



Université Libanaise
Faculté de Technologie

LICENCE II

Année universitaire 2023 - 2024

Génie des réseaux informatiques et de
télécommunications

Filtrage du signal ECG par apprentissage automatique

Préparé par HAMAD Batoul

du 1/7/2024 au 11/11/2024

À



Remerciements

Je tiens à exprimer ma sincère gratitude à toute l'équipe d'ENSEMBLE POUR CHEHIM pour l'opportunité de stage que j'ai pu réaliser, axée sur le filtrage du signal ECG.

Je remercie particulièrement Dr Nassib ABDALLAH et Dr Hassan SERHAL pour leur encadrement et leurs précieux conseils tout au long de cette expérience. Leur expertise et leur disponibilité m'ont grandement aidé à développer mes compétences. Je souhaite également remercier Dr Aya HAGE CHEHADE, Rana SAAB et Shakib YOUSSEF pour leur soutien et leur collaboration. Leurs échanges enrichissants et leur esprit d'équipe ont rendu ce stage encore plus agréable et formateur.

Cette expérience m'a permis d'acquérir des connaissances précieuses et de mieux comprendre les enjeux liés au traitement des signaux ECG. Merci à tous pour votre aide et votre bienveillance.

Je souhaite également exprimer ma profonde gratitude à Dr Abed El Salam Hajjar, chef de département de la Faculté de Technologie, pour son soutien inestimable tout au long de mon parcours universitaire. Son leadership, son engagement envers l'excellence académique et sa vision inspirante ont été des sources de motivation constantes pour moi et mes camarades. Je tiens également à remercier tous les docteurs de mon département en Génie des Réseaux Informatiques et Télécommunications. Chacun d'entre vous a joué un rôle essentiel dans ma formation, en partageant vos connaissances et en nous encourageant à toujours donner le meilleur de nous-mêmes. Vos cours captivants et votre disponibilité pour répondre à nos questions ont largement contribué à notre développement personnel et professionnel.

Votre passion pour l'enseignement et votre dévouement envers vos étudiants font de notre département un lieu d'apprentissage exceptionnel. Je me sens chanceux d'avoir pu bénéficier de votre expertise et de votre soutien.

Merci à tous pour cette belle expérience et pour avoir enrichi mon parcours académique.

Plan

Introduction	1
• Contexte	1
• Problématiques et motivation	1
• Objectifs à atteindre	1
1 Bilan du Stage	3
1.1 Analyse Fonctionnelle	3
1.1.1 Détermination de l'objectif : méthode QQQQCP	3
1.1.1.1 Quoi ?	3
1.1.2 Définition du besoin	4
➤ Diagramme bête à cornes.....	4
➤ Diagramme de pieuvre	5
1.2.1 Contexte.....	5
1.2.2 Objectifs	5
2 Etat de l'art.....	6
2.1 Logiciels Utilisés.....	6
2.1.1 Anaconda	6
2.1.2 Spyder	6
2.2 Intelligence artificielle	6
2.2.1 Apprentissage Supervisé.....	7
2.2.2 Apprentissage Non Supervisé	7
2.3 Langage.....	8
2.3.1 Les bibliothèques	8
2.4 Apprentissage automatique.....	8
2.5 Le cœur	8
2.6 Les électrodes.....	9

2.7 L'électrocardiogramme (ECG)	9
2.8 Le signal ECG.....	9
2.9 Dataset	10
2.9.1 Dataset iCentia11k.....	10
2.9.2 Dataset Fibrillation Auriculaire à Long Terme (LTAfDB)	10
2.10.2 Caractéristiques :	10
2.10.3 Complications et Symptômes :	10
2.11 Fibrillation Auriculaire à Long Terme (LTAf)	10
2.12 La circulation sanguine	11
2.12.1 Circulation dans un cœur normal :	11
2.14 Analyse du signal ECG.....	12
2.14.1 Installation des Données	12
2.14.1.1 Collecte des Données :	12
2.14.1.2 Transfert des Données :	12
2.15 Prétraitement des Données	12
2.15.1 Nettoyage des Données :	12
2.16 Segmentation du Signal.....	12
2.17 Extraction des Caractéristiques.....	12
2.20 Interprétation Clinique et Rapport des Résultats :	13
2.21 Fréquence des Ondes ECG :	13
2.22 Les Ondes du Signal ECG	13
2.22.1 Onde P :	13
2.22.2 Complexe QRS :	14
2.22.3 Onde T :	14
3 Travaux réalisés	14
• Pipeline.....	14

Analyse des Signaux Bruts des Patients	14
3.2 Prétraitement des Signaux ECG	16
3.2.1 Application des Filtres.....	16
3.2.2 Représentation multidimensionnelle :	18
3.2.3 Extraction des Caractéristiques	28
• Phase de Filtration et Objectifs	28
• Détection des Pics	28
3.2.4 Modélisation en Apprentissage Automatique.....	31
3.2.5 Métriques de Classification.....	32
Conclusion	41
Références.....	42

Liste des figures

Figure 1:L'IA au service du cœur(Gamkrelidze,2021).....	Error! Bookmark not defined.
Figure 2:Diagramme bête à cornes	4
Figure 3:Diagramme de pieuvre	Error! Bookmark not defined.
Figure 4:logo d'Anaconda(Rivas, Jesús A. Anaconda, 2020.)	6
Figure 5:ogo de spyder(Grace, Tom. Spyder web. Simon and Schuster, 2017)	6
Figure 6:Intelligence artificielle (Sadin, Éric,2018)	Error! Bookmark not defined.
Figure 7:Apprentissage supervisé(ETTARID, Pr Mohamed, et al,2021)	7
Figure 8: Apprentissage non supervise(Jacopin, Eliott, et al.,2021)	7
Figure 9:Logo de python(Dorling, M., & White, D. (2015, February)).....	8
Figure 10:schéma d'anatomie du cœur(Netter, F. H. (2023))	8
Figure 11:ECG électrode(Chlaihawi AA, Narakathu BB,	9
Figure 12:Électrocardiogramme(Berkaya, Selcan Kaplan, et al,2018)	9
Figure 13:Signal ECG(Houssein, Essam H., Moataz Kilany, and Aboul Ella Hassanien,2017)	Error! Bookmark not defined.
Figure 14:La circulation sanguine chez une personne normale et autre atteinte de la fibrillation auriculaire(MENIF, H., et al,2019)	Error! Bookmark not defined.
Figure 15: Rythme sinusal normal vs. fibrillation auriculaire (Dobrev, D., et al,2018)	Error! Bookmark not defined.
Figure 16:Pipeline du travail.....	Error! Bookmark not defined.
Figure 17:Détection d'artefacts dans un signal ECG.....	15
Figure 18:Signal ECG simulé avec des artefacts	16
Figure 19:Signal ECG filtré chez le premier patient en utilisant le filtre passe bande	16
Figure 20:Signal cardiaque complexe PQRST filtré à l'aide d'un filtre passe-bande	17
Figure 21:Signal cardiaque complexe PQRST filtré à l'aide d'un filtre passe-bas.....	17
Figure 22:Signal cardiaque complexe PQRST filtré à l'aide d'un filtre passe-haut.....	18
Figure 23:Visualisation 3D des résultats de l'Analyse en Composantes Principales (ACP)	18
Figure 24:Visualisation 2D des résultats de l'Analyse en Composantes Principales (ACP)	19
Figure 25:Composantes Indépendantes Extraites par (ACI) des Signaux ECG	20
Figure 26:Spectre de Fréquence du Signal ECG Filtré par Transformée de Fourier	21
Figure 27:Densité Spectrale de Puissance du Signal ECG-Pic Dominant à 1 Hz	21
Figure 28:Spectrogramme des Résultats de la STFT chez le Premier Patient	22
Figure 29:Scalogramme 3D des Résultats de la Transformée en Ondelette Continue (CWT)	23
Figure 30:Scalogramme 2D des Résultats de la Transformée en Ondelette Continue (CWT)	23
Figure 31:Visualisation 3D des Coefficients de la Transformée en Ondelette Discrète (DWT)	24
Figure 32:Représentation des coefficients d'ondelettes pour 5 signaux	25
Figure 33:Représentation 3D des Coefficients de la Transformée en Ondelette Multi-résolution	26
Figure 34:Représentation 2D des Coefficients de la Transformée en Ondelette Multi-résolution	26
Figure 35:Scalogramme 3D de la Transformée en Ondelette à Borne.....	27
Figure 36:Scalogramme 2D de la Transformée en Ondelette à Borne.....	28
Figure 37:Signal filtré avec passe-bande et pics détectés.....	29
Figure 38:Figure 37:Séparation des classes normales et anormales par KNN dans un espace 3D	31
Figure 39:Séparation des classes normales et anormales par SVM dans un espace 3D.....	33
Figure 40:Arbre de décision pour la classification avec une précision de 91%.....	35
Figure 41: Matrice de confusion en utilisant le GBM	36

Figure 42:Performance du modèle du Courbe ROC	38
---	----

Liste des tableaux

Tableau 1:Diagramme de Gantt	6
Tableau 2:Équations des Métriques Clés	13
Tableau 3:Évolution temporelle d'un phénomène oscillatoire	Error! Bookmark not defined.
Tableau 4:Résultats Généraux des Performances du Modèle	32
Tableau 5:Rapport de Classification des Classes	32
Tableau 6:Moyenne des Scores de Validation Croisée	32
Tableau 7:Résultats Généraux des Performances du Modèle	Error! Bookmark not defined.
Tableau 8:Rapport de Classification des Classes	34
Tableau 9:Moyenne des Scores de Validation Croisée	34
Tableau 10:Résultats de performance des Modèles de Classification	35
Tableau 11:Métriques de performance et validation croisée du modèle XGBoost	37
Tableau 12:Meilleurs hyperparamètres Valeurs	38
Tableau 13:Meilleurs Hyperparamètres du Modèle	39
Tableau 14:Métriques de Performance par Classe	Error! Bookmark not defined.
Tableau 15: Résumé Global des Performances	40
Tableau 16: comparaison entre l'exactitude de chaque modèles de performance	40

Liste des abréviations

AA - Atténuation des Artefacts
 ACI - Analyse en Composantes Indépendantes
 ACP - Analyse en Composantes Principales
 AUC - Area Under the Curve
 CF - Caractéristiques du Signal Filtré
 CWT - Transformée en Ondelette Continue
 DWT - Transformée en Ondelette Discrète
 DSP - Densité Spectrale de Puissance

EA - Efficacité de l'Approche

ECG - Électrocardiogramme

F1 - Fréquence de coupure 1 (0,5 Hz)

F1-score - Mesure de la performance (harmonique de précision et rappel)

F2 - Fréquence de coupure 2 (15 Hz)

FN - Faux Négatifs

FP - Faux Positifs

ICENTIA11K - Dataset contenant des enregistrements de signaux ECG de 11 000 patients

KNN - K-Nearest Neighbors

LTAF - Fibrillation Auriculaire à Long Terme

LTAFDB - Long Term Atrial Fibrillation Database

MRWT - Transformée en Ondelette Multi-résolution

ROC - Receiver Operating Characteristic

SVM - Machines à Vecteurs de Support

STFT - Transformée de Fourier à Court Terme

TF - Transformée de Fourier

TN - Vrais Négatifs

TP - Vrais Positifs

Résumé

Nos travaux ont exploré diverses approches pour améliorer l'analyse des signaux ECG dans le dataset LTAfDB (Long Term Atrial Fibrillation Database). Un défi majeur est la présence d'artefacts, comme les interférences musculaires et les bruits de mouvement, qui compromettent la qualité des données. Pour remédier à ces limitations, nous avons d'abord effectué une segmentation des signaux, permettant d'identifier des périodes spécifiques pour une analyse plus fine. Ensuite, un processus de filtrage a été mis en place pour éliminer ces artefacts, garantissant ainsi une classification plus précise.

Après le filtrage, nous avons appliqué des représentations multidimensionnelles, en utilisant des techniques comme l'Analyse en Composantes Principales (ACP) et la transformation en ondelettes. Cela a permis d'extraire des caractéristiques significatives du signal, telles que les pics RR, l'amplitude et la fréquence. Nous avons ensuite utilisé plusieurs modèles d'apprentissage automatique, dont KNN et SVM, qui ont obtenu les meilleures précisions, respectivement 0,97 et 0,99. Les performances de ces modèles ont été évaluées à l'aide de métriques appropriées, soulignant leur efficacité dans la classification des épisodes de fibrillation auriculaire.

Mots clés :

LTAfDB - Signal ECG - Prétraitement - Représentation multidimensionnelle - Transformée en ondelettes - Classification - Extraction des caractéristiques - Apprentissage automatique.

Introduction

- Contexte

Le projet vise à analyser des signaux ECG (électrocardiogrammes) afin d'améliorer la détection et la classification des anomalies cardiaques. Les ECG, essentiels pour le diagnostic des troubles cardiaques, peuvent contenir divers bruits qui compliquent leur interprétation. En appliquant des techniques de filtrage, nous cherchons à améliorer la qualité des signaux ECG, ce qui pourrait favoriser une prise en charge plus efficace des patients.

- Problématiques et motivation

Bien que les signaux ECG soient essentiels pour le diagnostic des anomalies cardiaques, ils sont souvent altérés par des problèmes tels que le bruit électrique, les interférences et les mouvements du patient, ce qui peut compromettre leur qualité et la précision des analyses. Ces perturbations rendent difficile la détection des irrégularités et peuvent mener à des interprétations erronées. Enfin, la question principale que ce rapport cherche à aborder est : **Dans quelle mesure peut-on atténuer ces bruits indésirables et améliorer la qualité des signaux ECG pour assurer des analyses plus précises et fiables ?**

- Objectifs à atteindre

L'objectif principal de ce projet est d'améliorer la qualité des signaux ECG en atténuant les bruits indésirables qui peuvent compromettre leur analyse. En intégrant des méthodes de filtrage avancées et en appliquant des techniques d'apprentissage automatique, nous visons à garantir une interprétation plus précise des données ECG, ce qui est crucial pour un diagnostic rapide et fiable des anomalies cardiaques. Cela inclut :

- **Analyse des sources de bruit** : Identifier les bruits (électriques, interférences) affectant les signaux ECG.
- **Filtrage des signaux ECG** : Appliquer des techniques de filtrage (passe-bande, adaptatifs) pour réduire le bruit tout en conservant les informations importantes.
- **Évaluation des filtres** : Mesurer l'efficacité des filtres via des critères quantitatifs et visuels.
- **Modèles de classification** : Développer des modèles d'apprentissage automatique pour détecter les anomalies dans les signaux ECG filtrés.
- **Comparaison des modèles** : Comparer les performances des modèles à l'aide de métriques comme la précision et le rappel pour une meilleure interprétation clinique.



Figure 1:L'IA au service du cœur(Gamkrelidze,2021)

Plan du Rapport

Le rapport s'articule autour de plusieurs grandes sections. Il débute par une présentation de l'association, qui met en avant son engagement dans la recherche et les initiatives culturelles. Ensuite, le bilan du stage est détaillé, comprenant une analyse fonctionnelle, la définition des besoins et un cahier des charges qui précise les tâches à réaliser. Suit une section sur l'état de l'art, qui explore les méthodes et technologies pertinentes pour l'analyse des signaux ECG. Ensuite, les travaux réalisés sont présentés, mettant en lumière les résultats obtenus et les défis rencontrés. Enfin, le rapport se conclut par une conclusion qui résume les enseignements tirés et propose des recommandations.

Présentation de l'association

L'association "**Ensemble pour Chehim**" s'engage pleinement dans la recherche académique et scientifique, tout en étant richement impliquée dans des initiatives culturelles. Nous sommes également actifs dans des activités philanthropiques, en nous concentrant sur le don de ressources médicales et académiques aux individus et institutions de santé dans le besoin. De plus, nous organisons une gamme variée d'ateliers scientifiques et éducatifs, de stages, de sessions de formation et d'événements collaboratifs. Ces activités couvrent non seulement diverses associations locales, mais s'étendent également à des partenariats avec des entreprises internationales.

L'association "Ensemble pour Chehim" comprend plusieurs comités, dont le **Comité de Recherche et de Projets Collaboratifs (CRPC)**, où j'ai effectué mon stage. Ce comité se consacre à la recherche scientifique, notamment dans les domaines de l'analyse de données, de la prédiction et de la classification grâce à l'intelligence artificielle.

1 Bilan du Stage

1.1 Analyse Fonctionnelle

1.1.1 Détermination de l'objectif : méthode QQQQCP

1.1.1.1 Quoi ?

1.1.1.1.1 Qu'est-ce que c'est ?

Le projet se concentre sur le filtrage des signaux ECG pour éliminer les bruits indésirables, permettant ainsi un diagnostic précis des anomalies cardiaques.

1.1.1.1.2 Objectif du travail :

Développer des techniques de filtrage avancées et utiliser des modèles de classification pour améliorer l'interprétation des signaux ECG.

1.1.1.2 Qui ?

1.1.1.2.1 Qui est impliqué ?

Le projet est mené par une équipe pluridisciplinaire comprenant des chercheurs en ingénierie biomédicale, des professionnels de la santé, des data scientists et des spécialistes en traitement du signal.

1.1.1.3 Où ?

1.1.1.3.1 Où est-ce utilisé ?

La recherche se déroule dans des laboratoires universitaires, avec une application des méthodes dans des établissements de santé, notamment pour des dispositifs médicaux et la surveillance cardiaque.

1.1.1.4 Quand ?

1.1.1.4.1 Quand est-ce utilisé ?

Les techniques seront utilisées pour le diagnostic et la surveillance des patients, avec un développement prévu à court et moyen terme pour une application rapide des résultats.

1.1.1.5 Comment ?

1.1.1.5.1 Comment ça fonctionne ?

Le projet applique des techniques de filtrage pour réduire le bruit et utilise des modèles d'apprentissage automatique pour détecter les anomalies, afin d'améliorer le diagnostic et les résultats cliniques.

1.1.1.6 Pourquoi ?

1.1.1.6.1 Pourquoi est-ce important ?

Améliorer la qualité des signaux ECG est crucial pour la détection précoce des maladies cardiaques, réduisant ainsi les erreurs de diagnostic et optimisant les soins aux patients.

1.1.2 Définition du besoin

1.1.2.1 Contexte

Dans le domaine de la cardiologie, l'analyse précise des signaux ECG est cruciale pour le diagnostic et la surveillance des maladies cardiaques. Cependant, les signaux ECG sont souvent affectés par des bruits indésirables, tels que les interférences électriques et les artefacts de mouvement, qui peuvent altérer l'interprétation des données. Face à ces défis, il est essentiel de développer des méthodes robustes pour filtrer ces signaux et améliorer leur qualité.

