

# Détection des artefacts dans les signaux de photopléthysmographie

par

Clara MACABIAU

MÉMOIRE PAR ARTICLES PRÉSENTÉ À L'ÉCOLE DE TECHNOLOGIE  
SUPÉRIEURE COMME EXIGENCE PARTIELLE À L'OBTENTION DE LA  
MAÎTRISE AVEC MÉMOIRE  
M. Sc. A.

MONTRÉAL, LE 14 DÉCEMBRE 2023

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE  
UNIVERSITÉ DU QUÉBEC



Clara Macabiau, 2023



Cette licence Creative Commons signifie qu'il est permis de diffuser, d'imprimer ou de sauvegarder sur un autre support une partie ou la totalité de cette oeuvre à condition de mentionner l'auteur, que ces utilisations soient faites à des fins non commerciales et que le contenu de l'oeuvre n'ait pas été modifié.

**PRÉSENTATION DU JURY**

CETTE THÈSE A ÉTÉ ÉVALUÉE

PAR UN JURY COMPOSÉ DE:

Mme. Rita Noumeir, directrice de mémoire  
Département de génie électrique, École de technologie supérieure

M. Philippe Jovet, codirecteur  
Unité des soins intensifs pédiatriques, CHU Sainte-Justine

M. Luc Duong, président du jury  
Département de génie logiciel et des TI, École de technologie supérieure

M. Jean-Marc Lina, membre du jury  
Département de génie électrique, École de technologie supérieure

ELLE A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC

LE 13 NOVEMBRE 2023

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE



## REMERCIEMENTS

Je tiens à exprimer ma profonde gratitude envers toutes les personnes qui ont contribué de près ou de loin à la réalisation de ce mémoire.

Je souhaite tout d'abord remercier mes directeurs de mémoire, Philippe Jovet et Rita Noumeir, pour leur confiance et leur soutien inestimable tout au long de ce projet. Votre expertise et vos conseils m'ont permis de mieux comprendre le sujet et de progresser dans ma recherche.

Je tiens également à remercier chaleureusement tous les collègues du LATIS et ceux du CHU Sainte-Justine pour leurs précieux commentaires et discussions stimulantes. Votre aide a grandement enrichi ce travail.

Mes remerciements s'adressent aussi à ma famille et mes amis, dont la valeur de leur présence a été précieuse, que ce soit à proximité ou malgré la distance qui nous séparait.

À mes parents, pour votre amour, votre confiance, vos encouragements et votre soutien qui a été une source inspiration et m'a permis de persévérer dans les moments de doute.

Finalement, je tiens à remercier l'ENSEEIH pour m'avoir offert l'opportunité de réaliser un double diplôme au Canada, et l'ETS pour m'avoir permis de faire cette maîtrise.



# Détection des artefacts dans les signaux de photopléthysmographie

Clara MACABIAU

## RÉSUMÉ

L'objectif principal de ce projet est d'assurer une détection précise des artefacts de mouvement dans les signaux photopléthysmographiques (PPG), une étape cruciale pour garantir la fiabilité et l'exactitude des analyses cliniques et de recherche basées sur ces données. Parmi les nombreux types d'artefacts, nous nous sommes particulièrement intéressés aux artefacts de mouvement, car ils ont le potentiel de perturber significativement les signaux PPG et de compromettre l'interprétation des paramètres physiologiques.

Pour atteindre cet objectif, nous avons entrepris une approche basée sur l'apprentissage machine. Notre mémoire détaille l'ensemble du processus, débutant par le prétraitement initial du signal PPG, puis se poursuivant avec l'annotation des données et l'exploration de diverses méthodes de rééquilibrage des données, pour finalement aboutir à la comparaison des performances des classificateurs supervisés avec celles de l'algorithme semi-supervisé de propagation d'étiquettes.

Les résultats obtenus montrent que l'algorithme d'apprentissage machine semi-supervisé de propagation d'étiquettes a obtenu une précision de 91%, un rappel de 90% et un score F1 de 90%. Bien que l'algorithme supervisé KNN (K-Nearest Neighbors) ait également obtenu des résultats solides, avec une précision de 89%, un rappel de 95% et un score F1 de 92%, l'algorithme de propagation d'étiquettes s'est avéré plus efficace pour détecter avec précision les artefacts de mouvement.

En résumé, cette recherche contribue à l'avancement des connaissances dans le domaine de la détection des artefacts dans les signaux PPG, en mettant en évidence l'efficacité de l'algorithme semi-supervisé de propagation d'étiquettes dans la détection des artefacts de mouvement.

**Mots-clés:** Photopléthysmographie, artefacts de mouvement, traitement du signal, apprentissage machine, classificateurs supervisés





## Detecting artifacts in photoplethysmogram signals

Clara MACABIAU

### ABSTRACT

The main objective of this project is to ensure accurate detection of motion artifacts in photoplethysmographic (PPG) signals, a crucial step in guaranteeing the reliability and accuracy of clinical and research analyses based on these data. Among the many types of artifacts, we were particularly interested in motion artifacts, as they have the potential to significantly disrupt PPG signals and compromise the interpretation of physiological parameters.

To achieve this goal, we undertook a machine learning-based approach. Our dissertation details the entire process, starting with the initial pre-processing of the PPG signal, continuing with data annotation and the exploration of various data re-balancing methods, and finally culminating in a comparison of the performance of the supervised classifiers with that of the semi-supervised label propagation algorithm.

The results obtained show that the semi-supervised label propagation machine learning algorithm achieved a precision of 91%, a recall of 90% and an F1 score of 90%. Although the supervised KNN (K-Nearest Neighbors) algorithm also achieved strong results, with a precision of 89%, a recall of 95% and an F1 score of 92%, the label propagation algorithm proved more effective at accurately detecting motion artifacts.

In summary, this research contributes to the advancement of knowledge in the field of artifact detection in PPG signals, by highlighting the effectiveness of the semi-supervised label propagation algorithm in detecting motion artifacts.

**Keywords:** Photoplethysmography, motion artifacts, signal processing, machine learning, supervised classifiers



## TABLE DES MATIÈRES

	Page
INTRODUCTION .....	1
CHAPITRE 1 REVUE DE LITTÉRATURE .....	5
1.1 Présentation du signal PPG .....	5
1.2 Approches traditionnelles .....	7
1.2.1 Analyse statistique .....	7
1.2.2 Filtrage adaptatif .....	8
1.3 Analyse temps-fréquence .....	10
1.3.1 Transformée en ondelettes .....	10
1.3.2 Décomposition en modes empiriques .....	12
1.4 Apprentissage machine .....	13
1.5 Lien avec l'article .....	14
CHAPITRE 2 LABEL PROPAGATION TECHNIQUES FOR ARTIFACT DE- TECTION IN IMBALANCED CLASSES USING PHOTOPLE- THYSMOGRAM SIGNALS .....	17
2.1 Abstract .....	17
2.2 Introduction .....	18
2.3 Materials and Methods .....	22
2.3.1 Data Collection .....	22
2.3.2 Preprocessing .....	24
2.3.3 Dataset annotation .....	26
2.3.3.1 Expert labeling .....	26
2.3.3.2 Statistical analysis .....	26
2.3.3.3 Imbalanced dataset .....	29
2.3.3.4 Label Propagation .....	30
2.3.4 Classification .....	32
2.4 Experimental Implementation .....	34
2.5 Results and Discussion .....	36
2.6 Conclusion .....	41
CONCLUSION ET RECOMMANDATIONS .....	45
LISTE DE RÉFÉRENCES .....	47



## LISTE DES TABLEAUX

	Page
Tableau 2.1	Results for the class "with artifacts" for different proportions of the dataset ..... 36
Tableau 2.2	Summary table containing the labeling portion and the imbalance rate ..... 37
Tableau 2.3	Results for the class "with artifacts" for different sampling methods ..... 37
Tableau 2.4	A comparison of the performance of different classifiers for the class "with artifact" ..... 40



## LISTE DES FIGURES

	Page
Figure 1.1	Présentation de deux signaux PPG : avec et sans artefacts de mouvement ..... 6
Figure 1.2	Filtre adaptatif des moindres carrés ..... 9
Figure 1.3	Transformée en ondelettes discrète - Filtres en cascade ..... 11
Figure 2.1	Workflow of the proposed method for detecting motion artifacts in PPG signals ..... 22
Figure 2.2	Example of a PPG signal with motion artifacts in three different forms : raw signal, filtered signal and segmented signal ..... 25
Figure 2.3	Example of a raw PPG signal segment with and without motion artifacts ..... 27
Figure 2.4	Confusion matrix and ROC curve for the Label Propagation algorithm ..... 38
Figure 2.5	Precision evaluation for different traditional Machine Learning classifiers ..... 39
Figure 2.6	Accuracy curve for the FCN model ..... 42





## LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

ETS	École de Technologie Supérieure
CHUSJ	Centre Hospitalier Universitaire Sainte-Justine
PICU	Pediatric Intensive Care Unit
CDSS	Clinical Decision Systems
PPG	Photopléthysmographie
ECG	Électrocardiogramme
ABP	Arterial Blood Pressure
ECMO	Extracorporeal Membrane Oxygenation
HRDB	High-resolution Research Database
ACC	Accélération
SVD	Singular Value Decomposition
STD	Standard Deviation
LP	Label Propagation
SVM	Support Vector Machine
SVC	Support Vector Classification
KNN	K-Nearest Neighbors
RBF	Radial Basis Function
DT	Decision Tree
NB	Naive Bayes
RF	Random Forest
LR	Linear Regression
MLP	Multi-Layer Perceptron
BERT	Bidirectional Encoder Representations from Transformers

FCN	Fully Convolutional Network
ROS	Random OverSampling
RUS	Random UnderSampling
SMOTE	Synthetic Minority Over-Sampling Technique
ADASYN	Adaptive Synthetic algorithm
FN	False Negative
FP	False Positive
TN	True Negative
TP	True Positive
ROC	Receiver Operating Characteristic
AUROC	Area Under the Receiver Operating Characteristic

## LISTE DES SYMBOLES ET UNITÉS DE MESURE

$X$	Variable aléatoire d'une distribution
$\beta_2$	Coefficient d'acuité d'une distribution
$\gamma_1$	Coefficient d'assymétrie d'une distribution
$\mu$	Moyenne d'une distribution
$\sigma$	Écart-type d'une distribution
$x$	Signal temporel
$g$	Réponse impulsionnelle d'un filtre passe-bas
$h$	Réponse impulsionnelle d'un filtre passe-haut
$U_{max}$	Enveloppe temporelle des valeurs maximales
$U_{min}$	Enveloppe temporelle des valeurs minimales
$s$	Secondes
$m$	Signal temporel résultant de la moyenne de deux signaux
$th_l$	Seuil inférieur
$th_u$	Seuil supérieur
$u$	Nombre de points non-étiquetés
$l$	Nombre de points étiquetés
$N$	Nombre total de points
$k$	Nombre de classes
$t$	Nombre d'étapes
$T$	Matrice de transition des probabilités
$T_{ll}$	Sous-matrice de transition des probabilités (noeud étiqueté vers noeud étiqueté)
$T_{lu}$	Sous-matrice de transition des probabilités (noeud étiqueté vers noeud non-étiqueté)

$T_{ul}$	Sous-matrice de transition des probabilités (noeud non-étiqueté vers noeud étiqueté)
$T_{uu}$	Sous-matrice de transition des probabilités (noeud non-étiqueté vers noeud non-étiqueté)
$Y$	Matrice des probabilités
$Y_l$	Sous-matrice des probabilités pour les noeud étiquetés
$Y_u$	Sous-matrice des probabilités pour les noeuds non-étiquetés
$I$	Matrice identité
$D$	Matrice des degrés
$A$	Matrice d'adjacence

## INTRODUCTION

La photopléthysmographie (PPG) est une technique non invasive largement utilisée pour évaluer la variabilité de la fréquence cardiaque et l'oxygénation du sang. Les signaux PPG sont acquis à partir d'un oxymètre de pouls qui comporte un capteur, placé à la surface de la peau, et qui est généralement situé au niveau du doigt ou du lobe de l'oreille. Ces appareils permettent de surveiller en temps réel les fluctuations du volume sanguin dans les vaisseaux sanguins superficiels. En effet, la variation de volume provoquée par l'onde de pression est détectée en illuminant la peau avec la lumière émise par une diode électroluminescente puis en mesurant la quantité de lumière transmise ou réfléchiée par une photodiode. Étant donné que l'afflux sanguin vers la peau peut être modulé par plusieurs autres systèmes physiologiques, le signal PPG peut également être utilisé pour surveiller la respiration, l'hypovolémie et d'autres conditions circulatoires. La forme de la courbe PPG diffère d'un sujet à l'autre et varie en fonction de l'emplacement et de la manière dont l'oxymètre de pouls est fixé.

Cette approche prometteuse a trouvé de multiples applications cliniques, allant du suivi de la santé cardiovasculaire au diagnostic de maladies chroniques, en passant par la surveillance étroite des patients pendant une opération.

Cependant, malgré les avantages indéniables de la photopléthysmographie, les signaux PPG sont souvent sujets à des perturbations et des artefacts qui peuvent entraîner des résultats incorrects ou trompeurs. Les artefacts dans les signaux PPG proviennent de diverses sources, notamment les mouvements du sujet, les interférences électromagnétiques, l'éclairage ambiant et les variations de la perfusion sanguine. Ces perturbations non désirées peuvent masquer les informations cruciales, altérer la précision des mesures et rendre les analyses incohérentes.

