

RAPPORT D'ENSEIGNEMENT D'APPROFONDISSEMENT

**Explicabilité des diagnostics cardiaques d'un
réseau de neurones profond**

Septembre 2020 - Décembre 2020

Maxence NOBLE, Gaspard SAGOT
Encadrants : Laurent RISSER, Jean-Michel LOUBES



ABSTRACT

La cardiologie fait partie des domaines dans lesquels l'utilisation de réseaux de neurones profonds (DNN) est rare. Une des explications à cette situation est la difficulté d'obtenir des jeux de données suffisamment cohérents et massifs d'électrocardiogrammes (ECG). En 2020, une équipe de chercheurs de l'Universidade Federal de Minas Gerais a entraîné un DNN sur plus de 2 millions d'ECG standards à 12 pistes et a obtenu un F-score supérieur aux cardiologues de l'étude dans la prédiction de 6 pathologies cardiaques courantes [1]. Cependant, le besoin d'explicabilité du réseau est grand puisque demeure la question de la différence entre les caractéristiques que reconnaît le DNN et celles que reconnaissent les cardiologues. Dans ce document, nous extrayons des caractéristiques numériques des signaux utilisés pour la phase de test du DNN qui ont un sens physique et qui sont exploitables en pratique par les médecins. D'un côté, nous comparons leur influence sur le diagnostic donné par le DNN, tentant ainsi de donner une explication de son fonctionnement. De l'autre, nous comparons ces influences avec celles sur le diagnostic des médecins qui considèrent moins les valeurs statistiques comme les moyennes ou l'asymétrie des distributions des signaux, mais plus la structure des motifs des ECG, ce qui s'avère être difficilement faisable de manière automatique. Nous proposons aussi d'autres perspectives futures pour expliquer le fonctionnement du réseau de neurones, comme la reconnaissance de motifs temporels dans les ECG, ou la reconstruction d'un signal de manière non-linéaire par réseau de neurones à partir d'une représentation fonctionnelle.

TABLE DES MATIÈRES

Introduction	4
1 Traitement des données	5
1.1 Structure des données et méthodologie	5
1.1.1 Structure des données	5
1.1.2 Extraction de features	6
1.2 Approche temporelle	6
1.3 Approche statistique	9
1.4 Approche représentative	12
1.4.1 Analyse de Fourier	12
1.4.2 Décomposition en ondelettes	14
2 Algorithmes d'explicabilité étudiés	20
2.1 Ethik AI	20
2.2 Local Interpretable Model-agnostic Explanations (LIME)	22
3 Résultats d'explicabilité	25
3.1 Pathologie 1dAVb	25
3.2 Pathologie RBBB	26
3.3 Pathologie LBBB	26
3.4 Pathologie SB	27
3.5 Pathologie AF	27
3.6 Pathologie ST	28
4 Discussion	29
4.1 Limites de notre travail	29
4.1.1 Limitations de la base de données	29
4.1.2 Limitations des méthodes explorées	30
4.2 La vision du réseau de neurones	30
4.3 Perspectives ouvertes	31
Conclusion	33
Références	34
Annexes	35
Tableaux complets d'explicabilité	35

INTRODUCTION

Au cours des dernières années les réseaux neuronaux profonds (DNN) ont obtenu des succès remarquables dans des tâches telles que la classification d'images et la reconnaissance vocale, et les attentes sont grandes quant à la manière dont cette technologie peut améliorer les soins de santé et la pratique clinique. En 2020 une équipe de chercheurs principalement de l'Universidade Federal de Minas Gerais à Belo Horizonte, Brésil, a publié un article sur le diagnostic automatique de pathologies cardiaques par réseau de neurones profond[1]. L'équipe a entraîné de manière supervisée le DNN sur pas moins de deux millions d'ECG standards à 12 pistes, l'étiquetage des données ayant été fait par des internes cardiologues ainsi que par des cardiologues confirmé dans le cas d'un doute.

L'ECG standard de courte durée à 12 pistes est l'examen complémentaire le plus couramment utilisé pour l'évaluation du cœur, étant employé dans tous les cadres cliniques, des centres de soins primaires jusqu'aux unités de soins intensifs. Cet examen peut fournir une évaluation complète de l'activité électrique cardiaque, ce qui inclut les arythmies, les troubles de la conduction, les syndromes coronariens aigus, l'hypertrophie et l'élargissement de la chambre cardiaque et même les effets des médicaments et les perturbations électrolytiques.

Le DNN résultant de l'article[1] surpasse les médecins résidents en cardiologie sur le diagnostic de 6 types d'anomalies dans les ECG à 12 pistes, avec des F-scores supérieurs à 80% et une spécificité supérieure à 99%. Ce résultat invite à dépasser le modèle «boîte noire» et à chercher à expliquer d'une part comment le DNN détecte une anomalie et d'autre part quelles caractéristiques des ECG sont observées de manière différente par le DNN et par les cardiologues.

Dans ce document nous tentons de rentrer dans les rouages du DNN et de mieux comprendre *a posteriori* ce qu'il détecte en extrayant des caractéristiques des signaux et en leur appliquant plusieurs méthodes d'explicabilité de réseaux de neurones. Cette approche permet d'expliquer le fonctionnement du DNN tout en gardant des caractéristiques qui ont du sens. Nous mettons cela en parallèle du diagnostic des médecins qui regardent moins les valeurs statistiques comme les moyennes ou l'asymétrie des distributions des signaux mais plus la structure des motifs des ECG, ce que nous montrons être difficilement faisable de manière automatique.

En première partie du rapport nous détaillerons la structure des données ainsi que les méthodes mises en œuvre pour en extraire des features. En seconde partie nous présenterons les algorithmes d'explicabilité étudiés pour poursuivre avec les résultats obtenus grâce à ces algorithmes en 3ème partie. Enfin en 4ème partie nous discuterons des limitations des données et des méthodes utilisées ainsi que des enseignements tirés de notre travail, tant sur la comparaison entre la vision des cardiologues et celle du réseau de neurones que sur les perspectives ouvertes sur l'explicabilité en intelligence artificielle. On pourra trouver en annexe les tableaux d'explicabilité complets.

1

TRAITEMENT DES DONNÉES

1.1 STRUCTURE DES DONNÉES ET MÉTHODOLOGIE

Afin de comprendre ce que reconnaît le réseau de neurones, il est essentiel de connaître la structure des données qu'il prend en entrée. Pour des raisons de confidentialité les données d'entraînement ne sont pas accessibles publiquement, cependant il est possible de récupérer sur le GitHub de l'équipe de l'article[1] le réseau pré-entraîné d'une part et les données de test d'autre part.

1.1.1 STRUCTURE DES DONNÉES

Les données de test sont structurées en tableau de taille (827, 4096, 12), ce qui correspond aux coordonnées (patient_ID, time_step, lead_number). Ainsi si l'on veut récupérer le vecteur des valeurs de la piste 6 du patient 17, il faut appeler le tableau aux coordonnées [17][:][6]. La «piste» ou «lead» correspond à une des pistes standard de l'ECG à 12 pistes, qui utilise 10 électrodes.

Tous les signaux sont échantillonnés à une fréquence d'échantillonnage standardisée $f_s = 400$ Hz, le pas de temps vaut donc $\tau = 1/400$ s. Les signaux étant constitué de 4096 pas de temps, l'ECG physique correspondant est enregistré sur une durée standard de 10, 24s. Bien sûr il est impossible en pratique d'avoir un jeu de données physiques aussi cohérent, par conséquent la plupart des pistes ont été rallongées artificiellement par des 0 afin d'atteindre 4096 pas de temps (*zero-padding*).

Les labels quant à eux se déclinent en 6 valeurs, dont chacune détermine le diagnostic d'une pathologie. Cette valeur vaut 0 si le DNN ou les cardiologues estiment que la probabilité que l'individu soit atteint de cette pathologie est insuffisante, et 1 sinon. Les pathologies sont les suivantes :

1. '**1dAVb**' pour *1st Degree AV Block* : Bloc atrio-ventriculaire (ou auriculo-ventriculaire) du premier degré
2. '**RBBB**' pour *Right Bundle Branch Block* : Bloc de branche droit
3. '**LBBB**' pour *Left Bundle Branch Block* : Bloc de branche gauche
4. '**SB**' pour *Sinus Bradycardia* : Bradycardie sinusale
5. '**AF**' pour *Atrial Fibrillation* : Fibrillation auriculaire
6. '**ST**' pour *Sinus Tachycardia* : Tachycardie sinusale

1.1.2 EXTRACTION DE FEATURES

Afin de mettre en lien les prédictions du réseau de neurone avec les données d'entrées de manière explicable, nous allons extraire un certain nombre de caractéristiques (features) des signaux d'entrées. Ces caractéristiques seront interprétables par un médecin et représenteront les signaux au moment d'utiliser les outils d'explicabilité de la section 2.

La structure du réseau de neurones utilise des concepts de traitement d'image et par conséquent noie toute information «lisible» sur les courbes en elles-mêmes. Si cette approche consistant en l'oubli de la structure des données en signal périodique (voire en signal temporel tout court) est la clé de la précision du réseau, nous avons choisi dans une certaine mesure de conserver l'approche de traitement du signal afin de conserver des features explicables. Ainsi nous avons choisi 3 approches différentes pour extraire des features du signal :

1. l'approche temporelle, qui extrait des features des signaux bruts,
2. l'approche statistique, qui condense le signal en un histogramme puis en extrait des features,
3. l'approche représentative du signal. Pour celle-ci nous avons choisi la représentation fréquentielle grâce à la transformée de Fourier ainsi que la représentation en ondelettes.

Cependant comme nous le verrons ci-dessous, la dernière approche se heurte rapidement à la non-linéarité du réseau. Le fait qu'il ait été entraîné sur des images brutes et non sur des signaux fait que sa non-linéarité se ressent d'abord sur l'écart de prédiction de diagnostic entre deux signaux proches. Deux signaux qui nous apparaissent proches d'un point de vue traitement du signal (par exemple relativement aux normes L^1 ou L^2) peuvent ne pas l'être du tout du point de vue du réseau.

1.2 APPROCHE TEMPORELLE

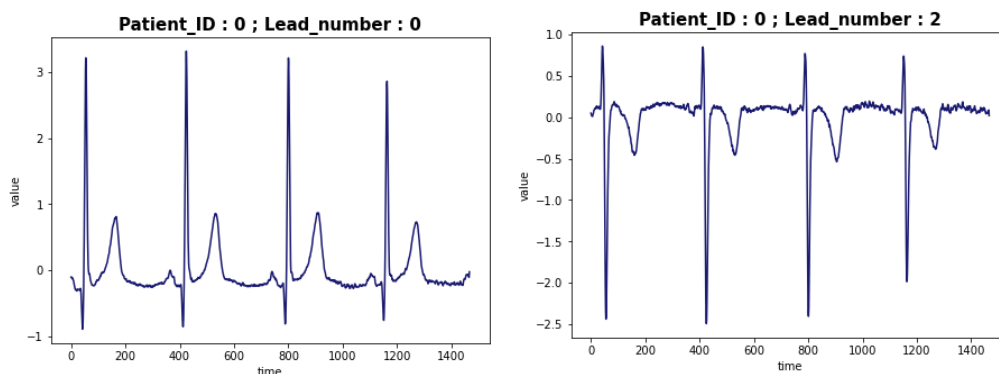


FIGURE 1 – Les pistes 0 et 2 de l'ECG du patient 0

Commençons par l'approche temporelle pour extraire des features. Deux exemples de pistes ECG pour le patient numéro 0 sont montrés en figure 1. On pourra trouver plus de détails sur ces features dans la plupart des textes de traitement du signal, par exemple [2] ou [3].

Dans toutes les features qui suivent nous adoptons des conventions de traitement du signal et notons $x = (x[k])_{k \in [0, N-1]} \in \mathbb{R}^N$ un signal temporel numérique. Ces fonctions, sauf précision contraire, s'appliquent à une unique piste ECG. L'intérêt de ce choix est de pouvoir identifier quelle piste a le plus de poids dans l'explication du diagnostique qu'une autre.

asynchrony (feature appliquée sur l'ensemble des ECG d'un patient) : Cette feature caractérise le transport optimal moyen entre toutes les pistes ECG d'un patient et la piste ECG classique d'un patient ne portant aucune pathologie et de rythme cardiaque égal à 75 bpm (la médiane de l'ensemble de test). Pour calculer cette feature, on utilise l'algorithme DTW (*Dynamic Time Warping*, soit la déformation temporelle dynamique), qui permet de mesurer la similarité entre deux séries temporelles. Dans notre cas, cet algorithme est appliqué au couple $(ECG_{\text{étudié}}, ECG_{\text{modèle}})$ pour chacun des 12 $ECG_{\text{étudié}}$ du patient considéré. Il donne en sortie l'appariement optimal entre ces deux séries temporelles sous forme d'une suite de couples d'indices $\{(i_1, j_1), (i_2, j_2), \dots, (i_{L'}, j_{L'})\}$ où les indices i correspondent à $ECG_{\text{étudié}}$, les indices j à $ECG_{\text{modèle}}$ et L' est une certaine longueur. On caractérise alors l'asynchronie du signal étudié en calculant la moyenne sur les 12 ECG du patient les quantités suivantes :

$$\begin{aligned} \text{--- } \mathbf{asynchrony1} &= \frac{1}{L'} \sum_{l=1}^{L'} j - i \\ \text{--- } \mathbf{asynchrony2} &= \frac{1}{L'} \sum_{l=1}^{L'} (j - i)^2 \end{aligned}$$

De la sorte, **asynchrony1** est positif lorsque les $ECG_{\text{étudié}}$ sont en avance sur $ECG_{\text{modèle}}$ et à l'inverse, est négatif lorsque les $ECG_{\text{étudié}}$ sont en retard sur $ECG_{\text{modèle}}$. Dans le cas de signaux très irréguliers, la quantité **asynchrony1** peut valoir 0 par compensation des retards et des avances. Pour pallier cela, on regarde **asynchrony2** qui permet de savoir si cette nullité est en effet justifiée par une vraie synchronie avec $ECG_{\text{modèle}}$ (dans ce cas, égal à 0) ou non.

auto-correlation : La corrélation du signal avec lui-même, évaluée au pas de temps correspondant à la période du signal physique et normalisée par la variance du signal. La corrélation de deux signaux (x, y) est généralement définie dans les textes de traitement du signal de la manière suivante :

$$C(x, y)[n] = \sum_k x[k+n] \overline{y[k]}$$

les signaux x et y ayant été complétés de 0 au besoin et la barre signifiant la conjugaison complexe. En figure 2 on peut voir comment évolue l'auto-corrélation pour la piste 0 du patient 0, qui ne présente aucune pathologie ni pour le DNN ni pour les cardiologues, et en figure 3 l'auto-corrélation pour le patient 120 qui présente une fibrillation atriale (fréquence cardiaque irrégulière et souvent très rapide).

