 99,71% sur l'ensemble de données PASCAL et 98,70% sur l'ensemble de données PhysioNet, dépassant plusieurs modèles d'apprentissage automatique de base [47]

D'autre part, dans cet article, un système de diagnostic assisté par ordinateur (CAO) est proposé, utilisant des réseaux de neurones convolutifs (CNN) pour classer les sons cardiaques dans des catégories normales ou anormales en fonction de leurs spectrogrammes. La méthodologie consistait à utiliser un ensemble de données comprenant des fichiers audio cardiaques normaux et anormaux au format WAV. Les résultats indiquent que la méthode basée sur CNN a atteint une précision de classification de 82,75% en distinguant les sons cardiaques normaux et anormaux. Cela suggère le potentiel des systèmes de CAO basés sur CNN pour aider les professionnels de la santé dans le diagnostic des maladies cardiaques basé sur l'analyse du son cardiaque [48].

L'article se concentre sur l'utilisation de l'apprentissage par transfert pour détecter les murmures cardiaques dans les enregistrements de phonocardiogramme (PCG). Il évalue la performance de trois modèles

d'apprentissage par transfert - VGG16, VGG19 et ResNet50 - dans la classification des sons cardiaques normaux et anormaux à partir de la base de données PASCAL CHSC. Parmi ces modèles, le pipeline d'apprentissage par transfert Spectrogram-ResNet50 a atteint la meilleure précision de classification de 87,65 %. Les résultats suggèrent que l'approche Spectrogram-ResNet50 peut faciliter la détection efficace des murmures cardiaques avec un temps d'entraînement réduit par rapport à l'entraînement d'un modèle à partir de zéro. Cette étude s'appuie sur des travaux antérieurs qui ont utilisé des techniques d'apprentissage automatique et de traitement du signal telles que les SVM, les réseaux de neurones et kNN pour la détection des murmures cardiaques à partir de données PCG. La contribution principale de cet article est d'évaluer l'efficacité de l'apprentissage par transfert, en utilisant des architectures CNN populaires, pour la détection automatisée des murmures cardiaques à partir d'enregistrements PCG [49].

Enfin, ce travail a présenté une nouvelle approche hybride pour la classification des battements cardiaques qui combine Transformer et les réseaux de neurones convolutifs multi-branches (CNN) pour surmonter les limitations observées dans les méthodes actuelles de classification ECG basées sur l'apprentissage profond. Cette méthode hybride aborde le défi de considérer simultanément les caractéristiques morphologiques et temporelles dans les signaux ECG. En utilisant une architecture CNN multi-branches, la méthode extrait diverses fonctionnalités du signal ECG, tandis qu'un module Transformer est utilisé pour capturer les dépendances temporelles. L'efficacité de la méthode proposée a été évaluée sur la base de données d'arythmie MIT-BIH à l'aide de trois protocoles de classification des battements cardiaques différents, y compris les protocoles intra-patient et inter-patient. Les résultats ont montré des performances remarquables, avec une précision globale de 99,5 % obtenue pour le protocole intra-patient, ainsi que des taux de sensibilité et de spécificité élevés pour différentes classes de battements cardiaques. De plus, pour les protocoles

inter-patients, la méthode a atteint des précisions globales de 98,8% et 97,2% respectivement. Les auteurs soulignent l'importance de cette approche hybride pour les applications cliniques dans la détection et la classification des arythmies basées sur l'ECG [50].

Les technologies médicales numériques permettent aux médecins d'accéder à l'information médicale à distance. Ils comprennent les systèmes d'information médicale, les systèmes d'imagerie médicale, les systèmes de télémédecine et les systèmes d'analyse de données. Notre approche est de fournir à la santé des outils d'information qui traitent le son, et il y a de nombreux sujets connexes comme nous l'avons vu à ce chapitre.

Puisque l'audio contient des informations importantes pour les médecins, ces informations nous aident à diagnostiquer la maladie.

# **C**hapitre 03

**Approche proposée**

Puisque l'audio peut révéler plusieurs informations importantes, nous allons essayer d'utiliser plusieurs techniques de traitement de son numériques afin de construire un système intelligent pour la classification des sons numérique en se basant sur leurs représentations spectrographiques.

L'approche proposée se focalise sur l'utilisation des spectrogrammes comme représentation des sons cardiaques, et l'application des modèles CNN pour effectuer la classification automatique des sons de battement de. Le système à développer est destiné pour l'aide au diagnostic des maladies cardiovasculaires à partir des sons cardiaques. Les détails d'implémentation et d'expérimentation de cette solution s'en suivra dans ce qui suit dans ce chapitre.

## 1-Outils et bibliothèques

- **Python** : Python est un langage de programmation destiné pour la recherche scientifique. Il nous a permis de travailler plus rapidement et d'intégrer nos systèmes plus efficacement.
- **Colab**: Google Colab ou Colaboratory est un service cloud, offert par Google (gratuit), destiné à la formation et à la recherche dans l'apprentissage automatique. Cette plateforme permet d'entraîner des modèles de Machine Learning et Deep Learning directement dans le cloud. La plateforme permet de manipuler un document exécutable, écrire, écouter et de partager le code dans google drive avec les collaborateurs [51].
- **Librosa**: est une puissante bibliothèque Python qui offre un large éventail d'outils et de fonctionnalités pour gérer les fichiers audio [52].
- **Tensorflow**: est une bibliothèque logicielle open source pour le calcul numérique de haute performance. Son architecture flexible permet un déploiement facile du calcul sur diverses plates-formes (CPUs, GPUs, TPUs), et des ordinateurs de bureau aux clusters de serveurs, aux périphériques

mobiles. Initialement développé par des chercheurs et des ingénieurs de l'équipe de Google Brain au sein de l'organisation de l'IA de Google, il s'appuie sur l'apprentissage automatique et l'apprentissage en profondeur. [53].

- **Keras**: est une API de réseaux neuronaux de haut niveau, écrite en Python et capable de s'exécuter sur TensorFlow, CNTK ou Theano. Il a été développé dans le but de permettre une expérimentation rapide. Être en mesure de passer de l'idée au résultat le plus rapidement possible, est la clé pour faire de la recherche :

- ✓ Permet un prototypage facile et rapide (grâce à la convivialité, à la modularité et à l'extensibilité).

- ✓ Prend en charge les réseaux convolutionnels et les réseaux récurrents ainsi que les combinaisons des deux.

- ✓ Fonctionne de manière transparente sur le processeur et le processeur graphique [54].

- **Heartpy**: HeartPy est une bibliothèque Python conçue pour l'analyse et le traitement des données de fréquence cardiaque. Il simplifie la manipulation des données physiologiques de divers dispositifs et capteurs, offrant une suite complète d'outils pour le prétraitement, l'analyse et la visualisation.

HeartPy, le Python Heart Rate Analysis Toolkit est un module pour l'analyse de la fréquence cardiaque en Python. Il a commencé comme implémentation pure-python pour analyser les données physiologiques prises dans des expériences de conduite et de cyclisme naturalistes [55]

Le module prend un signal de fréquence cardiaque discret et produit des mesures de domaine temporel et de domaine de fréquence souvent trouvées dans la littérature scientifique :

Domaine temporel :

- ✓ battements par minute, BPM
- ✓ intervalle interbeat, IBI
- ✓ écart type si intervalles entre battements adjacents, SDNN
- ✓ écart type des différences successives entre les intervalles R-R adjacents, SDSD
- ✓ quadrature moyenne des différences successives entre les intervalles adjacents et R-R, RMSSD
- ✓ proportion des différences entre les intervalles R-R supérieurs à 20 ms, 50 ms, pNN20, pNN50
- ✓ écart absolu médian, MAD
- ✓ Analyse de poincare (SD1, SD2, S, SD1/SD2)
- ✓ Tracé de poinçon

Domaine fréquentiel :

- ✓ composant très basse fréquence (0,0033â€“0,04â€ Hz), VLF
- ✓ composant basse fréquence (0,04â€“0,15â€ Hz), LF
- ✓ composant haute fréquence (0,15â€“0,4â€ Hz), HF
- ✓ Rapport lf/hf, LF/HF

**1-1 Distribution des Classes du jeux de données**

En cardiologie, les jeux de données désignent des ensembles de données cliniques et/ou de signaux cardiaques recueillis à partir de patients souffrant de diverses affections cardiaques, nous avons choisi les jeux de données CirCor DigiScope Phonocardiogram et Shiraz University Fetal Heart Sounds Database pour valider notre approche proposée.

**1-2 Distribution des Classes du jeu de données Shiraz**

Le jeu de données de l'Université de Shiraz sur les sons cardiaques fœtaux comprend des enregistrements de phonocardiogrammes de 109

femmes enceintes en grossesses simples et jumelles. Les enregistrements ont été réalisés à l'hôpital Hafez en Iran, avec un stothoscope numérique placé sur le cœur de la mère. La durée moyenne de chaque enregistrement est d'environ 90 secondes [44]. Le tableau ci-dessous illustre la distribution de deux classes.

Classes	Fetal	Maternal
Nombre	120	92
Pourcentage	57%	43%
Total	212	

**Table.3.1. Distribution des Classes du jeu de données Shiraz**

Selon une analyse de la distribution des classes « *Fetal / Maternal* » du jeu de données ont les taux de 57% pour la classe « Fetal » et 43% pour la classe « Maternal », les taux sont presque équilibrés pour les deux classes si pour ça on a préféré utilise la métrique d'exactitude (*accuracy*) dans notre étude de classification.

Les fournisseurs de ce jeu de données n'ont pas spécifié une division préalable, par conséquent on utilisera un scenario de division 70/30 % car le jeu de données Shiraz est petit ; ainsi il faut introduire une bonne partie du jeu de données au classifieur lors de la phase d'entraînement.