En générant des fausses alarmes, les artefacts de mouvement sont un obstacle à la bonne lecture des signaux PPG par le personnel hospitalier. Ils engendrent aussi des pertes dans des données, qui sont primordiales pour réaliser un suivi précis du patient. Cela rajoute une

dose de travail supplémentaire pour le personnel hospitalier qui doit vérifier l'équipement plus fréquemment ainsi que l'état général du patient. Les artefacts de mouvement dans les signaux PPG apparaissent pendant les mouvements et les périodes de faible perfusion périphérique. Lors de ces mouvements, de nombreux oxymètres de pouls ne parviennent pas à différencier le sang artériel pulsé du sang veineux en mouvement, ce qui entraîne une sous-estimation de la saturation en oxygène (Barker & Shah, 1996). Une saturation faible en oxygène dans le sang est un signe d'un mauvais état de santé mais les artefacts de mouvement induisent les infirmiers et infirmières en erreur. Les mouvements volontaires et involontaires, tels que les mouvements lors du transport, mais aussi les différentes gestuelles habituelles comme bouger le poignet, se gratter, utiliser ses doigts, ainsi que les crises chez les patients adultes et pédiatriques sont quelques-unes des sources courantes de mouvement du patient dans le cadre clinique. Pour les nouveau-nés les mouvements peuvent se traduire par des coups de pied, des étirements et des pleurs (Tobin, Pologe & Batchelder, 2002).

Ainsi, la détection précise des artefacts dans les signaux PPG est un défi crucial pour garantir l'exactitude et la validité des données acquises, et pour assurer des interprétations fiables et pertinentes dans les applications cliniques et de recherche. Mais cette détection est complexe. D'après Petterson, Begnoche & Graybeal (2007), le mouvement du patient a tendance à être composé d'épisodes discrets et aperiodiques se produisant étroitement les uns après les autres, de sorte que leur effet sur l'oxymétrie de pouls pourrait apparaître comme un mouvement continu, souvent d'une durée de plusieurs minutes à chaque fois. De plus, les artefacts de mouvement se superposent à la gamme de fréquences de la respiration ou de la fréquence cardiaque, ce qui rend leur détection délicate.

D'autres facteurs tels que la nature non stationnaire des signaux PPG et les interférences externes rendent la tâche de détection difficile. L'existence de ces différents paramètres rend difficile

la distinction entre les variations dues aux artefacts et les variations liées à la physiologie, nécessitant ainsi le développement de méthodes sophistiquées pour une détection précise.

Au fil des années, de nombreuses méthodes ont été développées pour détecter, réduire ou éliminer les artefacts dans les signaux PPG. Certaines de ces approches sont basées sur des algorithmes de traitement du signal, telles que la détection de mouvement, la suppression de bruit, ou la correction des variations d'amplitude. D'autres approches utilisent des techniques d'apprentissage automatique pour analyser les signaux et identifier automatiquement les artefacts.

La méthode présentée dans ce mémoire propose d'utiliser le traitement du signal en combinaison avec l'apprentissage machine pour détecter les artefacts. La partie traitement du signal permet de préparer le signal à travers différentes étapes : filtrage, segmentation, ré-échantillonnage et normalisation. Ces étapes sont cruciales pour enlever les hautes fréquences ainsi que le bruit parasite. L'objectif est de segmenter les signaux PPG de la base de données pour détecter les artefacts de mouvement dans les impulsions du signal. Après le pré-traitement du signal, les caractéristiques temporelles du signal sont extraites pour être utilisées en apprentissage machine. Plusieurs algorithmes d'apprentissage machine supervisés sont comparés avec l'algorithme semi-supervisé de propagation d'étiquettes, afin de conclure sur leur capacité à détecter les artefacts présents sur chaque impulsion des signaux de la base de données. L'apprentissage machine et l'intelligence artificielle sont des nouvelles approches qui permettent une détection plus précise et efficace. Avec leur capacité à reconnaître des modèles à partir de petits ensemble de données, ces algorithmes sont très utilisés dans le domaine médical. Les données médicales sont aussi caractérisées par des classes très déséquilibrées. Dans notre cas, la classe des impulsions sans artefacts de mouvement contient plus de données que la classe des impulsions avec artefacts de mouvement. Pour pouvoir travailler avec ces données malgré cet obstacle, plusieurs fonctions sont utilisées en apprentissage machine. Chaque fonction amène une façon de ré-équilibrer

différente, que ça soit en sur-échantillonnant, augmenter le nombre d'échantillons d'une classe ou en sous-échantillonnant, diminuer le nombre d'échantillons d'une classe.

Ce mémoire vise donc à contribuer à l'amélioration de la qualité des signaux PPG en proposant une méthode approfondie de détection des artefacts de mouvement. L'algorithme de propagation d'étiquettes permet de gagner du temps et des efforts en évitant la détection manuelle fastidieuse des artefacts. Il libère ainsi les chercheurs des tâches laborieuses de détection. Cela garantit une plus grande quantité d'informations fiables et pertinentes pour la prise en charge des patients. De plus, la comparaison des différentes méthodes d'équilibrage des données ainsi que la comparaison des différents classificateurs, apporte une évaluation objective de ces différentes approches. Cela permettra aux futurs chercheurs de choisir les méthodes à utiliser en fonction des données mises à leur disposition ainsi que d'autres facteurs.

Dans une première partie, la revue de littérature présentera les différentes méthodes mises en place pour la détection des artefacts, dans le domaine du traitement du signal et de l'apprentissage machine. Ensuite, une deuxième partie sera constituée de l'article rattaché à ce mémoire qui explique dans le détail les méthodes utilisées ainsi que les résultats obtenus. Une troisième et dernière partie présentera les recommandations et conclura ce mémoire.



# **CHAPITRE 1**

## **REVUE DE LITTÉRATURE**

Les méthodes de détection des artefacts dans les signaux PPG peuvent être divisées en plusieurs approches distinctes. Les méthodes traditionnelles, telles que l'analyse statistique ainsi que le filtrage des signaux, constituent la première approche. La deuxième approche se base sur une analyse temps-fréquence du signal. Enfin, l'apprentissage machine est la dernière approche présentée. Après une présentation du signal PPG, cette revue de littérature explore les diverses approches de détection des artefacts dans les signaux PPG pour ces catégories, en évaluant leur pertinence pour améliorer la fiabilité des analyses cliniques et de recherche.

### **1.1 Présentation du signal PPG**

Le signal PPG est présenté sous forme d'une courbe qui représente l'intensité lumineuse en fonction du temps. Chaque cycle cardiaque se traduit par un ensemble de variations pulsatives, avec des pics et des creux caractéristiques. Ces variations sont principalement dues aux changements de volume sanguin liés à la contraction et à la relaxation du cœur.

La forme d'onde du signal PPG est généralement divisée en trois composantes principales :

- composante pulsatile : c'est la composante du signal PPG qui résulte des changements de volume sanguin induits par la contraction du cœur. Elle est souvent utilisée pour déterminer la fréquence cardiaque et les variations de la fréquence cardiaque, qui sont des indicateurs importants de la santé cardiovasculaire et du bien-être général du patient ;
- composante continue : c'est la composante du signal PPG qui représente l'intensité lumineuse moyenne. Elle est influencée par des facteurs tels que l'intensité de l'éclairage ambiant, la pigmentation de la peau, la pression sanguine et la perfusion tissulaire ;
- composante de bruit : cette composante représente les variations aléatoires dans le signal PPG qui peuvent être causées par des artefacts de mouvement, des interférences électromagnétiques ou d'autres sources de bruit.

La figure 1.1 présente deux signaux PPG issus de l'ensemble de données mis à notre disposition. Le signal du haut est un signal sans artefacts de mouvement, tandis que celui du bas présente des artefacts de mouvement.

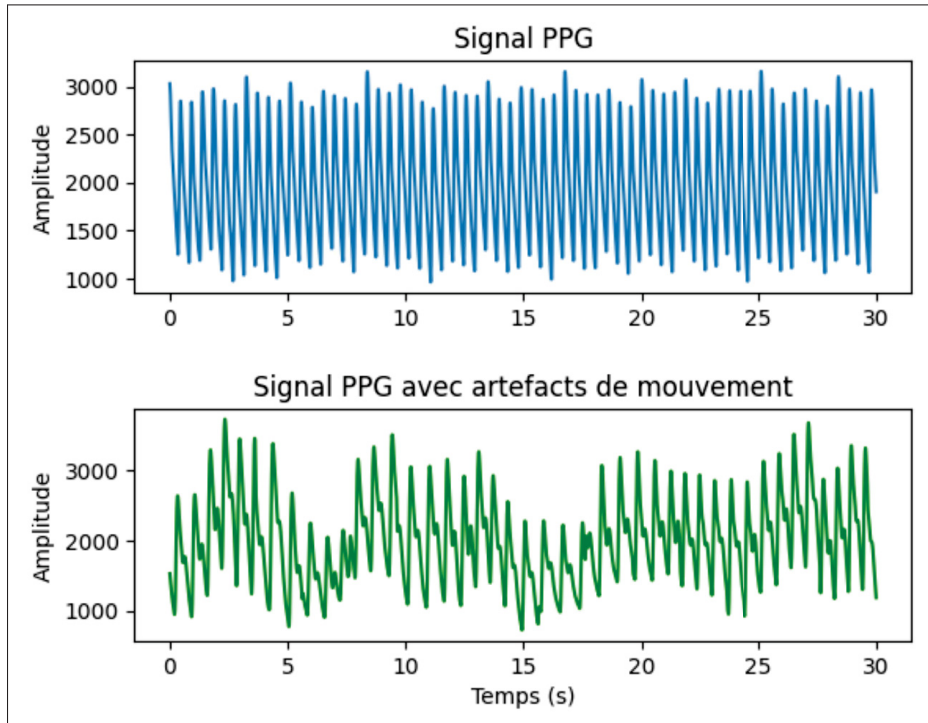


Figure 1.1 Présentation d'un signal PPG sans artefacts de mouvement (image du haut) et un signal PPG avec artefacts de mouvement

Dans la première figure, le signal comporte une composante continue stable. On remarque que les artefacts de mouvement induisent une modification de la composante continue du signal en la rendant instable avec des variations. Certaines impulsions sont aussi affectées par du mouvement ce qui augmente leur amplitude par rapport aux impulsions voisines. Ces amplitudes extrêmes rendent la lecture des paramètres tels que la fréquence cardiaque ou la fréquence respiratoire impossible.

## 1.2 Approches traditionnelles

Dans la recherche sur la détection d'artefacts de mouvement dans les signaux PPG, les méthodes traditionnelles sont les approches les plus basiques. Ces approches s'appuient sur des techniques de traitement de signal classiques et des calculs simples, qui ont été largement utilisés dans le domaine pour repérer les artefacts.

Dans certains cas, des données d'un accéléromètre sont disponibles en plus des données du signal PPG. Cela permet d'avoir un signal de référence des mouvements du patient, ce qui rend la détection des artefacts de mouvement plus simple. Pour information, dans notre étude, nous avons utilisé les données de l'oxymètre de pouls du Centre Hospitalier Universitaire Sainte-Justine (CHUSJ) de Montréal. Cet appareil, qui permet d'enregistrer le signal PPG, n'était pas couplé avec un accéléromètre.

### 1.2.1 Analyse statistique

La première approche pour détecter les artefacts de mouvement est d'utiliser les paramètres statistiques du signal. Dans cette optique, Hanyu & Xiaohui (2017) présentent une méthode qui utilise trois statistiques : le coefficient d'acuité ( $\beta_2$ ), le coefficient d'asymétrie ( $\gamma_1$ ) et l'écart type ( $\sigma$ ). Si on considère  $X$  une variable aléatoire avec  $\mu$  l'espérance de  $X$  et  $\sigma$  l'écart type de  $X$ , ces valeurs peuvent être calculées de la manière suivante :

$$\sigma(X) = \sqrt{E[(X - \mu)^2]} \quad (1.1)$$

$$\gamma_1(X) = E\left[\left(\frac{X - \mu}{\sigma}\right)^3\right] = \frac{E[(X - \mu)^3]}{\sigma^3} \quad (1.2)$$

$$\beta_2(X) = E\left[\left(\frac{X - \mu}{\sigma}\right)^4\right] = \frac{E[(X - \mu)^4]}{\sigma^4} \quad (1.3)$$

Ces trois statistiques sont sensibles aux variations du signal PPG. Si il n'y a pas de mouvement, les valeurs sont stationnaires, autrement dit elles sont constantes pour toute la durée du signal.

A l'inverse, si il y a du mouvement, les valeurs seront en dehors des seuils fixés et donc les segments du signal seront détectés comme des segments comportant des artefacts.

Un approche similaire est utilisée par Krishnan, Natarajan & Warren (2008). Les trois statistiques présentées ci dessus sont couplées avec une analyse statistique fréquentielle, qui inclut le calcul du coefficient d'acuité pour le spectre du signal, et une analyse bi-spectrale. L'analyse bi-spectrale est une technique avancée de traitement de signal qui examine comment différentes fréquences d'un signal interagissent entre elles. Contrairement à l'analyse spectrale traditionnelle, elle examine les relations croisées entre les amplitudes et les phases de différentes fréquences pour repérer des variations de fréquence non linéaires, des schémas de phase irréguliers, et des interactions harmoniques multiples.

Lorsque des relations entre les phases d'un signal existent, le coefficient de couplage quadratique est défini. En l'absence d'artefacts de mouvement, les signaux présentent un coefficient de couplage fort entre les fréquences fondamentales et le spectre. L'avantage de cette méthode est de lier l'analyse temporelle, fréquentielle et l'analyse bi-spectrale ce qui permet un examen plus approfondi du signal et un test de décision des artefacts plus robuste.