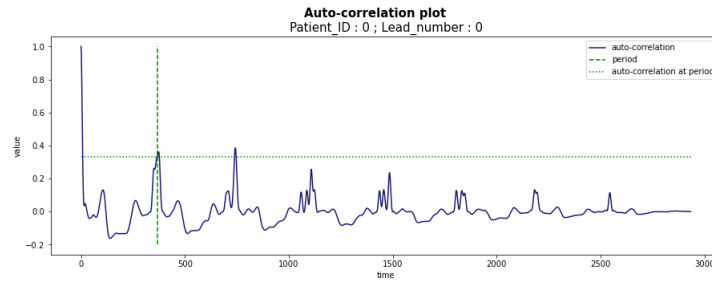


FIGURE 2 – Auto-corrélation de la piste 0 de l'ECG du patient 0

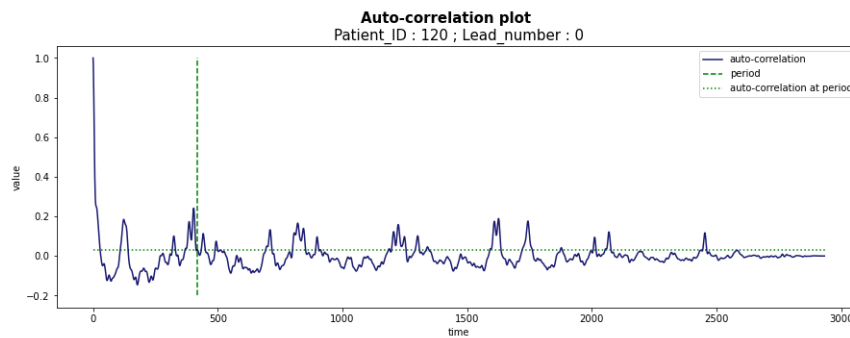


FIGURE 3 – Auto-corrélation de la piste 0 de l'ECG du patient 120

On voit que l'auto-corrélation est beaucoup plus faible pour le patient atteint de fibrillation. Et en effet, l'auto-corrélation donne une mesure de ressemblance entre le signal décalé d'un certain pas de temps et lui même. Si elle est faible, cela veut dire que le signal est irrégulier, c'est-à-dire que soit le motif du battement du cœur est irrégulier, soit les battements cardiaques ne font pas tous la même durée (arythmie cardiaque). En figure 4, le signal ECG du patient 120 ayant donné l'auto-corrélation ci-dessus.

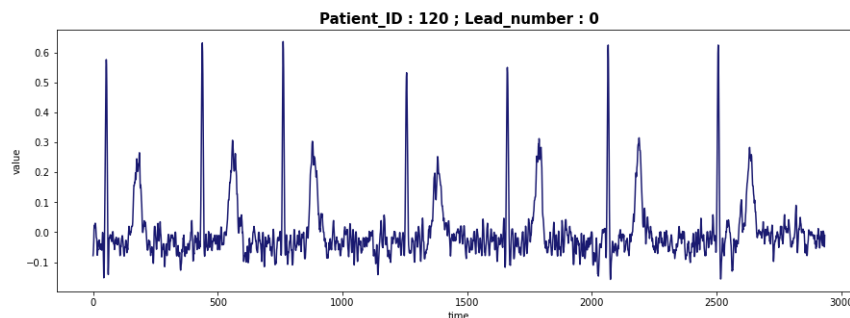


FIGURE 4 – La piste 0 de l'ECG du patient 120

average : La moyenne du signal donné en entrée.

$$\text{average}(x) = \frac{1}{N} \sum_k x[k]$$

energy : La moyenne du carré du signal :

$$\text{energy}(x) = \frac{1}{N} \sum_k x[k]^2$$

Cette feature correspond à l'énergie du signal physique qu'est l'ECG, c'est aussi le carré de sa valeur efficace. Un signal d'énergie nulle est nul, et un signal de grande énergie par rapport à la moyenne signifie soit que le cœur est rarement au repos, soit que l'amplitude de ses battements est très conséquentes.

frequency : La fréquence en battements par minutes. Peut aussi être calculée sur toutes les pistes à la fois, auquel cas la fonction renvoie la fréquence qui apparaît le plus parmi les 12 pistes. La méthode choisie pour la calculer sera exposée ci-dessous.

maximum : Le maximum du signal.

$$\text{maximum}(x) = \max_k x[k]$$

mid-range : La moyenne arithmétique du maximum et du minimum du signal, c'est donc une mesure de position.

$$\text{mid-range}(x) = \frac{\max_k x[k] + \min_k x[k]}{2}$$

minimum : Le minimum du signal.

$$\text{minimum}(x) = \min_k x[k]$$

period : La période en nombre de pas de temps (1 pas de temps = 1/400 s). Peut aussi être calculée sur toutes les pistes à la fois, auquel cas la fonction renvoie la période qui apparaît le plus parmi les 12 pistes. La méthode choisie pour la calculer sera exposée ci-dessous.

range : L'écart entre le maximum et le minimum du signal, c'est donc une mesure d'étalement.

$$\text{range}(x) = \max_k x[k] - \min_k x[k]$$

signal magnitude area : La somme de toutes les valeurs du signal. Cette valeur est liée à la moyenne par la formule :

$$\text{average}(x) = \frac{1}{N} \times \text{signal magnitude area}(x)$$

1.3 APPROCHE STATISTIQUE

Prenons ici l'approche statistique en passant par l'histogramme des signaux pour extraire des features. Les histogrammes du patient 0 pour les pistes 0 et 2 sont montrés en figure 5. La plupart de ces grandeurs statistiques sont définies dans *High-Dimensional Deep Learning* de B. Laurent-Bonneau et B. Guillouet[4].

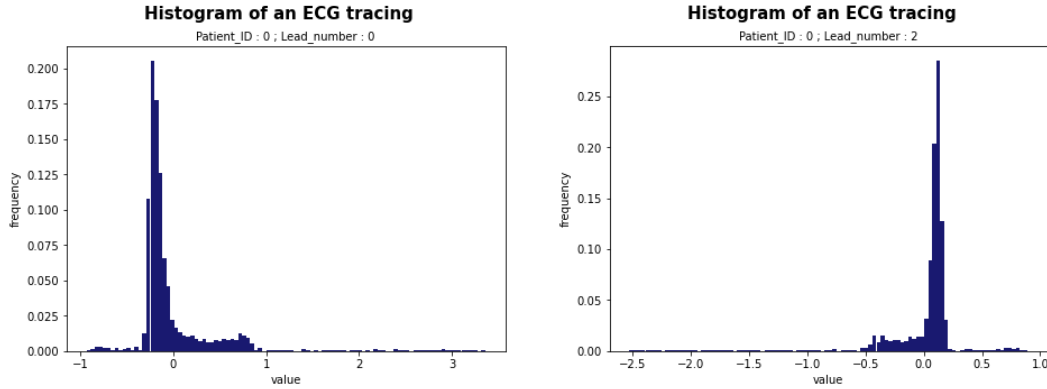


FIGURE 5 – Histogrammes des pistes 0 et 2 de l'ECG du patient 0

entropy : Donne l'entropie de Shannon de la distribution associée à l'histogramme de valeurs. Étant donné une distribution discrète de probabilité $(p_i)_i$, l'entropie associé est :

$$S(p) = - \sum_i p_i \ln(p_i)$$

On pourra trouver plus d'information dans *Information theory, inference and learning algorithms* de D. JC MacKay[5].

gm asymetry 1 : L'asymétrie de Groeneveld et Meeden. C'est une mesure d'asymétrie, et pour la calculer on considère la fonction

$$\gamma(F)(u) = \frac{F^{-1}(u) + F^{-1}(1 - u) - 2F^{-1}(1/2)}{F^{-1}(u) - F^{-1}(1 - u)}$$

où $u \in]1/2, 1[$ et F est la fonction de répartition d'une variable aléatoire ($F^{-1}(1/2)$ est donc la médiane du signal). On prend alors le supremum de $\gamma(F)$ sur $]1/2, 1[$. En pratique, nous considérons le maximum de cette fonction de $u = 55/100$ à $u = 99/100$ pour éviter les divisions par zéro.

gm asymetry 2 : Notre variante de l'asymétrie de Groeneveld et Meeden, dans laquelle nous ne prenons plus le supremum de $\gamma(F)$ mais sa valeur extrême. C'est une mesure de plus de l'asymétrie de l'histogramme d'un ECG, assez robuste puisque le calcul est effectué à partir des quantiles et non pas des valeurs du signal.

interpercentile range(p_1, p_2) : L'écart entre les percentiles p_1 et p_2 . Cette grandeur donne donc la largeur de la bande de valeurs contenant $p_2 - p_1$ des points, par exemple 50% des points pour $(p_1, p_2) = (25, 75)$. Ceci s'interprète de la manière suivante pour un ECG : un ECG standard d'une personne saine doit être relativement calme entre deux battements. Les battements étant courts, une grande quantité de points doit être localisée autour de la valeur médiane. Ainsi si l'écart-interpercentile d'un patient est faible, celui-ci devrait avoir plus de chance d'être en bonne santé, alors que s'il est grand, les valeurs «au repos» sont très disparates et par conséquent il a une probabilité plus élevée d'être malade. L'intérêt de cette feature, au delà d'être une mesure robuste de l'étalement d'une courbe autour de

sa médiane pour $(p_1, p_2) = (25, 75)$ (dans cette situation on parle d'écart interquartile), est qu'elle peut-être appliquée à de nombreux écarts différents. Nous prenons en compte les suivants :

- $(p_1, p_2) = (10, 90)$
- $(p_1, p_2) = (25, 75)$
- $(p_1, p_2) = (40, 60)$
- $(p_1, p_2) = (0, 50)$: dans ce cas, on mesure l'étalement de la moitié des points qui sont plus petits que la médiane, à savoir la distance : $\tilde{x} - \min(x)$.
- $(p_1, p_2) = (50, 100)$: dans ce cas, on mesure l'étalement de la moitié des points qui sont plus grands que la médiane, à savoir la distance : $\max(x) - \tilde{x}$

Enfin cette feature peut aussi être calculée normalisée, et dans cette situation la valeur observée est l'écart interpercentile divisé par l'écart entre valeur minimale et valeur maximale du signal. Cette option permet une meilleure comparaison de l'asymétrie de l'histogramme d'un signal entre les différents ECG.

kurtosis : La kurtosis de l'histogramme d'une piste ECG. C'est une mesure d'étalement de l'histogramme.

$$\text{Kurtosis}(X) = E \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right] - 3$$

La correction de -3 est introduite pour pouvoir comparer les histogrammes à celui d'une loi normale, dont la kurtosis vaut 3.

median : La médiane du signal, une mesure de localité.

median absolute deviation : L'écart médian absolu à la médiane. Ainsi si x est un signal et \tilde{x} sa médiane, alors

$$\text{median absolute deviation}(x) = \text{median} [|x[k] - \tilde{x}|_k]$$

Cette valeur est une autre mesure de l'étalement d'une distribution, mais elle est plus robuste que l'écart-type relativement aux valeurs aberrantes ou extrêmes.

midhinge : La moyenne entre les quartiles 25% et 75% du signal. C'est donc une mesure de position.

skewness : Le coefficient d'asymétrie de l'histogramme d'une piste ECG. Cette valeur correspond à une mesure de l'asymétrie de la distribution, et elle est définie pour toute variable aléatoire réelle X de moyenne μ et d'écart-type σ par :

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

standard deviation : L'écart-type du signal donné en entrée, c'est-à-dire l'écart quadratique moyen à la moyenne. C'est une mesure d'étalement assez sensible aux valeurs extrêmes.

trimean : La moyenne entre la médiane et le midhinge. Ainsi si l'on note Q_1, Q_2, Q_3 les 3 quartiles (Q_2 étant alors la médiane), c'est une moyenne pondérée de la manière suivante :

$$\text{trimean}(x) = \frac{1}{2} \left(Q_2 + \frac{Q_1 + Q_3}{2} \right) = \frac{Q_1 + 2Q_2 + Q_3}{4}$$

1.4 APPROCHE REPRÉSENTATIVE

L'objectif d'une modélisation mathématique des ECG autrement que par leur format d'origine est double. D'un côté, il s'agit de trouver un espace de représentation assez compact pour conserver en grande partie l'information : on s'attaque ainsi à la thématique de la réduction de la dimension, grand fléau des algorithmes d'apprentissage. Dans notre cas, l'idéal serait de passer de 4096 dimensions pour chaque piste à quelques centaines, tout en héritant des informations qu'elle contient. D'un autre côté, cette représentation doit être interprétable par l'observateur et utilisable en pratique d'un point de vue médical pour analyser les ECG de patients aux risques de pathologies cardiaques.

1.4.1 ANALYSE DE FOURIER

La première approche représentative que nous avons choisi est la représentation fréquentielle, à savoir la décomposition en série de Fourier. C'est la représentation la plus classique en traitement du signal et elle semble particulièrement adaptée aux signaux ECG puisqu'ils sont périodiques.