1.1.2.2 Besoins identifiés

- **Amélioration de la qualité des signaux ECG :**
 - **Problème :** Les bruits indésirables compromettent la clarté des signaux ECG, rendant difficile une analyse précise.
 - **Besoin :** Mettre en place des techniques efficaces pour atténuer les bruits et garantir des signaux de haute qualité.
- **Développement de techniques de filtrage :**
 - **Problème :** Les méthodes de filtrage actuelles ne parviennent pas toujours à éliminer les artefacts sans altérer les données pertinentes.
 - **Besoin :** Concevoir des algorithmes de filtrage avancés capables de préserver les caractéristiques essentielles des signaux tout en éliminant les interférences.
- **Intégration de modèles de classification :**
 - **Problème :** Les anomalies dans les signaux ECG peuvent passer inaperçues sans une analyse approfondie, ce qui retarde le diagnostic.
 - **Besoin :** Utiliser des modèles d'apprentissage automatique pour détecter et classifier automatiquement les anomalies, améliorant ainsi la réactivité des soins
- **Support aux professionnels de la santé :**
 - **Problème :** Les médecins manquent souvent d'outils fiables pour interpréter les signaux ECG de manière rapide et précise.
 - **Besoin :** Fournir des outils numériques qui permettent une interprétation assistée des signaux, facilitant ainsi la prise de décision clinique.

➤ Diagramme bête à cornes

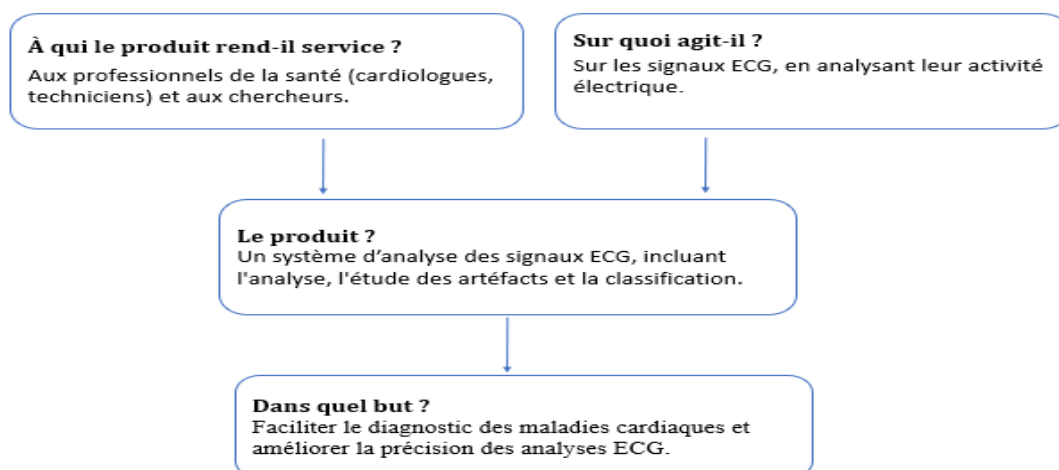


Figure 2: Diagramme bête à cornes

➤ Diagramme de pieuvre

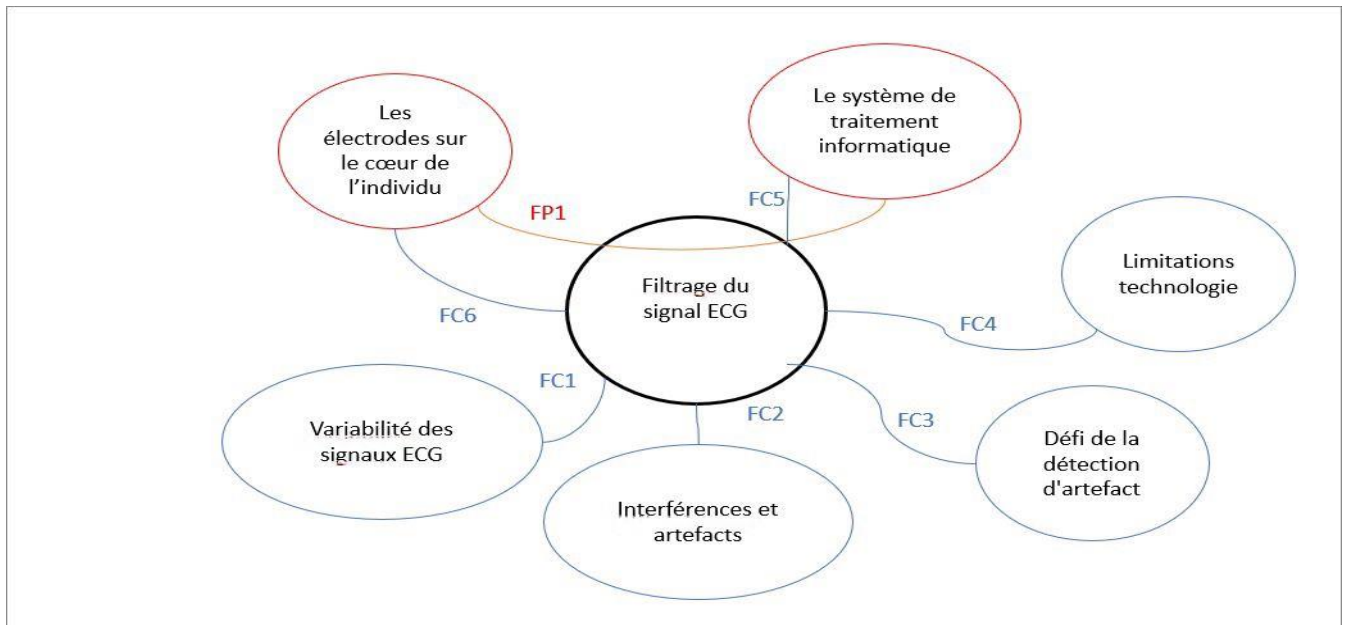


Figure 3: Diagramme de pieuvre

Cahier des charges

1.2.1 Contexte

Les signaux ECG (électrocardiogrammes) sont essentiels pour le diagnostic des pathologies cardiaques, mais ils sont souvent affectés par divers bruits et artefacts. Ces perturbations, qu'elles soient dues à des interférences musculaires ou à des mouvements, rendent l'interprétation des données difficile. Les méthodes de filtrage classiques montrent leurs limites face à cette complexité, ce qui souligne la nécessité d'approches plus avancées. L'apprentissage automatique émerge comme une solution prometteuse, capable de s'adapter aux particularités des signaux ECG tout en offrant une meilleure précision dans le filtrage des bruits.

1.2.2 Objectifs

Le projet vise à appliquer des techniques d'apprentissage automatique pour améliorer le filtrage des signaux ECG. L'objectif est de développer un modèle capable d'identifier et d'éliminer les artefacts tout en préservant les informations essentielles. Cela permettra d'améliorer la qualité des signaux ECG et de faciliter la détection d'anomalies cardiaques, contribuant ainsi à l'optimisation des diagnostics en cardiologie.

1.2.3 Tâches et Livrables

Le projet se compose de quatre tâches principales :

Installation et analyse initiale du jeu de données ECG avec une analyse statistique et des visualisations.

Étude des artefacts ECG et leur impact, accompagnée de recommandations de filtrage.

Représentation multidimensionnelle des signaux à travers des techniques d'exploration et d'évaluation de l'efficacité.

Construction de modèles de classification, développement et comparaison d'algorithmes pour sélectionner les meilleurs modèles.

Les ressources incluent Python et des bibliothèques comme NumPy, pandas, et scikit-learn, avec un focus sur la reproductibilité, les délais et la qualité des données, évalués par les rapports et la performance des modèles.

➤ Diagramme de Gantt

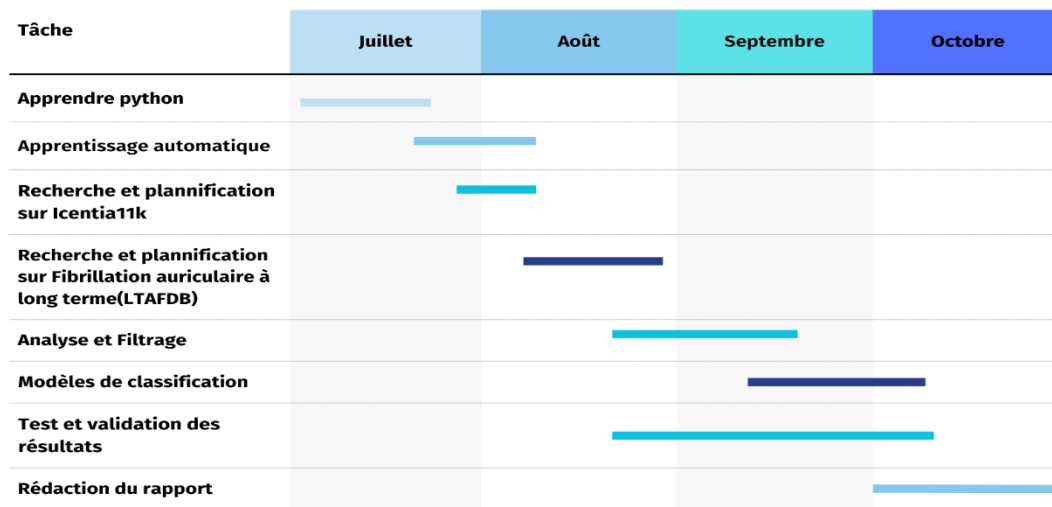


Tableau 1:Diagramme de Gantt

2 Etat de l'art

2.1 Logiciels Utilisés

2.1.1 Anaconda

Anaconda est une distribution open-source de Python facilitant la gestion des packages et des environnements virtuels pour la science des données. Elle inclut Jupyter Notebook, permettant d'écrire et d'exécuter du code de manière interactive.



Figure 4:logo d'Anaconda(Rivas, Jesús A. Anaconda, 2020.)

2.1.2 Spyder

Spyder est un environnement de développement intégré (IDE) pour Python, conçu pour la science des données. Il offre une interface conviviale avec des fonctionnalités d'édition de code, de débogage et d'analyse en temps réel, facilitant l'intégration des bibliothèques de calcul scientifique.



Figure 5:ogo de spyder(Grace, Tom. Spyder web. Simon and Schuster, 2017)

2.2 Intelligence artificielle

L'intelligence artificielle (IA) est un domaine de l'informatique visant à créer des systèmes imitant l'intelligence humaine, en apprenant, raisonnant et prenant des décisions. Elle inclut des techniques comme l'apprentissage automatique, le traitement du langage naturel et la vision par ordinateur. L'IA repose sur deux approches principales : l'une utilisant des données étiquetées pour prédire des résultats, l'autre explorant des données non étiquetées pour détecter des motifs cachés. Ces approches complètent l'IA et permettent de résoudre une variété de problèmes. Elles sont utilisées pour des tâches telles que la classification et la segmentation.

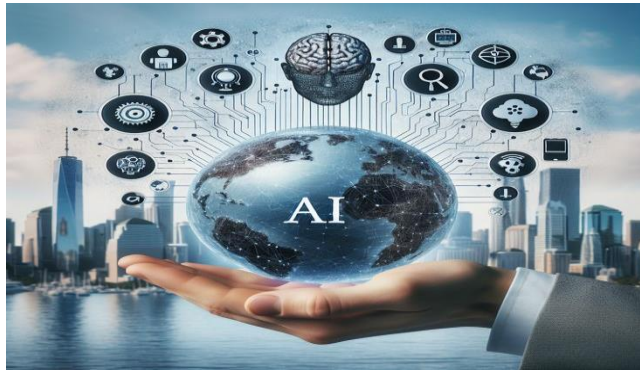


Figure 6: Intelligence artificielle (Sadin, Éric, 2018)

2.2.1 Apprentissage Supervisé

L'apprentissage supervisé consiste à entraîner un modèle sur un ensemble de données étiquetées, où chaque entrée est associée à une sortie connue. L'objectif est de prédire des résultats pour de nouvelles données en apprenant une fonction reliant les entrées aux sorties.

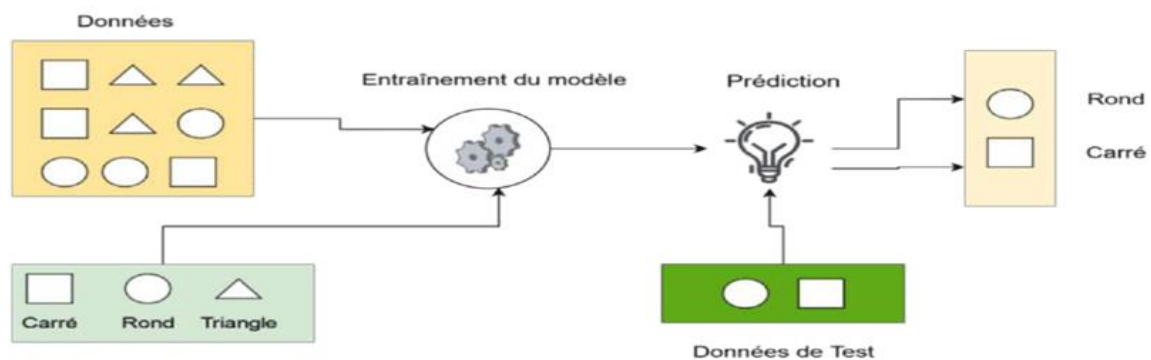


Figure 7: Apprentissage supervisé (ETTARID, Pr Mohamed, et al, 2021)

2.2.2 Apprentissage Non Supervisé

L'apprentissage non supervisé traite des données non étiquetées et vise à découvrir des motifs ou des structures sous-jacentes sans supervision. Les algorithmes de cette méthode, comme le clustering, cherchent à regrouper des données similaires ou à réduire la dimensionnalité.

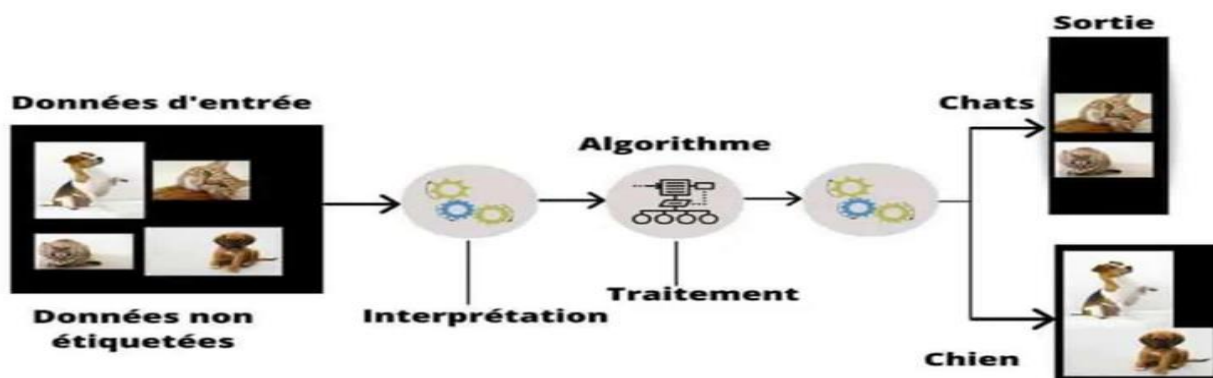


Figure 8: Apprentissage non supervisé (Jacopin, Eliott, et al., 2021)

2.3 Langage

Python est un langage de programmation interprété, haut niveau

et polyvalent, connu pour sa syntaxe claire et sa facilité d'apprentissage. Il est utilisé dans divers domaines, tels que le développement web, l'analyse de données, l'intelligence artificielle et l'automatisation. Python favorise une approche orientée objet et dispose d'une large bibliothèque standard, ce qui permet de simplifier le développement de programmes complexes.



Figure 9: Logo de python (Dorling, M., & White, D. (2015, February))

2.3.1 Les bibliothèques

2.3.1.1 NumPy : Outil essentiel pour les calculs numériques, NumPy permet de manipuler des tableaux multidimensionnels et d'effectuer des opérations mathématiques rapidement.

2.3.1.2 Pandas : Conçu pour la manipulation de données, Pandas facilite le traitement, l'analyse et la visualisation de données sous forme de tableaux via des DataFrames.

2.3.1.3 Matplotlib : Bibliothèque de visualisation puissante, Matplotlib permet de créer des graphiques variés pour représenter visuellement les données.

2.3.1.4 Scikit-learn : Scikit-learn fournit des outils d'apprentissage automatique pour la classification, la régression et le clustering, rendant l'implémentation de modèles prédictifs accessible et efficace.

2.4 Apprentissage automatique

L'apprentissage automatique, une sous-discipline de l'intelligence artificielle, permet aux systèmes d'apprendre et d'améliorer leurs performances sans programmation explicite. Il utilise des algorithmes pour analyser des données et est appliqué dans des domaines comme la reconnaissance vocale, la vision par ordinateur et la recommandation de contenus.

2.5 Le cœur

Le cœur est un organe musculaire central du système circulatoire, composé de quatre cavités : deux oreillettes (droite et gauche) et deux ventricules (droit et gauche). L'oreillette droite reçoit le sang désoxygéné provenant du corps, qui est ensuite pompé par le ventricule droit vers les poumons pour être oxygéné. Une fois le sang enrichi en oxygène, il revient vers l'oreillette gauche, puis est éjecté par le ventricule gauche vers le reste du corps.

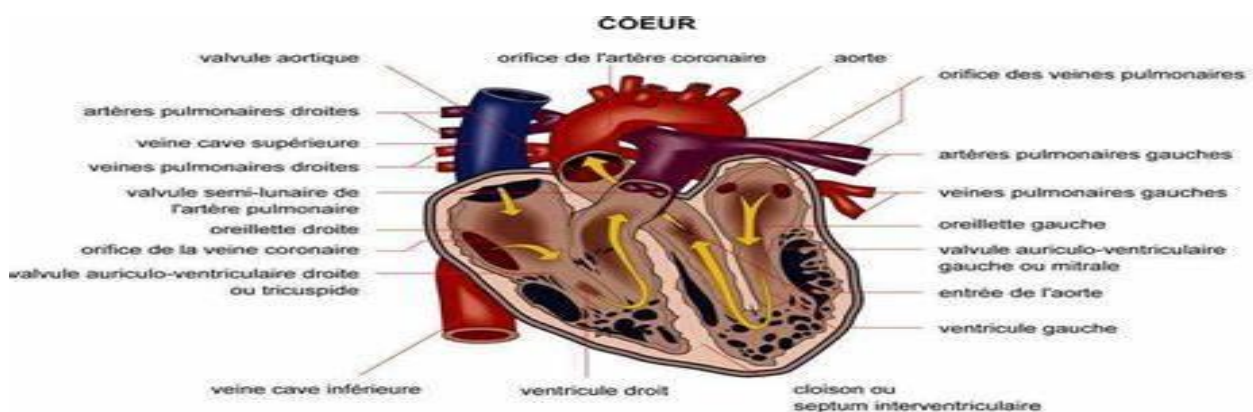


Figure 10: schéma d'anatomie du cœur (Netter, F. H. (2023))

2.6 Les électrodes

Dans le domaine médical, une électrode est un dispositif qui permet de mesurer ou de stimuler l'activité électrique des tissus biologiques. Par exemple, dans un électrocardiogramme (ECG), des électrodes sont placées sur la peau pour enregistrer les signaux électriques générés par le cœur, permettant ainsi de diagnostiquer diverses conditions cardiaques.