### 1-3 Distribution des Classes du jeu de données Circor22

Le CirCor DigiScope Phonocardiogram Dataset est un ensemble de données qui contient des enregistrements de phonocardiogramme (PCG) recueillis à l'aide du CirCor DigiScope, un dispositif utilisé pour enregistrer des sons cardiaques. Ces sons comprennent les sons cardiaques normaux ainsi que tous les sons anormaux (murmures, clics, frottements) qui peuvent

indiquer des problèmes cardiaques.

Ces ensembles de données sont utiles pour la formation et l'évaluation de modèles et d'algorithmes d'apprentissage automatique visant à aider les professionnels de la santé à diagnostiquer les maladies cardiaques à partir d'enregistrements sonores du cœur.

Le jeu de donnée comporte 5272 enregistrements sonores cardiaques provenant des principaux emplacements de stéthoscopie obtenus auprès de 1568 patients dans l'ensemble de données, âgés de 0 à 21 ans (MST moyenne = 6,1 4,3 ans), d'une durée comprise entre 4,8 et 80,4 secondes (MST moyenne = 22,9 7,4 s), totalisant plus de 33,5 heures d'enregistrement. L'ensemble de données téléchargé contient 3163 enregistrements de 942 patients, y compris des enregistrements sonores cardiaques obtenus à partir de quatre emplacements de stéthoscopie, le point de valvule pulmonaire (PV), le point de valvule tricuspide (TV), le point de valvule aortique (AV) et le point de valvule mitrale (MV); les positions de segmentation du son cardiaque étiquetées manuellement; et les descriptions de sujets correspondantes, comme le poids, l'âge et la taille.

La description des données du phonocardiogramme CirCor DigiScope peut inclure des paramètres tels que :

- Un fichier d'enregistrement d'ondes (format binaire .wav), qui contient les données sonores du cœur
- Un fichier d'en-tête (format texte .hea) décrivant le fichier .wav au format standard WFDB
- Un fichier de données de segmentation (format texte .tsv), qui contient des informations de segmentation concernant les points de début et de fin des sons cardiaques fondamentaux
- Un fichier texte de description du sujet (format texte .txt) par sujet, dont le nom correspond à l'ID du sujet. Des données démographiques telles que le

poids, la taille, le sexe, le groupe d'âge et une description détaillée des événements de murmure sont fournies dans ce fichier [56]. Le tableau ci-dessous illustre la distribution de deux classes.

Classe	Normal	Abnormal
Number	1580	1468
Persontage	52%	48%
Total	3048	

**Table.3.2. Distribution des Classes du jeu de données Circor22**

## 2- Méthodologie

Notre méthodologie propose une approche qui commence par utiliser des fichiers audios au format WAV de différentes longueurs. Pour normaliser la durée, nous limitons chaque audio aux 10 premières secondes afin d'obtenir des audios de même durée. Ensuite, nous convertissons les sons cardiaques en images spectrales en utilisant la transformation STFT et la bibliothèque Librosa de Python. Après cela, nous découpons les images pour se concentrer sur les zones d'intérêt, car les images de tailles différentes ne peuvent pas être utilisées directement dans un algorithme d'apprentissage automatique. Par conséquent, les nouveaux jeux de données générés sont prêts à être exploités par une classification basée sur le transfert d'apprentissage avec des modèles tels que ResNet50 et VGG19. Nous avons deux classes pour les deux jeux de données : Circor22 (normal/anormal) et pour Shiraz (fetal/maternel).

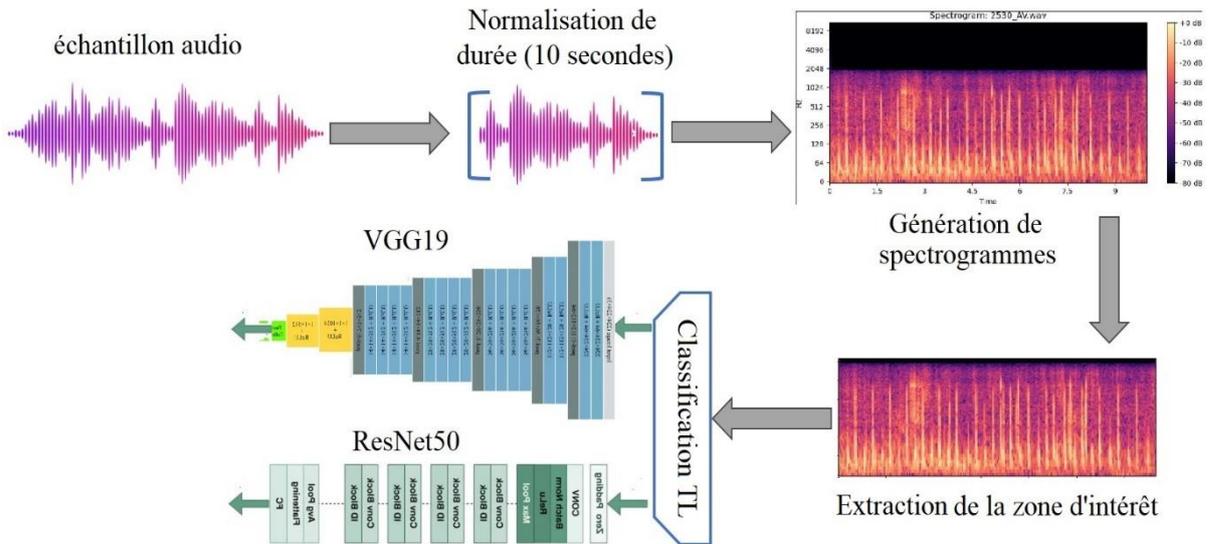


Figure.3.1. Méthodologie proposée

## 2-1 Prétraitement des données audios

### Étape 01:

- Chargement des fichiers audio : utiliser la bibliothèque Librosa pour charger les fichiers audios.

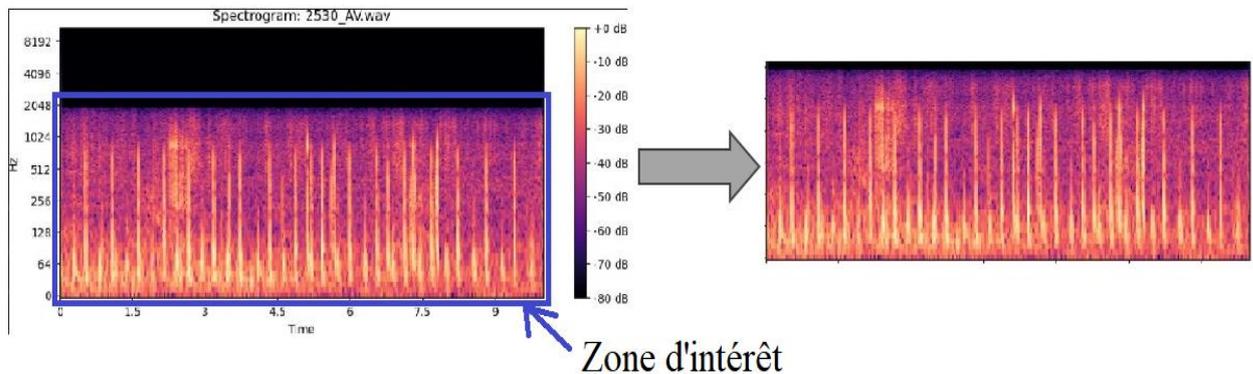
- **Normalisation de la durée des échantillons:** pour obtenir des audios de même durée, nous limitons la durée de chaque échantillon audio dans les 10 premières secondes. Le choix des 10 secondes revient à des directives expérimentales connues dans la littérature. Il est important de noter qu'une petite période moins de 10 secondes ne peut contenir un nombre de battements significatifs permettant de bien détecter la rythmie du cœur. Par conséquent, cette normalisation des durée garanti une normalisation dans la taille des images de spectrogrammes et les coordonnées de la zone d'intérêt.

- **Convertir l'échantillon audio en spectrogramme:** on convertit les échantillons audio contenant des sons de battements cardiaques en images spectrales par l'extraction de leurs propriétés spectrales en utilisant la transformation de Fourier à courte portée (STFT) de la bibliothèque Librosa.

**Étape 02:**

**-Extraction des zones d'intérêt des spectrogrammes:** les images de hauteur et de largeur différentes ne sont pas un élément valide à empiler dans un tableau ou à entrer dans un algorithme d'apprentissage automatique. Pour cela, nous devons limiter le spectrogramme pour obtenir tous les spectrogrammes la même durée et après redimensionner les images (recadrer) pour obtenir les zones d'intérêts.

Ainsi, la zone d'intérêt a été extraite en spécifiant les dimensions en hauteur et en largeur (voir les figures 3.2).



**Figure.3.2 Extraction de la zone d'intérêt.**

## 2-2 Classification basées Transfert Learning

La classification à travers des images comme on dit avant la transformation des données audio implique de travailler avec des images plutôt que d'écouter en raison de la nature même des informations audios. Ces informations audios peuvent être complexes et difficiles à analyser directement à l'oreille humaine, en particulier lorsqu'il s'agit de traiter de grandes quantités de données.

En transformant les données audios en images de spectrogramme, on peut bénéficier des techniques de traitement d'image qui permettent de visualiser et d'analyser les données de manière plus efficace.

Dans ce contexte notre classification basée sur le Transfert d'apprentissage est une méthode de classification qui emploie des modèles pré-entraînés sur de grandes quantités de données. Le transfert d'apprentissage permet d'exploiter les connaissances acquises précédemment grâce à des modèles avancés tels que ResNet50 ou VGG19 [57].