Cependant cette périodicité, que l'on constate facilement en observant un cœur battre, n'est pas immédiate d'un point de vue représentation discrète du signal. En effet ce dernier est encodé sur une durée de 4096 pas de temps ce qui ne correspond pas nécessairement (et même presque jamais) à un multiple entier de la période. Les algorithmes de transformée de Fourier numérique (DFT pour *Discrete Fourier Transform* et FFT pour *Fast Fourier Transform*) s'appliquent à des signaux discrets parfaitement périodiques, c'est-à-dire que si un signal x de longueur finie N est déclaré périodique de période n , cela veut dire qu'il existe un signal X infini de période n tel que $\forall k \in \mathbb{N}, X[k] = x[k \bmod N]$. Ainsi même si un ECG apparaît à l'écran comme étant périodique de période $n = 400$, la définition de la périodicité cyclique (ou circulaire) impose que le signal vérifie

$$x[400] = x[800] = \dots x[4000] = x[304] = \dots = x[3904] = x[208] = \dots$$

Ceci en pratique ne peut pas être vérifié avec des signaux à valeurs réelles, et par conséquent l'approche la plus pratique consiste à déclarer la «vraie» période de 400 comme étant secondaire, et de rendre le signal périodique de période N .

Une première méthode pour «périodiser» le signal consiste en l'ajout à la fin de celui-ci de la première valeur du signal. Cette étape intervient bien-sûr après avoir préalablement retiré les zéros ajoutés par *zero-padding*. Cependant imaginons que l'enregistrement de l'ECG commence sur un pic et finisse par une zone creuse. Alors cette méthode fait apparaître artificiellement un dernier pic, beaucoup plus abrupt. Cet ajout apparaît alors dans la transformée de Fourier comme un ensemble de pics de fréquences très élevées reflétant ce mouvement vertical abrupt du signal. Ce n'est donc pas une méthode idéale si l'on veut récupérer de l'information se trouvant dans les hautes fréquences.

Une seconde méthode, toute aussi courante, consiste en la régularisation du signal par parité. Cette méthode duplique le signal en 2 et donc double sa longueur, en revanche elle rend bien le nouveau signal périodique et presque aussi régulier que le premier. Cependant cette méthode influe sur le poids de la «vraie» fréquence dans le spectre final puisque le repliement n'en tient pas en compte.

Vient ensuite la question du filtrage du spectre afin de réduire la dimension. L'idée la plus simple est celle du seuil, à savoir : on extrait les couples (fréquences, valeur) dont la valeur dépasse une valeur seuil (par exemple $\frac{1}{2} \max(\text{Spectre})$). Cette méthode simple pose néanmoins un problème majeur dans le cas des ECG : elle retranscrit assez mal les subtilités des battements cardiaques. En effet, un signal cardiaque se caractérise par un enchaînement de zones de haute fréquence (pics reconnaissables) et de zones à très basse fréquence, qui sont toutes deux des zones où se combinent plusieurs signaux sinusoïdaux. Sur la transformée de Fourier, seuls les passages à haute fréquence sont reconnaissables, ce qui ne permet pas d'avoir une représentation complète du signal une fois un seuillage (*threshold*) raisonnable appliqué sur le module de la FFT, comme le montre la figure 6.

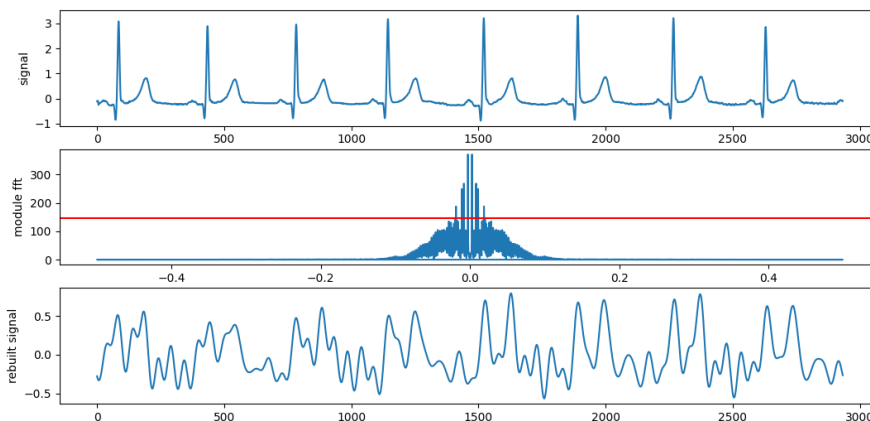


FIGURE 6 – Exemple de reconstruction d'un ECG par FFT (threshold sur le module)

Cependant même si la transformée de Fourier s'est révélée inefficace pour la simplification et le débruitage des signaux, nous avons tout de même réussi à en tirer quelque chose concernant l'identification de la «vraie» période du signal : la durée du battement cardiaque.

L'idée, simple, est que puisque cette fréquence est la fréquence fondamentale du signal, son amplitude par rapport au spectre doit être maximale. Cependant pour certaines pistes comme sur la figure 7, on peut voir que le signal est fortement bruité sur les basses fréquences, ce qui se voit dans l'ECG par un décalage fixe ainsi que par une lente ondulation. Sa fréquence cardiaque réelle est par conséquent dissimulée derrière ce bruit.

La solution que nous avons trouvée est simplement de filtrer le bruit. Nous avons choisi deux fréquences cardiaques extrêmes, 50 bpm et 120 bpm, et nous appliquons au signal un filtre de Butterworth de bande passante (5/6 Hz, 2 Hz). Ce choix de bande passante s'explique par le

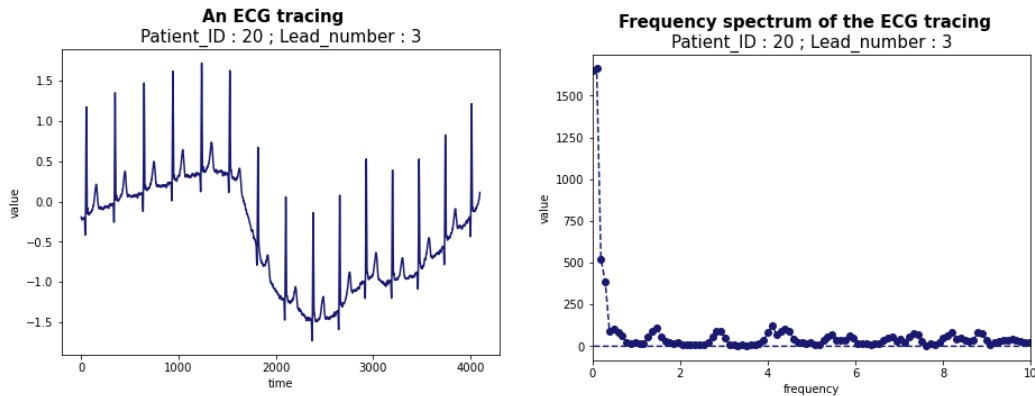


FIGURE 7 – Signal et spectre en fréquence de l'ECG 3 du patient 20

fait que nous avons la certitude que les fréquences cardiaque sont dans cet intervalle, puisque ce sont les limites couramment acceptées pour la bradycardie et la tachycardie respectivement, et le choix du filtre s'explique par le fait que les filtres de Butterworth sont les plus «plats» dans leur zone de fréquences admises[6], ce qui permet d'identifier précisément la fréquence cardiaque sans se tromper à cause d'une résonance sur une fréquence secondaire. Voici en figure 8 ce que donne le filtre appliqué au signal précédent, ainsi que la reconstruction du signal après filtrage.

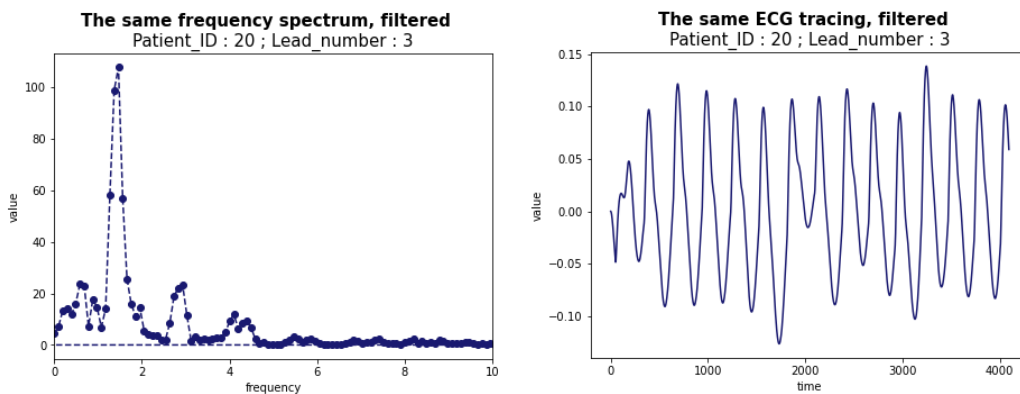


FIGURE 8 – Spectre et signal de l'ECG 3 du patient 20 après filtrage

Enfin en figures 9 et 10 : l'effet du filtre sur une personne atteinte de fibrillation cardiaque.

1.4.2 DÉCOMPOSITION EN ONDELETTES

Une fois la transformée de Fourier écartée, il reste la décomposition en ondelettes, elle aussi abondamment décrite dans la littérature mais plus difficile à maîtriser. Elle est notamment très utile pour le débruitage et la compression de signaux et d'images.

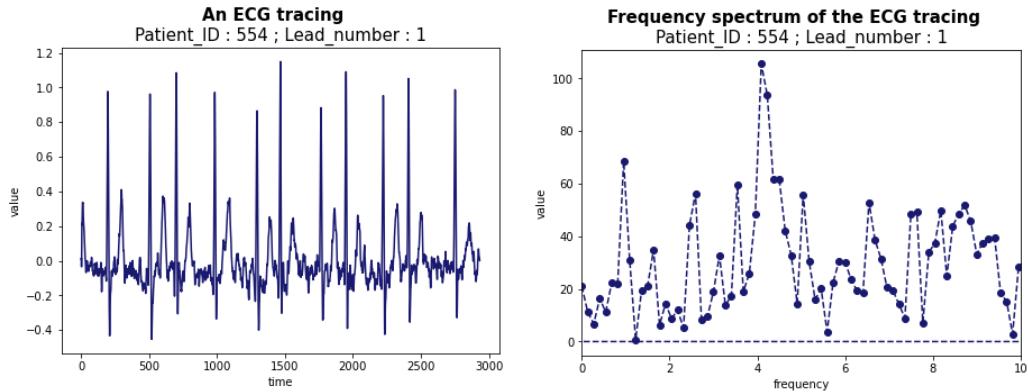


FIGURE 9 – Signal et spectre de l'ECG 1 du patient 554 avant filtrage

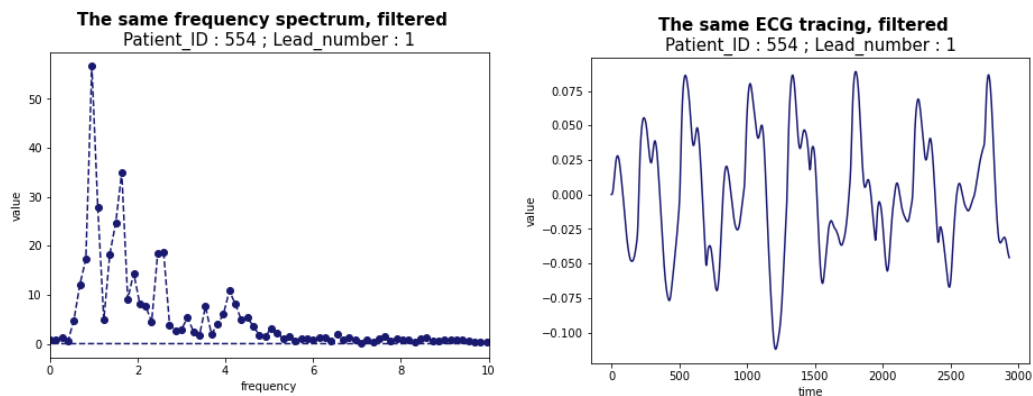


FIGURE 10 – Signal et spectre de l'ECG 1 du patient 554 après filtrage

Cette décomposition est avantageuse pour notre projet car d'un côté, les fonctions de base d'ondelettes ont une forme semblable aux battements cardiaques et de l'autre, leur utilisation nécessite un nombre de points d'échantillonnage égale à une puissance de 2 entière (contrainte satisfaite par les ECG fournis puisqu'ils durent tous 4096 pas de temps).[7]

Fonctionnement : Il existe plusieurs familles de décomposition d'ondelette, qui ont chacune des propriétés et des formes particulières (Haar, Daubechy, Symlet...). A chaque famille correspond plusieurs types d'ondelettes définis par un numéro. Plus le numéro est élevé, plus le niveau de complexité et le support de l'ondelette sont agrandis (par exemple, plusieurs ondelettes de type Daubechy sont présentées dans la figure 11).

On peut définir chaque ondelette par deux fonctions[4] :

- **une fonction mère** $\psi(t)$ (fonction d'ondelette à proprement parler) qui agit comme un filtre bande-passante dans l'espace des fréquences. A partir de cette fonction mère, on peut définir par opérations de dilatation dyadique un ensemble d'ondelettes dans lequel peut se décomposer le signal. Par exemple, en partant de la fonction $\psi(t)$, on considère $\psi_1(t) = \psi(2t)$, $\psi_2(t) = \psi(4t)$, $\psi_3(t) = \psi(8t)$, etc. qui correspondent à une couverture des fréquences de plus en plus élevées. En répétant un nombre infini de fois ces opérations,

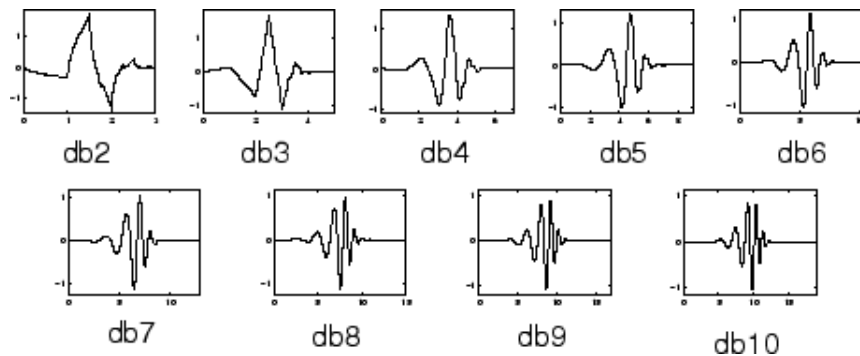


FIGURE 11 – Ondelettes de la famille Daubechies (2 à 10)

on peut couvrir tout le spectre des fréquences et décomposer le signal exactement.