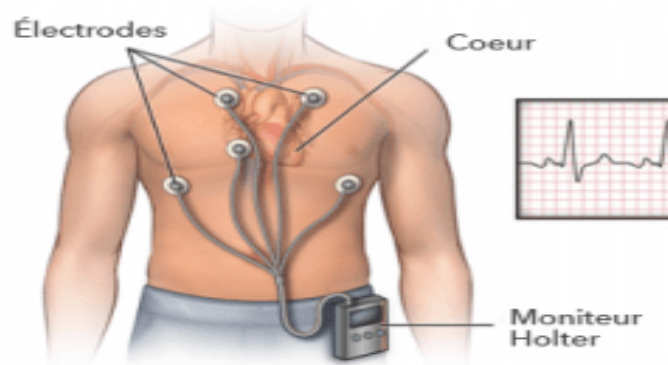


Figure 11:ECG électrode(Chlaihawi AA, Narakathu BB,

Les électrodes sont placées sur la peau du patient pour surveiller l'activité électrique du cœur ou du cerveau. Elles sont connectées à un appareil qui enregistre les signaux, permettant aux médecins d'analyser les données et d'évaluer l'état de santé du patient.

2.7 L'électrocardiogramme (ECG)

L'électrocardiogramme (ECG) est un test médical qui enregistre l'activité électrique du cœur sur une période donnée. Il permet de détecter des anomalies cardiaques, d'évaluer le rythme cardiaque et de diagnostiquer des conditions telles que l'infarctus du myocarde ou les arythmies.

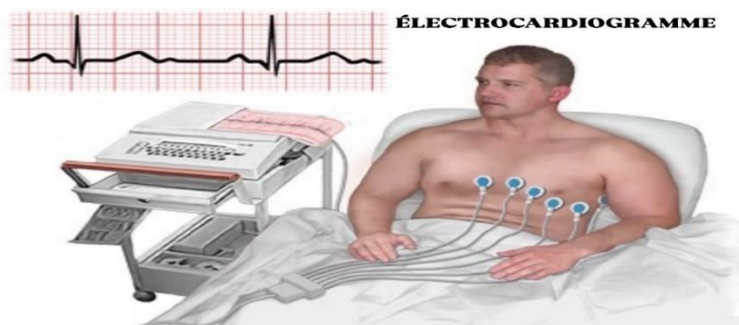


Figure 12:Électrocardiogramme(Berkaya, Selcan Kaplan, et al,2018)

2.8 Le signal ECG

Le signal ECG est une représentation graphique de l'activité électrique du cœur, composée de plusieurs ondes caractéristiques :

- **Onde P** : Elle représente la dépolarisation des oreillettes, indiquant leur contraction.
- **Complexe QRS** : Il correspond à la dépolarisation des ventricules, montrant leur contraction. C'est la partie la plus élevée et la plus large du tracé.
- **Onde T** : Elle indique la repolarisation des ventricules, signalant la phase de récupération après contraction.

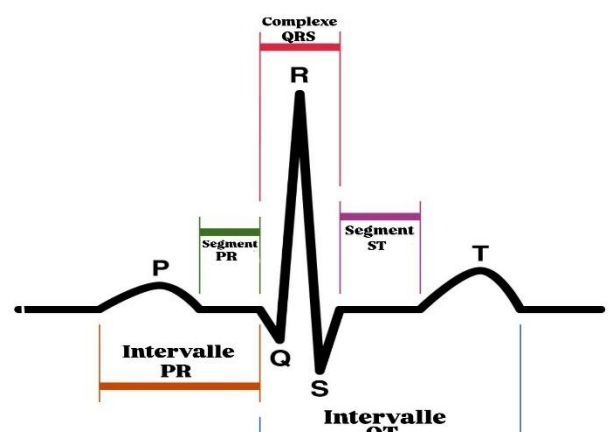


Figure 13:Signal ECG(Houssein, Essam H., Moataz Kilany, and Aboul Ella Hassanien,2017)

2.9 Dataset

2.9.1 Dataset iCentia11k

- **Composition du dataset** : 11 000 patients, 2 milliards de battements ECG étiquetés.
- **Détails techniques** : Signaux enregistrés à 16 bits, 250 Hz, avec une sonde à une dérivation, pendant jusqu'à 2 semaines.
- **Informations sur les patients** : Âge moyen des patients de $62,2 \pm 17,4$ ans.
- **Annotation des battements** : Vingt techniciens ont annoté chaque battement, en précisant son type (normal, prématurée) et le rythme (sinusal normal, fibrillation, etc.).
- **Objectif** : Destiné aux recherches en cardiologie.

2.9.2 Dataset Fibrillation Auriculaire à Long Terme (LTAfDB)

- **Volume de données** : Le dataset LTAfDB comprend des enregistrements ECG de 84 patients atteints de fibrillation auriculaire paroxystique ou soutenue.
- **Durée des enregistrements** : Les signaux capturent des épisodes prolongés de fibrillation, pouvant durer plusieurs heures ou jours.
- **Fréquence d'échantillonnage et résolution** : Les enregistrements sont échantillonnés à 128 Hz avec une résolution de 12 bits.
- **Annotations détaillées** : Contient des informations sur le début, la fin, la durée des épisodes de fibrillation auriculaire et les rythmes cardiaques.
- **Utilisation** : Idéal pour des études longitudinales, le développement de modèles prédictifs et la recherche clinique sur la fibrillation auriculaire.

2.10 Fibrillation Atriale (Atrial Fibrillation - AF)

2.10.1 Définition : La fibrillation atriale est un trouble du rythme cardiaque où les oreillettes battent rapidement et irrégulièrement, perturbant le flux sanguin vers les ventricules.

2.10.2 Caractéristiques :

- **Fréquence cardiaque** : Très irrégulière, souvent au-dessus de 100 bpm, parfois bien plus rapide.
- **Origine** : Les impulsions électriques proviennent de foyers multiples dans les oreillettes, perturbant la contraction coordonnée.
- **ECG** : Absence d'ondes P régulières, tracé irrégulier avec complexes QRS irréguliers, fréquence atriale > 300 bpm.

2.10.3 Complications et Symptômes :

- **Symptômes fréquents** : Palpitations, fatigue.
- **Risques associés** : Augmentation du risque d'AVC et d'insuffisance cardiaque si non traitée.

2.11 Fibrillation Auriculaire à Long Terme (LTAf)

Définition : La fibrillation auriculaire à long terme (LTAf) est une forme persistante d'arythmie cardiaque, fréquente chez les personnes âgées et celles souffrant de maladies cardiovasculaires.

Symptômes : Les patients peuvent ressentir des palpitations, de la fatigue, des essoufflements et parfois des douleurs thoraciques, avec une variation d'intensité.

Causes : L'hypertension, le diabète, les antécédents familiaux, le tabagisme, l'obésité et d'autres conditions cardiaques, comme les valvulopathies, sont des causes courantes de LTAF.

2.12 La circulation sanguine

2.12.1 Circulation dans un cœur normal :

La circulation sanguine est fluide et coordonnée, permettant une oxygénation efficace du sang dans les poumons et sa distribution au corps.

2.12.2 Circulation en cas de fibrillation auriculaire :

Les oreillettes battent de manière chaotique, perturbant le remplissage des ventricules, ce qui rend la circulation moins efficace.

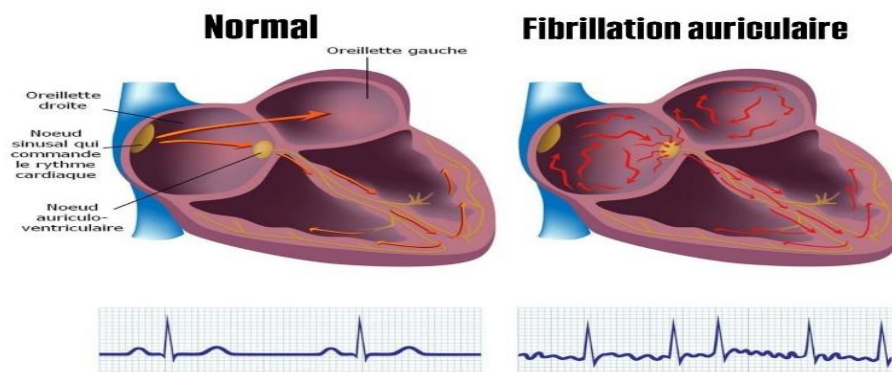


Figure 14: La circulation sanguine chez une personne normale et autre atteinte de la fibrillation auriculaire (MENIF, H., et al, 2019)

Diagnostic des Anomalies Cardiaques

Artefacts

- Définition : Les artefacts sont des perturbations non liées à l'activité cardiaque qui altèrent le signal ECG.
- Origine : Ils proviennent de mouvements du patient, de problèmes de contact avec les électrodes ou d'interférences électriques.
- Impact : Ils peuvent masquer ou déformer les caractéristiques du signal ECG, rendant son interprétation difficile.

Bruit

- Définition : Le bruit est une variation aléatoire dans le signal ECG, souvent perçue comme un fond indésirable.
- Origine : Il résulte de fluctuations électriques, d'interférences radio ou d'autres sources de perturbations aléatoires.

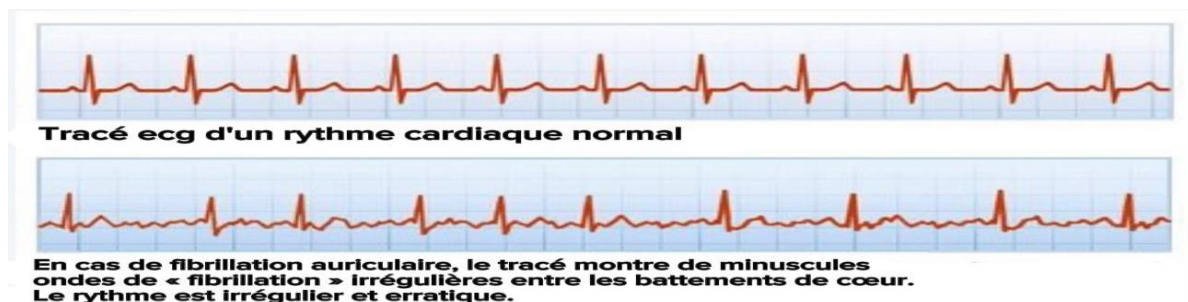


Figure 15: Rythme sinusal normal vs. fibrillation auriculaire (Dobrev, D., et al, 2018)

2.14 Analyse du signal ECG

2.14.1 Installation des Données

2.14.1.1 Collecte des Données :

Les données ECG ont été collectées à l'aide d'électrodes et enregistrées au format .dat.

2.14.1.2 Transfert des Données :

Le fichier .dat a été transféré vers un ordinateur pour une analyse ultérieure.

2.14.1.3 Configuration de l'Environnement :

L'analyse a été réalisée dans Spyder avec l'installation de bibliothèques telles que NumPy, Pandas, Matplotlib et SciPy.

2.14.1.4 Exploration des Données :

La structure des données a été explorée, en examinant les colonnes et les types de valeurs présents (temps, amplitude, etc.).

2.15 Prétraitement des Données

2.15.1 Nettoyage des Données :

Le nettoyage a consisté à identifier et traiter les valeurs manquantes ou aberrantes pour assurer la précision des données.

2.15.2 Filtrage :

Des filtres ont été appliqués, incluant des filtres Passe-Bas, Passe-Haut et Passe-Bande pour réduire les bruits et artefacts.

2.15.3 Représentation multidimensionnelle :

Plusieurs techniques comme l'ACP, l'ACI, et la Transformée en Ondelettes ont été utilisées pour visualiser et analyser les relations entre les données.

2.16 Segmentation du Signal

Le signal ECG a été divisé en segments pour faciliter l'analyse de chaque cycle cardiaque.

2.17 Extraction des Caractéristiques

2.17.1 Importance de l'Intervalle RR :

L'intervalle RR a été utilisé comme paramètre clé pour analyser le rythme et la santé cardiaque.

2.17.2 Amplitudes des Ondes :

P, QRS et T ont été analysées pour extraire des informations essentielles sur le signal ECG.

2.17.3 Analyse et Visualisation :

Les caractéristiques extraites ont été visualisées pour faciliter l'interprétation du signal ECG.

2.18 Classification

2.18.1 Types de Classification :

Des classifications binaires et multiclasse ont été utilisées pour distinguer entre différents types d'anomalies cardiaques.

2.18.2 Méthodes de Classification : Plusieurs algorithmes ont été appliqués, y compris KNN, SVM, Arbres de Décision, Forêts Aléatoires, XGBoost et Light GBM.

2.19 Métriques de Performance

- 2.19.1 Précision** : Mesure la proportion de prédictions correctes.
2.19.2 Rappel (Sensibilité) : Mesure la proportion de cas positifs correctement identifiés.
2.19.3 F1-Score : Moyenne harmonique de la précision et du rappel pour équilibrer les deux.
2.19.4 AUC-ROC : Évalue la capacité du modèle à distinguer entre les classes.
2.19.5 Accuracy (Exactitude) : Proportion d'exemples correctement classés.
2.19.6 Vrais Positifs (VP) : Nombre de cas où l'anomalie a été correctement prédite.
2.19.7 Vrais Négatifs (VN) : Nombre de cas où l'absence d'anomalie a été correctement prédite.
2.19.8 Faux Positifs (FP) : Nombre de cas où une anomalie a été incorrectement prédite.
2.19.9 Faux Négatifs (FN) : Nombre de cas où une anomalie n'a pas été identifiée.

➤ Pour approfondir l'analyse, il faut évaluer le modèle avec des métriques telles que :

Taux d'erreur	$\frac{FP+FN}{TP+FP+TN+FN}$
Précision	$\frac{TP}{TP + FP}$
Rappel	$\frac{TP}{TP + FN}$
F1-score	$2 * \frac{precision*rappel}{precision+rappel}$
Spécificité	$\frac{TN}{TN + FP}$

Tableau 2:Équations des Métriques Clés

Ces métriques offriront une vue d'ensemble de la performance du modèle et indiqueront des axes d'amélioration.

2.20 Interprétation Clinique et Rapport des Résultats :

Les résultats obtenus après l'analyse doivent être examinés par les médecins pour une interprétation clinique appropriée. Cela permet de poser un diagnostic et d'orienter les recommandations pour un suivi ou traitement. Un rapport détaillant les conclusions, diagnostics et recommandations sera rédigé pour documenter le processus et les résultats de l'analyse.

2.21 Fréquence des Ondes ECG :

Chez une personne en bonne santé, la fréquence des ondes ECG varie généralement entre 60 et 100 battements par minute, avec des valeurs typiques pour chaque onde.

2.22 Les Ondes du Signal ECG

2.22.1 Onde P :

Représente la dépolarisation des oreillettes, avec une fréquence typique entre 0.5 Hz et 5 Hz. Cette onde est importante pour détecter les anomalies de la conduction auriculaire.

2.22.2 Complexe QRS :

Correspond à la dépolarisation ventriculaire, avec une fréquence typique entre 5 Hz et 15 Hz. Cette fréquence est cruciale pour identifier les problèmes de conduction ventriculaire et les anomalies du rythme cardiaque.

2.22.3 Onde T :

Reflète la repolarisation ventriculaire, avec une fréquence oscillante entre 0.5 Hz et 5 Hz, similaire à celle de l'onde P. Elle est importante pour évaluer la récupération du cœur après une contraction.

3 Travaux réalisés

J'ai initialement utilisé ICENTIA11K pour mon étude, mais en raison de sa grande taille, j'ai décidé de me concentrer sur LTAADB pour faciliter la gestion des données et approfondir l'analyse du filtrage des signaux ECG.

- Pipeline

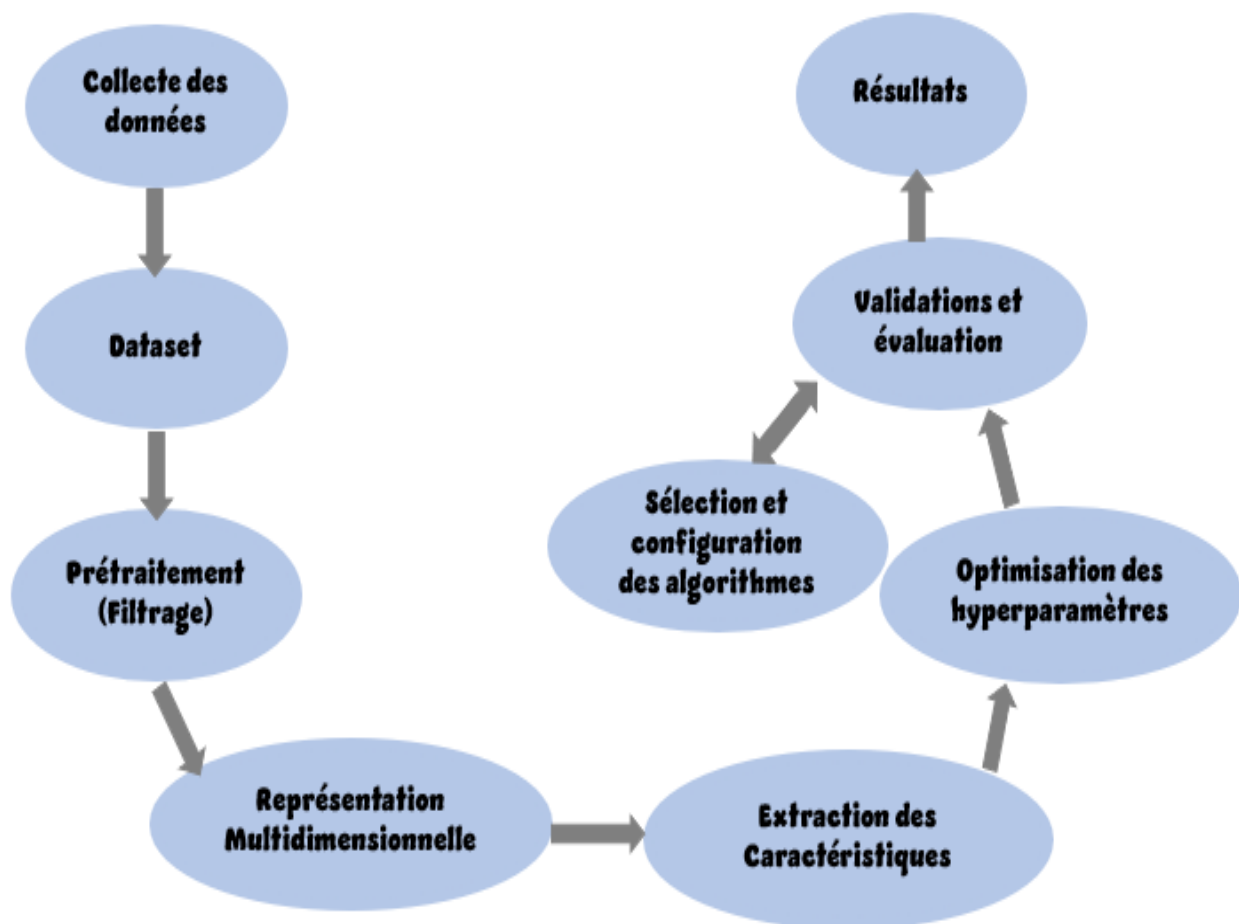


Figure 16: Pipeline du travail

Analyse des Signaux Bruts des Patients

3.1.1 : Introduction à l'Analyse des Données

Après avoir installé l'ensemble de données et effectué une analyse initiale, l'objectif est d'examiner le fichier .dat contenant les informations de 11 000 patients.