### **2-2-1 ResNet50**

ResNet50 est un réseau résiduel à 50 couches, qui est un type d'architecture de réseau neuronal convolutif (CNN). Il a été développé par des chercheurs de Microsoft en 2015 et fait partie de la famille de modèles ResNet (Residual Network). Il est largement utilisé pour la classification d'images, la détection d'objets et d'autres tâches de vision par ordinateur en raison de ses excellentes performances [57].

### **2-2-2 VGG19**

VGG19 est un modèle de réseau neuronal convolutif (CNN) à 19 couches développé par des chercheurs de l'Université d'Oxford en 2014. Il s'agit d'une extension du modèle VGG16 précédent, avec 19 couches au lieu de 16.

Les modèles VGG sont connus pour leur architecture relativement simple et uniforme, utilisant uniquement 3x3 filtres de convolution et 2x2 pooling maximum, VGG19 a un total de 19 couches de poids - 16 couches convolutives, 5 couches de pooling maximum et 3 couches entièrement connectées et il a été l'un des premiers modèles d'apprentissage profond très réussis pour la classification d'images, obtenant d'excellentes performances sur l'ensemble de données ImageNet [57].

## 2-3 Résultats

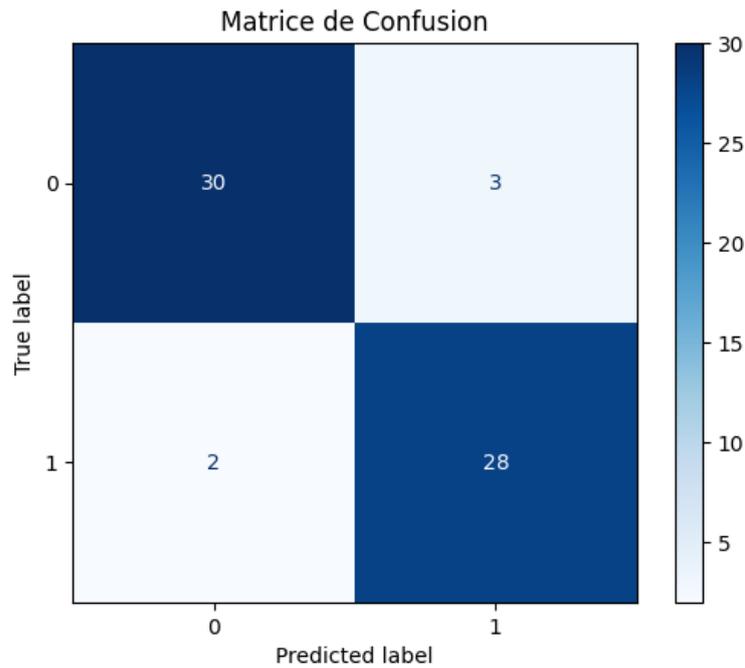
Nous avons divisé le jeu de données en une partition d'entraînement de 70 % qui permet à l'algorithme de s'entraîner et une partition de test de 30% qui est inconnu pour l'algorithme, qui consiste à tester l'algorithme pour sa capacité de prédiction sur des nouvelles données.

Concernant le jeu de données Shiraz, la classification des spectrogrammes sur les deux classes (Fetal / Maternal) avec ResNet50 a été conclu sur les résultats suivants :

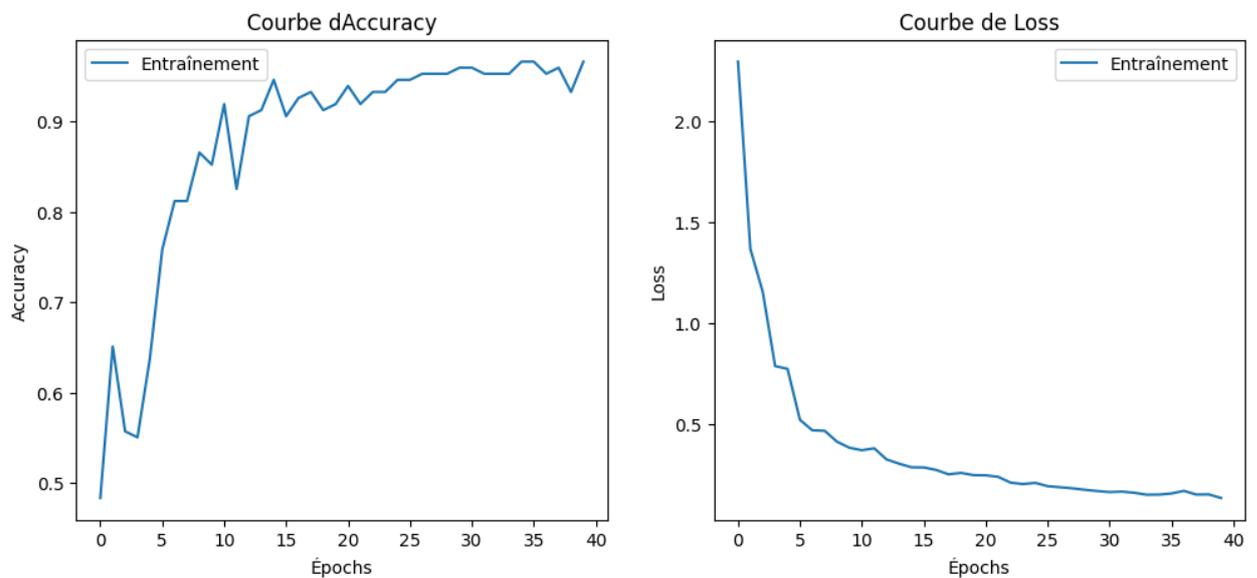
Métriques	Scenario Random (70/30)
Loss	0,21
Accuracy	0,9206
F1-Score	0.9207
Precision	0.9212
Recall	0.9206

**Tableau.3.3 Métrique de la classification en utilisant ResNet50**

Par évidence, on présente la matrice de confusion et les deux courbes de l'exactitude (Accuracy) et celle du Loss.



**Figure.3.3 Matrice de confusion de la phase test en utilisant ResNet50**



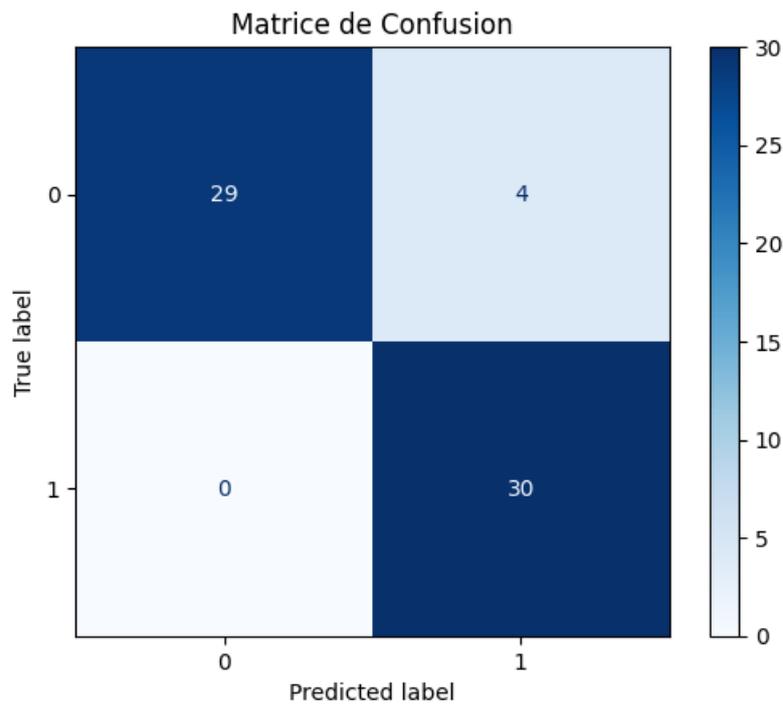
**Figure.3.4. Courbes de l'Accuracy et Loss en utilisant ResNet50**

Par la suite, on présente les résultats de classification en utilisant VGG19 avec le jeu de données Shiraz dans ce qui suit :

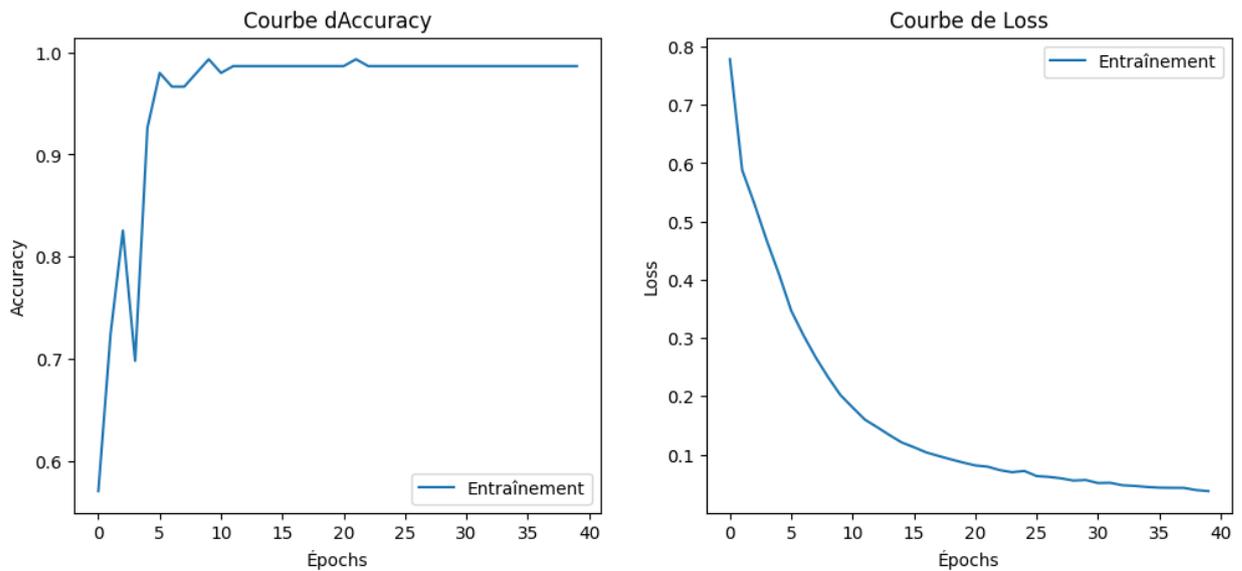
Métriques	Scenario Random (70/30)
Loss	0,1030
Accuracy	0,9365
F1-Score	0.9364
Precision	0.9440
Recall	0.9365

**Tableau.3.4. Métrique de la classification en utilisant VGG19**

En outre, on présente la matrice de confusion et les deux courbes de l'exactitude (Accuracy) et celle du Loss.