Ces méthodes présentent l'avantage de leur simplicité, car elles impliquent des calculs de base, cependant l'utilisation des statistiques pour la détection d'artefacts contraint de définir des seuils empiriques. Le choix de ces seuils peut être délicat et dépend de la spécificité des données, ce qui peut rendre la méthode moins générale.

Nous utiliserons la méthode de l'analyse statistique, qui est facile à mettre en place, pour l'annotation des impulsions du signal PPG par la suite. L'objectif sera de créer une base de vérité pour la suite du projet.

### **1.2.2 Filtrage adaptatif**

Comme expliqué précédemment, dans certains cas, en plus du signal PPG, un accès à des données d'un accéléromètre peut être disponible. Cet accès permet de rendre plus rapide la

détection des artefacts car l'accéléromètre enregistre tous les mouvements du patient. Dans le cas des artefacts de mouvement, le contenu spectral du bruit se superpose à celui du signal. Les filtres statiques, dont les coefficients restent constants, ne peuvent pas être utilisés. La section suivante présente donc les filtres adaptatifs, plus adaptés à la détection des artefacts de mouvement.

Le principe fondamental du filtrage adaptatif est de minimiser l'erreur entre le signal filtré et le signal désiré. Le filtre des moindres carrés (voir Figure 1.2) et le filtre des moindres carrés récurrents sont deux types de filtres adaptatifs.

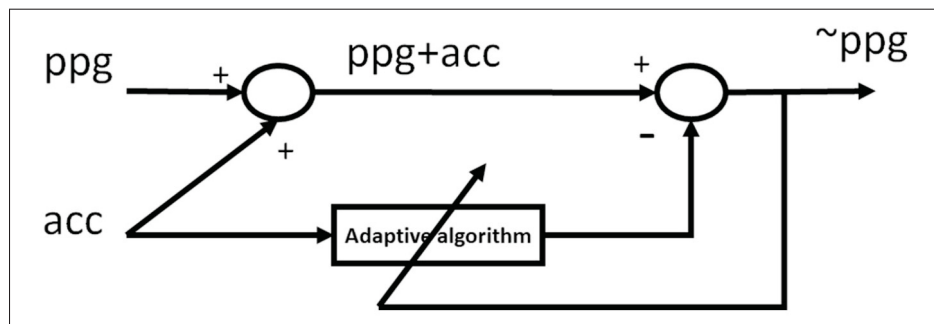


Figure 1.2 Présentation d'un filtre adaptatif des moindres carrés. PPG est le signal de photopléthysmographie et ACC est le signal issu de l'accéléromètre  
Tiré de Ismail *et al.* (2021)

Le filtre adaptatif utilise un algorithme qui met à jour en continu ses coefficients en fonction des différences entre le signal filtré et le signal cible. Sur la figure précédente, le signal cible est le signal  $\tilde{\text{PPG}}$  qui est donc un estimateur du signal PPG sans artefacts de mouvement. Cela permet au filtre de s'ajuster progressivement pour réduire ou éliminer les composantes indésirables du signal, tout en préservant les informations importantes.

Dans un premier temps, le signal est filtré par un filtre passe-bande. Un signal de référence des artefacts de mouvement, issu des données de l'accéléromètre, est généré. Le filtre des moindres carrés est ensuite utilisé pour adapter la valeur des coefficients du filtre en essayant de minimiser l'erreur entre le signal original (PPG) et le signal de sortie ( $\tilde{\text{PPG}}$ ).

Wu, Chen & Fang (2017) ont développé un algorithme qui enlève la composante continue du signal et utilise un algorithme des moindres carrés récursifs. Ce filtre élimine le décalage du signal PPG, ce qui permet à l'algorithme de filtrage des moindres carrés récursifs de converger plus rapidement par rapport à un algorithme simple des moindres carrés.

Les approches traditionnelles ont pour avantage leur facilité de mise en oeuvre mais présentent plusieurs limitations, notamment une sensibilité aux seuils empiriques et une adaptabilité limitée. Elles peuvent ne pas fonctionner aussi bien avec des mouvements complexes ou des scénarios inattendus.

### **1.3 Analyse temps-fréquence**

L'analyse temps-fréquence est une technique puissante utilisée pour étudier les signaux qui varient dans le temps et la fréquence, comme le signal PPG. Contrairement à une analyse classique à une dimension (souvent le temps), cette analyse se réalise en deux dimensions. Elle permet de décomposer un signal complexe en ses composantes fréquentielles à différents moments. Deux méthodes sont présentées par la suite : la transformée en ondelettes et la décomposition en modes empiriques. Pour certaines techniques, un signal de mouvement de référence doit être généré.

#### **1.3.1 Transformée en ondelettes**

La méthode de transformation en ondelettes discrète est la méthode en ondelette la plus utilisée pour détecter et enlever les artefacts de mouvement. Le signal PPG est un signal qui n'est pas stationnaire car la composition fréquentielle du signal varie au cours du temps. Ainsi, la transformée de Fourier n'est pas aussi performante contrairement à la transformée en ondelettes. Cette méthode ne requiert pas de signal de référence.

Joseph, Joseph, Titus, Thomas & Jose (2014) expliquent que la transformée en ondelettes divise le signal en une composante de basse fréquence et une composante de haute fréquence, chacune ayant la moitié de la longueur originale du signal. La composante de basse fréquence est

également connue sous le nom d'information lisse et le processus est répété sur cette composante, la divisant en sous-composantes basse et haute, et cela est répété plusieurs fois.

Plus précisément, on considère d'abord un filtre passe bas avec une réponse impulsionnelle  $g$ , le signal original  $x$  et  $n$  les échantillons discrets. Dans ce cas le signal résultant d'une convolution  $y$  peut s'écrire :

$$y[n] = (x * g)[n] = \sum_{k=-\infty}^{\infty} x[k]g[n - k] \quad (1.4)$$

De la même manière, le signal  $x$  est décomposé en utilisant un filtre passe-haut avec une réponse impulsionnelle  $h$ . Après chaque convolution, la moitié des fréquences du signal original sont enlevées, c'est le processus de décimation. Il est donc nécessaire de sous-échantillonner par 2.

La transformée en ondelettes discrète d'un signal  $x$  est calculée en utilisant deux filtres en cascade. De ces deux filtres passe haut et passe bas en quadrature, deux types de coefficients sont obtenus en sortie. Les coefficients obtenus en sortie du filtre passe bas sont appelés les coefficients d'approximation et ceux obtenus en sortie du filtre passe haut sont les coefficients de détail. Ces convolutions sont répétées en cascade sur la composante de basse fréquence jusqu'à ce que l'augmentation de la résolution en fréquence soit satisfaisante. (voir Figure 1.3).

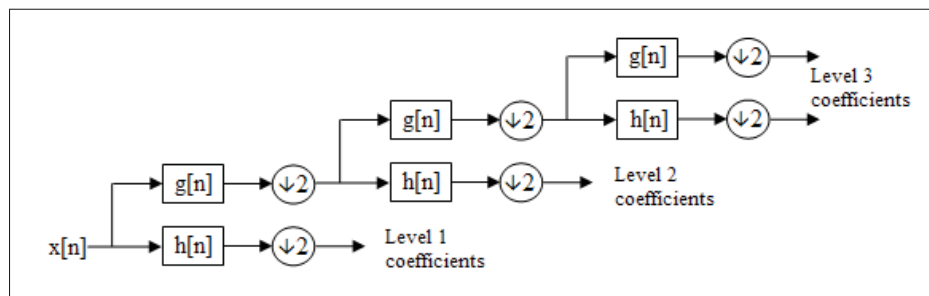


Figure 1.3 Exemple de filtres en cascade pour une décomposition discrète en ondelettes de niveau 3  
Tiré de Wikipedia (2023)

Une fois le signal décomposé, l'analyse des coefficients est effectuée pour identifier les artefacts. Les artefacts de mouvement se manifestent généralement par des composantes à haute fréquence et à petite échelle dans les coefficients. Des seuils sont définis pour identifier les coefficients qui

dépassent un certain niveau de sensibilité aux artefacts. Les coefficients qui dépassent ces seuils sont considérés comme des artefacts potentiels.

### 1.3.2 Décomposition en modes empiriques

Cette méthode consiste à décomposer un signal sur une base de fonctions. Dans ce cas, les fonctions ne sont pas données mais calculées à partir des propriétés du signal.

L'objectif de cette méthode est d'extraire les composantes intrinsèques appelées "modes empiriques" qui constituent un signal. Dans un premier temps, les valeurs maximales et minimales du signal sont identifiées et reliées entre elles pour former deux enveloppes : une enveloppe supérieure ( $U_{max}(t)$ ) et une enveloppe inférieure ( $U_{min}(t)$ ). La moyenne  $m(t)$  de ces deux enveloppes est calculée :

$$m(t) = \frac{U_{max}(t) + U_{min}(t)}{2} \quad (1.5)$$

Cette fonction moyenne est ensuite soustraite du signal original  $x(t)$  pour obtenir  $x_1(t)$  :

$$x_1(t) = x(t) - m(t) \quad (1.6)$$

Ce processus est itératif, en considérant à chaque nouvelle boucle que  $x_1(t) = x(t)$ . Les étapes de localisation des valeurs extrêmes, calcul et soustraction de la moyenne au signal sont répétées jusqu'à ce que  $x_1(t)$  rentre dans les critères d'un mode empirique. Ainsi,  $x_1(t)$  devient un mode empirique. Ce processus est répété jusqu'à ce que tous les modes empiriques du signal soient extraits (Raghuram, Sivani & Reddy, 2016).

D'après Wang, Yang & Zhang (2010), lorsque les modes empiriques du signal PPG sont obtenus, l'idée est de calculer la fréquence instantanée pour chaque mode afin de détecter les modes qui ont une fréquence proche des harmoniques d'un signal PPG et les modes qui sont caractéristiques des artefacts de mouvement.



Rojano & Isaza (2016) ont comparé l'utilisation des ondelettes avec la décomposition en mode empiriques et la décomposition en valeurs singulières pour estimer la fréquence cardiaque issue du signal PPG, des patients réalisant des activités sportives. En utilisant la décomposition en valeurs singulières, la première composante correspond à la composante de fréquence principale du signal PPG. Les autres composantes comportent donc les artefacts de mouvement, c'est l'avantage de cette décomposition. Les deux autres méthodes, la décomposition en mode empirique et la transformée en ondelettes, nécessitent des seuils ce qui est plus contraignant car cela induit des paramètres supplémentaires. En effet, en utilisant ces deux méthodes, il est courant de rendre nuls tous les coefficients qui sont inférieurs à un seuil pour réduire le bruit du signal.

L'analyse temps-fréquence a l'avantage de préserver l'information temporelle et fréquentielle du signal PPG, ce qui rend l'analyse du signal plus complète. En revanche, ces méthodes peuvent être plus lentes, ce qui est un obstacle lors de l'application en temps-réel ou pour le traitement d'une grande quantité de données.

#### **1.4 Apprentissage machine**

L'apprentissage machine est une méthode nouvelle pour la détection des artefacts. L'avantage de l'apprentissage machine est de pouvoir s'adapter automatiquement aux données sans nécessiter de seuils empiriques ou de paramètres manuels. Le fonctionnement des modèles débute par le choix des caractéristiques. Cette étape est primordiale car ces caractéristiques vont ensuite servir au modèle pour la reconnaissance des motifs. Une fois que ces caractéristiques sont sélectionnées, différents classificateurs peuvent être mis en place. Les classificateurs s'entraînent sur les données fournies pour prédire les segments qui comportent des artefacts et ceux qui n'en comportent pas. Un segment se définit comme une portion du signal généralement délimitée par des points spécifiques ou caractéristiques. Ces segments sont définis en fonction de critères pertinents à l'analyse du signal.

Les classificateurs tirent des modèles du nombre d'exemples données dans la partie entraînement. Dans le cas de données médicales, les données sont souvent limitées mais plusieurs algorithmes arrivent à avoir une bonne précision. Les classificateurs les plus connus sont basés sur un apprentissage machine supervisé, autrement dit l'apprentissage se fait à partir d'exemples annotés. Par exemple, les modèles SVM (Support Vector Machine), Decision Tree, KNN (K-Nearest Neighbors) peuvent être cités. Ces modèles se basent sur des principes simples, par exemple le modèle KNN classe un segment selon le résultat majoritaire des statistiques des classes de  $k$  voisins.

Chong *et al.* (2014) ont utilisé le modèle SVM pour détecter les artefacts. Ce modèle prend les paramètres des segments bruités et propres des signaux PPG et adopte une décision pour la classification de ces segments. C'est un outil de classification qui cherche à trouver la meilleure séparation entre différentes classes en maximisant l'espace entre celles-ci.

D'autres modèles utilisent des réseaux de neurones, c'est le cas des algorithmes MLP (Multi-Layer Perceptron) ou des Transformers. Les réseaux de neurones comportent plusieurs couches avec des noeuds, et chaque noeud ou neurone est connecté avec les autres. Ces algorithmes sont plus utilisés dans le cas de très grands ensembles de données. Goh *et al.* (2020) ont développé un modèle de convolution de réseaux de neurones à une dimension pour la détection des segments sans artefacts de mouvement sur les signaux PPG. Ce modèle est présenté comme robuste aux variations des signaux dues aux paramètres morphologiques spécifiques à chaque patient, qui modifient la forme d'onde du signal PPG.