- **une fonction père** $\phi(t)$ (fonction de scaling) qui correspond à un filtre passe-bas dans l'espace des fréquences. Généralement, la décomposition en hautes fréquences du signal se fait assez bien et ne requiert pas une quantité trop importante de vecteurs de ce côté du spectre. Cependant, du côté des basses fréquences, la décomposition nécessite un nombre infini de vecteurs issus de l'ondelette. Pour avoir une représentation simple du signal, on définit ainsi la fonction père qui couvre toute la bande de fréquence précédant celle de $\psi(t)$.

A partir d'une ondelette type, on obtient ainsi la décomposition d'un signal avec la famille de vecteurs $(\phi(t), \psi(2^k t), k \in \mathbb{N})$.

Cas d'un signal discret :

En pratique, la décomposition d'un signal discret selon un certain type d'ondelette ne requiert pas une infinité d'éléments dans la base. On définit en effet un **niveau d'affinement du maillage**, un nombre entier qui traduit la précision de la discrétisation. Ce niveau est limité par la taille du signal considéré et il est propre à chaque type d'ondelette en fonction de sa complexité. De la sorte, chaque type d'ondelette admet un niveau de d'affinement maximal par rapport au signal traité.[4]

D'un point de vue calculatoire, on peut décomposer un signal discret $f(t)$ sur un type d'ondelette par décomposition multi-niveaux. On donne en entrée un niveau souhaité (inférieur au niveau maximal). On obtient alors en sortie :

1. les «scaling coefficients» : coefficients de la fonction ϕ du niveau demandé
2. les «detail coefficients» : coefficients des fonctions ψ_k du niveau demandé et des niveaux supérieurs jusqu'au niveau maximal.

Au vu du nombre de coefficients en sortie, il est alors nécessaire d'appliquer un seuillage (threshold) sur le module des coefficients obtenus. On obtient alors une décomposition de f sous la forme suivante :

$$f(t) = \sum_{k \in \Lambda(L)} \alpha_k^L \phi_k^L(t) + \sum_{l=L}^{L_{max}} \sum_{k \in \Lambda(l)} \beta_k^l \psi_k^l(t)$$

Utilisation pour le projet :

Dans le cadre de notre projet, l'objectif est le suivant : **trouver l'ondelette type et les métaparamètres (L , le niveau de décomposition initial et ϵ , la valeur du threshold appliqué aux coefficients) qui représentent au mieux le signal.** Une fois ces paramètres établis, chaque signal serait défini par une centaine de features, reliés à des dilatations de la fonction mère ou à la fonction père et donc visuellement exploitables.

Pour atteindre cet objectif, nous avons défini comme fonction de coût à minimiser une combinaison linéaire positive du nombre de coefficients non nuls issus de la décomposition et de l'erreur quadratique après reconstruction avec ces coefficients. Cette fonction de coût correspond à un tradeoff entre exhaustivité et fidélité de l'information.

Notre étude des fonctions d'ondelette s'est portée sur 4 types d'ondelette : *db5*, *db6*, *sym7*, *sym8* dont les fonctions mère présentent des motifs identiques à ceux des ECG. Du fait de leur similarité, ces ondelettes ont un niveau d'affinement maximal identique, égal à 8 étant donné la longueur des ECG. Trois étapes de sélection s'ensuivent.

1. *Choix de L* : afin d'obtenir un nombre de coefficients après décomposition quasi-identique à la taille de départ, nous avons choisi $L = 5$.

2. *Choix de l'ondelette type* : comparaison des composantes de la fonction de coût pour les différentes familles d'ondelettes à threshold fixe pour tous les types d'ECG. Les figures 12 et 13 montrent par exemple une étape de cette comparaison pour les ondelettes *db5* et *db6* au niveau de la piste 0 des ECG. C'est finalement la famille *sym7* qui a été choisie.

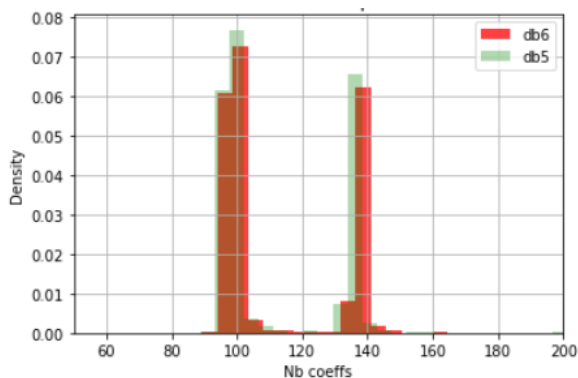


FIGURE 12 – Distribution du nombre de coefficients non nuls dans la décomposition des ECG

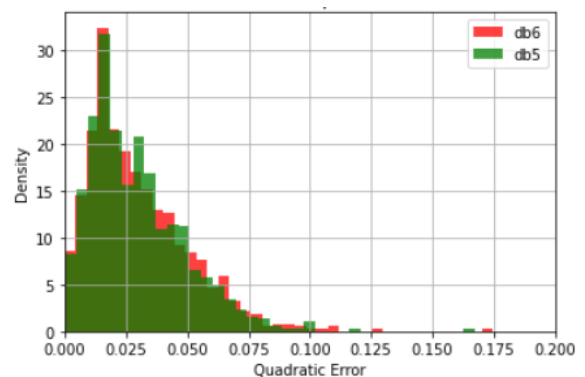


FIGURE 13 – Distribution de l'erreur quadratique dans la reconstruction des ECG

3. *Choix du threshold dans l'obtention des coefficients* (threshold s'appliquant au moment de la décomposition).

A l'issue de ces trois phases, les ECG sont donc représentés dans un espace d'environ 4160 dimensions pour chaque piste ECG, avec en moyenne 2% à 4% de ces dimensions portant des coefficients non nuls. Cependant, ces zéros ne sont pas forcément situés au même endroit selon les ECG. Il convient donc d'appliquer un nouveau seuillage uniformisant la sélection des coefficients de la décomposition en ondelette sur tous les ECG.

Pour cela, un threshold a été appliqué au regard de la distribution de la valeur des coefficients sur tous les types d'ECG. L'objectif était d'aboutir à une valeur de threshold identique à tous les niveaux de la décomposition. Par exemple, les figures 14 et 15 montrent l'effet de l'application d'un threshold égal à 0.01 au niveau de décomposition 6 sur les coefficients de la décomposition.

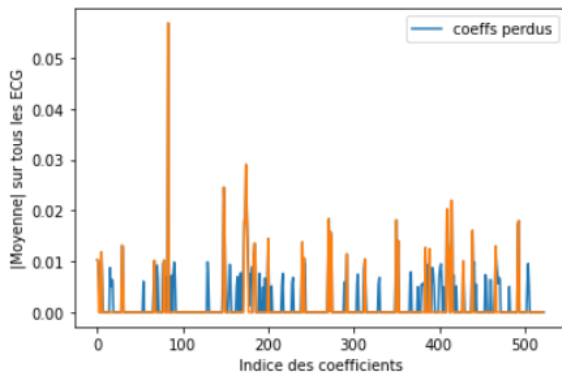


FIGURE 14 – Moyenne de la valeur absolue coefficients d'ondelette sur tous les ECG avant/après threshold

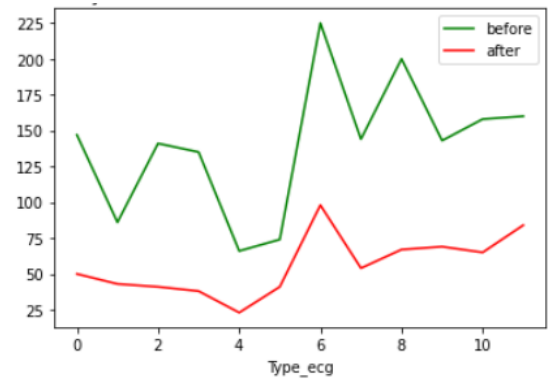


FIGURE 15 – Moyenne du nombre de coefficients non nuls avant/après threshold par type d'ECG

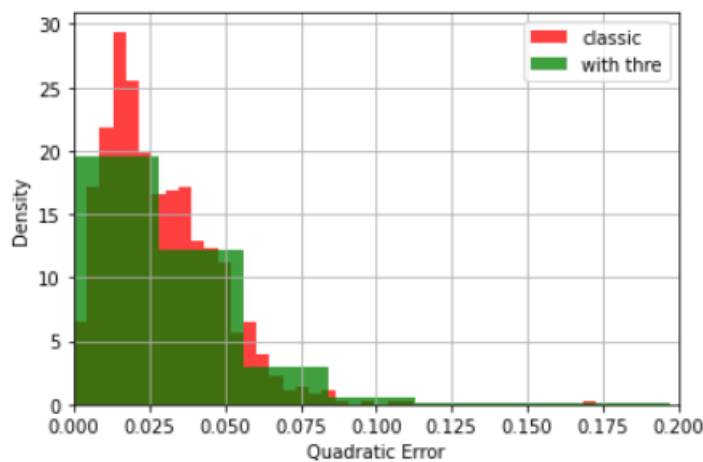


FIGURE 16 – Distribution de l'erreur quadratique avec/sans threshold global

Ce gain au niveau des coefficients est à mettre en perspective avec l'erreur quadratique comme le montre la figure 16 qui montre l'évolution de cette erreur sur la piste 0 des ECG avec utilisation du threshold. On remarque que mis-à-part quelques effets de bord, l'erreur quadratique est globalement conservée.

Conclusion : La représentation par ondelettes des signaux ECG permet d'obtenir un espace de dimension bien définie pour chaque piste ECG. En moyenne, un ECG est représenté par environ 400 coefficients pour une erreur quadratique proche de 3.10^{-2} . En considérant cette erreur comme le tradeoff nécessaire à une compression de l'espace de départ, on obtient ainsi une nouvelle famille de features potentiellement explicatives pour les ECG. Pour cela, on représente un patient par l'ensemble des coefficients de ses 12 pistes ECG, soit près de 5000 dimensions contre les quelques 50000 de départ.

2

ALGORITHMES D'EXPLICABILITÉ ÉTUDIÉS

2.1 ETHIK AI

Ethik est un module Python qui applique une méthode d'explicabilité globale de modèles de machine learning utilisant une projection entropique des variables[8]. En introduisant un biais dans les données par changement de mesure de probabilité, on peut quantifier la réponse du modèle à ce biais sans générer de nouvelles données et ainsi déterminer une forme d'influence des features des données sur la prédiction du modèle.

La théorie

Notons $X = ((X_i^1, X_i^2, \dots, X_i^p))_{i=1 \dots N} \in \mathbb{R}^p$ les données en entrée du modèle, $Y \in \mathbb{R}$ leur label et $Y' \in \mathbb{R}$ le label prédit par le modèle. En théorie, $(X, Y, Y') \sim \mathbb{P}$ où \mathbb{P} est une mesure de probabilité continue. En pratique, X étant de taille finie, \mathbb{P} est de la forme $\mathbb{P}^N = \frac{1}{N} \sum_{i=1}^N \delta_{X_i, Y_i, Y'_i}$.

On considère $\phi : (X, Y, Y') \in \mathbb{R}^{p+2} \rightarrow \phi(X, Y, Y') \in \mathbb{R}^k$. Cette fonction détermine le biais par lequel on souhaite comprendre le modèle.

Il existe $t_0 \in \mathbb{R}^k$ tel que $t_0 = \int_{\mathbb{R}^k} \phi d\mathbb{P} = \frac{1}{N} \sum_{i=1}^N \phi(X_i, Y_i, Y'_i)$, c'est à dire que t_0 est la moyenne de ϕ sous la mesure de probabilité \mathbb{P} . L'idée de la méthode est la suivante : pour $t \in \mathbb{R}^k$ fixé, on cherche une nouvelle mesure de probabilité \mathbb{Q}_t tel que :

$$\mathbb{Q}_t = \arg \inf_{\mathbb{Q} \in \Lambda_t} KL(\mathbb{Q}, \mathbb{P}) \quad \text{avec } KL \text{ la distance de Kullback-Leiler}$$

où Λ_t est l'ensemble des mesures de probabilités \mathbb{Q} sur \mathbb{R}^{p+2} tels que $\int_{\mathbb{R}^k} \phi d\mathbb{Q} = t$.

L'article [8] garantit l'existence et l'unicité d'un tel \mathbb{Q}_t . Par ailleurs, avec l'écriture $\mathbb{P} = \mathbb{P}^N$, on obtient une écriture explicite de \mathbb{Q}_t sous la forme $\mathbb{Q}_t^N = \frac{1}{N} \sum_{i=1}^N \lambda_i(t) \delta_{X_i, Y_i, Y'_i}$ où les $\lambda_i(t)$ sont des coefficients de pondération positifs qui dépendent des $\phi(X_i, Y_i, Y'_i)$.

De cette façon, on a opéré un changement de mesure de probabilité. Sous la nouvelle mesure \mathbb{Q}_t , $\phi(X, Y, Y')$ vaut en moyenne t : sans introduire de nouvelles données et de nouveaux labels, on crée un biais dans les données. Il ne reste plus qu'à étudier Y' sous \mathbb{Q}_t pour comprendre l'impact du biais.

Application d'Ethik AI à notre projet

Dans notre cas, on ne considère pas directement X_i sous sa forme d'origine (tableau de dimensions 4096x12). Ici, on décrit une observation par l'ensemble des features abordés en 1.2 et 1.3 : X_i prend ses valeurs dans $\mathbb{R}^{12 \times P}$ où P est le nombre total de features considérés s'appliquant sur une piste ECG.

Par ailleurs, on sous-traite le problème initial puisque l'on passe d'une classification à 2 outputs possibles pour 6 classes (qui plus est avec perméabilité entre les classes) à 6 problèmes de classification binaire. Autrement dit, on étudie le comportement du DNN maladie par maladie avec tous les ECG représentés par leurs features statistiques et temporels.