**Figure.3.5. Matrice de confusion de la phase test en utilisant VGG19**



**Figure.3.6. Courbes de l'Accuracy et Loss en utilisant VGG19**

L'expérience avec les deux modèles pré-entraînés sur le jeu de données Shiraz ont démontré la supériorité du VGG19 sur le ResNet50, avec un taux de 1,59% et un Loss réduit de 10,3 % comparé à 21% du ResNet50.

Concernant le jeu de données Circor22, plusieurs expériences ont été menées, malheureusement les taux d'Accuracy ont été stagnés dans les alentours de 50 % avec un taux de Loss très élevé dépassant les 75%. Ceci nous a poussé à entamer une étude analytique profonde expliquant la cause de ces résultats.

### **3- Analyse et Filtrage des jeux de données basée sur le balayage rythmique**

L'examen audio de plusieurs échantillons des deux jeux de données nous a permis d'observer que les débuts des enregistrements comportent des bruits et des anomalies sonores. Ceci nous a poussé à appliquer une étude dont le but est d'extraire les morceaux audio contenant des séquences de battement de cœur, ou autrement dit les morceaux significatifs pour notre méthodologie.

Pour les jeux de données Shiraz et Circor22, on peut remarquer les détails dans le tableau suivant.

Shiraz		Circor22	
Durée	Nombre d'échantillons	Durée	Nombre d'échantillons
25-36	89	10-20	1002
36-47	11	20-30	1408
47-57	2	30-40	575
57-68	6	40-50	16
68-79	4	50-60	6
79-90	5	60-70	3
90-101	91		
101-111	1		
111-122	1		
122-133	2		

**Tableau.3.5. Statistiques Histogrammiques des durées d'enregistrements**

En premier lieu, pour le jeu de données Shiraz les durées des enregistrements ne sont pas identiques. Selon le tableau, on constate qu'un ensemble constituant 41 % du jeu de donnée a une durée comprise entre 25 et 36 secondes. Un deuxième ensemble comporte plus de 43% du jeu de donnée contient des enregistrements dont la durée est entre 90 et 101 secondes. Le reste du jeu de donnée est distribuée sur les autres espaces. Pour le jeu de donnée de Circor22, on remarque que plus de 46% des échantillons ont des durées entre 20 et 30 secondes, 33% des échantillons entre 10 et 20 secondes, et le reste est distribué sur les autres espaces comme c'est illustré dans le tableau ci-dessus. Il faut noter que ces statistiques négligent les échantillons

dont les durées sont inférieures de 10 secondes (la durée expérimentale de notre méthodologie).

Le balayage des échantillons afin de détecter les zones ayant un rythme a été réalisé par la fonction de librairie HeartPy : *heartpy.process\_segmentwise*. La fonction permet d’analyser le vecteur du fichier audio, en cherchant un rythme bien défini, en balayant une fenêtre de 10 secondes sur le long de l’échantillon. La fonction aura comme résultat une suite de valeurs décrivant la rythmie de chaque fenêtre balayée. Le pas de balayage (*overlap*) est un paramètre expérimental qu’on doit spécifier pour déterminer le taux d’intersection entre fenêtres. Cette technique a été appliqué sur les deux jeux de données dans le but de découvrir les zones ayant une rythmie acceptable pour notre étude.

La figure suivante illustre un exemple des résultats obtenus par la technique de balayage rythmique. Le résultat obtenu est une carte Bpm de chacun des segments balayés pour tous les échantillons (*voir le cadre bleu en pointillés*).

A	C	D	E	F	G	H	I	J
ID	Classe	File name	durée(sec)	sgmnt1	sgmnt2	sgmnt3	sgmnt4	sgmnt5
0	f	1.wav	38,01002268	10025.76	10827.826	nan	728.12328	651.4178
1	f	10.wav	90,09002268	[nan	169.24651	320.44243	315.02307	170.6546
2	f	100.wav	95,7	[51606.38	1291.3399	677.58262	677.58262	677.5826
3	f	101.wav	90,28503401	[2214.967	668.83963	668.83963	739.22413	13354.00
4	f	102-1.wav	100,62	[nan	154.03422	nan	nan	nan
5	f	102.wav	131,2950113	[374.8583	nan	98.029045	244.23112	974.2268
6	f	103-1.wav	90,24	[nan	3999.1109	3999.1109	nan	7236.098
7	f	103.wav	91,86	[128.3220	129.59153	126.68773	124.82898	124.9527
8	f	104.wav	90,21002268	[485.8611	1181.7775	1181.7775	1181.7775	592.2444
9	f	105-1.wav	92,65501134	[nan	nan	683.74448	323.34764	143.5313
10	f	105.wav	90,18	[296.6035	150.67908	154.29171	165.97666	2477.012
11	f	106.wav	99,09002268	[nan	169.70241	124.02737	161.95372	124.0273
12	f	107.wav	90,36	[nan	1092.4855	1165.3179	98.782946	1029.571
13	f	108.wav	99,85501134	[nan	nan	nan	nan	72.34251
14	f	109.wav	90,54	[1056.287	1299.7414	1250.1443	160.57774	684.0744
15	f	11.wav	70,89002268	[6129.729	7151.3513	7151.3513	7151.3513	194.4634
16	f	110.wav	90,3	[367.4927	193.25153	187.93948	487.26717	292.2545

### Figure.3.7. Résultats du balayage rythmique de quelques échantillons du jeu de donnée Shiraz

#### 3-1 Catégorisation des Segments

L'étude des segments obtenus a exprimé trois catégories de segments :

1. Segments dans les normes : ce sont les segments dont la rythmie est dans l'espace de la rythmie cardiaque humaine entre 40 bpm et 300 bpm.
2. Segments hors normes : ce sont les segments dont la rythmie est en dehors de l'espace de la rythmie cardiaque humaine (*généralement supérieur*).
3. Segments indéfinis : ce sont les segments qui n'ont pas eu de rythmie, qu'elle soit dans les normes ou hors normes. Ils portent la valeur « nan ».

Cette catégorisation nous a permis de découvrir et expliquer les résultats de la méthodologie proposée dans son premier scénario. Le tableau suivant illustre plusieurs informations importantes sur le jeu de données Shiraz.

Colonne Segment	Échantillons	nans	Vide	Segments rythmés	Bbm min	Bpm max	Segments dans la norme (SN)	Pourcentage des SN	Pourcentage des Bruits
1	212	30	0	182	79	300	95	44.81 %	<b>55.19%</b>
2	212	32	0	180	58	300	103	48.58 %	51,41 %
3	212	32	0	180	60	300	106	<b>50.00 %</b>	50.00 %
4	212	31	0	181	77	300	110	<b>51.89 %</b>	48.11 %
5	212	26	0	186	70	300	110	51.89 %	48.11 %

**Tableau.3.6. Résultats du balayage rythmique et Histogramme sur le jeu de donnée Shiraz**

Le parcours successif des colonnes de la carte Bpm du jeu de donnée, nous a offert la possibilité de créer pour chacune des colonnes un histogramme (*voir colonne rouge sur la figure 3.7*). On constate alors que la première colonne du premier segment représente les 10 premières secondes des échantillons étudiés. De ce fait, après l'élimination des morceaux indéfinis (nan), on trouve qu'il y a 182 échantillons rythmés, avec 95 seulement qui obéissent à la norme du rythme du cœur humain, ce qui constitue seulement 44,81 % du jeu de données avec 55,19 % des morceaux bruités. Ce pourcentage très élevé des morceaux bruités devient un problème

crucial pour n'importe quelle forme de classification éventuelle. On remarque aussi que le nombre des morceaux rythmés dans normes devient très faible comparé aux taux d'échantillons initial. Ainsi, l'analyse histogramique des colonnes suivantes ont permis de produire le tableau du balayage rythmique sur le jeu de donnée Shiraz.

En avançant sur les colonnes de la carte Bpm, on observe clairement une diminution des pourcentages du bruit, et une augmentation dans le pourcentage des morceaux rythmés dans les normes. Par conséquent, on peut extraire facilement les échantillons et leurs segments rythmés dans les normes. Cette technique de sélection permet de constituer un jeu de données obéissant aux normes des rythmes cardiaques humains d'une manière purement aléatoire. Toutefois, on ne pourra pas garantir un bon équilibre du jeu de données sur les classes étudiées.