3.1.2 : Focus sur le Premier Patient

L'analyse commence par le premier patient. En utilisant Spyder, nous allons extraire et visualiser le signal brut enregistré sur une période de 24 à 25 heures.

3.1.3 : Identification des Artefacts

Pour ce premier patient, l'attention se porte sur les 500 premières secondes du signal pour détecter les artefacts. Cette étape est essentielle pour garantir la qualité des données et préparer l'analyse pour l'ensemble des patients.

3.1.4 : Analyse Approfondie

Après l'examen des 500 premières secondes, l'analyse se poursuit avec les 300 secondes suivantes, après le pic du signal, pour affiner la compréhension des données collectées.

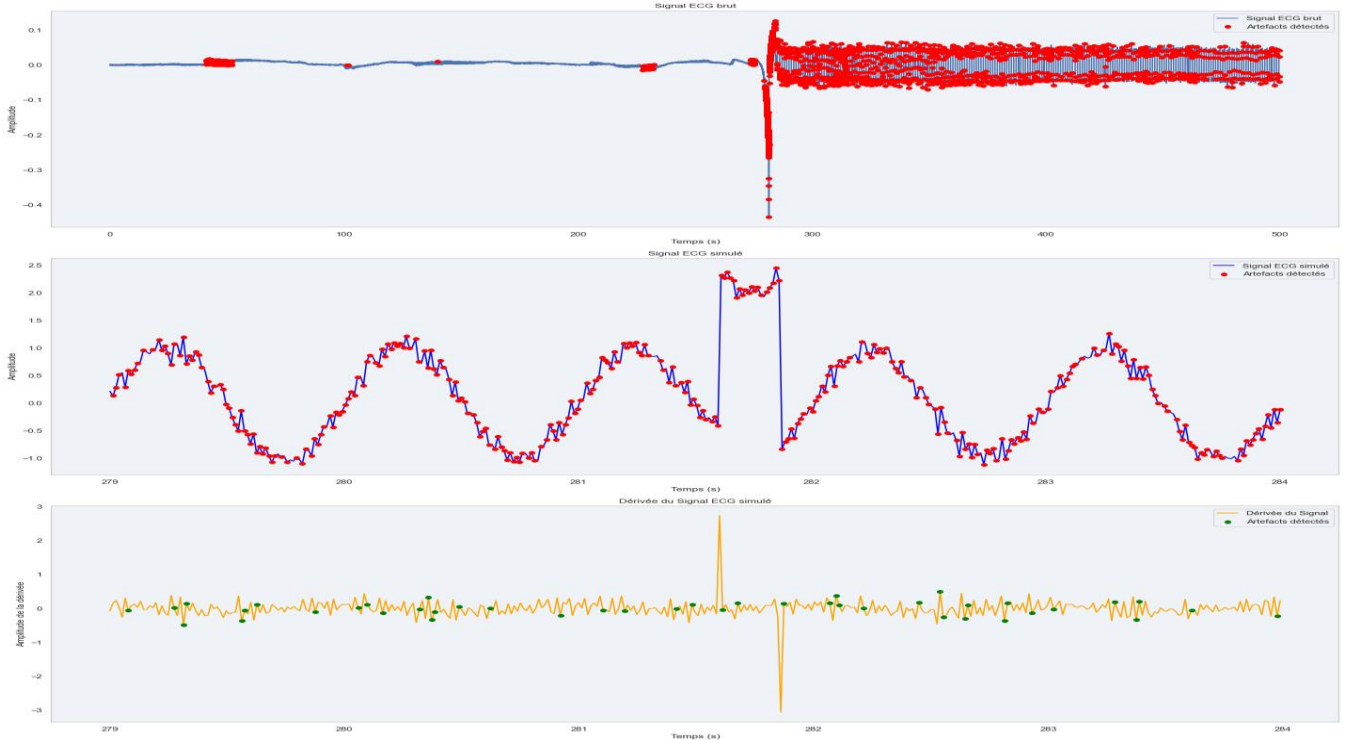


Figure 17: Détection d'artefacts dans un signal ECG

- **Signal ECG Brut** : Ce graphique montre un enregistrement de l'activité cardiaque, qui est initialement horizontal en raison de nombreux artefacts présents. Le bruit, provenant de mouvements et d'interférences, complique l'interprétation du signal sur une durée de 500 secondes.
- **Signal ECG Simulé** : Ce graphique présente une portion d'un signal ECG, créé avec une onde sinusoïdale et du bruit pour simuler les interférences réelles. L'analyse de 279 à 284 secondes permet d'identifier des artefacts, avec des événements marqués en rouge.
- **Dérivée du Signal ECG Simulé** : La dérivée met en évidence les variations rapides du signal, en se concentrant sur les complexes QRS liés à la contraction des ventricules sur une durée de 0,059 seconde.
- **Détection des Artefacts** : Des seuils sont définis pour distinguer les perturbations non cardiaques. Pour le signal brut, le seuil est à 0,01, et pour le simulé, à 0,3, afin d'optimiser la détection des anomalies.

Ces graphiques illustrent l'analyse des signaux ECG et soulignent l'importance de détecter les anomalies pour le diagnostic cardiaque.

3.2 Prétraitement des Signaux ECG

3.2.1 Application des Filtres

Il est crucial de déterminer lequel des filtres restitue le mieux la forme du signal ECG, afin de l'utiliser de manière optimale dans mon étude, tout en mettant en évidence les ondes P, QRS et T.

➤ Observation des Artefacts avant le filtrage durant 300 secondes:

Au cours de cette simulation, des artefacts se sont manifestés sous forme de pics anormaux dans le signal.

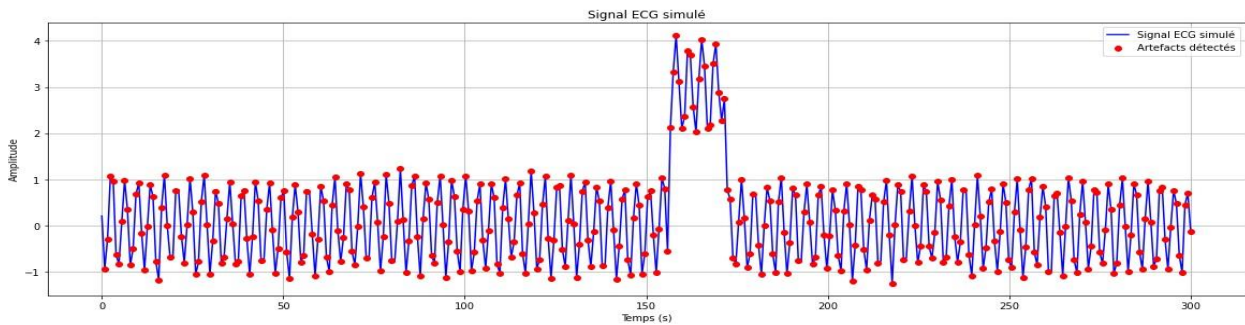


Figure 18: Signal ECG simulé avec des artefacts

➤ Après l'application du Filtre Passe-Bande:

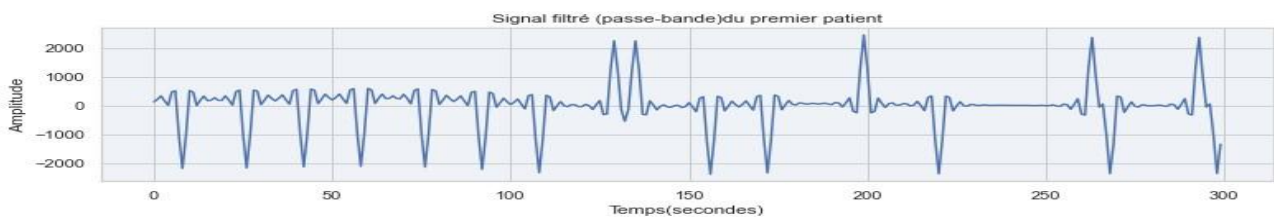


Figure 19: Signal ECG filtré chez le premier patient en utilisant le filtre passe bande

- **Atténuation des Artefacts** : Le filtre passe-bande a atténué les artefacts du signal ECG, visibles en rouge sur le tracé (figure 19).
- **Amélioration du Signal** : La filtration a clarifié le signal en éliminant les perturbations et en mettant en évidence les caractéristiques essentielles du rythme cardiaque.
- **Efficacité de la Filtration** : L'approche a prouvé son efficacité dans la purification du signal, rendant l'analyse plus fiable.
- **Caractéristiques du Signal Filtré** :
 - Le signal filtré montre des oscillations régulières caractéristiques des signaux biologiques.
 - L'utilisation de NumPy et matplotlib a permis un traitement numérique efficace.
 - Le filtrage passe-bande cible des fréquences spécifiques, isolant les ondes P, T et le complexe QRS.
 - Les fréquences de coupure ($F1 = 0,5 \text{ Hz}$ et $F2 = 15 \text{ Hz}$) ont permis de capturer efficacement les ondes et de réduire le bruit.
 - Ce filtrage fournit une représentation claire de l'activité ventriculaire et des ondes P et T, essentielles pour le diagnostic cardiaque.

➤ Focalisation sur une Plage Temporelle Spécifique :

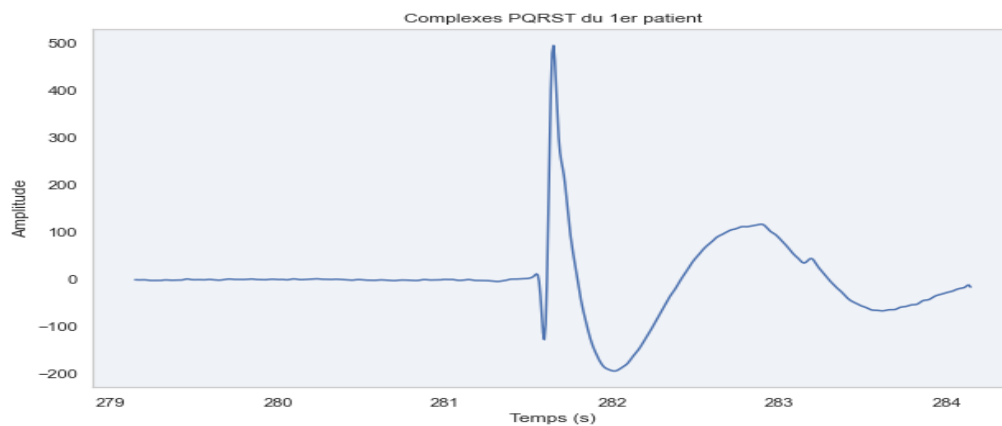


Figure 20: Signal cardiaque complexe PQRST filtré à l'aide d'un filtre passe-bande

- **Figure** : Illustration du signal ECG après l'application d'un filtre passe-bande avec des fréquences de coupure à 0,5 Hz et 15 Hz, mettant en évidence les complexes PQRST.
- **Complexe P** : Représente la dépolarisation des oreillettes, visible par de petites oscillations avant le pic principal.
- **Complexe QRS** : Caractérisé par un pic principal et des oscillations autour de celui-ci, symbolisant la dépolarisation des ventricules, généralement le plus prononcé.
- **Onde T** : Apparaît après le complexe QRS, sous forme d'une oscillation plus large et plus lente, représentant la repolarisation des ventricules.
- **Impact du Filtre** : Le filtre a réduit le bruit de fond en éliminant les fréquences indésirables, améliorant ainsi la clarté des signaux pertinents.
- **Importance de l'Analyse** : Cette approche permet une analyse plus précise de l'activité cardiaque, cruciale pour le diagnostic médical.

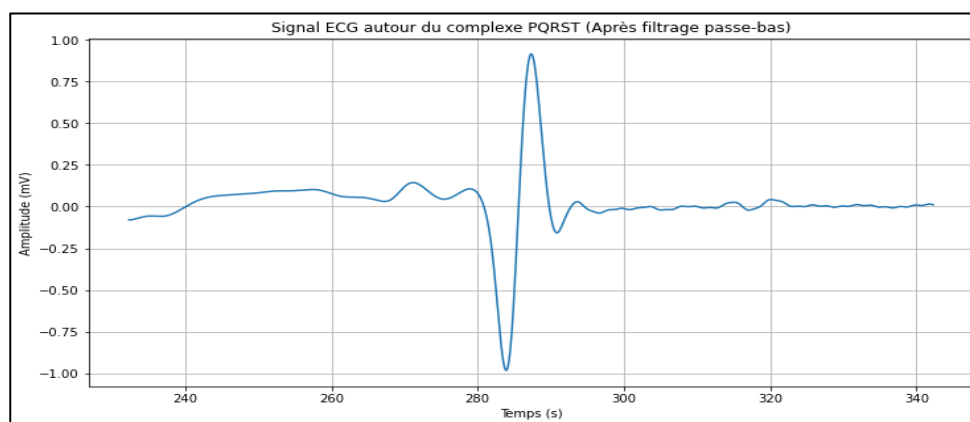


Figure 21: Signal cardiaque complexe PQRST filtré à l'aide d'un filtre passe-bas

- **Présentation du Graphe** : Le graphe montre une portion d'un ECG filtré à 0,5 Hz.
- **Identification des Ondes** : Les ondes P, Q, R, S et T correspondent à des activités électriques spécifiques du cœur :
 - **Onde P** : Dépolarisation des oreillettes.
 - **Complexe QRS** : Dépolarisation des ventricules.
 - **Onde T** : Repolarisation des ventricules.
- **Utilisation de 0,5 Hz** : Cette fréquence est cruciale pour isoler les variations lentes et éliminer les bruits de fond.
- **Difficultés d'Identification** : L'échelle de temps et le filtrage rendent l'identification des ondes P et Q difficile.
- **Observations Notables** : L'onde R est très prononcée, suivie d'une onde T normale.

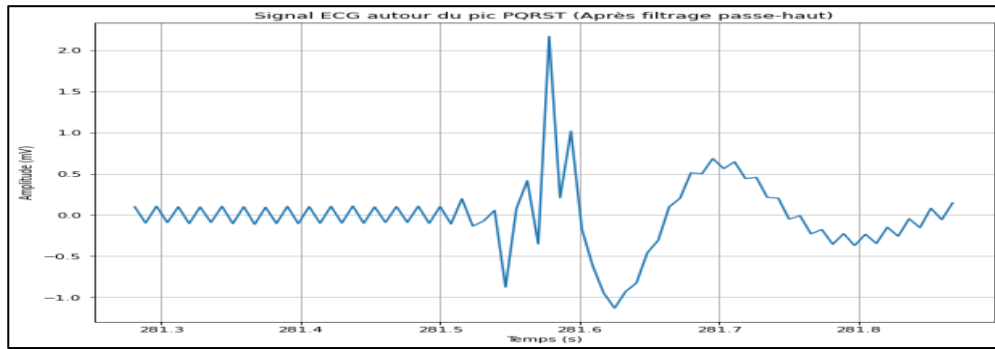


Figure 22: Signal cardiaque complexe PQRST filtré à l'aide d'un filtre passe-haut

■ Caractéristiques du Graphe :

- Les ondes P, QRS et T sont clairement visibles, représentant l'activité électrique du cœur.
- Le filtre a éliminé le bruit de basse fréquence, permettant une meilleure visualisation des ondes.
- Le filtrage a accentué les pics du signal, les rendant plus nets.

■ Justification de la Fréquence de Coupure à 5 Hz :

- Le filtre passe-haut élimine les bruits de basse fréquence (mouvement musculaire, activité respiratoire), améliorant la qualité du signal.
- Les ondes P, QRS et T, ayant des fréquences supérieures à 5 Hz, sont préservées.
- La fréquence de coupure de 5 Hz équilibre bien la résolution temporelle et l'atténuation du bruit.

3.2.2 Représentation multidimensionnelle :

La représentation multidimensionnelle des signaux ECG, après filtrage, améliore la visualisation et l'analyse des ondes cardiaques, facilitant ainsi la détection d'anomalies pour un diagnostic précis.

3.1.2.1 Analyse en Composantes Principales (ACP):

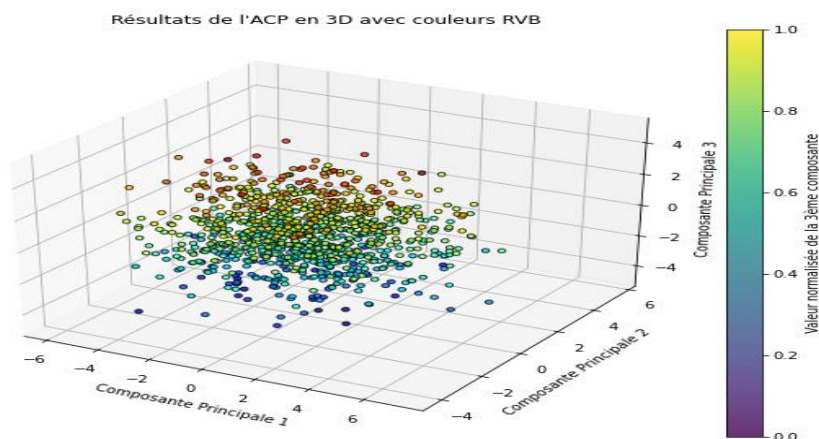


Figure 23: Visualisation 3D des résultats d'ACP

Le graphique en figure 24 est une représentation 3D obtenue par Analyse en Composantes Principales (ACP), avec trois axes représentant les premières composantes principales (CP1, CP2, CP3).