## **1.5 Lien avec l'article**

L'article écrit dans le chapitre suivant permet d'utiliser de nouveaux concepts pour détecter les artefacts de mouvements. L'apprentissage machine est le coeur du sujet et le traitement du signal est utilisé dans la partie pré-traitement. L'émergence des techniques d'apprentissage machine pour le classification des signaux, qui sont basés sur les modèles utilisés pour le traitement des images par exemple, permet d'avoir une base solide de comparaison des méthodes. L'analyse

statistique, présentée en première partie de la revue de littérature, sera utilisée comme outil pour annoter les impulsions des signaux PPG. En effet, les annotations seront comparées avec les annotations d'un expert. Cet algorithme agit donc comme un second expert.

Les méthodes présentées dans la revue de littérature, de filtrage et d'analyse temporelles/spectrales, bien que longtemps utilisées pour détecter les artefacts dans les signaux PPG, présentent des limitations liées à leur degré de complexité dans la modélisation. En effet, ces méthodes s'appuient fréquemment sur des modèles préétablis pour la détection des artefacts. Cependant, cette approche peut s'avérer problématique lorsqu'il s'agit de gérer la diversité des artefacts dans des situations réelles. En comparaison, les approches basées sur l'apprentissage machine se distinguent par leur capacité d'adaptation et de généralisation aux variations complexes des signaux, ainsi que par leur aptitude à reconnaître les caractéristiques spécifiques des artefacts. L'apprentissage machine offre également une mise à jour continue des modèles et une meilleure gestion du bruit.

L'article se base sur l'utilisation de l'algorithme de propagation d'étiquettes, qui est un algorithme d'apprentissage machine semi-supervisé. L'efficacité de cet algorithme pour annoter une grande taille de données est démontrée, dans un premier temps. En annotant seulement 5% des données initiales grâce à l'aide d'un expert, l'algorithme montre des résultats de 90% pour la précision et le score F1 ainsi que 91% pour le rappel. Ces résultats sont très prometteurs pour mettre d'annoter une grande quantité de données automatiquement. Ensuite, cet algorithme est évalué en tant que classificateur pour être comparé à des classificateurs conventionnels d'apprentissage machine mais aussi des modèles de réseau de neurones, comme cités dans la dernière partie de la revue de littérature. Cet article permet ainsi de présenter l'algorithme de propagation d'étiquettes comme un algorithme fiable pour l'annotation des impulsions d'un signal PPG mais aussi comme un classificateur pour la détection des artefacts de mouvement. Il permet aussi de dresser une comparaison des principaux classificateurs d'apprentissage machine. Les classes de données étant déséquilibrées, elles comportent plus d'impulsions propres que d'impulsions avec artefacts, cet article évalue aussi les résultats des méthodes d'équilibrages de données pour des données médicales.



## CHAPITRE 2

### LABEL PROPAGATION TECHNIQUES FOR ARTIFACT DETECTION IN IMBALANCED CLASSES USING PHOTOPLETHYSMOGRAM SIGNALS

Clara Macabiau<sup>1</sup> , Thanh-Dung Le<sup>1</sup> , Kévin Albert<sup>2</sup> , Mana Shahriari<sup>2</sup> , Philippe Jouvét<sup>2</sup> , Rita Noumeir<sup>1</sup>

<sup>1</sup> Département de Génie Electrique, École de Technologie Supérieure,  
1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

<sup>2</sup> Centre de Recherche du CHU Sainte-Justine, CHU Sainte-Justine,  
3175 Chemin de la Côte-Sainte-Catherine, Montréal, Québec, Canada H3T 1C5

Article soumis au journal « IEEE Access », novembre 2023

#### 2.1 Abstract

This study aimed to investigate the application of label propagation techniques to propagate labels among photoplethysmogram (PPG) signals, particularly in imbalanced class scenarios and limited data availability scenarios, where clean PPG samples are significantly outnumbered by artifact-contaminated samples. Our research compares the performance of supervised classifiers, such as conventional classifiers and neural networks (Multi-Layer Perceptron (MLP), Transformers, Fully Convolutional Network (FCN)), with the semi-supervised label propagation algorithm for artifact classification in PPG signals. The results indicate that the label propagation algorithm achieves a precision of 91%, a recall of 90%, and an F1 score of 90% for the "artifacts" class, showcasing its effectiveness in annotating a medical dataset, even in cases where clean samples are rare. Although the K-Nearest Neighbors (KNN) supervised model demonstrated good results with a precision of 89%, a recall of 95%, and an F1 score of 92%, the semi-supervised algorithm excels in artifact detection. In the case of imbalanced and limited pediatric intensive care environment data, the semi-supervised label propagation algorithm is promising for artifact detection in PPG signals. The results of this study are important for improving the accuracy of PPG-based health monitoring, particularly in situations in which motion artifacts pose challenges to data interpretation.

## 2.2 Introduction

Machine learning, a sub-field of artificial intelligence (Helm *et al.*, 2020), has emerged as a transformative technology in various domains, including healthcare. With its ability to analyze large amounts of data (Dash, Shakyawar, Sharma & Kaushik, 2019) and finding important patterns, machine learning is a robust technique. It has the potential to improve healthcare outcomes, help doctors make better decisions (Gutierrez, 2020), and revolutionize medical research. Using advanced algorithms and computing power, machine learning can extract all necessary information from various types of healthcare data, such as electronic medical records (Ho, Ledbetter, Aczon & Wetzel, 2018), medical images, and physiological signals.

In clinical context, machine learning techniques can improve decision making. Support systems help with things like medication alerts and reminders for patient safety. Indeed, the model uses patient medical records and the vast amount of information available to develop models that aim to predict injuries (Sanchez-Pinto & Khemani, 2016), detect heart disease earlier (Choi, Schuetz, Stewart & Sun, 2017) and mortality (Huang *et al.*, 2019). Based on clinical experience and physiological principles, machine learning provides scores that enable hospital staff to make faster, more accurate decisions. Additionally, machine learning algorithms can contribute to drug discovery and development by expediting the identification of promising drug targets, optimizing drug efficacy, and predicting potential adverse reactions (Dara, Dhamercherla, Jadav, Babu & Ahsan, 2022).

Despite its potential, the integration of machine learning into healthcare comes with challenges and considerations. Privacy and ethical implications must be taken into account when applying machine learning to healthcare (Ngiam & Khor, 2016). The data acquired must respect patient privacy and confidentiality. This data is often recorded by several devices, so it is important to link the recordings so that the data is standardized and can be more easily managed. In order to be able to link these data, they need to be collected in a single database, and thus be consistent with each other. In other words, the data must be harmonized (Johnson *et al.*, 2016). Data security

and accountability should be prioritized, especially when using cloud-based technologies that require extensive storage and computing power.

One major concern is the availability of high-quality data for training and testing these algorithms (Habehh & Gohel, 2016). To be able to evaluate the performance of the algorithms implemented, it is necessary to have access to a ground truth. This ground truth is often non-existent or requires the help of experts, and is often time-consuming. Moreover, it's crucial to have a large enough sample size to ensure accurate and reliable predictions. However, obtaining such labeled data can be challenging, as it may be limited and incomplete. Furthermore, the imbalance in classes during the algorithm's training phase is a common consequence of medical data predominantly representing the most frequent cases present in patients. Consequently, a significant challenge arises in rebalancing the data whereas ensuring the retention of its crucial medical value. Additionally, the amount of erroneous, missing or imprecise data is an obstacle to model prediction. Erroneous data are data which have been altered by abnormal occurrences. Continuous waveform data captured by sensors is vulnerable to high-frequency artifacts that often result from patient motion or clinical interventions. Even periodic clinical measurements are not immune to contamination, as data collection and coding practices can introduce issues (Johnson *et al.*, 2016).

The main objective of this work is to detect motion artifacts in photoplethysmogram (PPG) signals obtained from the Pediatric Intensive Care Unit (PICU) database of the CHU Sainte-Justine hospital (CHUSJ). PPG signals are frequently captured during different types of movements which introduces motion noise and interferes with the accuracy of the signals. This noise is irregular and causes high-amplitude fluctuations within the PPG signals (Pollreisz & TaheriNejad, 2022). Motion artifacts can result in the pulse oximeter either misinterpreting movement as the actual signal or masking the true signal with unwanted interference, which can lead to incorrect readings, false alarms, and missed important alarms (Pettersen *et al.*, 2007). The cleaned PPG signals will be used in the construction of Clinical Decision Systems (CDSS) at the PICU of CHUSJ. Specifically, annotated signals will be used in screening and identifying various health-related concerns on children. As an example, changes in blood pressure in children serve

as significant indicators for identifying patients who require immediate care and admission to the PICU. Invasive methods, like catheter insertion for continuous blood pressure monitoring, offer precise real-time data but come with significant risks such as bleeding and infection (Kim *et al.*, 2014). On the other hand, conventional cuff-based measurements, though less invasive, provide only intermittent readings and may not capture sudden clinical changes effectively. Therefore, predicting blood pressure from PPG waveforms has emerged as a successful approach (Hill *et al.*, 2021) for comprehensive CDSS applications. In this project, we used a semi-supervised labeling algorithm to annotate PPG signals to effectively remove artifacts from them in real-time. A small proportion of data was annotated by an expert and a statistical analysis algorithm was then used to validate the annotations. To annotate the entire dataset, a machine learning algorithm that annotates all data using only a small proportion of previously annotated data, was used. The Label Propagation (LP) algorithm is a machine learning technique used for semi-supervised learning tasks. This algorithm is effective in our scenario because the availability of labeled data is limited and there is a large amount of unlabeled data (Bünger, Gondos, Peroche & Stoll, 2022).

Our medical data are very unbalanced, with around 80% of pulses free of artifacts and only 20% with artifacts. This means that to have an accurate labeling algorithm, a rebalancing of the classes in the training part need to be done. Several methods are available for this : oversampling, undersampling and, both oversampling and undersampling. It must be remembered that medical data is being worked with, so using sampling methods must make medical sense, whether by randomly duplicating data or by removing it. Medical data involves intricate relationships among different data elements, such as patient demographics, medical history, symptoms, diagnoses, treatments, and outcomes (Mazurowski *et al.*, 2008).

Another aim of this project is to compare classifiers to the Label Propagation algorithm, used as a classifier, to accurately detect artifacts. Xu & Funaya (2015) explore the use of semi-supervised models to classify temporal data. These models are based on a graphical approach, such as the label propagation algorithm. The results of the algorithm are evaluated on different datasets of varying lengths, including electrocardiogram (ECG) signal data. The results show that semi-supervised models are accurate for classifying time series data. However, these algorithms have



not been applied to artifact detection. Semi-supervised learning is widely used as a classification algorithm in cases where not all data is annotated. Active learning is also a very powerful semi-supervised classification method that has proven its effectiveness for temporal data (Shin *et al.*, 2021).

Machine learning classifiers are designed to categorize data into distinct classes or groups based on specific features or patterns. The use of classifiers has witnessed significant growth across different domains, especially in health care, where classifiers are a real help in decision-making (Robbins, Gurupur & Tanik, 2011). By leveraging statistical analysis and optimization techniques, classifiers enable accurate predictions and information retrieval. The spectrum of classifiers is very wide : from traditional classifiers like K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree (DT), Naive Bayes classifier (NB) (Aljaaf *et al.*, 2015), to classifiers using neural networks, such as Multi-layer Perceptron classifier (MLP) or Transformers. A comparison of the results of each type of classifier with the semi-supervised Label Propagation algorithm will be dressed. The effectiveness of these two streams is analyzed by the experimental results (in section 2.5) from the comparative analysis of semi-supervised label propagation (label propagation with KNN kernel) and fully-supervised learning, including conventional machine learning classifiers (KNN, Support Vector Classification (SVC), DT, Random Forest (RF), GaussianNB, MultinomialNB, and Logistic Regression (LR)), MLP, and Transformers. Then, the best classification method will be presented, followed by a conclusion on artifact detection.

The paper is structured as follows. In section 2.3, data characteristics, preprocessing and methodology : labeling and classification, are introduced. In section 2.4, the implementation of experiments is presented. Section 2.5 is used to evaluate the results with different metrics and present a comparison of experimental result tables. In section 2.6, the results are interpreted and the limitations are discussed.

## 2.3 Materials and Methods

This study was conducted following ethical approval from the research ethics board at CHUSJ (protocol number 2023-4556, accepted January 18, 2023). The detailed workflow of the various work stages is shown in the Figure 2.1.

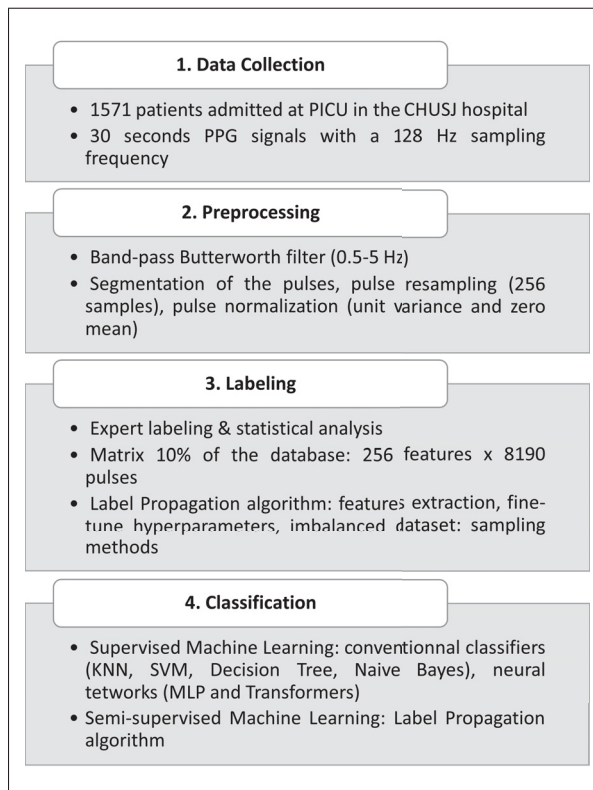


Figure 2.1 Workflow of the proposed method for detecting motion artifacts in PPG signals

### 2.3.1 Data Collection

The aim of this project is to detect motion artifacts in PPG signals. The eligible study population includes all children aged 0 to 18 years, admitted between September 2018 and July 2022 inclusive, for whom electrocardiogram (ECG), PPG and Arterial Blood Pressure (ABP) waveform records are available. In this population, specific exclusion criteria have been established to avoid bias. Data collected beyond the fourth day of hospital stay will be disregarded to prevent

potential bias from a few patients who may have prolonged stays with arterial lines. Patients on Extracorporeal Membrane Oxygenation (ECMO) treatment will also be excluded from the analysis. Furthermore, if a patient is readmitted to the PICU multiple times, only data from the first stay will be considered for analysis.