Pareil pour le jeu de données Circor22, l'étude de balayage rythmique a permis de tracer le tableau suivant :

Colonne Segment	Échantillons	nans	Vide	Segments rythmés	Bpm min	Bpm max	Segments dans la norme (SN)	Pourcentage des SN	Poucentage Bruits
1	3010	368	0	2648	61	300	1244	41,33 %	<b>58.67%</b>
2	3010	229	187	2594	51	300	1507	53,38 %	46,62 %
3	3010	222	375	2413	46	300	1412	<b>53,59 %</b>	46,41 %
4	3010	197	615	2198	52	300	1295	<b>54,07 %</b>	45,93 %
5	3010	166	1050	1794	51	300	1032	52,65 %	47,35 %

**Tableau.3.7. Résultats du balayage rythmique et histogramique sur le jeu de donnée Circor22.**

La première colonne (*les 10 premières secondes des échantillons*) de la carte Bpm du jeu de donnée Circor22, démontre clairement que le taux de bruit est très élevé atteignant 58,67%, ce qui explique clairement les résultats de classification du premier scénario. On constate qu'avec l'avancement sur les colonnes de la carte Bpm, le taux des morceaux rythmés dans normes augmente d'une manière remarquable. Par conséquent, l'étude analytique par

balayage rythmique et histogramme des échantillons des deux jeux de données a permis de découvrir les zones bruitées et les éliminer de la phase de classification du deuxième scénario ci-dessous.

### 3-2 Statistiques des jeux de données après filtrage des bruits

Après la sélection des échantillons élus de la phase d'analyse et de filtrage des segments bruités, on a pu conclure un nouveau jeu de donnée Shiraz réduit de 212 échantillons à 106. La même procédure a été appliquée sur le jeu de donnée Circor22 pour arriver à un nouveau jeu de données réduit de 3010 échantillons à 1412. Les deux tableaux suivants présentent les nouvelles distributions en termes de classes étudiées.

Classes	Fetal	Maternel
Nombre d'échantillons	26	80
Pourcentage	25%	75%
Total	106	

**Tableau.3.8. Distribution des classes après filtrage de bruits (Shiraz)**

Classes	Normal	Anormal
Number	773	639
Persontage	55%	45%
Total	1412	

**Tableau.3.9. Distribution des classes après filtrage de bruits (Circor22)**

---

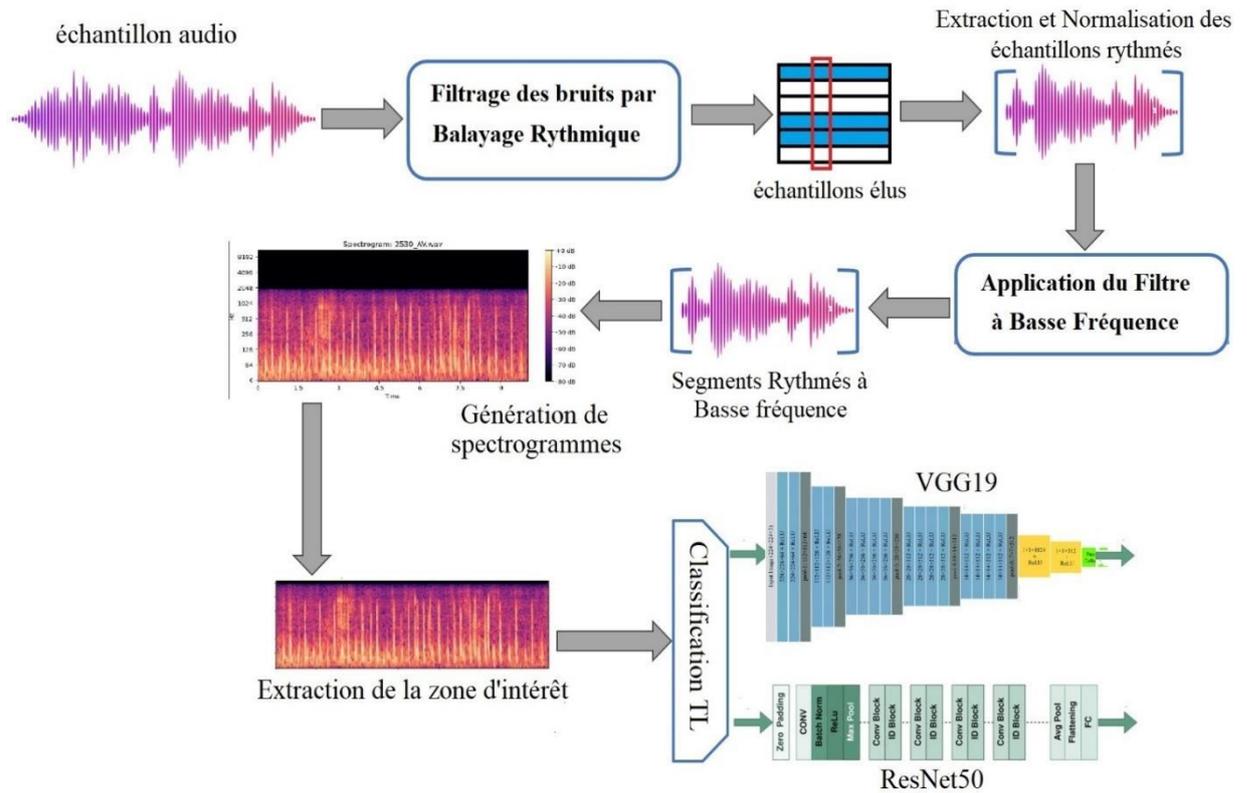
### 3-3 Prétraitement des données audios après filtrage des bruits

#### Étape 01: Limitation de la durée des audios et filtrage de basses fréquences

- Chargement des fichiers audio : utiliser Librosa pour charger les fichiers audios.
- extraction des morceaux audio sous les normes : en utilisant la même durée (10 secondes) l'extraction a débuté à partir de la 5ème seconde.
- Filtrage de basse fréquence : on applique la fonction de filtrage des basses fréquences pour améliorer la qualité des extraits audios vu que les battements de cœur sont de nature de basse fréquence. Le protocole proposé s'est limité sur trois niveaux de basses fréquences : 100, 200, 300 Hz.

#### Étape 02 : Conversion en spectrogramme et extraction des zones d'intérêts

- Convertir les segments contenant des sons cardiaques au format (wav), en images spectrales de type .jpg. on utilise alors la transformation STFT de la bibliothèque Librosa pour extraire les propriétés spectrales attendues.



**Figure.3.8. Méthodologie proposée avec Filtrage de bruits et Filtre à basse fréquence**

- Extraction des zones d'intérêt des spectrogrammes : Les images de dimensions différentes ne peuvent pas être empilées dans un tableau ni intégrées dans un algorithme d'apprentissage automatique. Nous allons donc recadrer les spectrogrammes pour maintenir l'intégrité de l'information spectrale, en les ajustant à des dimensions uniformes de 769 pixels de hauteur et 115 pixels de largeur pour échantillons à filtre de basse fréquence 100 Hz,  $769 \times 150$  pour les échantillons à filtre de basses fréquences 200 Hz et  $769 \times 170$  pour échantillons à filtre de basses fréquences 300 Hz. Ainsi, les spectrogrammes recadrés sont ensuite regroupés et préparés pour la phase de classification suivante.

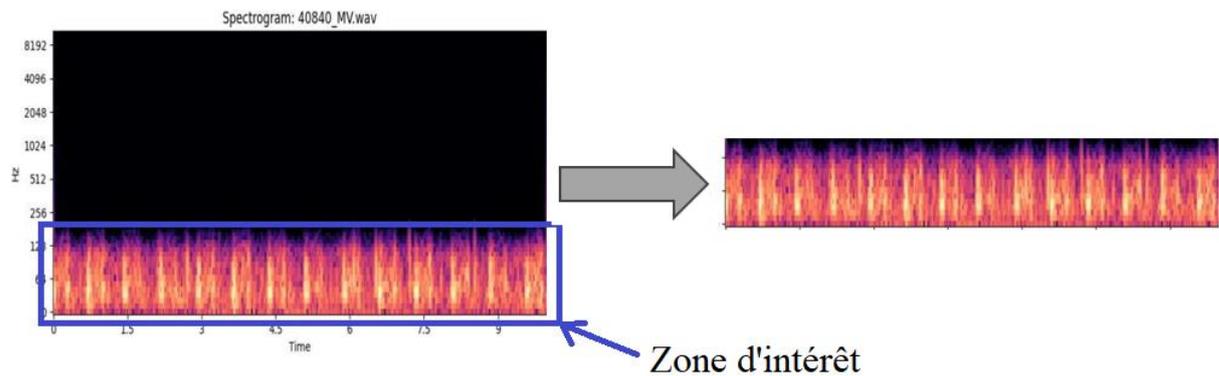


Figure.3.9. Extraction de la zone d'intérêt après filtre à basse fréquence

#### Étape 04: Classification basée Transfert Learning (VGG19 et ResNet50)

La phase de classification TL a été menée avec les deux modèles pré-entraînés VGG19 et ResNet50, comme c'était le cas pour le premier scénario expérimental de notre méthodologie. Ainsi, on peut observer les améliorations possibles entre les deux scénarios étudiés.

La classification TL a été appliquée aussi sur les zones de spectrogrammes issues de segments filtrés à basse fréquence (100, 200, 300 Hz). De cette façon, on peut observer l'utilité du filtre à basse fréquence sur les segments rythmés. Par conséquent, trois lots de jeux de données Shiraz ont été testés.