Considérons une maladie M . Pour une observation X_i , Y_i et Y'_i prennent leur valeur dans $\{0, 1\}$. Ayant cherché à donner une explication simple du diagnostic du DNN, nous avons utilisé Ethik avec les fonctions $\phi_{k,j}$ définies par $\phi_{k,j}(X, Y, Y') = X^{(k,j)} \in \mathbb{R}$ pour $k = 1 \dots 12$ et $j = 1 \dots P$. De cette façon, $\phi_{k,j}$ accentue le biais sur le j -ième feature de la i -ème piste de X pour obtenir en moyenne la valeur t . Pour t variant de $\min_{i=1 \dots N} X_i^{k,j}$ à $\max_{i=1 \dots N} X_i^{k,j}$, on peut calculer $P_1^{k,j}(t)$ la proportion de ($Y' = 1$) avec la formule suivante :

$$P_1^{k,j}(t) = \frac{1}{N} \sum_{i=1}^N \lambda_i^{k,j}(t) Y'_i$$

où les coefficients $\lambda_i^{k,j}(t)$ sont associés au changement de mesure de probabilité défini par $\phi_{k,j}$.

Pour chaque maladie M et chaque feature j , on peut ainsi tracer $P_1^{1,j}(t), P_1^{2,j}(t), \dots, P_1^{12,j}(t)$ (probabilité moyenne d'être malade pour une évolution du feature j sur les 12 pistes ECG) comme le montre la figure 17. Ce type de graphique permet de voir directement le couple (feature, piste) qui impacte le diagnostic du DNN.

Sur le graphique présenté, le domaine de définition des $\{P_1^{k,j}(t)\}_{k=1, \dots, 12}$ est uniformisé. Cette uniformisation s'effectue en deux temps. Pour chaque piste k , $[t_{min}^k, t_{max}^k]$ (de pas constant) est d'abord divisé en 2 intervalles $[t_{min}^k, t_{moy}^k]$ et $[t_{moy}^k, t_{max}^k]$, où t_{moy}^k est la moyenne du feature sur tous les ECG considérés. Ensuite, ces deux intervalles sont ramenés à deux domaines de distribution : $[t_{min}^k, t_{moy}^k]$ devient l'intervalle de quantiles $\tau \in [-1, 0]$ alors que $[t_{moy}^k, t_{max}^k]$ devient $\tau \in [0, 1]$. En concaténant ces deux intervalles, on obtient le domaine de définition uniformisé. De cette façon, $\tau = 0$ veut dire qu'on se place à la moyenne du feature, c'est à dire que l'on n'introduit aucun biais. On donne ainsi accès à la probabilité moyenne d'être malade, qui est donc identique pour toutes les pistes et tous les features. À $\tau > 0$, on se place au-dessus de la moyenne du feature ; à $\tau < 0$, on se place en-dessous. Toutefois, les écarts numériques à partir de $\tau = 0$ ne sont pas comparables entre les domaines positifs et négatifs car leur transformation est indépendante. On travaille plutôt sur des tendances.

Cette utilisation des quantiles permet ainsi d'uniformiser la lecture de l'influence de features vivant sur des gammes de valeurs différentes. Pour comparer l'influence entre plusieurs variables d'intérêt $X^{k,j}$, nous nous plaçons donc à un certain τ (soit positif, soit négatif) et étudions les

$P_1^{k,j}(\tau) = P_1^{k,j}(t^{k,j})$ où $t^{k,j}$ est la valeur associée au quantile τ pour la distribution correctement tronquée des $(X_i^{k,j}, Y_i, Y'_i)_{i=1\dots N}$ (respectivement au-dessus ou en-dessous de la moyenne). En prenant par exemple $\tau = 0.7$ et $\tau = -0.7$, on s'écarte assez de la moyenne sans considérer une valeur extrême (maximale ou minimal) par ailleurs. L'explication des prédictions pourra ainsi avoir un poids statistique raisonnable.

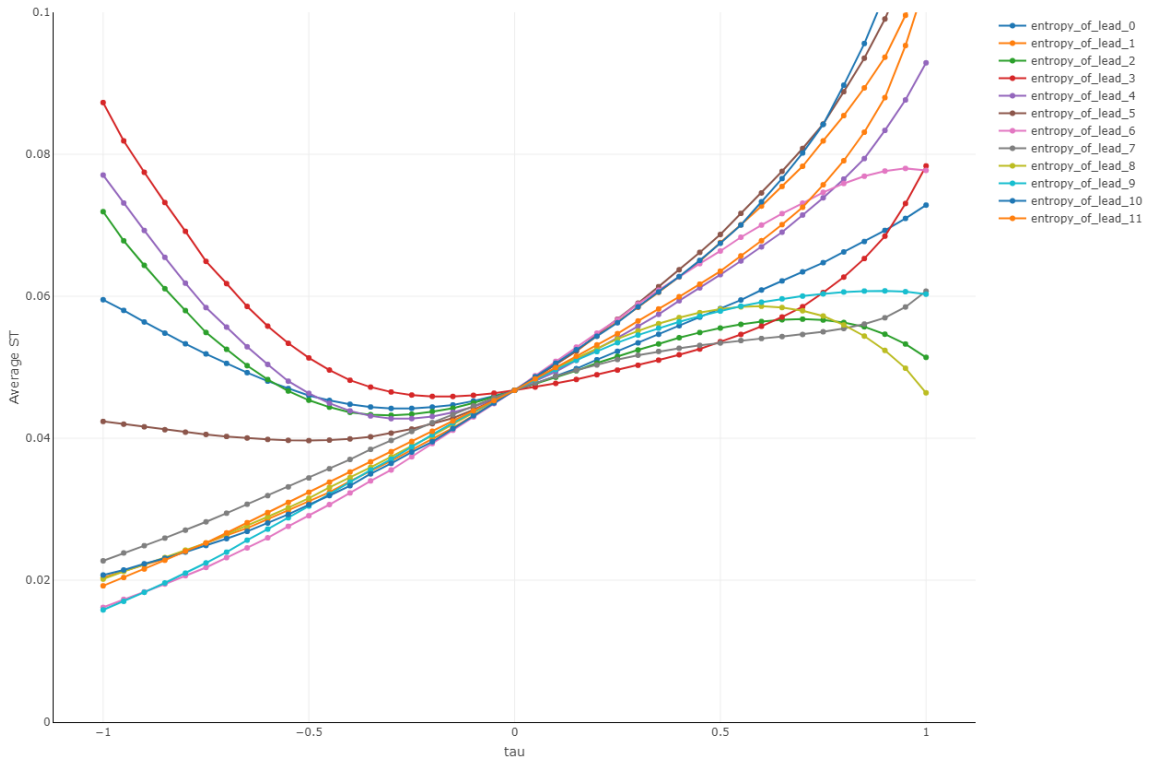


FIGURE 17 – Influence de l'entropie des pistes 1 à 12 sur la prédiction de la pathologie ST

2.2 LOCAL INTERPRETABLE MODEL-AGNOSTIC EXPLANATIONS (LIME)

Dans le cadre des features issus de la représentation fonctionnelle des ECG sous forme d'on-delettes (coefficients uniformément seuillés de la décomposition des 12 pistes ECG), la réduction de dimension est environ de facteur 10. Cependant, il ne serait pas raisonnable d'utiliser cette représentation avec Ethik puisque cela reviendrait à comparer de façon globale sur quelques milliers de features l'évolution de la probabilité d'être malade. Dans le cas de données aux grandes dimensions, l'approche locale du modèle est une piste intéressante (notons qu'Ethik AI apporte une compréhension globale) qu'il faut explorer. La méthode LIME en est un très bon exemple : elle consiste à comprendre localement le modèle «black-box» du DNN grâce à un modèle beaucoup plus simple (modèle «glass-box»)[9].

Fonctionnement

1. On définit tout d'abord un point d'intérêt $x \in V$ dans les données autour duquel on souhaite expliquer le comportement du modèle. Dans notre cas, il s'agit par exemple d'un patient malade, représenté par la concaténation des coefficients issus de la décomposition par ondelette de ses 12 pistes ECG (dimension proche de 5000).
2. Pour approcher le DNN au voisinage du point x , LIME transporte l'espace d'origine dans un espace de dimension beaucoup moins importante (typiquement, on passe de données réelles à des données catégoriques). Soit h cette fonction de mapping, on note $x' = h(x)$ et $V' = h(V)$.
3. Une fois x' obtenu, LIME génère N points au voisinage de x' dans V' . Cette génération de données (qui n'existent pas à l'origine) est le plus souvent obtenue à partir de perturbations gaussiennes autour de x' . Cette étape permet ainsi de s'affranchir de la complexité de la représentation d'origine. On note z'_1, z'_2, \dots, z'_N ces points, y'_1, y'_2, \dots, y'_N les sorties du DNN qui leur sont associées et $w'_1 = d_{V'}(x', z'_1), w'_2 = d_{V'}(x', z'_2), \dots, w'_N = d_{V'}(x', z'_N)$ les distances de x' à ces points.
4. LIME procède ensuite à une régression de type $K-LASSO$ sur (Z', Y') avec les poids W' , c'est à dire une régression linéaire qui fait intervenir les K features les plus explicatifs. Ce sont ces features que l'on peut alors désigner comme marque d'explicabilité de la pathologie du patient x .

Cette méthode comporte plusieurs points positifs. Tout d'abord, elle ne tient pas compte d'a priori sur le fonctionnement du modèle, nécessitant uniquement la fonction de régression du modèle. Elle permet aussi d'avoir une représentation interprétable des données : par exemple, si $A(x)$ est un feature défini sur \mathbb{R} , LIME opte généralement pour une représentation catégorielle de ce feature par le couple de features $(A(x) > b, A(x) \leq b)$ qui est tout à fait compréhensible par l'utilisateur. Enfin, cette méthode est adaptée à notre projet, puisqu'on s'intéresse au comportement du DNN aux points d'intérêt rares et localisés que sont les patients malades.

D'un point de vue pratique, LIME réalise la sélection des points dans le voisinage de x' de manière automatique. Cette sélection peut s'avérer biaisée dans le cas de données de dimension très grande. Par ailleurs, le résultat final de l'algorithme n'est qu'une simple approximation et l'erreur de régression n'est pas forcément uniforme au voisinage de tout patient x .

Application de Lime pour notre projet

L'utilisation de LIME s'est révélée être très compliquée dans le cas de notre projet. Tout d'abord, la classification du DNN sur les pathologies cardiaques est multi-classe sans être pour autant unique dans son affectation (un patient peut être porteur de deux pathologies différentes). Nous avons donc de raisonner par maladie en changeant le problème de classification à 6 classes "Malade / Non malade" par 6 problèmes de classification à 2 classes : "Malade / Non malade" imperméables.

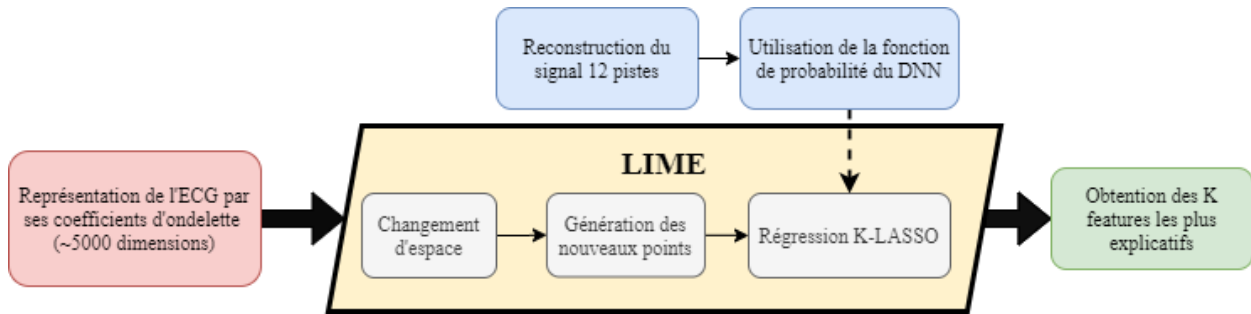


FIGURE 18 – Algorithme adapté de LIME autour d'un certain patient

On se donne une pathologie et un patient x porteur de ce maladie, autour duquel on lance l'algorithme de LIME, comme le montre la figure 18. En appliquant ce processus à tous les patients malades, on peut alors faire une distribution statistique des features les plus explicatifs pour cette maladie.

3

RÉSULTATS D'EXPLICABILITÉ

Les résultats présentés dans cette section ont été obtenus avec Ethik AI. En suivant la démarche présentée en 2.1, nous avons déterminé l'impact des features (calculés pour les 12 pistes ECG) sur le diagnostic du DNN et du Gold Standard. En reprenant les notations utilisées en 2.1, nous avons calculé deux valeurs pour chaque couple (k, j) :

- $\frac{P_1^{k,j}(0.7) - P_1^{k,j}(0)}{P_1^{k,j}(0)}$: écart relatif de la probabilité moyenne d'être malade si le feature a une valeur plus élevée que la moyenne,
- $\frac{P_1^{k,j}(-0.7) - P_1^{k,j}(0)}{P_1^{k,j}(0)}$: écart relatif de la probabilité moyenne d'être malade si le feature a une valeur plus petite que la moyenne,

avec k le numéro de la piste ECG et j le numéro du feature. Les seuils 0.7 et -0.7 ont été choisis de sorte à s'écarter de la moyenne du feature (0) sans toutefois atteindre les valeurs maximales et minimales (respectivement -1 et 1), où l'on peut trouver des effets de bord ou du sur-apprentissage. Les résultats numériques obtenus (intégralement présentés en annexe) donnent ainsi une idée de l'accroissement de la probabilité d'être malade en fonction d'une valeur de feature plus petite ou plus élevée que la moyenne.

En accord avec l'objectif de notre projet de recherche, nous avons décidé de présenter dans cette partie du rapport une version synthétisée et «physiquement» exploitable de ces résultats numériques. Celle-ci met en comparaison des indicateurs visuels et calculatoires des pistes ECG des patients diagnostiqués malades par le DNN avec la piste ECG «standard» correspondante. Cette démarche est réalisée pour chaque maladie et met en relief des irrégularités («plus petit» ou «plus élevé») pour une ou plusieurs pistes.

3.1 PATHOLOGIE 1dAVB

Les caractéristiques ci-dessous ont été identifiées comme étant les plus pertinentes pour expliquer la prédiction de bloc atrio-ventriculaire (ou auriculo-ventriculaire) du premier degré (1dAVb). Les résultats numériques et complets pourront être trouvés en annexe.