A PPG signal is recorded using a sensor called the pulse oximeter. This device is placed on a patient's skin, for children on a fingertip or earlobe. A PPG sensor emits light into the skin, which is partially absorbed by the blood vessels. Changes in blood flow during the cardiac cycle cause variations in light absorption. The sensor detects the reflected light, measuring its intensity modulated by blood volume changes. This varying intensity is converted into an electrical signal, creating the PPG waveform. Blood pressure signals are recorded using an invasive and continuous method, i.e., the catheter, and a non-invasive and discontinuous method, i.e., the blood pressure cuff. ECG is continuously recorded by placing electrodes on the patient's chest. The Sainte-Justine University Hospital Pediatric Intensive Care Unit (PICU) utilizes a High-Resolution research Database (HRDB) (Brossier *et al.*, 2018; Roumeliotis *et al.*, 2018) that has been approved by the ethical committee. The HRDB links biomedical signals extracted from the different devices, displayed through patient monitors, to the electronic patient record continuously throughout their stay in the unit (Mathieu, Sauthier, Jouvet, Emeriaud & Brossier, 2021).

Between 2018 and 2022, 1,571 patients met the inclusion criteria. For each patient, four physiological signals were extracted : ECG, PPG, blood pressure from the catheter, and blood pressure from the cuff. Each signal was extracted over a period of 96 hours (4 days). Signal values are grouped together in a table with the date and time of acquisition. For the PPG signal, 640 values are acquired every 5 seconds, corresponding to a sampling frequency of 128 Hz. For blood pressure and ECG signals, 2,560 values are acquired every 5 seconds, with a sampling frequency of 512 Hz. For the duration of the extraction, a fixed 30-second window of PPG signals will be used for the further processing.

### 2.3.2 Preprocessing

The raw PPG signal is preprocessed to increase its quality, remove unwanted noise, and make it more suitable for subsequent processing steps (Lim *et al.*, 2018). The different steps are described below :

1. **Filtering** : each signal window is filtered using a band-pass Butterworth filter, the cut-off frequencies are 0.5 and 5 Hz, corresponding to a heart rate between 30 and 300 bpm. A forward-backward filtering is used to avoid phase distortions. The objective is to remove baseline wander and high-frequency noise.
2. **Pulse segmentation** : a function to find all local minima by a comparison between samples is used. The aim is to divide the preprocessed PPG signal into smaller segments or windows to be able to detect the artifacts present for each pulse. In our case, a segment is a pulse. The size of each segment may vary depending on the characteristics of the PPG signal and the specific application the signal pulses. A pulse is considered to lie between two minima.
3. **Resampling** : the duration of a cardiac cycle for children is between 0.3 and 1 second. A pulse is representing a cardiac cycle. Therefore, not all pulses have the same number of samples. Each pulse is uniformly oversampled in time to contain 256 samples, which corresponds to a heart cycle of 1s. A linear interpolation function (Li & Clifford, 2012) is used to create the missing points for each pulse. Linear interpolation is favored for signals due to its simplicity, computational efficiency, and ability to estimate values between known data points. It maintains signal continuity and linearity, making it suitable for signals with relatively smooth and linear variations.
4. **Normalization** : the data are normalized to have a unit variance and zero mean. This normalization ensures that all features or variables in the data have the same scale, preventing certain features from dominating the learning process simply because they have larger numerical values.
5. **Data transformation** : each PPG pulse, which is essentially a waveform representing blood volume changes over time, can be represented as a data point in a column that contains 256 values. These values are equally spaced points obtained using step 3 of the preprocessing.

At the end of preprocessing, a vector of 256 points is obtained, representing a pulse of the PPG signal. The number of vectors depends on the number of pulses. This method allows to work with PPG data in a structured manner, suitable for a range of applications, from statistical analysis to machine learning.

The Fig. 2.2 shows the first 10 seconds of a raw PPG signal, then when the signal has been filtered and finally when the pulses have been segmented. The effect of the bandpass filter can be seen in the second figure. The filter has smoothed the signal by removing the extreme frequency components. The signal waveform is preserved, and the filter does not introduce resonances or significant ripples in the desired frequency range. Note that the first pulse has not been segmented. This is because the function could not detect the two minima that make up a pulse and therefore could not segment it. The signal does not start at the first low point of the pulse.

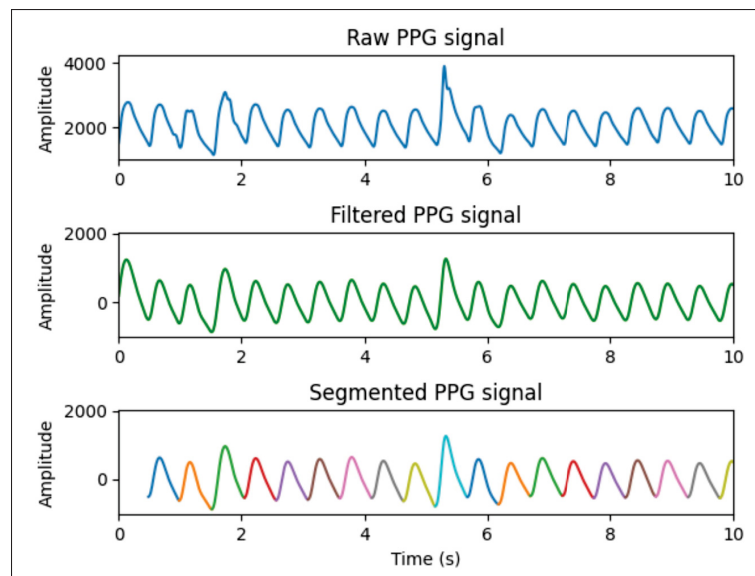


Figure 2.2 Example of a 10s segment of a 30s raw PPG signal in the top image, filtered signal in the middle image and segmented signal in the bottom image

### **2.3.3 Dataset annotation**

First, after preprocessing the data, the aim is to build a ground truth for future evaluation of the classification algorithms. In this objective, one human expert annotated 10% of the database. Annotation is visual, comparing pulses with each other and binary classifying each pulse as good or as containing artifacts. To avoid involving another human expert, in consideration of time and specialist resources, we implemented an automated algorithm to handle additional annotations. This algorithm, acting as a surrogate expert, was developed to recheck the entirety of the 10% annotated data by the human expert, identifying similarities in the process. Employing a statistical approach, the algorithm determines if the values of a given pulse lie within standard parameters or deviate from the norm. We subsequently cross-validated the algorithm's annotations against those from the professional expert to determine the algorithm's accuracy. It was decided to annotate a maximum of 10% of the database and then use the label propagation algorithm to annotate the rest of the data.

#### **2.3.3.1 Expert labeling**

PPG signals already segmented are presented to the expert. By analyzing each pulse, the expert classifies each pulse as artifact or artifact-free. A pulse is defined as artifact-free if its morphology is typical i.e. if its characteristics - amplitude, width, shape - are the same as those of adjacent signals. A pulse is defined with artifacts if its characteristics differ from those of adjacent pulses (see Fig. 2.3). To recheck the annotations of one expert, an algorithm that acts as a second expert was set up, allowing all impulses to be reannotated to see similarities. This algorithm uses a statistical approach to assess whether the statistical values of a pulse are normal or outside the norm.

#### **2.3.3.2 Statistical analysis**

For each cardiac cycle, which corresponds to a pulse, if the waveform is similar, statistics such as skewness, standard deviation (std) and kurtosis are approximately constant for each cycle. It is

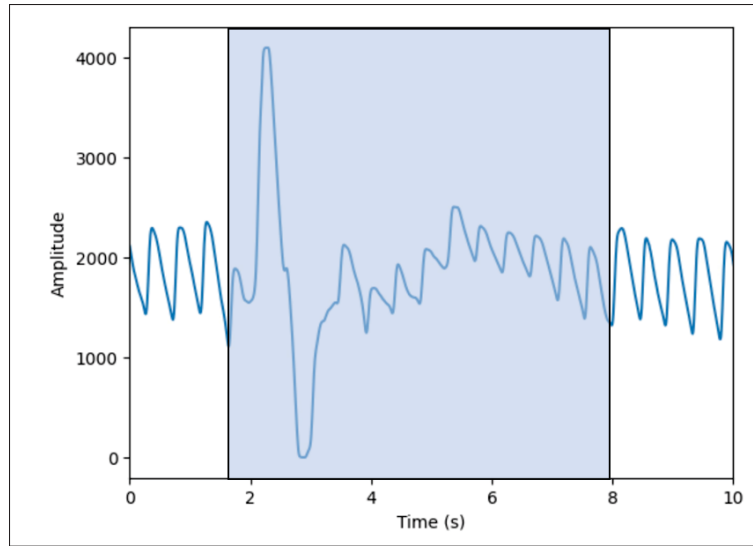


Figure 2.3 Example of a 10s segment of a 30s raw PPG signal - Inside the blue box are all the pulses containing motion artifacts

therefore possible to detect motion artifacts by using the value of these statistics to differentiate a pulse without artifacts from a pulse with motion artifacts (Hanyu & Xiaohui, 2017). Skewness indicates the degree of asymmetry in the probability distribution of a random variable around its mean. It can take on positive, zero, negative, or undefined values, reflecting the shape and symmetry of the distribution. Kurtosis is the sharpness of the peak of a frequency-distribution curve and standard deviation reflects the dispersion degree of a data set. If  $X$  is considered a variable with  $\mu$  and  $\sigma$  the mean and standard deviation respectively, statistical values are calculated as follows :

$$\text{Kurt}[X] = E \left[ \left( \frac{X - \mu}{\sigma} \right)^4 \right] = \frac{E [(X - \mu)^4]}{\sigma^4} \quad (2.1)$$

$$\text{Skew}[X] = E \left[ \left( \frac{X - \mu}{\sigma} \right)^3 \right] = \frac{E [(X - \mu)^3]}{\sigma^3} \quad (2.2)$$

$$\text{std}[X] = \sqrt{E [(X - \mu)^2]} \quad (2.3)$$

These values are calculated for each pulse of a signal. So in our case, the variable  $X$  represents a vector of all the samples in a pulse. If the shape of the cycle changes, then these statistical values will no longer be constant. To be able to detect outliers, thresholds that detect skewness, kurtosis and standard deviation values that are not normal, i.e. values for artifact-free cycles, were set up. For this reason, the distribution of each of these three statistics over a pulse can be estimated using a normal distribution (Krishnan *et al.*, 2008). The aim is to reduce the risk of a pulse being incorrectly annotated. To do this, a wide confidence interval is taken to ensure that the probability that the value of the corresponding statistic is not unnecessarily rejected. If  $X$  is considered to be a variable which can be approximated by a normal distribution  $\mathcal{N}(\mu, \sigma^2)$ , the probability that this variable lies within the chosen confidence interval can be written as follows :

$$\mathbb{P}(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.9545 \quad (2.4)$$

After several experiments, this 95% confidence interval gives the best results, as it reduces the risk of poor detection. So a lower and an upper thresholds can be defined as follows :

$$th_l = \mu - 2\sigma \quad (2.5)$$

$$th_u = \mu + 2\sigma \quad (2.6)$$

With the mean and standard deviation calculated for each statistic measure by taking the set of values for each pulse of a signal. To effectively detect motion artifacts, a waveform segment is classified as containing motion artifacts if at least one of the three statistics falls outside the defined thresholds. The result of this first step is a small proportion of the annotated dataset, with a binary value for each pulse : pulse with artifact or without artifact. The annotations given by the algorithm are then compared with the expert's annotations and found to have 80% similarity. After examining the annotations with the expert, the function chosen to segment the pulses did not always correctly segment a pulse that was formed by a distinct diastole and systole curve. In a cardiac cycle, diastole is the relaxation phase when the heart fills with blood, and systole is the contraction phase when the heart pumps blood out to the body or lungs. In this case, two



pulses were detected instead of one. This segmentation error partly explains the 20% difference in annotation between the expert and the algorithm. The percentage of similarity is considered high enough to validate the expert's annotations.