Concernant, le jeu de données Circor22

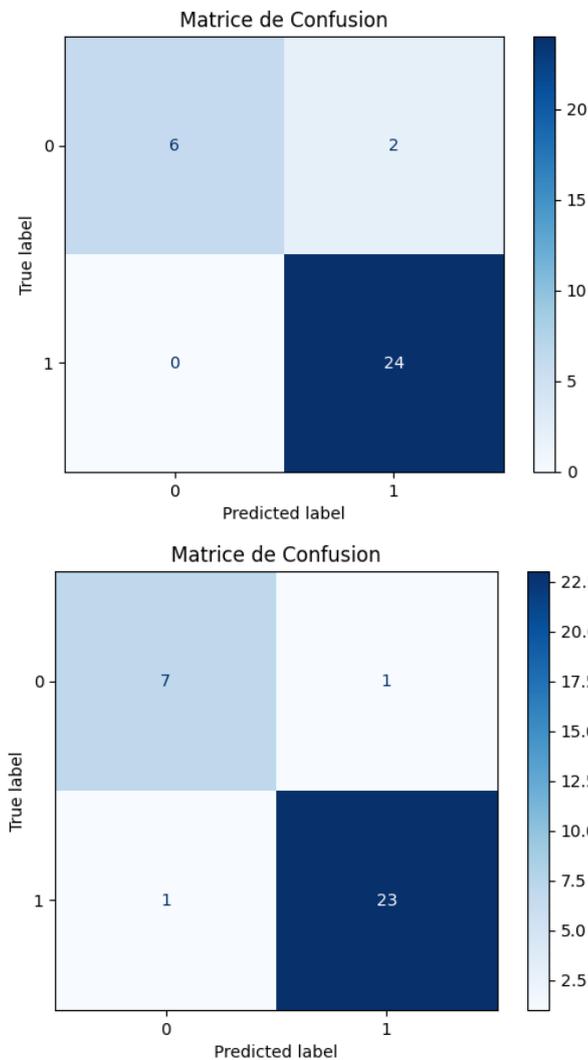
### 3-4 Résultats

Une première expérience a été menée avec les deux modèles pré-entraînés VGG19 et ResNet50 sur la sélection appliquée après le filtrage des échantillons bruités sur le jeu de données Shiraz. Les résultats ont démontré une amélioration même avec la réduction du volume du jeu de données. Le tableau suivant résume les résultats avec VGG19 et ResNet50.

Métriques	VGG19	ResNet50
Loss	0,1218	0.2049
Accuracy	0.9375	0.9375
F1-Score	0.9343	0.9375
Precision	0.9323	0.9375
Recall	0.9375	0.9375

**Tableau.3.10. Résultats de la classification TL sur le jeu de données Shiraz après filtrage de bruits**

Ainsi, on présente les deux matrices de confusion :



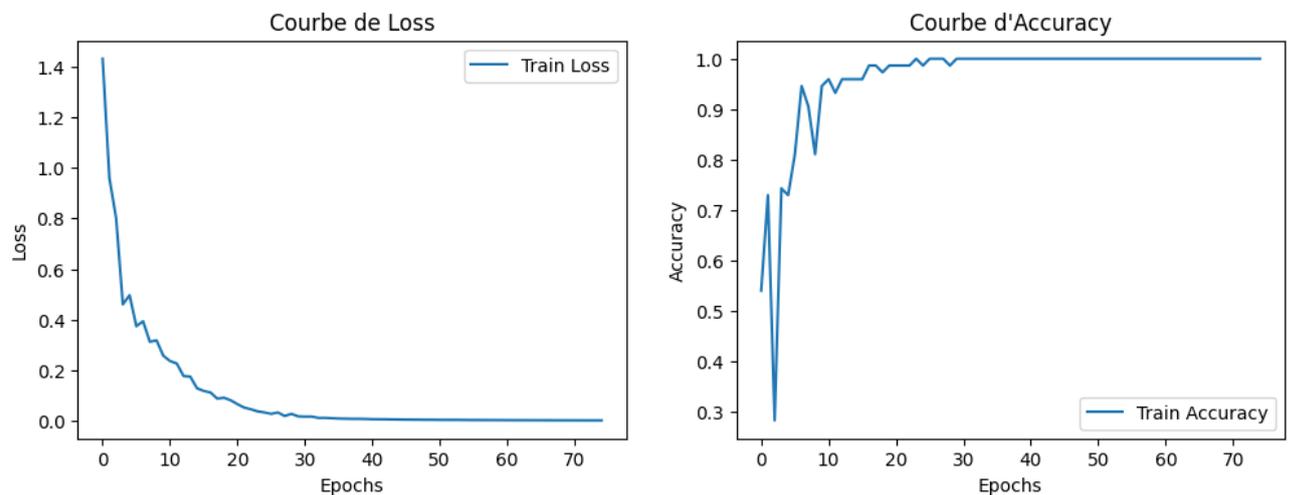
**Figure.3.10 Matrice de confusion de la classification VGG19 et ResNet50**

La suite expérimentale est de lancer la classification sur les trois lots de jeux de données Shiraz après l'application du filtre à basse fréquence. Les résultats sont résumés dans le tableau suivant :

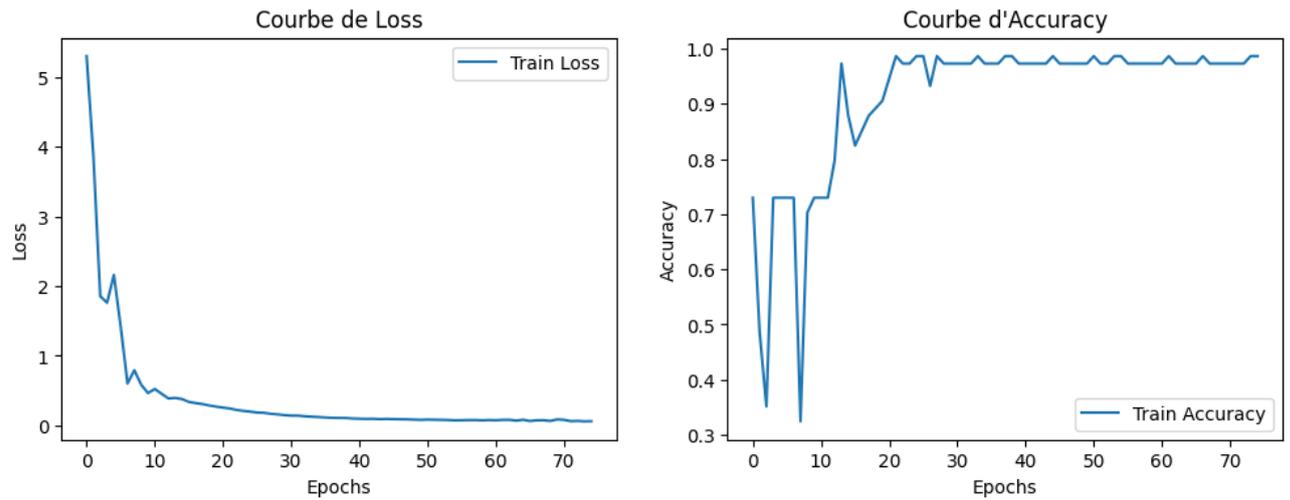
Modèle TL	Métriques	Loss	Accuracy	F1-Score	Precision	Recall
VGG19	Shiraz LFF 100	0,1224	0,9688	0,9676	0,9699	0,9688
	Shiraz LFF 200	0.0773	0.9688	0.9676	0.9699	0.9688
	Shiraz LFF 300	<b>0.0367</b>	<b>0.9688</b>	0.9676	0.9699	0.9688
ResNet50	Shiraz LFF 100	0,1323	0,9688	0,9676	0,9699	0,9688
	Shiraz LFF 200	0.1602	0.9375	0.9324	0.9420	0.9375
	Shiraz LFF 300	<b>0.0996</b>	<b>0.9688</b>	0.9676	0.9699	0.9688

**Tableau.3.11. Résultats de la Classification en utilisant le VGG19 et ResNet50 sur les trois lots après l'application du filtre de basse fréquence (LFF).**

Aussi, on présente les courbes d'exactitude (*Accuracy*) et du Loss des meilleurs résultats avec VGG19 et ResNet50 :



**Figure.3.11. Courbes VGG19 avec Shiraz LFF 300**



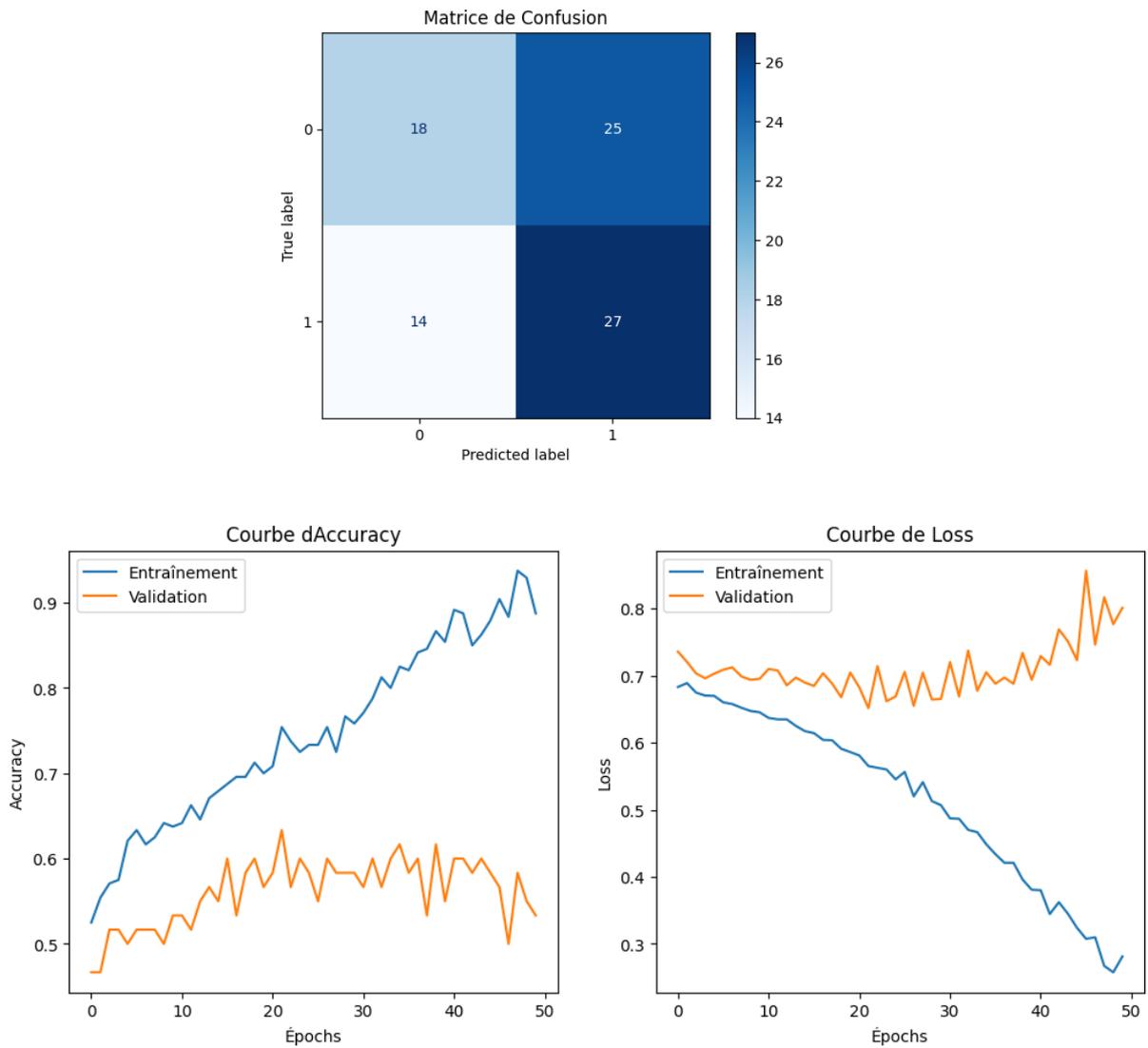
**Figure.3.12. Courbes ResNet50 avec Shiraz LFF 300**