• Axes :

- **Axe X** : CP1, direction de la variance maximale.
- **Axe Y** : CP2, direction orthogonale à CP1.

- **Axe Z** : CP3, orthogonal aux deux autres axes.
- **Visualisation** : Chaque point représente une observation, projetée sur les composantes principales. La couleur des points, basée sur CP3, montre la distribution des données.
- **Importance** : L'ACP aide à identifier des structures et tendances dans les données, essentielle pour l'analyse des signaux ECG.
- **Échelle de Couleurs** : Elle va du rouge/jaune (haute amplitude) au bleu/violet (faible amplitude), permettant de visualiser des groupes ou tendances en fonction de CP3.

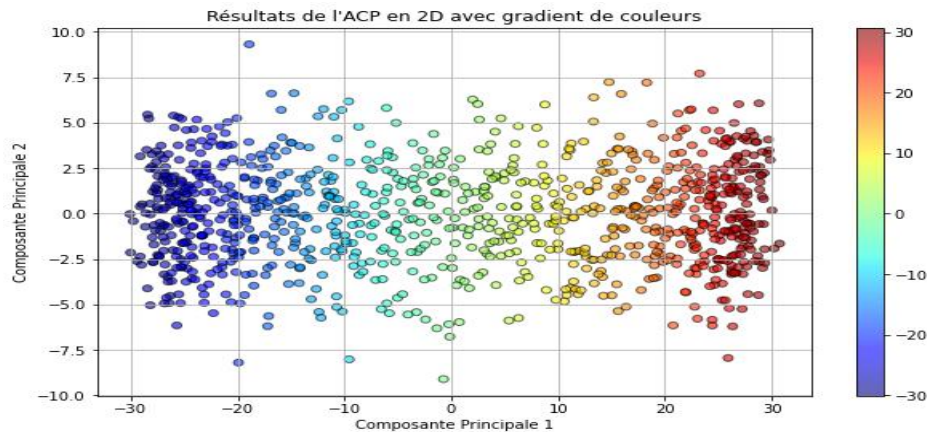


Figure 24: Visualisation 2D des résultats de l'Analyse en Composantes Principales (ACP)

- **Objectif et Fonction de l'Analyse** : La figure montre les résultats 2D d'une Analyse en Composantes Principales (ACP) appliquée à des données ECG simulées. L'ACP réduit la dimensionnalité des données tout en conservant l'essentiel de l'information.
- **Données et Méthodologie** :
 - **Génération des Signaux** : 750 signaux simulés, chacun avec 6000 échantillons à 128 Hz.
 - **Caractéristiques des Signaux** : Chaque signal est une sinusoïde de 1 Hz, avec du bruit gaussien ajouté pour simuler des variations ECG.
 - **Importance du Nombre d'Échantillons** : 6000 échantillons (environ 46,9 secondes) sont choisis pour analyser les caractéristiques temporelles des signaux.
- **Application de l'ACP** : L'ACP transforme les signaux en deux composantes principales qui capturent la variance maximale, simplifiant les données tout en conservant l'information significative.
- **Visualisation des Résultats** :
 - **Graphique 2D** : Les résultats sont affichés en 2D, avec un codage couleur selon la première composante principale (bleu pour faible amplitude, rouge pour forte amplitude).
 - **Interprétation** : Le gradient de couleurs aide à interpréter les relations entre les observations.
- **Importance de l'Analyse** : L'ACP facilite l'identification des tendances, des clusters et améliore la compréhension des signaux ECG simulés.

3.1.2.2 Analyse en Composantes Indépendantes (ACI) :

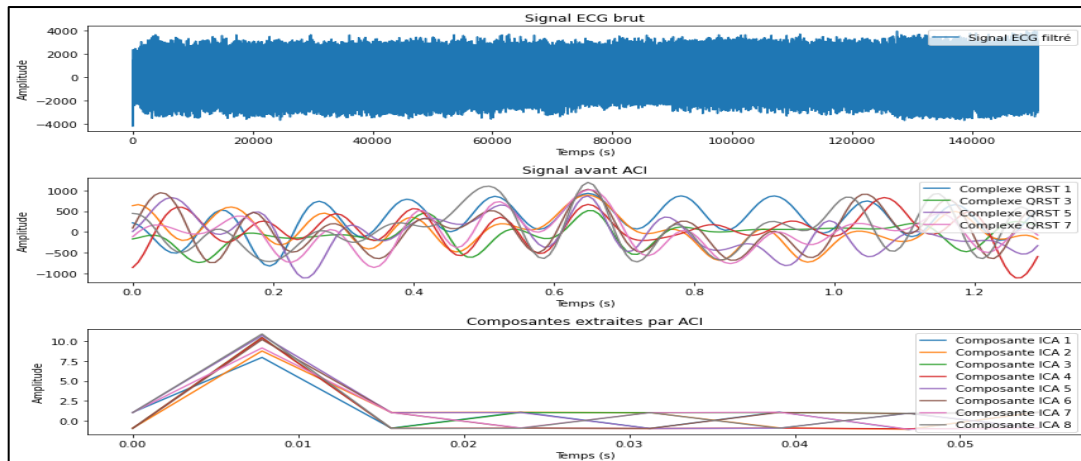


Figure 25: Composantes Indépendantes Extraites par (ACI) des Signaux ECG

■ Présentation des Graphiques :

- Le premier graphique montre le signal ECG brut, souvent bruité et rempli d'artefacts.
- Le second graphique représente le signal avant l'Analyse en Composantes Indépendantes (ACI), où les complexes QRS sont segmentés et mieux alignés.
- Le troisième graphique montre les composantes extraites par l'ACI, qui dissocient les signaux mélangés en sources indépendantes.

■ Détails Méthodologiques :

- **Taille de la Fenêtre** : Environ 1,3 seconde pour capturer un segment complet du complexe QRS.
- **Sélection des Pics R** : Seuls les huit premiers pics R sont retenus pour l'analyse.

■ Analyse des Résultats :

- **Signal Brut** : Le signal brut présente des fluctuations aléatoires dues au bruit et aux artefacts.
- **Composantes 1 à 3** : Elles ressemblent au complexe QRS, suggérant qu'elles capturent l'activité cardiaque principale.
- **Composantes 4 à 8** : Elles montrent des formes variées et des amplitudes faibles, probablement dues aux artefacts ou à d'autres sources de signal, comme la respiration ou des interférences électriques.

3.1.2.3 Transformée de Fourier :

Nous avons utilisé la Transformée de Fourier pour obtenir le spectre de fréquence du signal ECG filtré.

3.1.2.3.1 Spectre de fréquence:

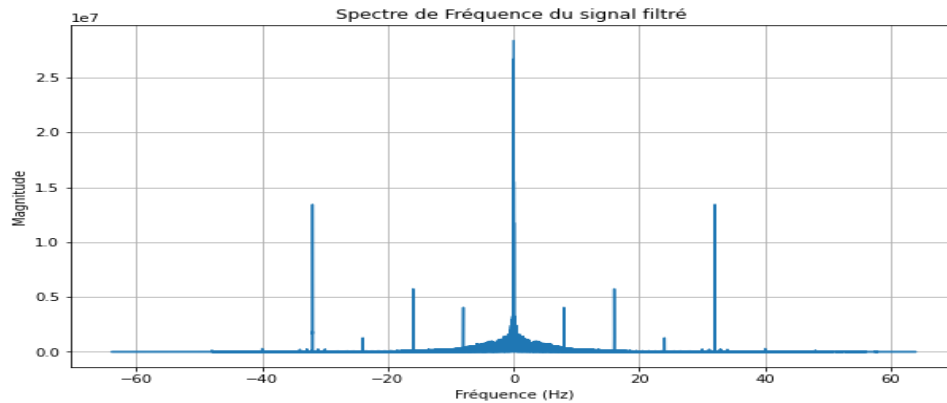


Figure 26: Spectre de Fréquence du Signal ECG Filtré par Transformée de Fourier

- **Analyse du Spectre de Fréquence :**
Le spectre de fréquence du signal ECG montre des pics correspondant à la fréquence fondamentale et ses harmoniques, représentant l'activité cardiaque.
- **Identification des Pics Significatifs :**
Le pic principal, entre 0,5 et 3 Hz, correspond à la fréquence cardiaque. Le spectre est tracé de -60 Hz à 60 Hz, incluant des fréquences négatives qui sont des artefacts.
- **Visualisation des Amplitudes :**
L'échelle logarithmique de l'axe des ordonnées permet de mieux visualiser la large gamme d'amplitudes.
- **Interprétation :**
Le spectre fournit des informations sur la fréquence cardiaque (BPM), le rythme (sinusal ou arythmique), et la qualité du signal. Des pics irréguliers ou du bruit peuvent indiquer des artefacts ou des interférences.

3.1.2.3.2 Signal ECG avec la Densité Spectrale de Puissance (DSP)

Pour approfondir l'analyse du signal ECG, nous avons appliqué la Densité Spectrale de Puissance (DSP) après la transformée de Fourier.

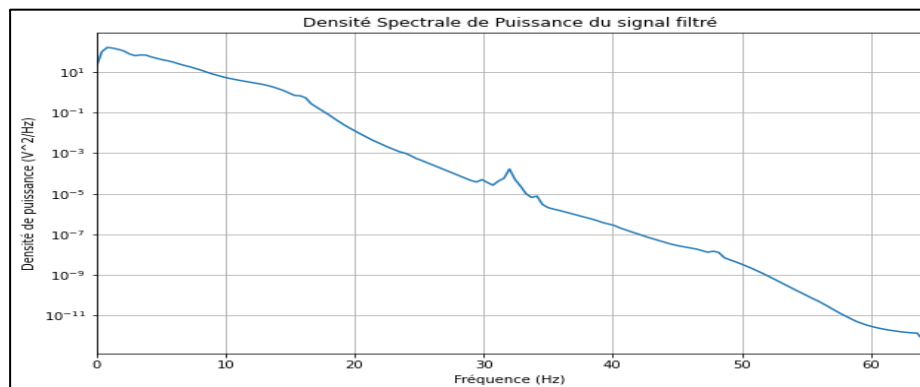


Figure 27: Densité Spectrale de Puissance du Signal ECG-Pic Dominant à 1 Hz

- **Distribution de Puissance (DSP) :**
Le graphique montre une distribution de puissance avec un pic autour de 1 Hz, correspondant à la fréquence cardiaque fondamentale.
- **Estimation de la Fréquence Cardiaque :**
Ce pic dominant permet d'estimer la fréquence cardiaque en battements par minute (BPM), par exemple 1 Hz correspond à 60 BPM.
- **Composants Additionnels :**
Des pics à des fréquences plus élevées représentent des éléments du signal ECG, comme les ondes P et T.
- **Niveau de Bruit de Fond :**
Le bruit de fond affecte la clarté du signal, un faible bruit de fond indiquant un signal plus propre.

- **Plage de Fréquences :**
L'axe des abscisses couvre de 0 à 60 Hz, une plage pertinente pour l'analyse ECG.
- **Facteurs d'Interprétation :**
L'interprétation de la DSP dépend de l'âge et de l'état de santé du patient, influençant le spectre cardiaque.
- **Conclusion :**
La DSP permet d'estimer la fréquence cardiaque, d'évaluer la qualité du signal et d'identifier des anomalies cliniques.

3.1.2.4 Transformée de Fourier à Court Terme (STFT) :

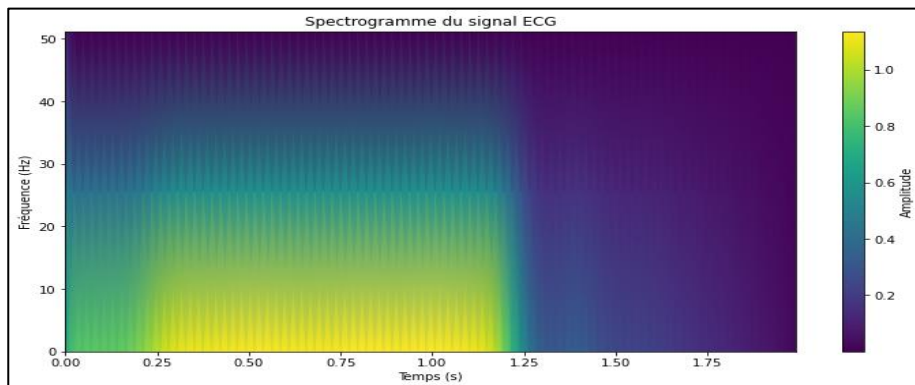


Figure 28: Spectrogramme des Résultats de la STFT chez le Premier Patient

- **Présentation du Spectrogramme :**
Le spectrogramme montre la distribution de l'énergie du signal ECG en fonction du temps et de la fréquence.
- **Interprétation des Couleurs :**
Les couleurs indiquent l'intensité du signal : les teintes chaudes (rouge, jaune) représentent une amplitude élevée, tandis que les teintes froides (bleu, violet) montrent des amplitudes faibles.
- **Composantes Cardiaques :**
 - Basses fréquences : liées à l'onde T et à l'isoélectrique.
 - Fréquences intermédiaires : reflètent le complexe QRS.
 - Hautes fréquences : souvent associées à du bruit ou des artefacts.
- **Évolution Temporelle :**
Les variations de couleur permettent d'observer l'évolution de l'activité cardiaque et détecter des événements significatifs, comme des arythmies.

3.1.2.5 Transformer en ondelette

3.1.2.5.1 Transformée en Ondelette Continue (CWT) :

L'utilisation de la transformée en ondelettes continue (CWT) en 2D et 3D vise à analyser les signaux ECG de manière détaillée en capturant à la fois les variations temporelles et fréquentielles, ce qui permet une meilleure identification des caractéristiques cardiaques et des anomalies. Voici les résultats :

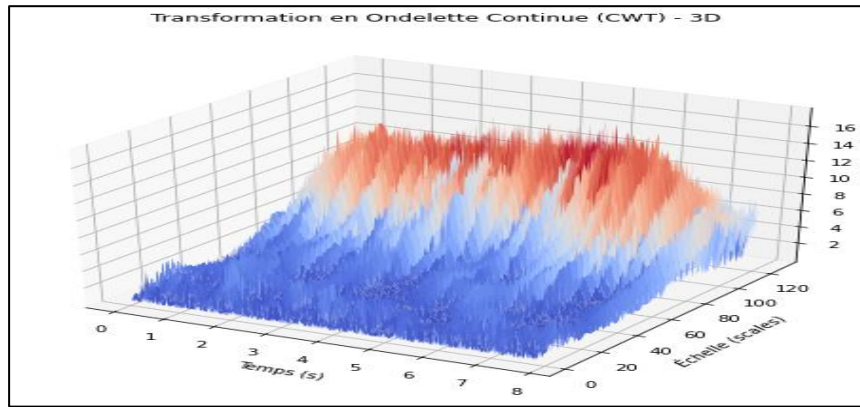


Figure 29: Scalogramme 3D des Résultats de la Transformée en Ondelette Continue (CWT)

- Le **scalogramme** représente la Transformée en Ondelette Continue (CWT) avec les axes suivants :
 - **Axe des temps (x)** : domaine temporel du signal.
 - **Axe des échelles (y)** : échelle de fréquence (les échelles élevées correspondent aux basses fréquences et vice versa).
 - **Axe d'amplitude (z)** : amplitude des coefficients d'ondelettes.
- **Observations clés :**
 - **Localisation temps-fréquence** : montre où l'énergie du signal est concentrée, avec des pics représentant des zones de forte énergie à des moments et fréquences spécifiques.
 - **Intensité des couleurs** : les couleurs chaudes (rouge, orange) indiquent une énergie élevée, tandis que les couleurs froides (bleu) indiquent une énergie faible.
 - **Évolution du contenu fréquentiel** : un déplacement vers des échelles plus élevées montre une diminution de la fréquence.
 - **Événements transitoires** : des pics pointus signalent des changements soudains ou des explosions d'énergie dans le signal.

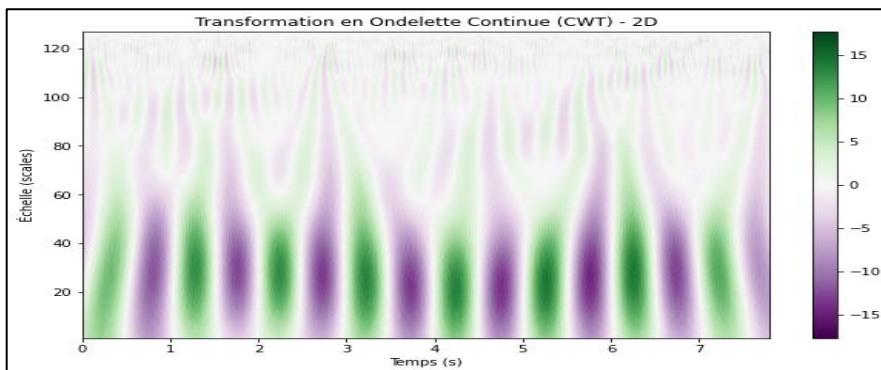


Figure 30: Scalogramme 2D des Résultats de la Transformée en Ondelette Continue (CWT)

- **Axes:**
 - **Axe horizontal** : évolution du signal dans le temps.
 - **Axe vertical** : fréquences analysées, avec les échelles élevées représentant les basses fréquences et les échelles basses les hautes fréquences.

- **Couleurs** : illustrent l'amplitude des coefficients d'ondelettes, les couleurs chaudes (vert clair) indiquant une forte énergie, tandis que les couleurs froides (violet foncé) signalent une faible énergie.

Analyse Visuelle:

- **Bandes verticales régulières** : indiquent une composante périodique forte. La distance entre elles est liée à la période.
- **Variabilité** : les bandes varient en intensité et position, indiquant des changements dans l'amplitude ou la fréquence.
- **Régions claires** : montrent une concentration d'énergie à certaines fréquences (par exemple, une bande horizontale claire pour les basses fréquences).
- **Variations brusques** : signalent des événements transitoires comme des changements rapides d'amplitude.

Interprétation des Couleurs:

- **Vert clair** : forte énergie à une fréquence et un moment spécifique.
- **Violet foncé** : faible énergie ou absence de signal.

Conclusion : Le scalogramme révèle que le signal est principalement constitué de composantes fréquentielles périodiques, avec des variations légères d'amplitude et de fréquence au fil du temps.