### **2.3.3.3 Imbalanced dataset**

The two classes of annotated data are unevenly distributed. In fact, the annotation includes many more pulses without motion artifacts, approximately 80% and 20% of pulses with motion artifacts. For accurate results with the algorithms, the data needs to be resampled. The complex characteristics of our clinical data : small training size, many features, and correlations between the features, makes the task more complicated. Understanding the interconnectedness of these variables is crucial for accurate analysis and prediction (Mazurowski *et al.*, 2008). Oversampling and undersampling methods are the most frequently used. Under-sampling reduces the majority class examples, achieving a balanced dataset, with random under-sampling (RUS) being a well-known method. However, under-sampling may lead to the loss of valuable information from the majority class. On the other hand, over-sampling increases the minority class examples. Random over-sampling (ROS) replicates existing minority examples, but it may result in overfitting. Synthetic minority over-sampling technique (SMOTE) generates artificial minority examples by interpolating between selected examples and their nearest neighbors. Modifications such as adaptive synthetic sampling (ADASYN) adjust the number of artificial minority examples based on the density of majority examples surrounding the original minority example (Fujiwara *et al.*, 2020). Also, it is concluded that there is no clear winner between oversampling and undersampling to compensate for the class imbalance if factors such as class distribution, class prevalence, and features correlations in medical decision-making (Mazurowski *et al.*, 2008) are not taken into consideration. In the section 2.5 the different results obtained with the sampling methods will be presented, to conclude on the best method for our study.

### 2.3.3.4 Label Propagation

The Label Propagation algorithm is an iterative algorithm that assigns labels to unlabeled data points by propagating labels through the dataset (Zhu & Ghahramani, 2002). In graph-based semi-supervised learning methods, a graph where each node is represented by a vector of features is created. The edges between nodes are weighted based on how similar the features are. When the weights of the edges are high, it means that the connected nodes are likely to have the same label. This idea is based on the assumption that samples close to each other in the graph are part of the same group or category (Song, Yang, Xu & King, 2022). At the start of the algorithm, only a small proportion of the data is already labeled, corresponding here to the proportion of data annotated in the previous step. In our case, considering that we have around 51 pulses per signal and that we have annotated 10% of the entire database of 1571 signals, we therefore have 8000 pulses, and thus 8000 nodes in the graph. This algorithm is based on the hypothesis that if two nodes are connected that means they carry a similarity. Usually, the Euclidean distance between nodes is calculated to establish the graph. Depending on the kernels chosen for the algorithm's operation, this distance measurement may be different. Consider the following notations :

$u$  : number of unlabeled points

$l$  : number of labeled points

$k$  : number of classes

The aim of this algorithm is to look, in the final state, at all the probabilities a node has of belonging to a certain class and to take the largest.  $Y$  a matrix with rows containing the probabilities that a node belongs to a certain class is considered. This matrix  $Y$  is a  $N \times k$  matrix where  $N = l + u$ . Also considered  $T$ , a  $N \times N$  probability transition matrix. This matrix  $T$  is obtained by calculating the degree matrix ( $D$ ) and the adjacency matrix ( $A$ ). It defines the probability of jumping from one node to another in  $t$  steps. This number  $t$  can tend towards infinity (Bodo & Csató, 2015). The matrix  $Y$  contains two sub-matrices :  $Y_l$  and  $Y_u$ , respectively for the known and unknown labels. The same applies to the  $T$  matrix, which contains 4 sub-matrices :

- $T_{ll}$  : probability to get from labelled nodes to labelled nodes. This matrix will be an identity matrix.
- $T_{lu}$  : probability to get from labelled nodes to unlabelled nodes. This will be a zero matrix because labelled nodes are absorbing states, it means you are in a self-loop and can't move in any direction.
- $T_{ul}$  and  $T_{uu}$  : probability to get from unlabelled nodes to labelled and unlabelled nodes respectively.

Consider  $\hat{Y}$ , the probability matrix of annotations obtained in the final state. The matrix  $T$  is set to the infinite power and  $Y_0$  represents the initial annotations of the nodes. The equation for the final state of this algorithm can be expressed as :

$$\hat{Y} = T^{t \rightarrow \infty} Y_0 \quad (2.7)$$

In 2.7, the matrix  $T$  is set to the power  $t$  with  $t$  tending to infinity. It can be written :

$$\lim_{t \rightarrow \infty} T^t = \begin{bmatrix} I & 0 \\ (\sum_{t=0}^{\infty} T_{uu}^t) T_{ul} & T_{uu}^{\infty} \end{bmatrix} \quad (2.8)$$

The sum between the brackets is similar, when  $t$  tends to infinity, to a geometric series that has an argument that is less than 1 in modulus. And if  $T_{uu}$  is multiplied by itself a large number of times, knowing that the values are less than 1, it will become very close to 0. Therefore, a conclusion on the limit of the transition matrix for a very large number of steps, is :

$$\lim_{t \rightarrow \infty} T^t = \begin{bmatrix} I & 0 \\ (I - T_{uu})^{-1} T_{ul} & 0 \end{bmatrix} \quad (2.9)$$

The equation 2.7 can, therefore, be rewritten :

$$\begin{bmatrix} \hat{Y}_l \\ \hat{Y}_u \end{bmatrix} = \begin{bmatrix} I & 0 \\ (I - T_{uu})^{-1} T_{ul} & 0 \end{bmatrix} \begin{bmatrix} Y_{l0} \\ Y_{u0} \end{bmatrix} \quad (2.10)$$

For unknown labels, the following formula can be written :

$$\hat{Y}_u = (I - T_{uu})^{-1} T_{ul} Y_{l0} \quad (2.11)$$

This matrix contains the new labels and is the output of the algorithm. To sum up, the various stages of the algorithm can be summarized as follows :

1. Creation of a graph with nodes labeled and unlabeled.
2. Calculation of the probability transition matrix  $T$ . This matrix is linked to the degree matrix  $D$ , it is a diagonal matrix where each diagonal element corresponds to the sum of edge weights connected to that node. And also linked to the adjacency matrix  $A$ , it is a square matrix where each row and column corresponds to a node, and the value at the intersection indicates whether there's an edge (value 1, otherwise 0) connecting those nodes. The formula is :  $T = D^{-1} \cdot A$ . This matrix is the same throughout the algorithm.
3. Calculation of the new labels for each  $t$  iteration :

$$Y^{t+1} = T^t Y^t \quad (2.12)$$

4. Repeat step 3 until convergence.

While performing label propagation, groups of closely linked nodes quickly reach a consensus on a single label, causing many labels to vanish. This leaves only a few labels remaining after propagation. When nodes end up with the same label after convergence, it signifies that they are part of the same group.

### 2.3.4 Classification

Once the ground truth has been established, the aim is to classify the pulses and compare the results obtained with the annotations. Machine learning classifiers are used for classification. These automatic algorithms categorize data into the two classes of our problem. They operate as mathematical models, utilizing statistical analysis and optimization techniques to detect patterns within the data. By identifying these patterns, classifiers can assign each instance to a

specific class or category. There are a wide variety of traditional classifiers, both supervised and unsupervised. To process medical data, which is also temporal data, supervised classifiers have been chosen to be utilized. Here are 4 examples (Aljaaf *et al.*, 2015) :

1. **K-Nearest Neighbor (KNN)** : this is a supervised method where k represents the number of neighbors. For classification, when given a new input data point, the algorithm identifies the k nearest neighbors from the training dataset based on their feature similarity. The class label of the majority of these k neighbors is then assigned to the new data point ;
2. **Support Vector Machine (SVM)** : in SVM, data points are mapped as vectors within a high-dimensional space. The algorithm aims to identify the optimal hyperplane that distinctly categorizes the classes. In binary classification, a hyperplane can be considered as a boundary delineating two distinct data classes. While numerous hyperplanes might achieve this separation, the algorithm selects the one that provides the most effective separation. For a specific classification purpose, as is the case for this project, we have subsequently used SVC (Support Vector Classification) which is a type of SVM specialized for classifications ;
3. **Decision Tree** : each internal node represents a feature or attribute, and each branch represents a decision rule based on that feature. The leaf nodes of the tree represent the final class label or predicted value. During training, each value is separated based on the attribute. When making predictions, new data points traverse the decision tree by following the decision rules at each node until reaching a leaf node, which then provides the predicted class label or value ;
4. **Naive Bayes classifier** : this classifier use probability to predict whether an input will fit into a certain category. It builds a statistical model based on these probabilities. Naive Bayes calculates the likelihood of the data point belonging to each class using the previously estimated probabilities.

Traditional classifiers have big advantages for small or medium datasets that require simpler or linear models. They have few layers in their architecture, on the other hand, deep learning (MLP and Transformers) architectures are comprised of multiple layers of neural networks. Deep architectures take advantage of unsupervised pre-training at layer level, which facilitates

efficient tuning of the deep networks and enables them to extract intricate structures from input data. These extracted features at higher levels contribute to improved predictions and overall performance (Miotto, Wang, Wang, Jiang & Dudley, 2018). For classification, MLP and Transformers are neural networks classifiers :

- **Multilayer Perceptron (MLP)** : it consists of multiple layers of nodes (neurons) that are interconnected through weighted connections. MLP employs a feedforward mechanism, where information flows from the input layer through the hidden layers to the output layer. Each node in the network applies an activation function to the weighted sum of its inputs to produce an output. Through a process called backpropagation, the MLP classifier adjusts the weights to minimize the error between predicted and actual labels during training ;
- **Transformers** : it relies on the attention mechanism. The attention-mechanism looks at an input sequence and decides at each step which other parts of the sequence are important. Transformer is an architecture for transforming one sequence into another one with the help of two parts (Encoder and Decoder).

The objective is to apply all these classifiers to the PPG signal pulses so that a comparison of the classifiers on our medical data can be built. In addition to being compared with each other, these classifiers will also be compared with the Label Propagation semi-supervised algorithm, which not only annotates our database but also classifies artifacts in PPG signals.

## 2.4 Experimental Implementation

First, as a reminder, in the Label Propagation algorithm, the two input matrices are the annotation matrix, which is a binary vector, and a matrix containing the features for each pulse. Each pulse represents a node in the algorithm's graph. For the choice of features, the signal from a temporal perspective has been considered. Therefore, an input matrix for the algorithm of size  $256 \text{ samples} \times \text{the number of pulses}$ , can be obtained.

To be able to evaluate our results different metrics have been chosen. The negative state (0) is a pulse without motion artifacts, whereas the positive state (1) is a pulse with motion artifacts. All

these measures are based on the evaluation of false negatives (FN), pulse with artifact incorrectly identified as a clean pulse, false positives (FP), clean pulse incorrectly identified as a pulse with artifact, true negatives (TN), clean pulse correctly identified as a clean pulse, and true positives (TP), pulse with artifact correctly identified as a pulse with artifact. The following metrics are defined :

- **confusion matrix** : a table with two rows and two columns that reports the number of true positives, false negatives, false positives, and true negatives ;
- **precision, recall & F1** : these three scores give a more general idea of how the algorithm works, rather than just looking at the algorithm's accuracy, which can be biased in certain situations. They are defined as follows :

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

- **AUROC** : the AUROC (Area Under the Receiver Operating Characteristic curve) represents the probability that the model correctly ranks a randomly chosen positive instance higher than a randomly chosen negative instance. The ROC curve is created by plotting the TP rate against the FP rate.

Using these metrics, the hyperparameters of our model for the Label Propagation algorithm need to be defined. The first parameters defined are the parameters of the function used. The choice of kernel is between KNN or RBF (radial basis function), and depending on this choice, two other associated parameters could be modified. The maximum number of iterations and the algorithm's convergence tolerance remained the default values : 1,000 iterations and  $10^{-3}$  for the threshold of convergence. The first parameter to be defined was the choice of kernel. For this, a cross-validation was carried out on the data. This involves dividing the data into several parts, then running the two algorithms using different values for the parameters on each part, keeping

one part aside for performance testing. Then a calculation of the average performance can be done over all the test parts for each value and choose the one that gives the best performance. A KNN kernel with a number of neighbors of 7 has been chosen. The data are separated as follows : 70% training and 30% testing. The data from the training part are redivided evenly to obtain 50% of unlabeled data and 50 % of labeled data.

## 2.5 Results and Discussion

To optimize the Label Propagation algorithm to achieve the best performance on automatic labeling, different proportions of the dataset were tried for annotation. The aim is to annotate as few pulses as possible. 2.5%, 5%, 7.5%, and 10% of the dataset were annotated, given that the entire database contains 1,571 signals, and the proportion that gave the best results was evaluated. For each proportion of the dataset the precision, recall and F1 values were analyzed to decide. The results for the class 1 (class "pulse with artifacts") of these metrics are presented in the table 2.1.

Tableau 2.1 Results for the class  
"with artifacts"  
for different proportions of the dataset

Dataset proportion	Precision	Recall	F1
2.5%	0.89	0.88	0.89
5%	0.91	0.90	0.90
7.5%	0.84	0.88	0.86
10%	0.83	0.90	0.86

Because the data are imbalanced, the results for the class 0 (class "pulse without artifacts") remain consistently good and don't change much with different parameters. The best results are obtained for a proportion of 5% of the dataset. Indeed, as the proportion of annotated data increases, the distribution of classes becomes even more disparate. For 2.5% there are 17.3% of pulses with artifacts, and for 5% the proportion of pulses with artifacts is 18.1%. As the size of the annotated dataset increases, for 7.5% there are 16.4% of pulses with artifacts. For 10% of the dataset, 17.7% of pulses contain artifacts. All these values are summarized in the table 2.2.



Tableau 2.2 Summary table containing the labeling portion and the imbalance rate

Dataset proportion	Artifacts (%)	Non-artifacts (%)
2.5%	17.3	82.7
5%	18.1	81.9
7.5%	16.4	83.6
10%	17.7	82.3

If the classes are more unbalanced, this may have an influence on the algorithm, which will have greater difficulty in finding a constant pattern for propagating the labels. In the case of 10%, the proportion of pulses with artifacts is high, but the number of annotated pulses increases, and this may induce new data that is less representative of the overall data distribution, leading to poor generalization on unseen data.

The various resampling methods presented in section 2.3.3.3 were applied. The results are shown in table 2.3.