Concernant le jeu de données Circor22, on a pu réaliser quelques expériences avec trois modèle pré-entraîné (VGG19, ResNet50 & MobilNet) sur un seul lot après l’application du filtre à basse fréquence (100 Hz). Les résultats sont résumés dans le tableau suivant :

Modèle TL	Métriques	Loss	Accuracy	F1-Score	Precision	Recall
VGG19	Circor22 LFF 100	0.7592	<b>0.5357</b>	<b>0.5291</b>	<b>0.5414</b>	<b>0.5357</b>
ResNet50	Circor22 LFF 100	<b>0.7396</b>	0.5119	0.3466	0.2620	0.5119
MobilNet	Circor22 LFF 100	0.8759	0.5238	0.3601	0.2744	0.5238

**Tableau.3.12. Résultats de la Classification en utilisant le VGG19, ResNet50 et MobilNet après l’application du filtre de basse fréquence (LFF 100) sur les échantillons élus du Circor22.**

Aussi, on présente la matrice de confusion et les courbes d'exactitude (*Accuracy*) et du Loss des meilleurs résultats avec VGG19 :



**Figures.3.13 Matrice de confusion & Courbes VGG19 avec Circor22 LFF**  
100

## 4- Discussion des Résultats

En prenant en considération les résultats du premier scénario (avant le filtrage de bruits) avec les résultats du deuxième scénario (après filtrage de bruits), on remarque principalement qu'il y eu une amélioration en termes de l'Accuracy et la métrique du Loss. En ce qui concerne le jeu de données Shiraz, l'Accuracy de la classification était meilleur avec VGG19 dans les trois expériences. La classification des images spectrogrammes en premier lieu a été conclu sur un taux d'Accuracy dépassant les 93,65 % lié à un Loss de 10 % comparé aux résultats du ResNet50 qui se sont avérés inférieurs. Après l'application du filtrage du bruit basé sur le balayage rythmique, les échantillons élus ont démontré une hausse dans l'Accuracy. Cette amélioration était remarquable avec les mêmes échantillons mais sous le filtre à basse fréquence à 300 Hz qui a prouvé sa supériorité par rapport aux autres filtres à basse fréquence 100 et 200 Hz. Le taux d'Accuracy a atteint les 96,88 % avec un taux de Loss remarquable réduit à 3,67 %. En parallèle on constate qu'avec le ResNet50 les échantillons du jeu de données Shiraz sous filtre à basse fréquence 300 Hz, donne le meilleur résultat par rapport aux autres lots d'échantillons. Ce qui prouve d'un côté que l'opération de filtrage de bruits suivi par un filtrage à basse fréquence peut être considéré comme une meilleure solution pour la classification des audios de battements de cœur, en utilisant les images de spectrogrammes. Aussi, on peut noter que les spectrogrammes utilisés peuvent porter des caractéristiques discriminatoires pouvant participer à la classification des sons de battements de cœur. De là, on peut dire que le jeu de données Shiraz à pu être un support de validation de notre méthodologie, même avec son petit volume d'échantillons.

Malgré le bon volume du jeu de données Circor22, on a trouvé du mal à gérer les bruits et les anomalies d'enregistrement qui ont produit les échantillons sonores. L'examen audio de plusieurs échantillons a démontré

que la qualité sonore des battements de cœurs dans la plupart des échantillons était difficile à cerner, comparé aux échantillons du jeu de données Shiraz, qui étaient plus claires à écouter. Cet inconvénient a été le principal obstacle vis-à-vis l'opération de classification TL. Toutefois, ce dernier a été un point de pousse pour entamer l'étude analytique des deux jeux de données, qui a été conclu par la découverte expérimentale de la méthodologie finale, qui a intégré le filtrage à basse fréquence et le filtrage de segments audio bruités. Finalement, le jeu de donnée Circor22 nécessite plus de travail de recherche exploitant d'autres formes de prétraitement des échantillons audios afin d'augmenter les taux de classifications.

# **C**onclusion général

## Conclusion général

---

La santé cardiaque est l'un des domaines de santé qui nécessite toutes les formes d'intention et d'investissement surtout en ce qui concerne l'introduction des technologies informatiques et de l'intelligence artificielle. Ce travail a été dirigé dans un cadre de recherche exploitant et mitigeant les techniques de traitement audios avec ceux du traitement d'images. Le but principal était de fournir une solution efficace et rapide alimentée par l'IA pouvant identifier avec précision les sons cardiaques normaux/ anormaux, sinon faire la différence entre les battements de cœurs de fœtus et celui de sa maternelle, en particulier dans les endroits où l'accès aux soins cardiaques spécialisés est limité.

Une méthodologie initiale était proposée et validée sur le jeu de données Shiraz, avec la réussite d'avoir des taux de classification TL acceptable. Puis, pour des raisons analytiques, la méthodologie a été étendue avec deux opérations importantes. Une première qui avait pour but d'exclure les extraits sonores n'obéissant pas aux normes de la rythmie cardiaque humaine, et une deuxième opération qui avait pour fin de filtrer le signal sonore et ne laisser que les basses fréquences. Les expériences avec cette dernière version de l'approche proposée a prouvé son efficacité et sa supériorité en atteignant un taux de classification de 96,88 % et un taux de Loss réduit à 3,67% en utilisant les Transfert Learning avec VGG19, alors que la méthodologie initiale n'a eu qu'un taux de 93,65 % et un Loss de 10 %.

Ce domaine de recherche a démontré l'utilité de travailler sur les représentations spectrographiques des sons de battements de cœur, ainsi on peut citer plusieurs perspectives de recherche qui peuvent être une succession à ce travail. Ce travail a beaucoup d'axes expérimentaux à exploiter. On cite alors un axe ciblant d'autres formes de spectrogrammes, un axe exploitant d'autre forme de filtrage de bruits et de nettoyage des échantillons audios, un axe exploitant des modifications sur les modèles de TL pré-entraîné, un axe

## **Conclusion général**

---

cherchant à bâtir son propre CNN, un axe exploitant d'autres techniques de prétraitement comme l'amplification des basses fréquences.

# **B**ibliographie

## **Bibliographie**

- [1] Sonix. (n.d.). History of Speech Recognition. Retrieved from <https://sonix.ai/history-of-speech-recognition>
- [2] Condec Software. (n.d.). A History of Voice Recognition Technology. Retrieved from <https://www.condecsoftware.com/blog/a-history-of-voice-recognition-technology>
- [3] Soundproofing Company. (n.d.). What is Sound?. Retrieved from [https://www.soundproofingcompany.com/soundproofing\\_101/what-is-sound](https://www.soundproofingcompany.com/soundproofing_101/what-is-sound)
- [4] Physics Classroom. (n.d.). The Speed of Sound. Retrieved from <https://www.physicsclassroom.com/class/sound/Lesson-2/The-Speed-of-Sound>
- [5] Jagran Josh. (n.d.). What are the Characteristics of Sound Waves?. Retrieved from <https://www.jagranjosh.com/general-knowledge/what-are-the-characteristics-of-sound-waves-1525678871-1>
- [6] Glenat, M., & Chareyron, D. (2012, June). Principe du passage de l'analogique au numérique. CultureSciences Physique - ISSN 2554-876X. Retrieved from <https://culturesciencesphysique.ens-lyon.fr/ressource/principe-numerisation.xml>
- [7] Sonart. (n.d.). Conversions. Retrieved from <http://sonart.free.fr/cours/Conversions.pdf>
- [8] Multipower. (n.d.). Signaux Analogiques et Numériques. Retrieved from <https://www.multipower.fr/ressources/blog-2/signaux-analogiques-numeriques/>
- [9] Chareyron, D. (2013, October). Numérisation d'un signal acoustique : effet de la quantification (partie 1). CultureSciences Physique - ISSN 2554-876X. Retrieved from <https://culturesciencesphysique.ens-lyon.fr/ressource/numerisation-acoustique-Chareyron1.xml>
- [10] Fillion. (n.d.). Format Fichiers Audio. Retrieved from <https://fillion.ca/blogs/guides-dachats/format-fichiers-audio>
- [11] Harrif, R., & Hamadou, R. (n.d.). Tatouage numérique des fichiers audio échangés en télé-médecine.
- [12] Repensko, D. (n.d.). Methods of Audio Analysis. Retrieved from <https://medium.com/@repenko.d1504/methods-of-audio-analysis-c296f781970c>