ECG 11

- médiane/moyenne du signal plus élevée
- signal plus dense autour de sa médiane
- minimum du signal plus élevé

ECG 9

- médiane/moyenne du signal plus petite
- maximum du signal plus petit

ECG 0

- proportion des valeurs du signal sous la médiane plus élevée
- fréquence du signal plus petite

ECG 7

- fréquence du signal plus petite
- portée verticale du signal plus petite

3.2 PATHOLOGIE RBBB

Les caractéristiques ci-dessous ont été identifiées comme étant les plus pertinentes pour expliquer la prédiction de bloc de branche droit (RBBB). Les résultats numériques et complets pourront être trouvés en annexe.

Sur les ECG 0,1,2,4,10,11

- proportion des valeurs du signal sous la médiane plus élevée

Sur les ECG 5,6,7,8

- proportion des valeurs du signal sous la médiane plus petite

3.3 PATHOLOGIE LBBB

Les caractéristiques ci-dessous ont été identifiées comme étant les plus pertinentes pour expliquer la prédiction de bloc de branche gauche (LBBB). Les résultats numériques et complets pourront être trouvés en annexe.

ECG 9 et 10

- proportion des valeurs du signal sous la médiane plus élevée
- signal plus irrégulier (horizontal et vertical)
- ECG 9 : valeurs du signal plus regroupées vers le bas par rapport à la moyenne
- ECG 10 : valeurs du signal plus largement étalées autour de sa moyenne

ECG 0 et ECG 11

- proportion des valeurs au-dessus de la médiane plus élevée
- valeurs du signal plus largement étalées autour de sa moyenne
- signal plus irrégulier (horizontal et vertical)

3.4 PATHOLOGIE SB

Les caractéristiques ci-dessous ont été identifiées comme étant les plus pertinentes pour expliquer la prédiction de bradycardie sinusale (SB). Les résultats numériques et complets pourront être trouvés en annexe.

Sur la plupart des ECG

- fréquence du signal plus petite
- asynchronie moyenne plus marquée avec l'ECG standard (en retard)

ECG 6

- proportion des valeurs du signal plus élevée dans la fenêtre d'observation des quantiles (25-75)
- plus d'irrégularité dans les motifs du signal
- médiane du signal plus élevée

ECG 3

- proportion des valeurs du signal sous la médiane plus élevée
- valeurs du signal plus regroupées vers le bas par rapport à la moyenne

3.5 PATHOLOGIE AF

Les caractéristiques ci-dessous ont été identifiées comme étant les plus pertinentes pour expliquer la prédiction de fibrillation auriculaire (AF). Les résultats numériques et complets pourront être trouvés en annexe.

Sur la plupart des ECG

- irrégularité des motifs du signal plus élevée
- proportion des valeurs du signal sous la médiane plus élevée

ECG 5

- valeurs du signal plus regroupées autour de la moyenne
- portée verticale du signal plus petite

ECG 9

- valeurs du signal plus regroupées vers le bas par rapport à la moyenne
- médiane du signal plus élevée

ECG 3

- valeurs du signal plus regroupées vers le bas par rapport à la moyenne
- portée verticale du signal plus petite

ECG 0

- valeurs du signal plus regroupées autour de la moyenne
- portée verticale du signal plus petite

3.6 PATHOLOGIE ST

Les caractéristiques ci-dessous ont été identifiées comme étant les plus pertinentes pour expliquer la prédiction de tachycardie sinusale (ST). Les résultats numériques et complets pourront être trouvés en annexe.

Sur la plupart des ECG

- fréquence du signal plus élevée
- signal plus irrégulier (horizontalement et verticalement)

ECG 6

- valeurs du signal plus largement étalées autour de la moyenne

4 DISCUSSION

4.1 LIMITES DE NOTRE TRAVAIL

4.1.1 LIMITATIONS DE LA BASE DE DONNÉES

L'utilisation de DNN en diagnostic cardiaque est encore balbutiante. L'un des facteurs contribuant à cette situation est le manque de bases de données numériques complètes d'ECG standards à 12 pistes, car la plupart des enregistrements sont encore enregistrés uniquement sur papier, archivés sous forme d'images ou stockés au format PDF. La plupart des bases de données disponibles comprennent quelques centaines de tracés et aucune annotation systématique de la liste complète des diagnostics ECG, ce qui limite leur utilité en tant qu'ensembles de données d'entraînement dans un contexte d'apprentissage supervisé. L'équipe du projet [1] a comblé ce manque grâce au Réseau de télésanté du Minas Gerais (TNMG), un système public de télésanté qui aide 811 des 853 municipalités de l'état du Minas Gerais, au Brésil. Cependant avec une telle quantité de données il est fortement probable d'avoir des erreurs.

Même dans les données de test, les 827 ECG à 12 pistes sur lesquels nous avons travaillé, un nombre important d'enregistrements sont inattendus voire problématiques, comme l'ECG du patient 20 dont nous avons parlé plus haut (ici en figure 19).

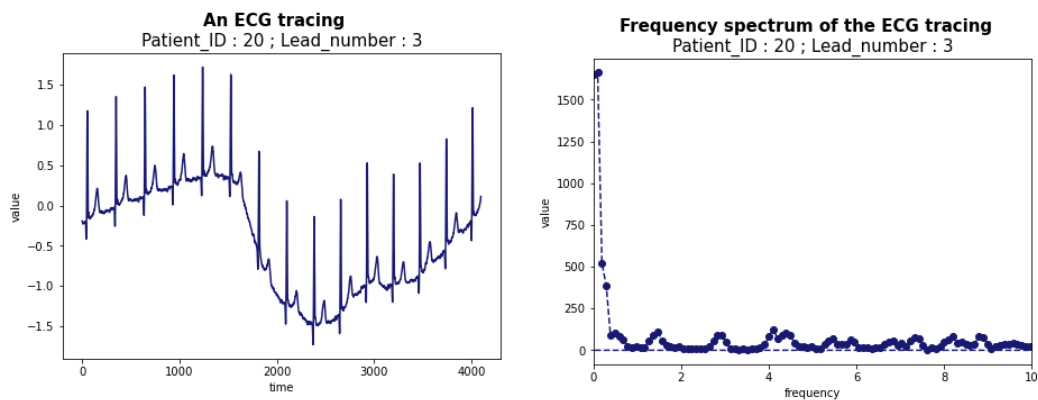


FIGURE 19 – Signal et spectre en fréquence de l'ECG 3 du patient 20

Nous ignorons si le décalage atypique de cette courbe résulte d'un problème d'enregistrement, d'un problème de numérisation, ou encore si cet ECG peut être considéré comme normal par un cardiologue, sachant que ce patient n'a pas été déclaré atteint d'une quelconque maladie. Un autre problème qui peut surgir est celui des pistes nulles : pour un certain nombre d'enregistrements, certains pistes voire toutes sont nulles, nous avons donc dû retirer ces patients.

4.1.2 LIMITATIONS DES MÉTHODES EXPLORÉES

En implémentant la méthode d'explicabilité issue de LIME, nous nous sommes heurtés à un problème de taille : la représentation des ECG sous forme d'ondelettes conserve très mal les prédictions réalisées par le DNN. Malgré toutes les précautions prises dans la sélection des coefficients d'ondelette (choix de la famille, du niveau de décomposition et de la valeur du threshold), les signaux issus de la reconstruction à partir des coefficients donnent en sortie du DNN une distribution de probabilité complètement bruitée. Même en diminuant la valeur de seuillage (ce qui augmente par ailleurs le nombre de coefficients représentatifs), l'output du DNN ne conserve pas la distribution de probabilité d'origine. Les figures 20 et 21 montrent en effet qu'un seuillage beaucoup plus fin n'améliore pas vraiment les prédictions pour la maladie *1dAVb* (*former* pour la prédiction d'origine et *new* pour la prédiction avec ondelette).

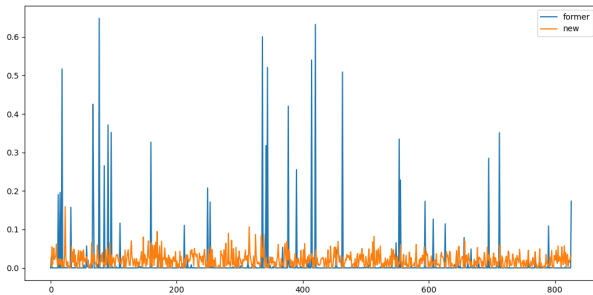


FIGURE 20 – Distribution de $\mathbb{P}(\text{malade})$ pour les 827 patients (threshold ϵ_1)

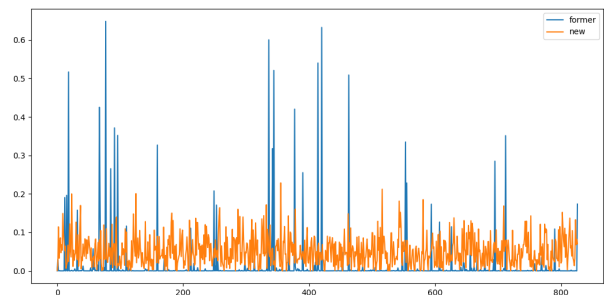


FIGURE 21 – Distribution de $\mathbb{P}(\text{malade})$ pour les 827 patients (threshold $\epsilon_2 < \epsilon_1$)

Ceci met ainsi en lumière l'extrême non linéarité du DNN, dont la sortie est davantage dépendante des ECG en tant qu'image plutôt qu'en tant que signal. Pour pallier ce problème, on pourrait éventuellement envisager la reconstruction du signal ECG à partir des ondelettes grâce à un autre réseau de neurones.

4.2 LA VISION DU RÉSEAU DE NEURONES

Au delà de l'identification des pistes les plus explicatives relativement aux features que nous avons choisi de calculer, les tableaux d'influence de la valeur des features sur la probabilité de présenter une pathologie permettent aussi de comparer la vision des cardiologues à celle du réseau de neurones.

Il est intéressant de noter que pour la plupart des features étudiés, le DNN et le Gold Standard (GS, i.e. le diagnostic des médecins) sont d'accord, dans le sens où l'influence statistique des valeurs extrêmes des features sur le diagnostic est du même ordre de grandeur. En annexe se trouvent les tableaux numériques d'influence des features, et on peut voir sur les pathologies LBBB et ST que le DNN et le GS se mettent d'accord (avec une particularité pour LBBB : les prédictions sur les données de test sont exactement les mêmes, il est donc naturel que les

influences soient les mêmes).

Cependant il est notable aussi que pour certaines maladies l'influence des features sur le diagnostic du DNN n'est pas corrélée à celle sur le diagnostic du GS. Dans les tableaux d'explicabilité en annexe nous avons noté en gras les cas où le DNN prend beaucoup plus en compte une feature que le GS. Par exemple, on pourra regarder la bradycardie (SB) en $\tau = +0.7$, où l'on note que le DNN donne une influence très positive aux écarts interpercentiles et à l'entropie, alors que le GS leur donne une influence négative. C'est la même situation que pour la pathologie RBBB en $\tau = -0.7$ pour l'entropie, cependant ici le GS prend tout de même en compte les écarts interpercentiles, mais deux fois moins que le DNN.

A l'inverse, il existe des features que le GS prend beaucoup plus en compte que le DNN. On peut regarder par exemple certains écarts interpercentiles pour la fibrillation (AF), où il y a un facteur 10 entre l'influence sur le GS et sur le DNN. Ainsi le DNN «décide» de les ignorer et effectue son diagnostic grâce à d'autres features. De même pour la tachycardie (ST), il y a un facteur 10 sur la fréquence de la piste 7, alors que les features les plus explicatifs pour le DNN sont tout de même la fréquence, simplement appliquée à d'autres pistes.

Ainsi il semble que le GS tienne compte des features de manière plus uniforme, alors que tout se passe comme si le DNN faisait un choix de features explicatives, en ignorant les autres. Ceci peut s'expliquer d'une certaine manière par l'algorithme de descente de gradient : plutôt que de chercher un groupe de features explicatives comme le fait le GS, le DNN choisit l'axe de pente maximale et le descend autant qu'il le peut, augmentant drastiquement l'influence de quelques features et ignorant quasiment le reste du groupe. Par ailleurs, on remarque que l'influence des features statistiques sur le diagnostic du DNN est beaucoup plus prononcée que pour le diagnostic des cardiologues, qui s'attacheraient davantage à des features plus simples à extraire comme la fréquence ou l'écart inter-quantile. Cela met en lumière une prise en compte plus globale de l'ECG du côté du DNN et pourrait être une façon d'améliorer le diagnostic des cardiologues.

4.3 PERSPECTIVES OUVERTES

La littérature concernant l'analyse automatique des ECG est grandissante et par conséquent il existe de nombreuses méthodes pour extraire des features des ECG que nous n'avons pas étudiées. Par exemple, pour effectuer un diagnostic, les cardiologues prêtent beaucoup d'attention à la structure des motifs de l'ECG, que l'on peut voir en figure 22 : les longueurs des différents pics, les ratios de leurs hauteurs, etc. Malheureusement et comme nous l'avons montré précédemment, il est difficile de débruiter les ECG de notre base de données sans détruire d'information. Plus généralement, il serait intéressant de rassembler des connaissances en électrocardiographie.

Une étape suivante de notre travail pourrait être le traitement des ECG afin d'extraire ces motifs. Plusieurs méthodes sont actuellement émergentes : soit par *dictionary learning*, soit

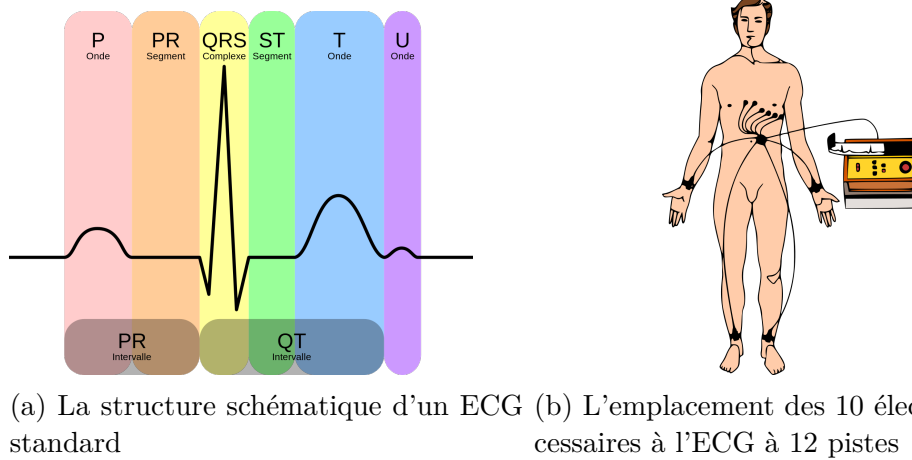


FIGURE 22 – Les rudiments de l'électrocardiographie

en utilisant un réseau de neurones chargé de classifier les zones définies par les cardiologues (comme présenté dans [7]).