Tableau 2.3 Results for the class "with artifacts" for different sampling methods

Sampling method	Precision	Recall	F1
None	0.96	0.82	0.89
RUS	0.87	0.90	0.88
ROS	0.91	0.87	0.89
SMOTE	0.91	0.90	0.90
ADASYN	0.88	0.91	0.90
ROS+RUS	0.89	0.91	0.90

The difference in results between undersampling and oversampling can be explained by the fact that undersampling will reduce the number of majority, this leads to loss of data and loss of information from this data. On the contrary, oversampling increases the number of values in the minority class, providing more data. In our case, SMOTE is the best oversampling method. SMOTE selects a minority class instance and identifies its k-nearest neighbors in the feature space. It then creates new synthetic examples along the line segments connecting the selected instance and its neighbors. By introducing these synthetic examples, SMOTE effectively

increases the size of the minority class, making it comparable to the majority class and improving the performance of classifiers in handling imbalanced datasets.

Given the correct sampling method and the appropriate proportion of the dataset to be selected, the results of the Label Propagation algorithm were evaluated. The confusion matrix is shown on the left and the ROC curve on the right, on the Figure 2.4.

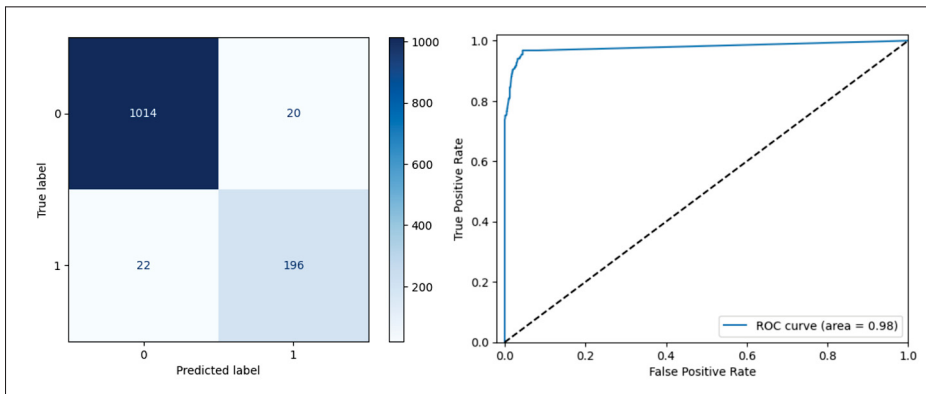


Figure 2.4 Confusion matrix and ROC curve for the Label Propagation algorithm with a KNN kernel with 7 neighbors, an oversampling method SMOTE and 5% of the dataset already labeled

The ROC curve is plotted for each decision threshold. In the case of the label propagation algorithm, this represents the probability assigned to each instance for each class. For a 5% dataset, the number of pulses for the validation part is 1,252 pulses. 1,034 belong to the "without artifacts" category and 218 to the "with artifacts" category. The number of true positives and true negatives is higher than the number of false positives or false negatives. This indicates that the algorithm is correctly able to understand the model and apply it to unlabeled signals. However, the number of pulses detected as clean but actually containing artifacts (false negatives) is higher than the reverse (false positives). This is because the Label Propagation algorithm uses neighborhood information to propagate labels through the data network. This means that labels for samples close in feature space tend to be similar. However, in the case of pulses containing artifacts, these artifacts may be similar to certain features of other clean pulses, leading to incorrect label propagation. As a result, pulses containing artifacts may be incorrectly labeled as

clean by the Label Propagation algorithm, leading to a higher number of falsely classified pulses. Having misclassified pulses is always a problem that is important in the medical field. This can lead to false alarms, if the pulse is not a clean pulse, or to misdetections. False alarms force hospital staff to make emergency visits due to outliers. These situations are exhausting, and not necessary as an additional burden on caregivers. The area under the ROC curve (AUROC) is 0.97. The closer the AUROC is to 1, the better the model's performance. A high AUROC indicates that our model is able to discriminate well between positive and negative classes.

After evaluating the Label Propagation algorithm, the performance of the different types of classifiers, presented in section 2.3.4, were assessed. First, dealing with the imbalanced classes by oversampling using Adaptive Synthetic (ADASYN) algorithm. Cross-validation was employed to ensure accurate model prediction and assess the reliability of the machine learning algorithms. The results of the 5-fold cross-validation on different classifiers are presented in Figure 2.5, which shows a precision comparison using a box plot. Each blue dot represents the performance of an individual fold in the cross-validation.

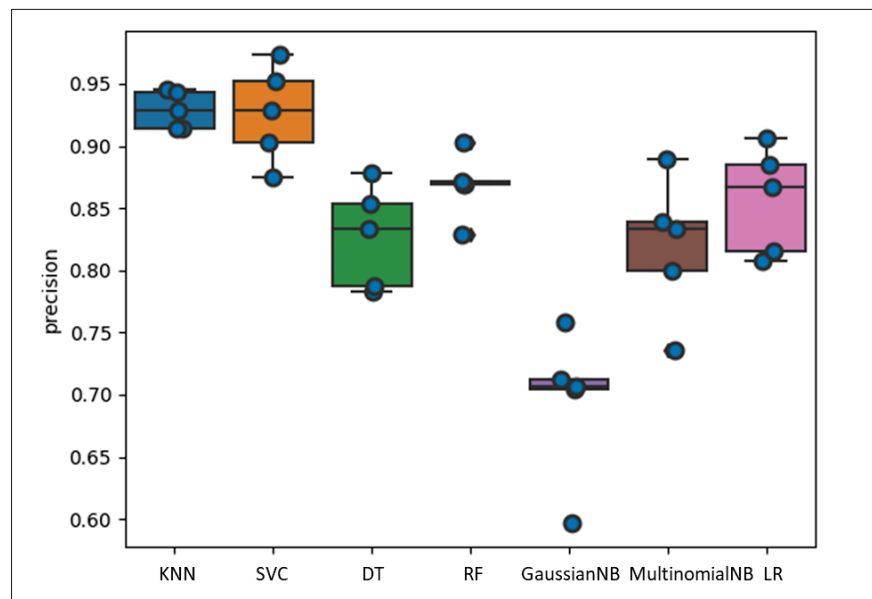


Figure 2.5 Precision evaluation for different traditional Machine Learning classifiers (in axis order : KNN, SVC, Decision Tree, Random Forest, Gaussian NB, Multinomial NB, Logistic Regression)

The figure indicates that KNN and SVC (with a kernel of 'rbf') are the top-performing classifiers, with median precision rates above 90%. However, KNN shows a slightly better and more consistent performance than SVC. This observation is highlighted by the broader range of variability for SVC, as indicated by the whiskers on the box plot, compared to KNN. To sum up, KNN and SVC are the top classifiers, but KNN is the more reliable and stable solution. KNN is also the best classifier compared to classifiers that use neural networks such as MLP and Transformers. Table 2.4 shows the different performances of the neural networks classifiers : MLP classifier, Transformer, ConvolutionNN VS KNN classifier.

Tableau 2.4 A comparison of the performance of different classifiers for the class "with artifact"

Model	Precision	Recall	F1	AUROC
KNN	0.89	0.95	0.92	0.98
MLP	0.76	0.97	0.85	0.99
Transformer	0.85	0.86	0.85	0.97
ConvolutionNN Baseline Classifier	0.86	0.83	0.84	0.95

MLP consists of 3 hidden layers with 500 neurons for each hidden layer. Its macro average accuracy (calculates the accuracy for each class individually and then computes the average accuracy across all classes) is 0.88, compared with 0.94 for KNN. In our case, using a complex model like MLP could lead to overfitting, as the model may have a high capacity relative to the amount of data available. In addition, training an MLP can be computationally expensive, especially with larger architectures and limited computational resources. Transformers, especially large ones like BERT (Bidirectional Encoder Representations from Transformers), have a high computational complexity and require significant computational resources for training and inference. Like the MLP classifier, Transformers works best on larger datasets because it needs a lot of data for the training part, otherwise the model has a greater capacity than the limited number of data and the risk is overfitting. Generally speaking, in the medical field, Transformers excel in natural language processing tasks (Yogarajan, Montiel, Smith & Pfahringer, 2021). They can learn complex relationships and patterns within the text, making them suitable for tasks involving medical text classification and understanding.

For the last classifier, experiments were conducted with a Fully Convolutional Network (FCN). FCN is a neural network architecture designed for semantic segmentation, producing dense pixel-wise predictions. It consists of convolutional layers without fully connected layers, enabling it to handle images of any size and preserve spatial information. Using FCN for time series classification involves adapting the fully convolutional architecture to process one-dimensional time series data. Instead of working with two-dimensional images, the FCN is applied to sequences of data points. The temporal convolutional layers capture temporal patterns and dependencies in the time series, and the decoding path with transposed convolutional layers helps to produce dense predictions for each data point in the sequence, enabling accurate time series classification (Wang, Yan & Oates, 2017). One key benefit is that FCN eliminate the need for manual feature engineering, as they can directly learn relevant features from raw time series data. This streamlines the classification process and saves time and effort in designing handcrafted features. Additionally, FCN enable end-to-end learning, optimizing feature representations and classification jointly, which can lead to improved performance. The flexibility of FCN with input size allows them to handle time series data of varying lengths without requiring resizing or padding, making them suitable for irregularly sized data. Moreover, FCN produce dense predictions for each time step, capturing fine-grained temporal patterns and enhancing the informativeness of classification results. Experimentally, during FCN training, it is evident that the process takes longer than other approaches. However, its performance is not comparable to those methods, mainly due to its lower accuracy. Moreover, the Fig. 2.6 shows that both the validation and training accuracy display fluctuations, indicating a lack of stability in its performance.

## **2.6 Conclusion**

In conclusion, this study delved into utilizing semi-supervised label propagation (LP) methods for artifact classification within PPG signals, especially in scenarios characterized by imbalanced class distributions. To validate the effectiveness of this approach, we compared the results from the semi-supervised LP method against those obtained using traditional supervised learning

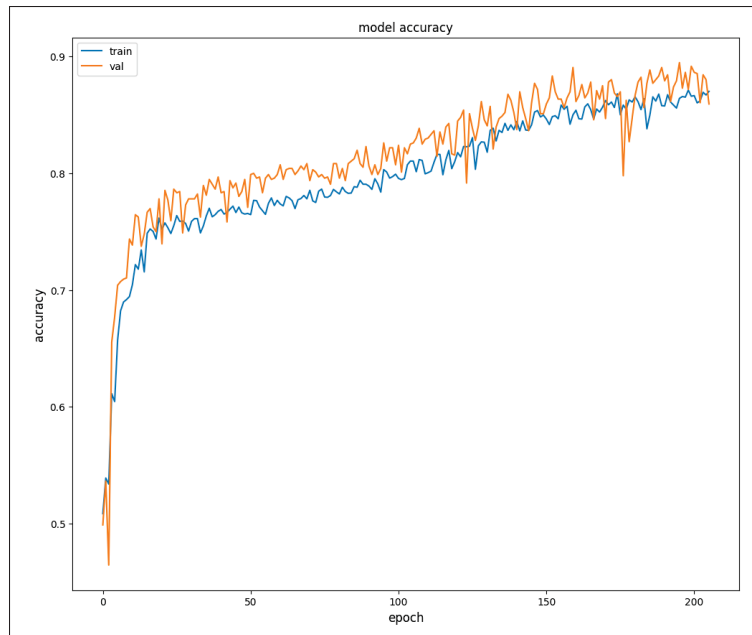


Figure 2.6 Accuracy curve during training and validation for the FCN model

algorithms, MLP, and Transformer-based models. Based on the experimental results, with oversampling method, the Label Propagation classifier performs better than neural network classifiers, MLP, FCN, and Transformers, achieving precision, recall, and F1 scores of 91%, 90%, and 90%, respectively. In particular, the performance of Label Propagation is comparable to that of KNN (k=3), which attains a higher recall score of 95% and an F1 score of 92%. However, the label propagation performs more consistently with a balanced precision-recall trade-off, whereas the KNN classifier struggles in that aspect. For improvements and future work, some directions can be given :

- improving the preprocessing : in section 2.3.3.2, the problem of segmentation has already been mentioned. Some steps can be added to the preprocessing part. First, adaptive filtering techniques can be used to attenuate artifacts without affecting the signal. Signal quality can be improved through noise reduction methods, such as singular value decomposition (SVD). To enhance the efficiency of the statistical analysis algorithm, alternative segmentation approaches can be employed. Peak or minima detection can be improved by employing

derivative-based algorithms. A CNN model can also be used to detect peaks, known for its pattern recognition capabilities ;

- machine learning approaches : adapting different methods for tackling labeling limitation like self-supervised or unsupervised learning approaches ;
- data augmentation : exploring data augmentation methods to increase data availability for improving the classifier's generalization.

Overall, the use of label propagation techniques for labeling and classifying artifacts detection in PPG signals presents a promising direction for more accurate and reliable healthcare monitoring systems.





## CONCLUSION ET RECOMMANDATIONS

En conclusion, la détection des artefacts dans les signaux PPG revêt une importance cruciale pour garantir la fiabilité et la précision des analyses cliniques et de recherche. Au cours de cette étude, nous avons utilisé différentes méthodes de détection des artefacts de mouvement, allant des techniques basées sur le traitement du signal aux approches avancées d'apprentissage automatique.

L'objectif de la méthode proposée était de comparer les classificateurs utilisés en apprentissage machine pour classifier les impulsions des signaux PPG. Dans un premier temps, nous avons pré-traité les signaux acquis en utilisant des méthodes traditionnelles en traitement du signal : filtrage, segmentation, normalisation et échantillonnage.