## **Bibliographie**

---

- [13] AltexSoft. (n.d.). Audio Analysis. Retrieved from <https://www.altexsoft.com/blog/audio-analysis/>
- [14] MathWorks. (n.d.). FFT. Retrieved from <https://www.mathworks.com/discovery/fft.html>
- [15] Jenkins, W. K. (2022). Fourier series, Fourier transforms and the DFT. In Mathematics for Circuits and Filters (pp. 83-111). CRC Press.
- [16] Gourisaria, M. K., Agrawal, R., Sahni, M. et al. (2024). Comparative analysis of audio classification with MFCC and STFT features using machine learning techniques. Discovery Internet of Things, 4, 1. <https://doi.org/10.1007/s43926-023-00049-y>
- [17] Jeon, H., Jung, Y., Lee, S., & Jung, Y. (2020). Area-efficient short-time Fourier transform processor for time–frequency analysis of non-stationary signals. Applied Sciences, 10(20), 7208.
- [18] PeoPeo. (n.d.). Filtres Audios et Égaliseurs. Retrieved from <https://www.peopeo.io/en/filtres-audios-et-egaliseurs/>
- [19] Adobe. (n.d.). Sound Equalizer. Retrieved from [https://www.adobe.com/lu\\_fr/creativecloud/video/discover/sound-equalizer.html](https://www.adobe.com/lu_fr/creativecloud/video/discover/sound-equalizer.html)
- [20] Aulart. (n.d.). Music Equalization: Why It Matters. Retrieved from <https://www.aulart.com/blog/music-equalization-why-it-matters/>
- [21] Adobe. (n.d.). Filter and Equalizer Effects. Retrieved from <https://helpx.adobe.com/no/audition/using/filter-equalizer-effects.html>
- [22] Emastered. (n.d.). What is Spectrogram. Retrieved from <https://emastered.com/fr/blog/what-is-spectrogram>
- [23] iZotope. (n.d.). Spectrogram Waveform Display. Retrieved from <https://s3.amazonaws.com/izotopedownloads/docs/rx6/07-spectrogram-waveform-display/index.html>
- [24] Fabien, M. (n.d.). Log-Frequency Power Spectrogram. Retrieved from <https://maelfabien.github.io/machinelearning/Speech10/#log-frequency-power-spectrogram>
- [25] Rabiner, L. R., & Juang, B. H. (2006). Speech recognition: Statistical methods. Encyclopedia of Language & Linguistics, 1-18.

## **Bibliographie**

---

- [26] Mamyrbayev, O., & Oralbekova, D. (2020). Modern trends in the development of speech recognition systems., (4), - 42-51.
- [27] Sehili, M. E. A. (2013). Reconnaissance des sons de l'environnement dans un contexte domotique (Doctoral dissertation, Evry, Institut national des télécommunications).
- [28] Liu, Z., Jiang, H., Zhang, F., Ouyang, W., Li, X., & Pan, X. (2023). Heart sound classification based on bispectrum features and Vision Transformer mode. Alexandria Engineering Journal, 85, 49-59.
- [29] CHUV. (n.d.). Fonctionnement du Coeur. Retrieved from <https://www.chuv.ch/fr/cardiologie/car-home/patients-et-famille/fonctionnement-du-coeur>
- [30] Zemouli, M. (2023). Un système intelligent pour améliorer la prédiction des maladies cardiovasculaires.
- [31] Jondeau, G. (2006). Insuffisance cardiaque et cardiomyopathies. Archives des Maladies du Cœur et des Vaisseaux, 99(2).
- [32] Mbelesso, P., Junior, S. G., & Gody, J. C. (2019). L'anencéphalie et le mythe des hommes sans tête: à propos d'un cas au Centre Hospitalier Universitaire Pédiatrique de Bangui. Revue Neurologique, 175, S49-S50.
- [33] Akoudad, H., & Benamer, H. (2004). Physiopathologie et étiopathogénie de l'infarctus du myocarde. EMC-Cardiologie-Angéiologie, 1(1), 49-67.
- [34] Balaka, A., Tchamdja, T., Djibril, M. A., Djagadou, K. A., Tchandana, M., Damorou, F., & Agbétra, A. (2015). Les valvulopathies cardiaques en milieu hospitalier à Lomé (Togo). Pan African Medical Journal, 20(1).
- [35] Le Guillou, S., Casalta, J. P., Fraisse, A., Kreitmann, B., Chabrol, B., Dubus, J. C., & Bosdure, E. (2010). Endocardite infectieuse sur cœur sain chez l'enfant: étude rétrospective de 11 cas. Archives de Pédiatrie, 17(7), 1047-1055.
- [36] Bastide. (n.d.). Stethoscope Cardio Prestige II. Retrieved from <https://www.bastideleconfortmedical.com/stethoscope-cardio-prestige-ii-double-pavillon-spengler-8095985.html>
- [37] AVF Biomedical. (n.d.). Durée d'un Électrocardiogramme. Retrieved from <https://www.avf-biomedical.com/blog/conseils/quelle-est-la-duree-dun-electrocardiogramme-et-comment-se-deroule-t-il/>

## **Bibliographie**

- [38] Hampton, J. R. (2013). What the ECG is about. In *The ECG Made Easy* (8th ed., pp. 22-97). Nottingham UK: Churchill Livingstone Elsevier.
- [39] Leatham, A. (1949). Phonocardiography. *Postgraduate Medical Journal*, 25(289), 568.
- [40] ResearchGate. (n.d.). Phonocardiography Signal (PCG Signal). Retrieved from [https://www.researchgate.net/figure/phonocardiography-signal-PCG-signal\\_fig1\\_338197056](https://www.researchgate.net/figure/phonocardiography-signal-PCG-signal_fig1_338197056)
- [41] Chamoux, A. (1999). Présentation du logiciel d'analyse de la fréquence cardiaque LAC 981. *Archives des Maladies Professionnelles et de Médecine du Travail*, 60(2), 145-147.
- [42] Moody, G. B., & Mark, R. G. (2001). The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, 20, 45-50. <https://doi.org/10.1109/51.932724>
- [43] Gomes, E. F., Bentley, P. J., Coimbra, M., Pereira, E., & Deng, Y. (2013). Classifying heart sounds: Approaches to the PASCAL challenge. In *Proceedings of the HEALTHINF 2013-Proceedings of the International Conference on Health Informatics* (pp. 337-340). Barcelona, Spain.
- [44] PhysioNet. (n.d.). CIRCOR Heart Sound Database. Retrieved from <https://physionet.org/content/sufhsdb/1.0.1/>
- [45] Balamurugan, A., Teo, S. G., Yang, J., Peng, Z., Xulei, Y., & Zeng, Z. (2019, May). Reshnet: Spectrograms based efficient heart sounds classification using stacked residual networks. In *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)* (pp. 1-4). IEEE.
- [46] Degirmenci, M., Ozdemir, M. A., Izci, E., & Akan, A. (2022). Arrhythmic heartbeat classification using 2D convolutional neural networks. *IRBM*, 43(5), 422-433.
- [47] Malik, H., Bashir, U., & Ahmad, A. (2022). Multi-classification neural network model for detection of abnormal heartbeat audio signals. *Biomedical Engineering Advances*, 4, 100048.
- [48] Yebda, T. (2022). Détection de situations à risques pour des personnes fragiles sur des données multimodales (Doctoral dissertation, Université de Bordeaux).

## **Bibliographie**

---

- [49] Almanifi, O. R. A., Ab Nasir, A. F., Razman, M. A. M., Musa, R. M., & Majeed, A. P. A. (2022). Heartbeat murmurs detection in phonocardiogram recordings via transfer learning. *Alexandria Engineering Journal*, 61(12), 10995-11002.
- [50] Wibawa, M. S., Maysanjaya, I. M. D., Novianti, N. K. D. P., & Crisnapati, P. N. (2018, August). Abnormal heart rhythm detection based on spectrogram of heart sound using convolutional neural network. In 2018 6th International Conference on Cyber and IT Service Management (CITSM) (pp. 1-4). IEEE.
- [51] Le Data Scientist. (n.d.). Google Colab: Le Guide Ultime. Retrieved from <https://ledatascientist.com/google-colab-le-guide-ultime/>
- [52] Librosa. (n.d.). Documentation. Retrieved from <https://librosa.org/doc/latest/index.html>
- [53] TensorFlow. (n.d.). Learn. Retrieved from <https://www.tensorflow.org/learn>
- [54] Keras. (n.d.). Documentation. Retrieved from <https://keras.io>
- [55] Almanifi, O. R. A., Ab Nasir, A. F., Razman, M. A. M., Musa, R. M., & Majeed, A. P. A. (2022). Heartbeat murmurs detection in phonocardiogram recordings via transfer learning. *Alexandria Engineering Journal*, 61(12), 10995-11002.
- [56] PhysioNet. (n.d.). CIRCOR Heart Sound Database. Retrieved from <https://physionet.org/content/circor-heart-sound>
- [57] Kamal, K., & Hamid, E. Z. (2023). A comparison between the VGG16, VGG19 and ResNet50 architecture frameworks for classification of normal and CLAHE processed medical images