Durant notre projet, nous nous sommes également intéressés à une méthode de visualisation sur les pistes ECG des zones activées sur par les dernières couches convolutionnelles du DNN. Il s'agit de méthodes «CAM» ou «Grad-CAM», dont un exemple appliqué à des données temporelles de type ECG est présenté en figure [10]. On y retrouve notamment un classifieur binaire qui donne un diagnostic sur des pathologies cardiaques. Une façon de confirmer les résultats obtenus avec Ethik consisterait à dessiner sur les pistes ECG des patients malades la heatmap d'activation et de mettre en lien les zones de forte activation avec les features explicitées en 3. Un exemple d'une telle heatmap d'activation appliquée à un ECG est montré sur la figure 23. En raison de la complexité de l'implémentation, nous n'avons malheureusement pas eu l'occasion d'explorer cette piste davantage.

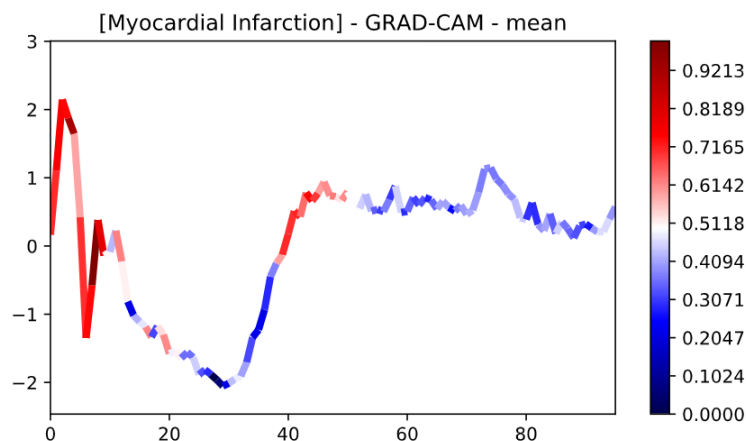


FIGURE 23 – Carte d'activation de type GRAD-CAM appliquée à une piste ECG d'un patient atteint d'un infarctus du myocarde

CONCLUSION

L'objectif de ce travail était de prendre du recul par rapport au modèle «black-box» mis en avant par l'équipe de recherche de l'article[1]. Si cette classification automatique de bout en bout apporte des performances supérieures aux réseaux de neurones usuels en électrocardiographie, qui prennent en entrée des features des signaux et non des signaux bruts, elle perd néanmoins en explicabilité puisqu'elle construit un modèle «black-box» opaque. Nos résultats d'explicabilité font ainsi un pas en arrière dans cette démarche puisqu'ils permettent d'identifier *a posteriori* les features des signaux qui influencent le DNN pour son diagnostic. De plus, notre travail permet de différencier ce que regardent les cardiologues de ce que regarde le réseau de neurones. Enfin nous montrons que la non-linéarité du DNN se constate notamment lorsque l'on cherche à appliquer le réseau de neurones sur des signaux reconstruits linéairement, par exemple à partir d'une base d'ondelettes. Finalement, on pourra approfondir l'étude de ce réseau grâce à d'avantage de connaissances en électrocardiographie ou encore en explorant l'identification de motifs d'intérêts ou la reconstruction d'un signal grâce à un second réseau de neurones.

RÉFÉRENCES

- [1] Antônio H. Ribeiro, M. H. Ribeiro, G. Paixão, D. Oliveira, P. R. Gomes, Jéssica A. Canazart, M. P. Ferreira, Carl R. Andersson, P. Macfarlane, W. Meira, Thomas Bo Schön, and A. Ribeiro. Automatic diagnosis of the 12-lead ecg using a deep neural network. *Nature Communications*, 11, 2020.
- [2] Alan V Oppenheim, John R Buck, and Ronald W Schafer. *Discrete-time signal processing. Vol. 2*. Upper Saddle River, NJ : Prentice Hall, 2001.
- [3] Simon Haykin and Barry Van Veen. *Signals and systems*. John Wiley & Sons, 2007.
- [4] Béatrice Laurent-Bonneau. High-dimensional-statistics, 2017. <https://github.com/Beatrice-Laurent/High-dimensional-statistics>.
- [5] David JC MacKay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.
- [6] Stephen Butterworth et al. On the theory of filter amplifiers. *Wireless Engineer*, 7(6) :536–541, 1930.
- [7] R. Dubois. *Application des nouvelles méthodes d'apprentissage à la détection précoce d'anomalies cardiaques en électrocardiographie : Chapitre 6-Modélisation du battement cardiaque*. PhD thesis, Laboratoire Signaux, Modèles et Apprentissage Statistique, Université Pierre et Marie Curie - Paris VI, 2004.
- [8] Francois Bachoc, Fabrice Gamboa, Max Halford, Jean-Michel Loubes, and Laurent Risser. Entropic variable projection for explainability and intepretability. *arXiv preprint arXiv :1810.07924*, 2018.
- [9] Przemyslaw Biecek and Tomasz Burzykowski. *Explore, Explain, and Examine Predictive Models*. Chapman & Hall, 2020.
- [10] AI Fast Track. Model interpretation for ecg200 dataset. https://colab.research.google.com/github/ai-fast-track/timeseries/blob/master/nbs/cam_tutorial_ECG200.ipynb.
- [11] Gregory A Roth, Degu Abate, Kalkidan Hassen Abate, Solomon M Abay, Cristiana Ababafati, Nooshin Abbasi, Hedayat Abbastabar, Foad Abd-Allah, Jemal Abdela, Ahmed Abdelalim, et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017 : a systematic analysis for the global burden of disease study 2017. *The Lancet*, 392(10159) :1736–1788, 2018.

ANNEXES

TABLEAUX COMPLETS D'EXPLICABILITÉ

Ci-dessous nous présentons les résultats numériques des calculs d'influence des features sur la probabilité de présenter une pathologie. Il y a donc 6 tableaux correspondants aux 6 pathologies.

Comment lire ces tableaux :

Pour chaque pathologie nous présentons deux tableaux. A chaque ligne correspond une certaine feature j pour un certain numéro de piste ECG. Les colonnes quant à elles sont :

- $\tau = -0.7$ (en %) :

$$\frac{P_1^{k,j}(-0.7) - P_1^{k,j}(0)}{P_1^{k,j}(0)}$$

à savoir l'écart relatif de la probabilité moyenne d'être malade si la feature a une valeur moins élevée que la moyenne, pour le DNN ainsi que pour le Gold Standard (GS, le diagnostic des médecins). La probabilité nominale ou moyenne d'être atteint de cette pathologie selon le DNN est donnée en première ligne.

- $\tau = +0.7$ (en %) :

$$\frac{P_1^{k,j}(0.7) - P_1^{k,j}(0)}{P_1^{k,j}(0)}$$

à savoir l'écart relatif de la probabilité moyenne d'être malade si la feature a une valeur plus élevée que la moyenne, pour le DNN ainsi que pour le GS. La probabilité nominale ou moyenne d'être atteint de cette pathologie selon le GS est donnée en première ligne.

Par exemple pour la première pathologie (1dAVb), le DNN prédit que la probabilité moyenne pour les patients d'en souffrir est de 0.036 (0.033 pour le GS), et avoir une moyenne de la piste 9 moins élevée que la moyenne (tableau de gauche) conduit à une augmentation de la probabilité de présenter la pathologie de 58% selon le DNN, alors qu'avoir un écart médian absolu à la médiane (median_absolute_value) plus élevé que la moyenne (tableau de droite) conduit à une augmentation de la probabilité de présenter la pathologie de 125% selon le GS.

MALADIE 1DAVB

		DNN (P=0.036)	GS (P=0.033)
ECG	Feature	$\tau=-0.7$ (en %)	$\tau=-0.7$ (en %)
lead_9	average	58	63
lead_9	signal_magnitude_area	58	63
lead_9	midhinge	47	51
lead_0	frequencies_via_fft	41	61
lead_7	frequencies_via_fft	41	62
lead_9	trimean	38	42
lead_9	maximum	36	41
lead_7	range	35	41
lead_7	mid_range	35	41
lead_9	median	35	38
lead_10	interpercentile_range_10_90	30	53
lead_11	interpercentile_range_10_90	30	41
lead_10	interpercentile_range_50_100	30	53
lead_11	interpercentile_range_50_100	30	41
lead_9	kurtosis	28	17
lead_6	interpercentile_range_0_50	27	-2
lead_11	entropy	26	38
lead_0	interpercentile_range_10_90	24	42
lead_0	interpercentile_range_50_100	24	42
lead_4	gm_asymetry2	23	23
lead_8	kurtosis	22	11
lead_5	standard_deviation	20	30
lead_10	interpercentile_range_0_50	20	34
lead_7	standard_deviation	19	30
lead_6	kurtosis	19	9
lead_5	skewness	19	4
lead_5	frequencies_via_fft	19	37
lead_8	frequencies_via_fft	19	29
lead_0	interpercentile_range_25_75	19	39
lead_4	range	17	25
lead_4	mid_range	17	25
lead_1	standard_deviation	16	33
lead_5	interquartile_range	16	32
lead_3	frequencies_via_fft	16	33
lead_8	maximum	15	19

		DNN (P=0.036)	GS (P=0.033)
ECG	Feature	$\tau=+0.7$ (en %)	$\tau=+0.7$ (en %)
lead_11	median_absolute_value	113	125
lead_11	interpercentile_range_25_75	96	99
lead_11	interpercentile_range_40_60	93	97
lead_11	interpercentile_range_10_90	69	71
lead_11	interpercentile_range_50_100	69	71
lead_0	interpercentile_range_0_50	57	46
lead_11	minimum	55	68
lead_11	interquartile_range	53	59
lead_11	average	48	56
lead_11	signal_magnitude_area	48	56
lead_11	median	46	54
lead_11	midhinge	46	55
lead_11	trimean	45	54
lead_0	skewness	40	68
lead_6	interpercentile_range_40_60	39	43
lead_8	interpercentile_range_40_60	38	44
lead_9	interpercentile_range_25_75	37	42
lead_3	interpercentile_range_25_75	36	-27
lead_6	interpercentile_range_10_90	34	32
lead_6	interpercentile_range_50_100	34	32
lead_11	skewness	28	50
lead_9	entropy	26	20
lead_4	interpercentile_range_0_50	25	21
lead_11	kurtosis	24	42
lead_0	minimum	23	33
lead_11	auto_correlation	23	16
lead_6	interpercentile_range_25_75	23	26
lead_2	gm_asymetry2	20	13
lead_0	average	19	24
lead_0	signal_magnitude_area	19	24
lead_5	gm_asymetry	19	-19
lead_1	interpercentile_range_0_50	18	2
lead_6	skewness	17	-9
lead_11	energy	17	20
lead_6	entropy	17	10
lead_8	auto_correlation	17	9
lead_9	interpercentile_range_10_90	17	15
lead_9	interpercentile_range_50_100	17	15
lead_2	interpercentile_range_10_90	16	-16
lead_2	interpercentile_range_50_100	16	-16
lead_11	entropy	15	1
lead_10	skewness	15	34
lead_0	median	15	20

MALADIE RBBB

		DNN (P=0.046)	GS (P=0.041)
ECG	Feature	$\tau=-0.7$ (en %)	$\tau=-0.7$ (en %)
lead_3	entropy	45	-17
lead_0	auto_correlation	16	-6
lead_4	entropy	58	-8
lead_2	minimum	39	3
lead_4	interpercentile_range_10_90	25	2
lead_4	interpercentile_range_50_100	25	2
lead_4	minimum	24	5
lead_4	interpercentile_range_40_60	19	5
lead_3	interpercentile_range_25_75	16	5
lead_4	interpercentile_range_25_75	25	10
lead_3	interpercentile_range_40_60	17	8
lead_2	interpercentile_range_10_90	30	15
lead_2	interpercentile_range_50_100	30	15
lead_2	interpercentile_range_40_60	26	14
lead_2	interpercentile_range_25_75	27	15
lead_9	skewness	15	11
lead_7	interquartile_range	16	13
lead_2	gm_asymetry2	97	81
lead_2	entropy	58	50
lead_0	maximum	34	30
lead_1	gm_asymetry2	217	192
lead_0	skewness	56	51
lead_4	gm_asymetry2	251	229
lead_11	skewness	103	96
lead_7	standard_deviation	30	28
lead_5	range	31	29
lead_5	mid_range	31	29
lead_5	interpercentile_range_0_50	154	149
lead_10	skewness	35	34
lead_1	skewness	48	47
lead_7	range	63	62
lead_7	mid_range	63	62
lead_4	skewness	121	122
lead_6	kurtosis	59	60
lead_7	interpercentile_range_0_50	149	152
lead_6	interpercentile_range_0_50	337	353
lead_1	gm_asymetry	131	138
lead_4	gm_asymetry	56	59
lead_8	interpercentile_range_0_50	95	102
lead_11	maximum	30	33
lead_1	maximum	46	51
lead_2	skewness	54	61
lead_5	kurtosis	84	98