Ensuite, un algorithme d'apprentissage machine semi-supervisé, l'algorithme de propagation d'étiquettes, a été mis en oeuvre pour annoter une base de données de signaux PPG. Étant donné le déséquilibre des données, différentes méthodes de rééquilibrage ont été comparées. La méthode de suréchantillonnage se révélant la plus appropriée pour les données médicales. Les résultats ont montré que malgré le déséquilibre initial, l'algorithme de propagation d'étiquettes a permis d'obtenir des performances satisfaisantes en minimisant le nombre de faux positifs et de faux négatifs.

Enfin, une comparaison des performances des classificateurs supervisés a été effectuée, mettant en lumière le modèle KNN comme le plus performant parmi les modèles traditionnels, en raison de sa compatibilité avec des ensembles de données de petite taille. Les modèles utilisant des réseaux de neurones tels que MLP et Transformers ont montré des résultats inférieurs, principalement en raison de la surcomplexité pour notre ensemble de données limité. Bien que le modèle KNN ait obtenu des scores plus élevés dans certaines mesures, la stabilité globale des scores favorise l'utilisation de l'algorithme de propagation d'étiquettes en tant que classificateur. Ainsi, l'algorithme de propagation d'étiquettes s'est révélé être le modèle le plus performant

pour la classification des artefacts de mouvement dans les signaux PPG, offrant une solution robuste malgré les défis posés par les données médicales.

Malgré tout, comme notre étude s'inscrit dans un cadre médical, l'objectif serait d'obtenir un taux de faux positifs et négatifs presque nul. En effet, cela signifie que certaines impulsions sont mal classifiées ce qui peut conduire à des fausses alarmes, par exemple une impulsion propre détectée comme comportant des artefacts peut amener certains paramètres vitaux à être aberrants et ainsi le personnel hospitalier à se déplacer en urgence au lit du patient. Pour de futures recherches, plusieurs axes d'amélioration se dessinent. Tout d'abord, il serait judicieux d'utiliser plusieurs méthodes de pré-traitement afin de réduire davantage la présence d'artefacts. Cette amélioration du pré-traitement pourrait contribuer à équilibrer les classes de données pour l'apprentissage machine. De plus, l'exploration de l'apprentissage auto-supervisé offre la possibilité de générer des étiquettes "fantômes" à partir des données elles-mêmes, éliminant ainsi le besoin d'un expert pour annoter les données. Enfin, l'augmentation des données à l'aide d'algorithmes appropriés pourrait permettre de disposer d'ensembles de données plus importants, favorisant ainsi les tests de modèles de réseaux de neurones.

En définitive, ce mémoire contribue à l'enrichissement des connaissances dans le domaine de la détection des artefacts dans les signaux PPG. En développant des méthodes efficaces pour permettre la classification ainsi que l'annotation des impulsions du signal PPG, la méthode proposée est intéressante. Nous pouvons assurer une interprétation plus précise des signaux PPG, ce qui aura un impact positif sur les soins de santé et sur la recherche médicale.

## LISTE DE RÉFÉRENCES

- Aljaaf, A. J., Al-Jumeily, D., Hussain, A. J., Fergus, P., Al-Jumaily, M. & Radi, N. (2015). Applied machine learning classifiers for medical applications : clarifying the behavioural patterns using a variety of datasets. *2015 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pp. 228–232.
- Barker, S. & Shah, N. (1996). Effects of motion on the performance of pulse oximeters in volunteers. *Anesthesiology*, 85(4), 774–781.
- Bodo, Z. & Csató, L. (2015). A note on label propagation for semi-supervised learning. *Acta Universitatis Sapientiae, Informatica*, 7.
- Brossier, D., El Taani, R., Sauthier, M., Roumeliotis, N., Emeriaud, G. & Jouvet, P. (2018). Creating a High-Frequency Electronic Database in the PICU : The Perpetual Patient. *Pediatric critical care medicine*, 19(4), 189-198.
- Bünger, D., Gondos, M., Peroche, L. & Stoll, M. (2022). An Empirical Study of Graph-Based Approaches for Semi-supervised Time Series Classification. *Frontiers in Applied Mathematics and Statistics*, 7.
- Choi, E., Schuetz, A., Stewart, W. F. & Sun, J. (2017). Using recurrent neural network models for early detection of heart failure onset. *Journal of the American Medical Informatics Association : JAMIA*, 24(2), 361-370.
- Chong, J. W., Dao, D. K., Salehizadeh, S., McManus, D. D., Darling, C. E., Chon, K. H. & Mendelson, Y. (2014). Photoplethysmograph signal reconstruction based on a novel hybrid motion artifact detection–reduction approach. Part I : Motion and noise artifact detection. *Annals of biomedical engineering*, 42, 2238–2250.
- Dara, S., Dhamercherla, S., Jadav, S. S., Babu, C. M. & Ahsan, M. J. (2022). Machine learning in drug discovery : a review. *Artificial Intelligence Review*, 55(3), 1947-1999.
- Dash, S., Shakyawar, S. K., Sharma, M. & Kaushik, S. (2019). Big data in healthcare : management, analysis and future prospects. *Journal of Big Data*, 6(54).
- Fujiwara, K., Huang, Y., Hori, K., Nishioji, K., Kobayashi, M., Kamaguchi, M. & Kano, M. (2020). Over- and Under-sampling Approach for Extremely Imbalanced and Small Minority Data Problem in Health Record Analysis. *Frontiers in Public Health*, 8.
- Goh, C.-H., Tan, L. K., Lovell, N. H., Ng, S.-C., Tan, M. P. & Lim, E. (2020). Robust PPG motion artifact detection using a 1-D convolution neural network. *Computer methods and programs in biomedicine*, 196, 105596.

- Gutierrez, G. (2020). Artificial Intelligence in the Intensive Care Unit. *Critical Care*, 24(1), 101.
- Habeheh, H. & Gohel, S. (2016). Machine Learning in Healthcare. *Current Genomics*, 22(4), 291-300.
- Hanyu, S. & Xiaohui, C. (2017). Motion artifact detection and reduction in PPG signals based on statistics analysis. *2017 29th Chinese Control And Decision Conference (CCDC)*, pp. 3114-3119.
- Helm, J. M., Swiergosz, A. M., Haeberle, H. S., Karnuta, J. M., Schaffer, J. L., Krebs, V. E., . . . & Ramkumar, P. N. (2020). Machine learning and artificial intelligence : definitions, applications, and future directions. *Current reviews in musculoskeletal medicine*, 13(1), 69-76.
- Hill, B. L., Rakocz, N., Rudas, Á., Chiang, J. N., Wang, S., Hofer, I., . . . & Halperin, E. (2021). Imputation of the continuous arterial line blood pressure waveform from non-invasive measurements using deep learning. *Scientific Reports*, 11(1), 15755.
- Ho, L. V., Ledbetter, D., Aczon, M. & Wetzel, R. (2018). The Dependence of Machine Learning on Electronic Medical Record Quality. *AMIA Symposium*, pp. 883–891.
- Huang, L., Shea, A. L., Qian, H., Masurkar, A., Deng, H. & Liu, D. (2019). Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records. *Journal of Biomedical Informatics*, 99, 103291.
- Ismail, S., Akram, U. & Siddiqi, I. (2021). Heart rate tracking in photoplethysmography signals affected by motion artifacts : a review. *EURASIP Journal on Advances in Signal Processing*, (5).
- Johnson, A. E., Ghassemi, M. M., Nemati, S., Niehaus, K. E., Clifton, D. A. & Clifford, G. D. (2016). Machine learning and decision support in critical care. *Proceedings of the IEEE*, 104(2), 444-466.
- Joseph, G., Joseph, A., Titus, G., Thomas, R. M. & Jose, D. (2014). Photoplethysmogram (PPG) signal analysis and wavelet de-noising. *2014 Annual International Conference on Emerging Research Areas : Magnetics, Machines and Drives (AICERA/iCMMD)*, pp. 1-5.
- Kim, S.-H., Lilot, M., Sidhu, K. S., Rinehart, J., Yu, Z., Canales, C. & Cannesson, M. (2014). Accuracy and precision of continuous noninvasive arterial pressure monitoring compared with invasive arterial pressure : a systematic review and meta-analysis. *Anesthesiology*, 120(5), 1080-1097.

- Krishnan, R., Natarajan, B. & Warren, S. (2008, 03). Analysis and detection of motion artifact in photoplethysmographic data using higher order statistics. pp. 613-616.
- Li, Q. & Clifford, G. (2012). Dynamic time warping and machine learning for signal quality assessment of pulsatile signals. *Physiological measurement*, 33(9), 1491–1501.
- Lim, P. K., Ng, S.-C., Lovell, N. H., Yu, Y. P., Tan, M. P., McCombie, D., . . . & Redmond, S. J. (2018). Adaptive template matching of photoplethysmogram pulses to detect motion artefact. *Physiological measurement*, 39(10), 105005.
- Mathieu, A., Sauthier, M., Juvet, P., Emeriaud, G. & Brossier, D. (2021). Validation process of a high-resolution database in a paediatric intensive care unit—Describing the perpetual patient’s validation. *Journal of Evaluation in Clinical Practice*, 27(2), 316–324.
- Mazurowski, M. A., Habas, P. A., Zurada, J. M., Lo, J. Y., Baker, J. A. & Tourassi, G. D. (2008). Training neural network classifiers for medical decision making : The effects of imbalanced datasets on classification performance. *Neural networks*, 21(2), 427-436.
- Miotto, R., Wang, F., Wang, S., Jiang, X. & Dudley, J. T. (2018). Deep learning for healthcare : review, opportunities and challenges. *Briefings in bioinformatics*, 19(6), 1236–1246.
- Ngiam, K. & Khor, I. (2016). Big data and machine learning algorithms for health-care delivery. *The Lancet. Oncology*, 20(5), 262–273.
- Petterson, M. T., Begnoche, V. L. & Graybeal, J. M. (2007). The Effect of Motion on Pulse Oximetry and Its Clinical Significance. *Anesthesia & Analgesia*, 105(6), 78-84.
- Pollreisz, D. & TaheriNejad, N. (2022). Detection and Removal of Motion Artifacts in PPG Signals. *Mobile Networks and Applications*, 27, 728–738.
- Raghuram, M., Sivani, K. & Reddy, K. (2016). Use of complex EMD generated noise reference for adaptive reduction of motion artifacts from PPG signals. *2016 international conference on electrical, electronics, and optimization techniques (ICEEOT)*, pp. 1816–1820.
- Robbins, D. E., Gurupur, V. P. & Tanik, J. (2011). Information architecture of a clinical decision support system. *2011 Proceedings of IEEE Southeastcon*, pp. 374–378.
- Rojano, J. & Isaza, C. (2016). Singular Value Decomposition of the Time-Frequency Distribution of PPG Signals for Motion Artifact Reduction. *International Journal of Signal Processing Systems*, 475-482.

- Roumeliotis, N., Parisien, G., Charette, S., Arpin, E., Brunet, F. & Jouvett, P. (2018). Reorganizing care with the implementation of electronic medical records : A time-motion study in the PICU. *Pediatric critical care medicine*, 19(4), 172–179.
- Sanchez-Pinto, L. N. & Khemani, R. G. (2016). Development of a prediction model of early acute kidney injury in critically ill children using electronic health record data. *Pediatric Critical Care*, 17(6), 508-15.
- Shin, Y., Yoon, S., Kim, S., Song, H., Lee, J.-G. & Lee, B. S. (2021). Coherence-based label propagation over time series for accelerated active learning. *International Conference on Learning Representations*.
- Song, Z., Yang, X., Xu, Z. & King, I. (2022). Graph-Based Semi-Supervised Learning : A Comprehensive Review. *IEEE Transactions on Neural Networks and Learning Systems*, 1-21.
- Tobin, R. M., Pologe, J. A. & Batchelder, P. B. (2002). A characterization of motion affecting pulse oximetry in 350 patients. *Anesthesia and analgesia*, 94(1), S54–S61.
- Wang, Q., Yang, P. & Zhang, Y. (2010). Artifact reduction based on Empirical Mode Decomposition (EMD) in photoplethysmography for pulse rate detection. *2010 annual international conference of the IEEE engineering in medicine and biology*, pp. 959–962.
- Wang, Z., Yan, W. & Oates, T. (2017). Time series classification from scratch with deep neural networks : A strong baseline. *2017 International joint conference on neural networks (IJCNN)*, pp. 1578–1585.
- Discrete wavelet transform — Wikipedia, The Free Encyclopedia. (2023). Dans *Wikipedia*. Repéré le 2023-08-03 à [https://en.wikipedia.org/w/index.php?title=Discrete\\_wavelet\\_transform&oldid=1166724423](https://en.wikipedia.org/w/index.php?title=Discrete_wavelet_transform&oldid=1166724423).
- Wu, C., Chen, I. & Fang, W. (2017). An implementation of motion artifacts elimination for PPG signal processing based on recursive least squares adaptive filter. *2017 IEEE biomedical circuits and systems conference (BioCAS)*, pp. 1-4.
- Xu, Z. & Funaya, K. (2015). Time series analysis with graph-based semi-supervised learning. *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 1-6.
- Yogarajan, V., Montiel, J., Smith, T. & Pfahringer, B. (2021). Transformers for multi-label classification of medical text : an empirical comparison. *International Conference on Artificial Intelligence in Medicine*, pp. 114-123.

Zhu, X. & Ghahramani, Z. (2002). Learning from labeled and unlabeled data with label propagation. *School of Computer Science*.