3.1.2.5.2 Transformée en Ondelette Discrète (DWT) :

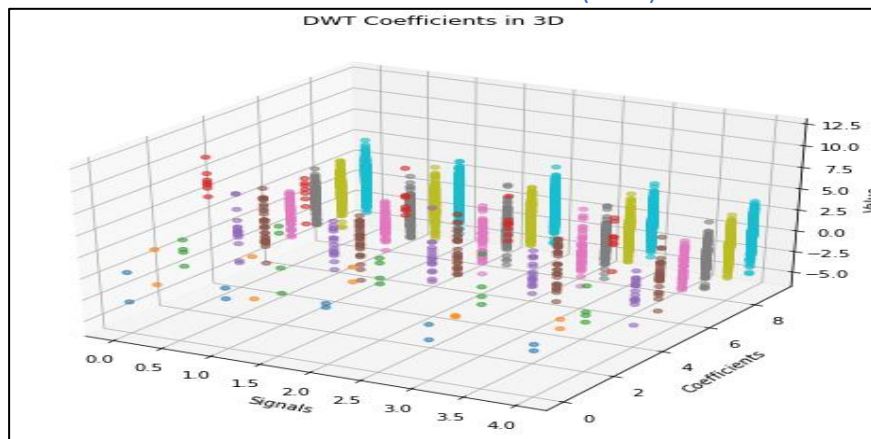


Figure 31: Visualisation 3D des Coefficients de la Transformée en Ondelette Discrète (DWT)

- **Axes:**
 - **Axe des signaux** : représente les différents signaux analysés.
 - **Axe des coefficients** : montre les coefficients d'approximation et de détail à différentes échelles.
 - **Axe des valeurs** : indique l'amplitude des coefficients.
- **Couleurs:**
 - **Chaud (rouge, orange, jaune)** : valeurs positives élevées, indiquant une forte énergie.
 - **Froid (bleu, violet)** : valeurs négatives élevées.
 - **Neutre (vert, gris)** : valeurs proches de zéro.

Importance de l'Analyse :

- La compréhension des types d'ondelettes utilisées et de l'analyse des signaux à différents niveaux de décomposition est essentielle pour une analyse précise.
- Les **types d'ondelettes** utilisés :
 - 'db2' (Daubechies) pour la transformation 2D.
 - 'sym2' (symlet) pour la transformation 3D.

Processus de Transformation :

- Les signaux sont reconstruits après transformation pour vérifier leur intégrité.

Analyse en Composantes Principales (ACP) :

- Appliquée pour réduire la dimensionnalité et permettre une visualisation efficace des résultats.

Conclusion : Cette approche combinée enrichit l'analyse des signaux ECG, offrant une vue détaillée des variations temporelles et fréquentielles, facilitant une meilleure compréhension des données.

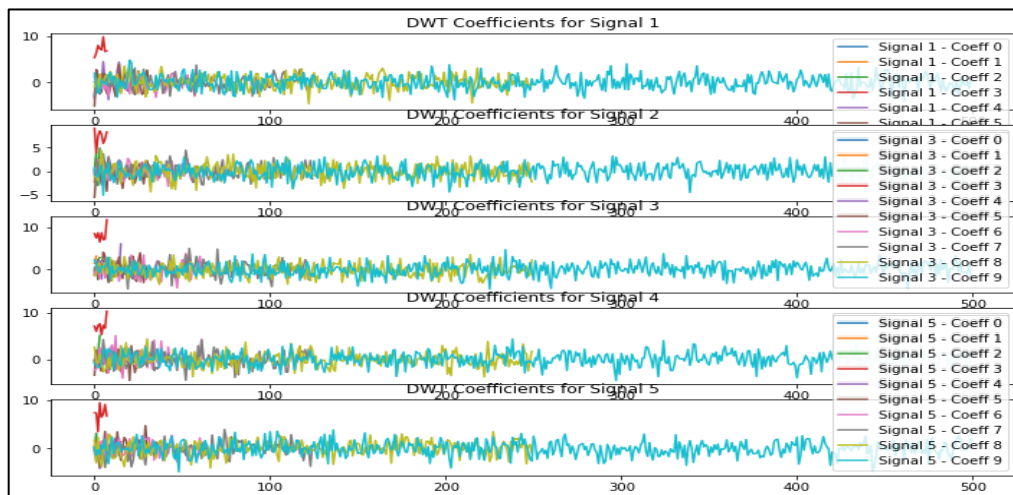


Figure 32: Représentation des coefficients d'ondelettes pour 5 signaux

Le **graphique** présente les résultats d'une **transformation en ondelettes discrètes (DWT)** appliquée à cinq signaux distincts, décomposant chaque signal en plusieurs sous-bandes de fréquences.

- **Caractéristiques des Coefficients:**
- **Variabilité des Amplitudes** : Les amplitudes des coefficients varient significativement d'un signal à l'autre et entre les sous-bandes, montrant des contenus fréquentiels variés.
- **Pics et Creux** : Les pics et creux dans les courbes des coefficients indiquent des détails ou des événements transitoires dans les signaux.
- **Évolution Temporelle:**
- L'analyse temporelle des coefficients permet de visualiser l'évolution des caractéristiques fréquentielles des signaux, offrant ainsi une perspective dynamique de leur structure.

3.1.2.5.3 Transformée en Ondelette Multi-résolution (MRWT) :

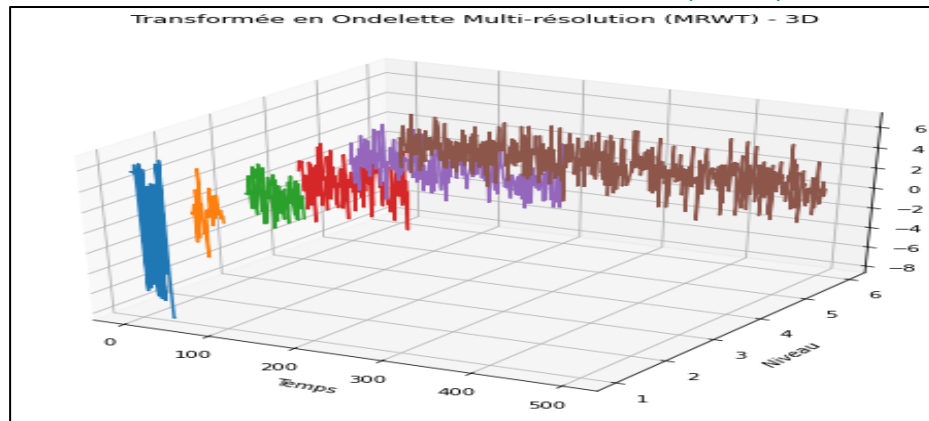


Figure 33: Représentation 3D des Coefficients de la Transformée en Ondelette Multi-résolution

- **Axes:**
 - **Axe des temps** : indique le moment précis du coefficient.
 - **Axe des niveaux** : représente l'échelle de résolution, où des niveaux plus élevés offrent une meilleure résolution en fréquence.
 - **Axe des valeurs** : correspond à l'amplitude numérique du coefficient.
- **Rôle des Couleurs:**
 - Les couleurs sont utilisées pour visualiser les variations d'amplitude des coefficients :
 - **Teintes vives** (rouge, bleu, etc.) : signalent des amplitudes élevées.
 - **Couleurs neutres** (gris, blanc) : indiquent des amplitudes faibles.
- **Interprétation des Couleurs:**
 - **Couleurs chaudes** (rouge, orange, jaune) : indiquent des valeurs positives élevées, reflétant une forte énergie.
 - **Couleurs froides** (bleu, violet) : signalent des valeurs négatives élevées.
 - **Couleurs neutres** (vert, gris) : montrent des valeurs proches de zéro, représentant une faible énergie.

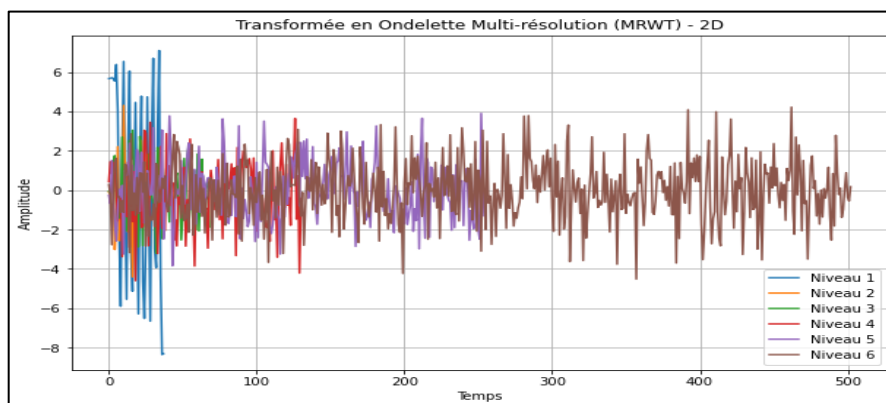


Figure 34: Représentation 2D des Coefficients de la Transformée en Ondelette Multi-résolution

La **Transformée en Ondlette Multi-résolution (MRWT) en 2D** est une méthode avancée pour analyser les signaux numériques, qui décompose un signal en différentes fréquences et échelles.

- **Axes :**

- **Axe des abscisses** : représente le **temps**, montrant l'évolution du signal.
- **Axe des ordonnées** : mesure l'**amplitude**, indiquant l'intensité du signal à un instant donné pour un niveau de résolution spécifique.
- **Visualisation des Niveaux de Résolution** :
 - **Niveaux supérieurs** : capturent les **détails fins** du signal.
 - **Niveaux inférieurs** : exposent les **tendances générales** du signal.
- **Interprétation des Résultats** :
 - **Variations d'Amplitude** : des **fluctuations importantes** indiquent des changements significatifs dans le signal, tandis que des **amplitudes faibles** révèlent des périodes de stabilité.
 - **Fréquences des Oscillations** : les courbes à **haute fréquence** montrent des détails subtils, tandis que celles à **basse fréquence** offrent une vue d'ensemble des tendances.
- **Conclusion** :

La **MRWT** est un outil puissant pour analyser et visualiser les signaux ECG, mettant en évidence l'importance des techniques d'ondelettes dans le traitement des données biologiques.

3.1.2.5.4 Transformée en Ondelette à Borne(TWB) :

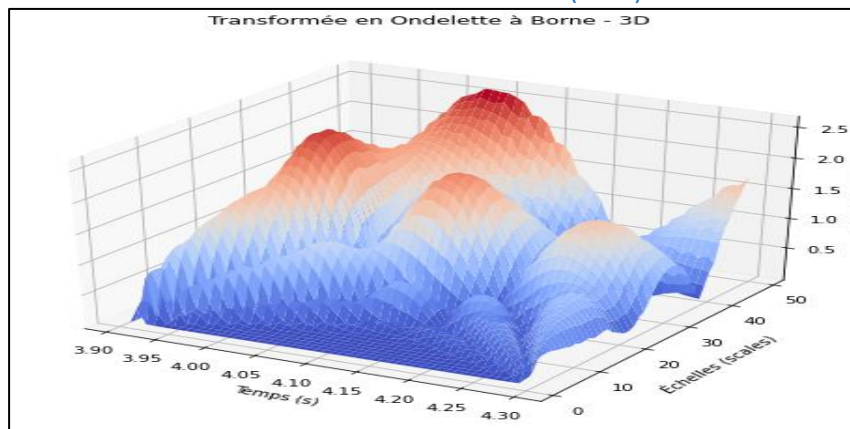


Figure 35: Scalogramme 3D de la Transformée en Ondelette à Borne

- **Axes**:
 - **Axe des x** : représente le **temps**.
 - **Axe des y** : indique l'**échelle**.
 - **Axe des z** : mesure la **magnitude** des coefficients.
- **Efficacité de la TWB**:
 - La TWB permet une **localisation précise** dans le temps et la fréquence. Les **coefficients concentrés** dans certaines zones du plan temps-échelle montrent des zones d'**énergie significative** à des moments et fréquences spécifiques.
- **Compromis entre Résolutions**:
 - La TWB établit un compromis entre **résolution temporelle** et **résolution fréquentielle** :
 - **Échelles basses** : offrent une **meilleure résolution temporelle**.
 - **Échelles élevées** : fournissent une **résolution fréquentielle plus précise**.
- **Analyse des Caractéristiques du Signal** :

- Les **pics et vallées** dans le graphique représentent des caractéristiques du signal.
- L'**amplitude des coefficients** indique la force de ces caractéristiques à un moment donné.
- L'analyse révèle que le signal contient des **variations complexes temps-fréquence**, suggérant des comportements signalétiques riches et variés.

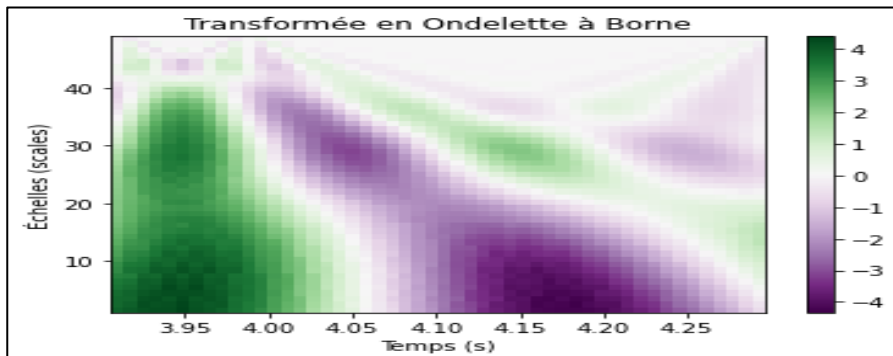


Figure 36: Scalogramme 2D de la Transformée en Ondelette à Borne

- **Axes:**
 - **Axe horizontal** : représente le **temps**, couvrant une courte durée de quelques secondes.
 - **Axe vertical** : indique l'**échelle**, inversement proportionnelle à la fréquence (les échelles plus élevées correspondent à des fréquences plus basses).
- **Échelle de Couleurs:**
 - **Couleurs froides** (bleu, violet) : indiquent des **valeurs faibles**, représentant une faible énergie du signal ou peu de détails, avec le violet pour les valeurs les plus faibles.
 - **Couleur verte** : représente des **valeurs intermédiaires à élevées**, signalant des zones d'**activité accrue** dans le signal, soit par amplitude, soit par variations fréquentielles.
- **Interprétation et Utilité :**

Les couleurs aident à visualiser et interpréter les **zones d'intérêt** du signal. Elles permettent une **quantification des amplitudes**, facilitant la comparaison entre scalogrammes et mettant en évidence les similitudes et différences des signaux analysés.

3.2.3 Extraction des Caractéristiques

- **Phase de Filtration et Objectifs**

Après la filtration, l'extraction des caractéristiques vise à simplifier et enrichir les données tout en conservant les informations pertinentes. Cette étape permet d'éliminer le bruit et de se concentrer sur les éléments significatifs pour les analyses.

- **Détection des Pics**

L'extraction des caractéristiques se concentre sur la détection des **pics** dans le signal ECG :

Amplitudes des Pics : Mesurent l'intensité des événements détectés.

Temps entre les Pics : Permet d'analyser la dynamique du signal.

Importance pour l'Évaluation Cardiaque

Les pics R dans un Signal ECG

Signal filtré avec passe-bande et pics détectés

Amplitude

Temps (secondes)

— Signal filtré
● Pics détectés

The graph displays a high-frequency signal with a clear periodic pattern. The signal is filtered, showing a clean waveform with sharp peaks. Red dots are placed at the top of each peak, indicating the detected peaks. The x-axis ranges from 0.0 to 20.0 seconds, and the y-axis ranges from -100 to 300 amplitude units.

- **Signal Filtré (ligne noire)** : Le signal filtré affiche une forme d'onde sinusoïdale régulière, avec une amplitude variant autour d'une valeur moyenne proche de zéro, résultant d'un processus de filtrage efficace.
- **Détection des Pics** : Les **pics détectés** (points rouges) marquent les maxima locaux du signal filtré, représentant les points de valeurs les plus élevées.
- **Filtrage**

- **Importance des Pics R**

- **Analyse Électrocardiographique** : Les pics R mettent en évidence les complexes PQRST, qui sont cruciaux pour l'évaluation des rythmes cardiaques.
- **Mesures Précises** : Cette détection permet de mesurer les amplitudes et les intervalles entre chaque complexe, facilitant ainsi l'identification d'éventuelles anomalies et améliorant l'interprétation clinique des dynamiques cardiaques.

Temps entre les complexes (en secondes)	[0.55055055, 0.55055055]
---	--

Temps des complexes (en secondes)	[279.55055055, 280.1011011, 280.65165165, 281.2022022, 281.75275275, 282.3033033, 282.85385385, 283.4044044, 283.95495495, 284.50550551, 285.05605606, 285.60660661, 286.15715716, 286.70770771, 287.25825826, 287.80880881, 288.35935936, 288.90990991, 289.46046046]
Amplitudes des pics détectés	[365.31320734, 354.98030909, 340.5843061, 332.6233126, 327.86929565, 336.04411658, 331.98310121, 336.02971381, 338.56359247, 343.97090284, 342.08831542, 339.23146846, 334.3473556, 329.85847432, 330.49928301, 327.00917898, 335.96010629, 355.81762869]
Fréquence cardiaque en BPM	108.98181818181813

3.2.3.2 Résultats d’entraînement

- **Observations générales:**
 - **Mesure des Intervalles et Interprétation**
 - Les **temps entre les complexes** sont mesurés à **0,5505 secondes**, constants sur 18 intervalles, indiquant un rythme cardiaque **régulier** et stable.
 - Cette constance suggère que le cœur fonctionne à un **rythme stable**, un signe de **santé cardiaque**.
 - La **fréquence cardiaque** correspond à **108,98 BPM**, ce qui est relativement élevé mais dans les limites normales pour un individu actif.
 - **Analyse des Complexes**
 - **Temps des complexes** : Varient entre **279,55 et 289,46 secondes**, montrant des moments de détection précis, sans fluctuation importante.
 - **Dynamique du rythme cardiaque** : L'écart constant entre les complexes suggère une activité cardiaque régulière, indicative d'une **réponse physiologique adéquate** à un exercice ou une stimulation.

Tableau 3:Évolution temporelle d'un phénomène oscillatoire

- Les **amplitudes des pics** varient entre **327,87 et 365,31** reflétant des fluctuations naturelles dans l'intensité des événements détectés.
 - Les pics plus élevés, dépassant **360**, suggèrent des **contractions cardiaques fortes**, tandis que des pics plus bas pourraient nécessiter une **évaluation cardiaque** supplémentaire.
 - **Fréquence Cardiaque et Conséquences Cliniques**
 - La fréquence cardiaque de **108,98 BPM** est significativement élevée et pourrait suggérer une **fibrillation auriculaire**, un trouble du rythme cardiaque, associé à un risque de **palpitations, accident vasculaire cérébral** et autres complications.

- Bien que dans une plage normale pour un adulte actif, cette fréquence élevée pourrait signaler un **déséquilibre électrique** nécessitant une attention médicale.