		DNN (P=0.046)	GS (P=0.041)
ECG	Feature	$\tau=+0.7$ (en %)	$\tau=+0.7$ (en %)
lead_3	maximum	40	-2
lead_5	gm_asymetry	41	32
lead_0	frequencies_via_fft	31	27
lead_0	interpercentile_range_0_50	133	121
lead_7	minimum	56	53
lead_7	skewness	80	76
lead_1	interpercentile_range_0_50	265	252
lead_11	interpercentile_range_0_50	94	92
lead_4	interpercentile_range_0_50	189	185
lead_5	skewness	57	56
lead_6	minimum	36	36
lead_9	frequencies_via_fft	20	20
lead_2	interpercentile_range_0_50	73	74
lead_5	gm_asymetry2	97	99
lead_10	interpercentile_range_0_50	68	70
lead_6	skewness	211	223
lead_5	minimum	81	86
lead_6	entropy	39	43
lead_11	average	32	36
lead_11	signal_magnitude_area	32	36
lead_11	median	32	36
lead_11	midhinge	32	36
lead_11	trimean	31	35
lead_11	frequencies_via_fft	54	61
lead_3	frequencies_via_fft	40	46
lead_11	median_absolute_value	84	98
lead_11	interquartile_range	37	44
lead_11	interpercentile_range_40_60	65	78
lead_11	interpercentile_range_25_75	73	88
lead_8	minimum	28	34
lead_8	skewness	30	37
lead_5	entropy	58	72
lead_10	interpercentile_range_10_90	45	56
lead_10	interpercentile_range_50_100	45	56
lead_7	entropy	38	48
lead_8	entropy	37	47
lead_11	entropy	52	67
lead_11	interpercentile_range_10_90	46	60
lead_11	interpercentile_range_50_100	46	60
lead_0	entropy	38	51
lead_1	entropy	23	35
lead_5	interpercentile_range_10_90	18	29
lead_5	interpercentile_range_50_100	18	29

MALADIE LBBB

		DNN (P=0.036)	GS (P=0.036)
ECG	Feature	$\tau=-0.7$ (en %)	$\tau=-0.7$ (en %)
lead_9	skewness	182	182
lead_0	kurtosis	111	111
lead_1	gm_asymetry2	96	96
lead_10	kurtosis	95	95
lead_11	kurtosis	93	93
lead_2	gm_asymetry2	68	68
lead_9	kurtosis	63	63
lead_4	gm_asymetry2	61	61
lead_5	kurtosis	59	59
lead_10	range	57	57
lead_10	mid_range	57	57
lead_10	maximum	56	56
lead_10	skewness	52	52
lead_9	maximum	41	41
lead_7	minimum	40	40
lead_6	kurtosis	40	40
lead_3	kurtosis	37	37
lead_8	kurtosis	34	34
lead_4	skewness	33	33
lead_11	skewness	30	30
lead_5	range	29	29
lead_5	mid_range	29	29
lead_0	maximum	27	27
lead_8	minimum	26	26
lead_8	skewness	26	26
lead_0	range	23	23
lead_0	mid_range	23	23
lead_10	standard_deviation	22	22
lead_4	interpercentile_range_40_60	21	21
lead_1	kurtosis	20	20
lead_3	gm_asymetry2	17	17
lead_11	maximum	16	16
lead_9	minimum	16	16
lead_7	kurtosis	15	15
lead_1	gm_asymetry	15	15

		DNN (P=0.036)	GS (P=0.036)
ECG	Feature	$\tau=+0.7$ (en %)	$\tau=+0.7$ (en %)
lead_9	interpercentile_range_0_50	284	284
lead_10	interpercentile_range_0_50	198	198
lead_0	interpercentile_range_10_90	189	189
lead_0	interpercentile_range_50_100	189	189
lead_11	interpercentile_range_10_90	156	156
lead_11	interpercentile_range_50_100	156	156
lead_8	interpercentile_range_0_50	142	142
lead_7	interpercentile_range_0_50	119	119
lead_0	entropy	95	95
lead_9	standard_deviation	88	88
lead_10	entropy	79	79
lead_9	interpercentile_range_10_90	79	79
lead_9	interpercentile_range_50_100	79	79
lead_9	entropy	78	78
lead_8	interpercentile_range_10_90	70	70
lead_8	interpercentile_range_50_100	70	70
lead_3	interpercentile_range_10_90	64	64
lead_3	interpercentile_range_50_100	64	64
lead_4	interpercentile_range_0_50	55	55
lead_0	auto_correlation	53	53
lead_8	auto_correlation	53	53
lead_11	auto_correlation	52	52
lead_5	entropy	51	51
lead_11	entropy	51	51
lead_7	auto_correlation	50	50
lead_9	auto_correlation	48	48
lead_2	interpercentile_range_0_50	47	47
lead_10	auto_correlation	44	44
lead_1	interpercentile_range_0_50	43	43
lead_8	entropy	39	39
lead_6	auto_correlation	39	39
lead_8	standard_deviation	37	37
lead_10	interpercentile_range_10_90	37	37
lead_10	interpercentile_range_50_100	37	37
lead_6	interpercentile_range_10_90	36	36
lead_6	interpercentile_range_50_100	36	36
lead_5	auto_correlation	32	32
lead_5	gm_asymetry	31	31
lead_9	energy	28	28
lead_2	auto_correlation	28	28
lead_3	interpercentile_range_25_75	26	26
lead_3	auto_correlation	25	25
lead_5	interpercentile_range_10_90	25	25

MALADIE ST

		DNN (P=0.047)	GS (0.046)
ECG	Feature	$\tau = -0.7$ (en %)	$\tau = -0.7$ (en %)
lead_6	kurtosis	46	42
lead_10	kurtosis	38	43
lead_3	entropy	31	32
lead_2	minimum	29	30
lead_9	kurtosis	27	35
lead_1	kurtosis	26	27
lead_11	kurtosis	26	32
lead_5	range	24	14
lead_5	mid_range	24	14
lead_5	kurtosis	22	24
lead_8	kurtosis	20	26
lead_1	range	20	14
lead_1	mid_range	20	14
lead_5	gm_asymetry	20	18
lead_4	entropy	18	20
lead_0	kurtosis	18	21
lead_11	range	18	18
lead_11	mid_range	18	18
lead_2	kurtosis	16	12
lead_4	kurtosis	15	12
lead_0	range	15	7
lead_0	mid_range	15	7

		DNN (P=0.047)	GS (0.046)
ECG	Feature	$\tau = +0.7$ (en %)	$\tau = +0.7$ (en %)
lead_11	frequencies_via_fft	247	247
lead_0	frequencies_via_fft	172	171
lead_10	frequencies_via_fft	169	168
lead_9	frequencies_via_fft	134	131
lead_8	kurtosis	112	114
lead_8	frequencies_via_fft	92	87
lead_6	frequencies_via_fft	81	76
lead_9	auto_correlation	74	66
lead_5	entropy	72	76
lead_1	frequencies_via_fft	72	66
lead_10	entropy	71	77
lead_5	auto_correlation	69	65
lead_1	entropy	67	69
lead_2	frequencies_via_fft	66	61
lead_3	frequencies_via_fft	64	59
lead_5	frequencies_via_fft	64	60
lead_0	auto_correlation	61	57
lead_8	auto_correlation	60	60
lead_4	frequencies_via_fft	57	53
lead_6	entropy	56	45
lead_1	auto_correlation	56	52
lead_11	entropy	54	63
lead_6	auto_correlation	54	51
lead_10	auto_correlation	54	56
lead_4	entropy	52	49
lead_4	auto_correlation	48	43
lead_7	auto_correlation	44	41
lead_11	auto_correlation	44	44
lead_8	skewness	39	36
lead_3	maximum	36	36
lead_0	entropy	35	38
lead_9	entropy	28	38
lead_3	entropy	25	22
lead_2	auto_correlation	24	21
lead_8	entropy	23	27
lead_2	entropy	21	14
lead_7	interpercentile_range_50_100	21	30
lead_7	interpercentile_range_10_90	21	30
lead_6	interpercentile_range_50_100	19	25
lead_10	interpercentile_range_25_75	19	21
lead_6	interpercentile_range_10_90	19	25
lead_7	frequencies_via_fft	18	170
lead_3	auto_correlation	17	14

MALADIE AF

		DNN (P=0.012)	GS (P=0.016)
ECG	Feature	$\tau=-0.7$ (en %)	$\tau=-0.7$ (en %)
lead_9	skewness	185	131
lead_0	standard_deviation	185	114
lead_5	standard_deviation	185	72
lead_3	auto_correlation	167	102
lead_0	auto_correlation	158	96
lead_5	auto_correlation	155	117
lead_6	auto_correlation	153	122
lead_1	auto_correlation	152	93
lead_3	gm_asymetry2	149	103
lead_11	auto_correlation	144	89
lead_8	auto_correlation	139	84
lead_4	auto_correlation	139	83
lead_2	auto_correlation	134	78
lead_0	range	129	109
lead_0	mid_range	129	109
lead_10	auto_correlation	128	92
lead_9	auto_correlation	120	69
lead_7	auto_correlation	100	75
lead_1	gm_asymetry	100	114
lead_5	range	95	98
lead_5	mid_range	95	98
lead_1	gm_asymetry2	86	75
lead_2	interpercentile_range_0_50	85	5
lead_3	range	82	59
lead_3	mid_range	82	59
lead_0	interquartile_range	73	52
lead_3	standard_deviation	71	49
lead_3	skewness	70	46
lead_3	gm_asymetry	67	52
lead_8	skewness	66	43
lead_0	median_absolute_value	65	45
lead_5	interquartile_range	61	42
lead_6	interpercentile_range_0_50	61	45
lead_8	entropy	56	24
lead_4	interpercentile_range_0_50	56	26
lead_8	interpercentile_range_10_90	55	26
lead_8	interpercentile_range_50_100	55	26
lead_4	gm_asymetry	52	67
lead_6	frequencies_via_fft	51	35
lead_0	maximum	50	48
lead_5	median_absolute_value	48	31
lead_8	interpercentile_range_25_75	47	16
lead_10	standard_deviation	44	18

		DNN (P=0.012)	GS (P=0.016)
ECG	Feature	$\tau=+0.7$ (en %)	$\tau=+0.7$ (en %)
lead_3	interpercentile_range_0_50	176	133
lead_9	interpercentile_range_0_50	154	112
lead_2	gm_asymetry	108	62
lead_8	interpercentile_range_0_50	102	84
lead_7	entropy	83	73
lead_4	gm_asymetry	78	37
lead_9	trimean	73	48
lead_9	midhinge	71	47
lead_9	average	70	46
lead_9	signal_magnitude_area	70	47
lead_7	interpercentile_range_0_50	70	46
lead_9	median	70	47
lead_7	minimum	69	37
lead_7	interpercentile_range_40_60	61	53
lead_9	interpercentile_range_25_75	57	34
lead_5	gm_asymetry2	54	49
lead_6	skewness	53	39
lead_4	skewness	49	18
lead_7	signal_magnitude_area	49	30
lead_1	interpercentile_range_0_50	44	32
lead_10	interpercentile_range_0_50	44	27
lead_1	gm_asymetry	41	8
lead_2	skewness	37	17
lead_7	interpercentile_range_25_75	33	463
lead_4	interpercentile_range_0_50	26	12
lead_7	interpercentile_range_10_90	25	310
lead_7	interpercentile_range_50_100	25	310
lead_5	gm_asymetry	25	39
lead_5	minimum	22	32
lead_9	energy	21	12
lead_8	skewness	20	4
lead_1	skewness	19	-7
lead_7	average	19	10
lead_9	median_absolute_value	18	9
lead_9	interquartile_range	18	8
lead_2	gm_asymetry2	17	5

MALADIE SB

		DNN (P=0.021)	GS (P=0.018)
ECG	Feature	$\tau=-0.7$ (en %)	$\tau=-0.7$ (en %)
lead_0	frequencies_via_fft	280	257
lead_9	frequencies_via_fft	270	274
lead_5	frequencies_via_fft	245	232
lead_11	frequencies_via_fft	239	228
lead_10	frequencies_via_fft	185	190
lead_1	frequencies_via_fft	167	168
lead_2	frequencies_via_fft	154	175
lead_8	auto_correlation	135	125
lead_3	frequencies_via_fft	116	73
lead_4	frequencies_via_fft	101	81
lead_7	auto_correlation	99	98
lead_6	frequencies_via_fft	99	132
all	asynchrony_l1	98	123
lead_3	gm_asymetry	62	72
lead_3	skewness	61	49
lead_6	auto_correlation	59	74
lead_6	skewness	55	71
lead_6	median	53	20
lead_6	trimean	49	19
lead_0	standard_deviation	48	72
lead_6	average	47	17
lead_6	signal_magnitude_area	47	17
lead_2	auto_correlation	45	40
lead_11	standard_deviation	45	60
lead_6	midhinge	44	17
lead_8	frequencies_via_fft	41	109
lead_4	auto_correlation	39	63
lead_3	gm_asymetry2	37	34
lead_1	interpercentile_range_25_75	37	42
lead_2	interpercentile_range_0_50	32	30
lead_1	auto_correlation	32	47
lead_3	maximum	28	35
lead_5	skewness	28	38
lead_3	kurtosis	28	47
lead_0	kurtosis	26	55
lead_3	standard_deviation	24	25
lead_6	interpercentile_range_25_75	24	30
lead_3	range	24	33
lead_3	mid_range	24	33
lead_0	range	24	56
lead_0	mid_range	24	56
lead_1	interpercentile_range_40_60	23	25
lead_9	kurtosis	23	37

		DNN (P=0.021)	GS (P=0.018)
ECG	Feature	$\tau=+0.7$ (en %)	$\tau=+0.7$ (en %)
lead_6	interpercentile_range_25_75	113	-24
lead_7	skewness	88	113
lead_11	kurtosis	69	79
all	asynchrony_l2	67	86
lead_3	interpercentile_range_0_50	64	72
lead_6	interpercentile_range_10_90	59	-18
lead_6	interpercentile_range_50_10	59	-18
lead_9	kurtosis	53	69
lead_0	auto_correlation	47	64
lead_8	skewness	43	78
lead_11	skewness	41	48
lead_9	auto_correlation	37	40
lead_6	interpercentile_range_40_60	35	-20
lead_1	skewness	32	45
lead_5	auto_correlation	31	48
lead_6	kurtosis	27	37
lead_3	auto_correlation	27	40
lead_9	skewness	27	44
lead_6	auto_correlation	24	13
lead_2	gm_asymetry	22	15
lead_11	auto_correlation	20	37
lead_10	skewness	19	26
lead_2	minimum	19	28
lead_1	auto_correlation	19	35
lead_6	entropy	18	-24
lead_7	frequencies_via_fft	17	14
lead_7	frequencies_via_fft	17	14
lead_7	frequencies_via_fft	17	14
lead_7	kurtosis	16	25
lead_3	interpercentile_range_40_60	16	25
lead_1	minimum	15	23