3.2.4 Modélisation en Apprentissage Automatique

- L'utilisation de **modèles d'apprentissage automatique** (comme KNN, SVM, forêts aléatoires, XGBoost, et GBM) est essentielle pour :
 - Obtenir des **prédictions précises**.
 - Gérer la **complexité des données** et les bruits.
 - Fournir une **interprétabilité** des résultats, facilitant ainsi l'analyse et la prise de décisions médicales éclairées.

3.2.4.1 K-Nearest Neighbors (KNN):

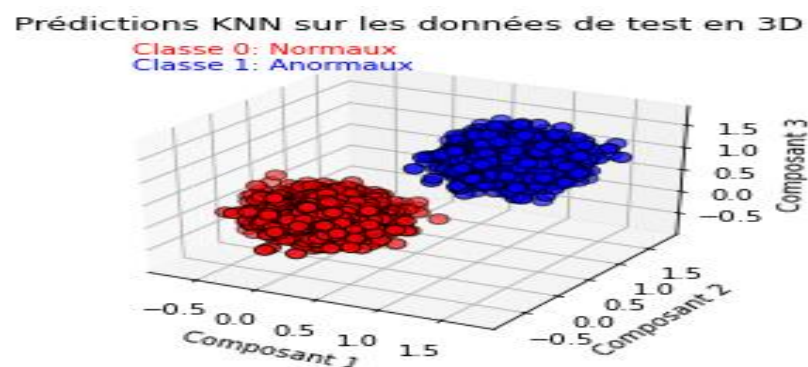


Figure 38: Séparation des classes normales et anormales par KNN dans un espace 3D

• Introduction au Graphique

Le graphique présente les prédictions d'un **classifieur KNN** (k-Nearest Neighbors) après une réduction de dimensionnalité via **Analyse en Composantes Principales (ACP)**.

• Rôle de l'ACP

L'**ACP** réduit la dimensionnalité des données tout en préservant leur structure principale, facilitant ainsi leur analyse et l'évaluation des résultats.

• Données et Classification

- Le graphique illustre deux classes de données : "**normales**" et "**anormales**", issues de **6000 échantillons** répartis presque équitablement.
- Le paramètre **k = 3** signifie qu'un échantillon est classé comme "**normal**" (Classe 0) si **2 des 3 voisins** les plus proches sont également classés comme normaux. Cela repose sur un mécanisme de **vote majoritaire**.

• Résultats de Classification

La séparation entre les classes est généralement claire, mais des **chevauchements** existent, représentant des cas difficiles à classer. Une gestion de la **variance** dans les données aide à améliorer cette distinction.

• Évaluation de la Performance

Des **métriques** comme l'**accuracy** et le **F1-score** sont utilisées pour évaluer l'efficacité du modèle. L'importance du choix du paramètre **k** dans le KNN est soulignée.

• Normalisation et Outils

La **normalisation des données** est recommandée pour améliorer la performance du modèle. Des fonctions spécifiques sont utilisées pour générer et visualiser les données de manière efficace.

3.2.5 Métriques de Classification

○ Résultats Généraux

Mesure	Valeur
Accuracy	0.97
Pourcentage de signaux Normaux	49.42%
Pourcentage de signaux Anormaux	50.58%

Tableau 4:Résultats Généraux des Performances du Modèle

○ Rapport de Classification

Classe	Précision	Rappel	F1-Score	Support
Normaux	0.97	0.97	0.97	588
Anormaux	0.98	0.97	0.97	612
Accuracy			0.97	1200
Macro avg	0.97	0.97	0.97	1200
Weighted avg	0.97	0.97	0.97	1200

Tableau 5:Rapport de Classification des Classes

○ Validation Croisée

Mesure	Valeur
Moyenne des Scores de Validation Croisée	0.97

Tableau 6:Moyenne des Scores de Validation Croisée

Les résultats montrent une performance exceptionnelle du modèle KNN avec une **accuracy** de 0,97, indiquant que 97 % des prédictions sont correctes. Voici les points clés :

- **Accuracy** : Score de 0,97, confirmant l'efficacité du modèle à distinguer les signaux normaux et anormaux.
- **Pourcentages des Classes** : 49,33 % des signaux sont normaux et 50,67 % sont anormaux, avec une distribution presque équilibrée, renforçant la robustesse du modèle.
- **Support pour Chaque Classe** : 588 échantillons pour les signaux normaux et 612 pour les anormaux. La légère prépondérance des signaux anormaux n'affecte pas les résultats.
- **Métriques de Classification:**
 - Pour les signaux normaux : précision de 96 %, rappel de 97 %, F1-score de 0,97.
 - Pour les signaux anormaux : précision de 97 %, rappel de 97 %, F1-score de 0,97.
- **Moyennes des Métriques** : Les moyennes macro et pondérées des scores sont de 0,97, indiquant une performance uniforme pour les deux classes.
- **Validation Croisée** : Score moyen de validation croisée de 0,97, prouvant la robustesse et la constance du modèle sur différentes itérations.

Conclusion : Le modèle KNN offre une excellente performance dans la classification des signaux normaux et anormaux, avec des métriques équilibrées et robustes.

3.2.4.2 Machines à vecteurs de support(SVM) :

Résultats SVM sur les données ACP en 3D (Normal vs Anormal)

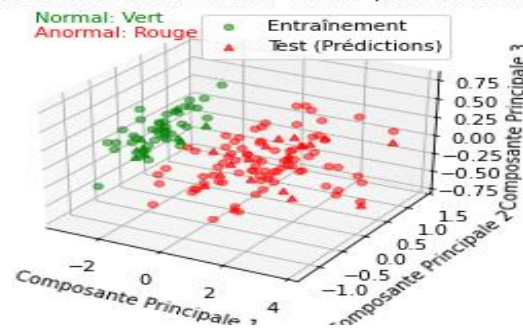


Figure 39: Séparation des classes normales et anormales par SVM dans un espace 3D

Le modèle SVM s'est montré efficace pour discriminer les données "normales" des données "anormales", avec l'ACP facilitant la visualisation et la simplification du modèle tout en conservant les informations essentielles. Une analyse plus approfondie des performances est nécessaire pour une évaluation complète.

- **Analyse en Composantes Principales (ACP)**
 - **Réduction de Dimensionnalité:**

Les données initiales, de haute dimension, ont été réduites à trois dimensions principales grâce à l'ACP, représentées par les axes du graphique.

- **Modèle SVM**

Entraînement et Prédiction : Le modèle SVM est entraîné sur un sous-ensemble de données (entraînement) et prédit les classes de nouveaux points de données (test).

- **Visualisation des Classes**
 - **Points Colorés :**
 - **Verts :** Données classées comme "normales".
 - **Rouges :** Données classées comme "anormales".
- **Formes :**
 - **Cercles:** Données d'entraînement.
 - **Triangles :** Données de test (prédictions du modèle).
- **Interprétation des Résultats**
 - **Séparation des Classes :** Une bonne séparation est observée entre les classes "normale" et "anormale", formant des clusters distincts.
- **Performance du Modèle :**
 - **Entraînement :** Le modèle généralise bien sur les données d'entraînement, avec une séparation claire des points.
 - **Test :** Les prédictions concordent majoritairement avec les données d'entraînement, indiquant de bonnes performances de généralisation.
- **Équilibre des Classes**

Échantillons : Le code génère 6000 échantillons, répartis équilibrément (3 000 normaux et 3 000 anormaux). Cette approche est essentielle pour éviter le biais, permettant un apprentissage équitable des deux classes.

- **Évaluation des Performances**
 - **Métriques à Calculer :** Pour évaluer quantitativement les performances, des métriques telles que précision, rappel et F1-score sont recommandées.
 - **Risques :** Des pourcentages trop faibles pour une classe pourraient indiquer un problème de classification ou un déséquilibre.
- **Importance de l'Ensemble de Données**

- **Séparation des Ensembles** : Les données sont divisées en 80 % pour l'entraînement et 20 % pour le test, crucial pour vérifier la capacité de généralisation du modèle.
- **Validation des Performances** : Cela permet de valider les performances du modèle en utilisant des métriques essentielles pour comprendre son comportement en situations réelles.
- **Résultats Généraux du Modèle**

Mesure	Valeur
Accuracy	0.99
Pourcentage de Signaux Normaux	49.42%
Pourcentage de Signaux Anormaux	50.58%

Tableau 7:Résultats Généraux des Performances du Modèle

- **Rapport de Classification**

Classe	Précision	Rappel	F1-Score	Support
Normaux	0.99	0.99	0.99	606
Anormaux	0.99	0.99	0.99	594
Accuracy			0.99	1200
Macro avg	0.99	0.99	0.99	1200
Weighted avg	0.99	0.99	0.99	1200

Tableau 8:Rapport de Classification des Classes

- **Validation Croisée**

Mesure	Valeur
Moyenne des Scores de Validation Croisée	0.97

Tableau 9:Moyenne des Scores de Validation Croisée

- **Performance Globale**

Accuracy (0.99) : Le modèle a correctement classé 99 % des échantillons, indiquant une excellente capacité de discrimination entre les classes normales et anormales.

- **Interprétation** : Les métriques élevées pour les deux classes montrent une performance exceptionnelle dans l'identification correcte des signaux.
- **Support:**
 - **Normal:** 606
 - **Anormal:** 594
- **Interprétation** : Le support similaire pour chaque classe renforce la fiabilité des métriques de performance.
- **Validation Croisée**
 - **Moyenne (0.97)** : Indique une performance stable du modèle à travers différentes itérations de validation, renforçant sa capacité à généraliser.

Conclusion

Ces résultats suggèrent que le modèle est très efficace pour classer les signaux normaux et anormaux. La précision élevée, l'équilibre des classes et les résultats de validation croisée soutiennent la robustesse du modèle, bien qu'il soit recommandé de le tester sur d'autres ensembles de données pour confirmer sa performance.

3.2.4.3 Random forest:

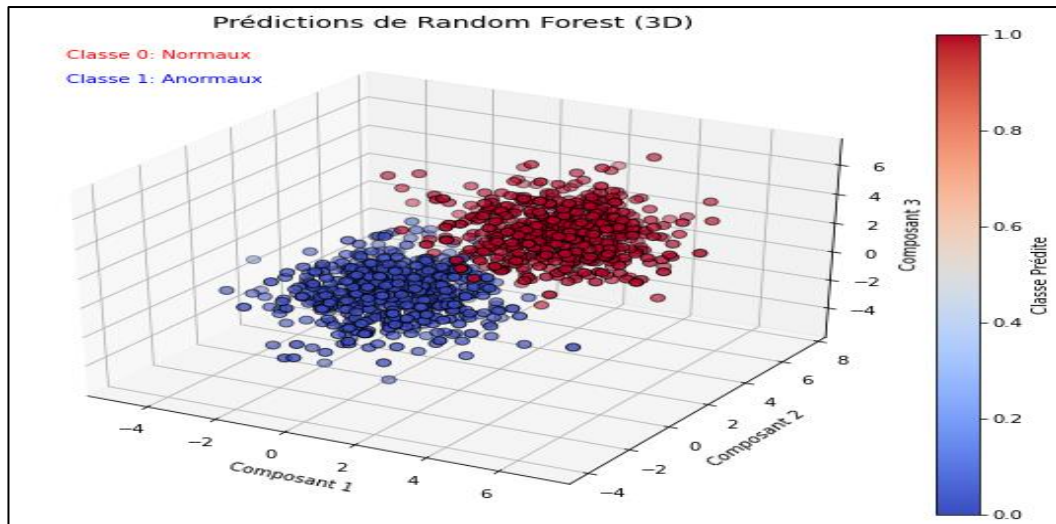


Figure 40:Arbre de décision pour la classification avec une précision de 91%

- **Composants (axes 1, 2, 3)** : Représentent les dimensions des données utilisées par le modèle.
- **Classe Prédite** : La couleur des points (rouge pour "Normaux", bleu pour "Anormaux") indique la classe, avec l'intensité de couleur montrant la confiance du modèle.
- **Séparation des Groupes** : Le groupe rouge est dense et bien séparé, tandis que le bleu est plus dispersé, indiquant une plus grande incertitude.
- **Applications** : Utilisé pour la détection d'anomalies, comme dans la sécurité réseau ou la fabrication pour identifier des comportements inhabituels.
- **Considérations** : L'interprétation dépend de la qualité du modèle et des données, ainsi que de la compréhension du domaine spécifique.

- **Explication du Code**

- **Génération des données** : Des données binaires simulées sont créées avec `make_blobs`, réparties autour de deux centres pour une meilleure séparation entre les classes "Normaux" et "Anormaux".
- **Application du modèle** : Un modèle Random Forest avec 100 arbres est entraîné sur les données d'entraînement et fait des prédictions sur l'ensemble de test.
- **Paramètres et division des données** : 6000 échantillons sont utilisés, avec un ratio 50% entre les classes et 20% des données réservées pour le test.
- **Évaluation du modèle** : L'accuracy du modèle est calculée, un rapport de classification est généré, et une validation croisée est effectuée pour tester la robustesse du modèle.
- **Visualisation 3D** : Les prédictions sont affichées en 3D à l'aide de `matplotlib`, avec des couleurs pour distinguer les classes prédites "Normaux" et "Anormaux"

- Résultats des Modèles de Classification

Modèle	Exactitude	MoyenneMacro	MoyennePondérée
Arbre de Décision	0.91	0.91	0.91
Forêt Aléatoire	0.98	0.98	0.98

Tableau 10:Résultats de performance des Modèles de Classification

- **Exactitude**

La Forêt Aléatoire, avec une exactitude de 0.98, offre une performance supérieure à l'Arbre de Décision (0.91), ce qui suggère qu'elle est plus efficace pour effectuer des prédictions précises.

- **Moyenne Macro et Moyenne Pondérée**

Les moyennes montrent que les modèles sont équitables et robustes :

- **Moyenne Macro** : Évalue la performance sans biais vers des classes plus grandes.
- **Moyenne Pondérée** : Tient compte du nombre d'échantillons par classe, confirmant l'absence de déséquilibre.
-
- **Interprétation Générale**
 - **Performance des Modèles** : Les résultats similaires des deux modèles montrent une bonne généralisation, essentielle pour des applications réelles.
 - **Simplicité et Robustesse** : L'arbre de décision est facile à interpréter, tandis que la forêt aléatoire offre une robustesse supplémentaire sans nécessité apparente dans ce cas.
- **Limitations et Suggestions**
 - **Évaluation Plus Approfondie** : Considérer d'autres métriques, comme la matrice de confusion, pour une meilleure analyse des erreurs de classification.
 - **Surveillance de l'Overfitting** : Tester la performance sur des ensembles de données non vus pour éviter le sur-apprentissage.
 - **Exploration des Caractéristiques** : Évaluer l'importance des caractéristiques pour mieux comprendre les décisions des modèles.

3.2.4.4_Extreme Gradient Boosting(XGBoost) :

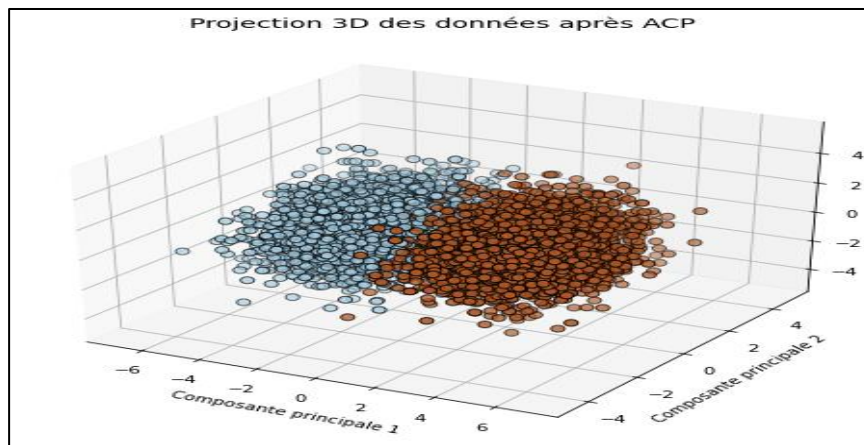


Figure 41: Matrice de confusion en utilisant le GBM

- **Graphique 3D après ACP** : Projection tridimensionnelle d'un ensemble de données après réduction de dimensionnalité par l'ACP, qui capture la variance principale des données.
- **Composantes Principales 1, 2, 3** : Représentent les trois axes qui expliquent la plus grande variance dans les données originales.
- **Séparation des Groupes** : Les points forment deux groupes distincts (bleu et marron), suggérant deux catégories ou groupes.
- **Applications** : Utilisé pour la réduction de dimensionnalité, l'extraction de caractéristiques, la détection d'anomalies et le clustering des données.
- **Considérations** : L'ACP suppose des données normales et linéaires ; l'interprétation des composantes peut être complexe et dépend du contexte des données.
- Explication du code :

- **n_components** : Le nombre de composantes principales à conserver après application de l'Analyse en Composantes Principales (ACP), fixé à 3.
- **num_samples** : Le nombre total d'échantillons générés pour la classification, défini à 6000.
- **normal_ratio** : Le ratio entre les classes normales et anormales dans les données simulées, défini à 0.5.
- **test_size** : La proportion des données utilisées pour les tests, fixée à 20% (0.2).
- **XGBoost Parameters** : L'entraînement du modèle utilise l'algorithme XGBoost avec un critère d'évaluation 'mlogloss' pour la classification binaire.

Métrique	Classe 0	Classe 1	Total
Précision	0.97	0.97	0.97
Rappel	0.97	0.96	0.97
F1-score	0.97	0.97	0.97
Exactitude			0.97

Tableau 11: Métriques de performance et validation croisée du modèle XGBoost

• Exactitude

L'exactitude générale du modèle est de **97 %**, ce qui signifie que 97 % des prédictions du modèle sont correctes. Cela témoigne d'une excellente performance globale.

• Rapport de Classification

Le rapport de classification fournit des métriques détaillées pour chaque classe, avec les valeurs de support.

• Classe 0 (Normaux)

- **Précision** : 0.97
97 % des instances prédites comme Classe 0 étaient effectivement de cette classe, montrant une très bonne confiance dans ces prédictions.
- **Rappel (Recall)** : 0.97
97 % des instances réelles de Classe 0 ont été correctement identifiées, ce qui signifie que le modèle est très efficace pour détecter cette classe, avec un faible taux de faux négatifs.
- **F1-Score** : 0.97
Ce score indique un excellent équilibre entre la précision et le rappel, montrant que le modèle fait peu d'erreurs pour cette classe.
- **Support** : 598
Cela indique qu'il y avait 598 exemples de Classe 0 dans l'ensemble de test, renforçant les bonnes performances du modèle dans cette catégorie.

• Classe 1 (Anormaux)

- **Précision** : 0.97
97 % des instances prédites comme Classe 1 étaient effectivement de cette classe, indiquant également une grande confiance dans ces prédictions.
- **Rappel** : 0.97
97 % des instances réelles de Classe 1 ont été correctement détectées. Ce chiffre indique une excellente capacité à identifier cette classe, avec peu de faux négatifs.
- **F1-Score** : 0.97
Un très bon score qui montre un équilibre parfait entre précision et rappel.
- **Support** : 602
Cela représente le nombre d'exemples de Classe 1 dans l'ensemble de test. Le support est légèrement supérieur à celui de la Classe 0, mais cela reste équilibré.
- **Moyennes**

- **Moyenne Macro : 0.97**
Cela indique la moyenne des scores pour chaque classe sans tenir compte du support, montrant que les performances sont équilibrées entre les deux classes.
- **Moyenne Pondérée : 0.97**
Cette moyenne prend en compte le support des classes et confirme que le modèle est performant dans l'ensemble, même si les classes ont des tailles légèrement différentes.

3.2.4.5 Light Gradient Boosting Machine (LightGBM) :

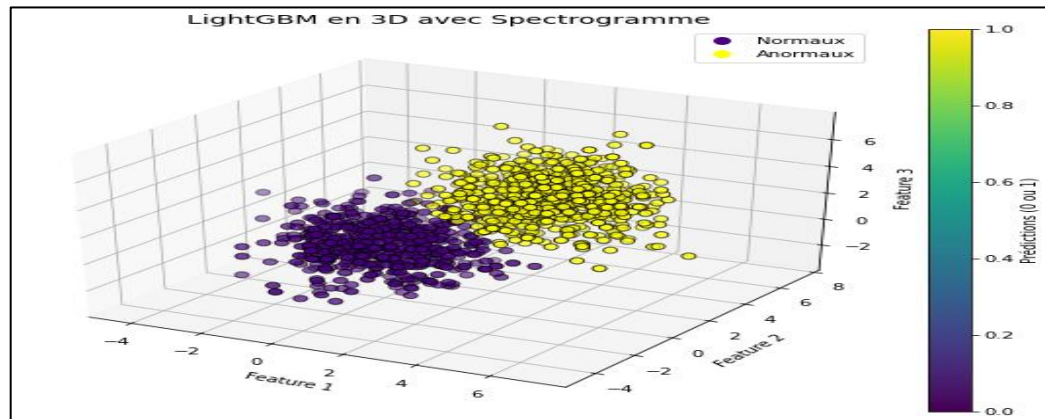


Figure 42: Performance du modèle du Courbe ROC

- **Visualisation 3D** : Le graphique montre les prédictions d'un modèle LightGBM pour une classification binaire entre les classes "Normaux" et "Anormaux". Les trois axes représentent les caractéristiques utilisées pour l'entraînement du modèle.
- **Axes et Couleurs** : Les axes du graphique correspondent aux trois caractéristiques d'entrée du modèle, tandis que la couleur des points indique la classe prédite. L'intensité de la couleur reflète la confiance du modèle dans sa prédiction.
- **Séparation des Classes** : Les classes "Normaux" et "Anormaux" sont généralement bien séparées, mais une superposition existe dans certaines zones denses, ce qui montre que la frontière de décision entre les deux classes n'est pas totalement claire.
- **Importance des Caractéristiques** : Les trois caractéristiques jouent un rôle clé dans la séparation des classes. Une analyse de l'importance des caractéristiques pourrait aider à identifier celles qui sont les plus discriminantes.
- **Évaluation et Sur-apprentissage** : Il est crucial d'évaluer la performance du modèle à l'aide de métriques comme la précision et de vérifier si le modèle présente des signes de sur-apprentissage.
- **Explication du Code**
 - **Paramètres de données** : 6000 échantillons, ratio 0,5 entre classes normales et anormales, 20% des données pour les tests.
 - **Données d'entraînement** : Les données et les étiquettes sont converties en un objet Dataset LightGBM.
 - **Paramètres du modèle** : Objectif binaire, métrique d'erreur binaire, type de boosting GBDT.
 - **Réglages supplémentaires** : Nombre de feuilles (31), taux d'apprentissage (0,05), fraction des caractéristiques (0,9).
 - **Entraînement** : Le modèle est entraîné pendant 100 itérations avec les paramètres définis

Métriques	Classe 0	Classe 1	Total
Précision	0.96	0.96	0.96
Rappel (Recall)	0.96	0.96	0.96
F1-Score	0.96	0.96	0.96
Exactitude			0.96

Tableau 12: Meilleurs hyperparamètres Valeurs

- Exactitude

L'exactitude globale du modèle est de 96 %, indiquant que 96 % des échantillons ont été correctement classés, ce qui montre une excellente performance en classification binaire.

- Détails par Classe :

- Classe 0 (Normaux) :

- Précision : 0.96
- Rappel : 0.96
- F1-score : 0.96 (très bon équilibre entre précision et rappel).

- Classe 1 (Anormaux) :

- Précision : 0.96
- Rappel : 0.96
- F1-score : 0.96 (bonne performance avec un équilibre similaire à celui de la classe 0).

- Proportions :

- Classe 0 : 51 % de l'ensemble de test

- Classe 1 : 49 % de l'ensemble de test

- **Performance du Modèle :**
Les métriques de précision, rappel et F1-score sont élevées pour les deux classes, ce qui indique une identification efficace. La répartition équilibrée des classes dans l'ensemble de test contribue à cette excellente performance.

- **Suggestions d'Amélioration :**
Explorer d'autres échantillons ou configurations de modèle pour optimiser davantage les performances.

Les résultats montrent que le modèle obtient une excellente performance avec des valeurs de précision, rappel et F1-score de 96 % pour les deux classes, assurant ainsi une très bonne identification des instances de chaque classe.

2.3.4.6 Sélection des Modèles Prometteurs pour un Réglage Plus Approfondi:

2.3.4.6.1 Recherche sur Grille

➤ Résultats des meilleurs Hyperparamètres du Modèle

Meilleurs hyperparamètres	Valeurs
C	100
gamma	auto
kernel	rbf

Tableau 13:Meilleurs Hyperparamètres du Modèle

- **C** : Valeur optimale de 100, régule la complexité du modèle.
- **Gamma** : Déterminé comme 'auto', influençant la portée des points de données.
- **Kernel** : Type 'rbf' (fonction de base radiale), utilisé pour transformer les données dans un espace de caractéristiques plus élevé.

➤ Résultats Métriques de Performance par Classe

Métriques de Performance	Précision	Rappel	F1-score	Support
Classe 0	0.97	0.98	0.97	593
Classe 1	0.98	0.97	0.98	607

Tableau 14:Métriques de Performance par Classe

- **Classe 0 :**
 - **Précision** : 0.97, indique que 97 % des prédictions positives étaient correctes.
 - **Rappel** : 0.98, montre que 98 % des échantillons réels de cette classe ont été identifiés.
 - **F1-score** : 0.97, équilibre entre précision et rappel.
 - **Support** : 593 échantillons dans l'ensemble de test.
- **Classe 1 :**

- **Précision** : 0.98, 98 % des prédictions positives étaient correctes.
- **Rappel** : 0.97, 97 % des échantillons réels de cette classe ont été détectés.
- **F1-score** : 0.98, un bon équilibre similaire à la Classe 0.
- **Support** : 607 échantillons dans l'ensemble de test.

➤ Résumé des résultats Global des Performances

Métrique	Valeur
Exactitude	0.97

Tableau 15: Résumé Global des Performances

- **Exactitude** : 0.97, indique que 97 % des échantillons ont été correctement classés.
- **Macro moyenne** : 0.97, reflète la moyenne des performances des classes sans tenir compte du support.
- **Moyenne pondérée** : 0.98, prend en compte la taille des classes, montrant une performance globale solide.
- **Nombre total d'échantillons** : 1200, total des échantillons utilisés pour l'évaluation.

➤ Analyse des performances des modèles

Les modèles	Exactitude
K-Nearest Neighbors (KNN)	0.97
Machines à Vecteurs de Support (SVM)	0.99
Random forest	0.98
Extreme Gradient Boosting(XG Boost).	0.96
Light Gradient Boosting Machine(GBM).	0.97

Tableau 16: comparaison entre l'exactitude de chaque modèle de performance

Les Machines à Vecteurs de Support (SVM) offrent la meilleure performance avec une précision de 0.99, indiquant qu'elles sont très efficaces pour des tâches complexes avec une séparation nette entre les classes.

2.3.4.6.2 Optimisation Bayésienne :

Optimisation des Hyperparamètres:

- Meilleurs hyperparamètres : **C = 0.05, penalty = 'l1'**
- **AUC-ROC** avec optimisation bayésienne : **0.91**

Performance Générale :

- Précision stable autour de **0.9278**, atteignant **0.9322** dans certaines itérations.
- Précision finale avec hyperparamètres optimisés : **0.9305**.

Hyperparamètres Optimaux :

- **C = 0.05** : Régularisation plus forte, utile contre le surajustement.
- **Penalty = 'l1'** : Régularisation L1 pour la sélection de caractéristiques.
- **AUC-ROC = 0.91** : Excellente discrimination entre les classes.
- Optimisation bayésienne a amélioré la performance du modèle (précision de 0.9305, AUC-ROC de 0.91).
- Hyperparamètres choisis (**C = 0.05, penalty = 'l1'**) assurent robustesse et interprétabilité.

Conclusion

Au cours de notre travail sur le filtrage des signaux ECG, nous avons identifié plusieurs points clés qui mettent en évidence l'importance d'un prétraitement rigoureux et de techniques d'analyse avancées.

Le premier constat majeur a été la présence significative d'artefacts dans les signaux ECG bruts, souvent causés par des mouvements des patients ou des interférences électromagnétiques. Ces artefacts peuvent sérieusement altérer la qualité des données et nuire à l'exactitude des diagnostics. Il est donc crucial de purifier les signaux avant toute analyse afin de garantir des résultats fiables.

Le deuxième point essentiel concerne l'application de divers types de filtres, notamment les filtres passe-bande, passe-bas et passe-haut. Chacun de ces filtres joue un rôle fondamental dans l'amélioration de la clarté des signaux ECG. Notamment, le filtre passe-bande a permis de conserver les ondes cardiaques pertinentes tout en éliminant le bruit indésirable, facilitant ainsi l'identification des complexes P, QRS et T.

Le troisième point que nous avons mis en lumière est l'utilisation de l'**Analyse en Composantes Principales (ACP)**. Cette méthode a non seulement facilité la visualisation multidimensionnelle des données, mais elle a également permis de détecter des structures et des anomalies significatives dans les signaux ECG. L'ACP aide à réduire la dimensionnalité des données tout en préservant les informations essentielles, ce qui est crucial pour identifier des tendances et des clusters. De plus, l'ACP a été utilisée dans les méthodes **KNN, XGBoost, LightGBM, SVM et Random Forest**, contribuant à améliorer la performance de ces modèles en simplifiant la structure des données tout en maintenant leur pertinence.

En conclusion, les résultats de notre étude illustrent l'importance d'intégrer des techniques avancées de filtrage et d'analyse dans le traitement des signaux ECG. Ces méthodes non seulement améliorent la qualité des données, mais elles ouvrent également la voie à des avancées potentielles dans le diagnostic et la surveillance de la santé cardiaque. L'intégration de ces analyses dans des systèmes de surveillance en temps réel pourrait révolutionner l'évaluation continue des patients, rendant les diagnostics plus rapides et plus précis. De plus, l'application de techniques telles que l'apprentissage automatique pourrait encore renforcer l'interprétation des données ECG, permettant ainsi de mieux identifier les anomalies et de répondre plus efficacement aux besoins cliniques. Enfin, l'étape suivante de notre étude sera la **prédiction des anomalies cardiaques**, en utilisant ces modèles et techniques pour anticiper de manière proactive les problèmes cardiaques chez les patients.

Références

Gamkrelidze, Tamari. *Des discours aux réalités de la conception, du déploiement et des usages des systèmes d'Intelligence Artificielle dans les situations de travail*. Diss. HESAM Université, 2022.

Grace, Tom. *Spyder web*. Simon and Schuster, 2017.

Rivas, Jesús A. *Anaconda: The Secret Life of the World's Largest Snake*. Oxford University Press, USA, 2020.

ETTARID, Pr Mohamed, et al. "Apport des algorithmes d'apprentissage supervisé pour la cartographie intelligente des zones irriguées à partir des images satellite à haute résolution spatiale." (2021).

Jacopin, Eliott, et al. "Détection automatique de plantes au sein d'images aériennes de champs par apprentissage non supervisé et approche multi-agents." *Revue Ouverte d'Intelligence Artificielle* 2.1 (2021): 123-156.

SAFI, Adila. *APPROCHE NEURONALE POUR LA CLASSIFICATION MORPHOLOGIQUE DES SIGNAUX ECG*. Diss. Université Ibn Khaldoun-Tiaret-, 2017.

Dorling, Mark, and Dave White. "Scratch: A way to logo and python." *Proceedings of the 46th ACM technical symposium on computer science education*. 2015.

Netter, Frank H. *Atlas Netter d'anatomie humaine*. Elsevier Health Sciences, 2023.

Chlahawi AA, Narakathu BB, Emamian S, Bazuin BJ, Atashbar MZ. Development of printed and flexible dry ECG electrodes. *Sensing and bio-sensing research*. 2018 Sep 1;20:9-15.

Berkaya, Selcan Kaplan, et al. "A survey on ECG analysis." *Biomedical Signal Processing and Control* 43 (2018): 216-235.

Houssein, Essam H., Moataz Kilany, and Aboul Ella Hassanien. "ECG signals classification: a review." *International Journal of Intelligent Engineering Informatics* 5.4 (2017): 376-396.

MENIF, H., et al. "LE TRALI, UN EFFET INDESIRABLE POST-TRANSFUSIONNEL MECONNU ET GRAVE: ETUDE DE 4 CAS ET REVUE DE LA LITTERATURE TRALI, A RARE AND SEVERE TRANSFUSION SIDE EFFECT: REPORT OF FOUR CASES AND A BRIEF REVIEW OF THE LITERATURE." *Journal de l'Information Médicale de Sfax* (2022): 35.

Dobrev, D., et al. "The G protein-gated potassium current IK, ACh is constitutively active in patients with chronic atrial fibrillation." *Circulation* 112.24 (2005): 3697-3706.

Allali, Abdenour, et al. "Revealing and evaluating the influence of filters position in cascaded filter: application on the ECG de-noising performance disparity." *Indonesian Journal of Electrical Engineering and Computer Science* 21.2 (2021): 829-838.

Taha, Luay Yassin, and Esam Abdel-Raheem. "Fetal ECG extraction using input-mode and output-mode adaptive filters with blind source separation." *Canadian Journal of Electrical and Computer Engineering* 43.4 (2020): 295-304.

Luo, S., & Johnston, P. (2010). A review of electrocardiogram filtering. *Journal of electrocardiology*, 43(6), 486-496.

Collen, Patrice. *Bandwidth extension tools for audio digital signals*. Diss. Télécom ParisTech, 2002.

Qartavol, Reza Maram. *Génération et traitement du signal optique basés sur le filtrage linéaire de phase-seule dans les domaines temporel et spectral*. Diss. Institut National de la Recherche Scientifique (Canada), 2016.

Wolkenstein, Fabio, and Christopher Wratil. "Multidimensional representation." *American Journal of Political Science* 65.4 (2021): 862-876.

- Malinowski, Elzbieta, and Esteban Zimányi. "Hierarchies in a multidimensional model: From conceptual modeling to logical representation." *Data & Knowledge Engineering* 59.2 (2006): 348-377.
- Jacquet, Nicolas, et al. "La steam explosion: application en tant que prétraitement de la matière lignocellulosique." *BASE* (2010).
- Andreao, Rodrigo Varejao, Bernadette Dorizzi, and Jérôme Boudy. "ECG signal analysis through hidden Markov models." *IEEE Transactions on Biomedical engineering* 53.8 (2006): 1541-1549.
- Liu, C. M., Liu, C. L., Hu, K. W., Tseng, V. S., Chang, S. L., Lin, Y. J., ... & Chen, S. A. (2022). A deep learning-enabled electrocardiogram model for the identification of a rare inherited arrhythmia: Brugada syndrome. *Canadian Journal of Cardiology*, 38(2), 152-159.
- Hadjem, M., & Nait-Abdesselam, F. (2015, July). A comparative study of supervised learning techniques for ECG T-wave anomalies detection in a WBS context. In *2015 International Conference on Protocol Engineering (ICPE) and International Conference on New Technologies of Distributed Systems (NTDS)* (pp. 1-6). IEEE.
- AOUN Fatima, ZERIFI Razika. *Classification of Normal/Abnormal Heart Sound Recordings*. Diss. Université de Echahid Cheikh Larbi Tébessi–Tébessa-, 2024.
- Tiwari, A., 2021. *Physiological features for mental state monitoring in real life conditions* (Doctoral dissertation, Institut National de la Recherche Scientifique (Canada)).
- Chettaoui, Hnia, Rabeb Mouaddeb, Tarek Moulahi, and Najoua Bennaji. "Classification of ECG signals using CNN models." (2024).
- Pillebout, Evangeline, et al. "Multicentre, randomised, economic evaluation of a web-based interactive education platform, simple or enhanced, for patients with end-stage renal disease: the PIC-R trial protocol." *BMJ open* 12.12 (2022): e062613.
- Yang, Min-Uk, Dae-In Lee, and Seung Park. "Automated diagnosis of atrial fibrillation using ECG component-aware transformer." *Computers in Biology and Medicine* 150 (2022): 106115.
- Tautz, Diethard. "Segmentation." *Developmental cell* 7.3 (2004): 301-312.
- Steinbach, Michael, and Pang-Ning Tan. "kNN: k-nearest neighbors." *The top ten algorithms in data mining*. Chapman and Hall/CRC, 2009. 165-176.
- Jakkula, Vikramaditya. "Tutorial on support vector machine (svm)." *School of EECS, Washington State University* 37.2.5 (2006): 3.
- Jiang, Yu, et al. "A pedestrian detection method based on genetic algorithm for optimize XGBoost training parameters." *IEEE Access* 7 (2019): 118310-118321.
- Rakotomalala, Ricco. "Arbres de décision." *Revue Modulad* 33 (2005): 163-187.
- Hindarto, Djarot. "Case Study: Gradient Boosting Machine vs Light GBM in Potential Landslide Detection." *Journal of Computer Networks, Architecture and High Performance Computing* 6.1 (2024): 169-178.
- McCarty, D. A., Kim, H. W., & Lee, H. K. (2020). Evaluation of light gradient boosted machine learning technique in large scale land use and land cover classification. *Environments*, 7(10), 84.