

Détection d'anesthésie générale par analyse de l'électrocardiogramme en chirurgie

Clément Fauchereau* Fanny Carimalo** Axelle Merienne** Marc Laffon** Emmanuel Godat** Jean-Christophe Aude*

*Université Paris-Saclay, CEA, CNRS
Institute for Integrative Biology of the Cell (I2BC) Gif-sur-Yvette
**Département d'Anesthésie-Réanimation, Hôpital Bretonneau
Centre Hospitalier Régional Universitaire de Tours, Tours

Résumé. Les outils de suivi de la profondeur d'anesthésie sont une aide essentielle pour les anesthésistes. Ils permettent de prévenir les sous-dosages et surdosages d'agents anesthésiques et les complications qui leur sont associées. Les dispositifs actuellement utilisés en salle d'opération reposent sur l'analyse de l'électroencéphalogramme. Dans de nombreux cas ces dispositifs ne peuvent pas être utilisés que ce soit lors de chirurgies céphaliques, en cas de brûlures ou à cause de la difficulté d'installer un électroencéphalogramme. Dans cet article, nous proposons d'exploiter l'électrocardiogramme, grâce à des techniques d'apprentissage, pour détecter l'état "anesthésié" de patients lors de chirurgies. Pour cela nous avons constitué une base de données annotées d'électrocardiogrammes en chirurgie. Nous montrons l'importance des indices de fragmentations pour cette tâche et le potentiel d'un indice de profondeur d'anesthésie reposant sur l'électrocardiogramme.

1 Introduction

Lors des interventions chirurgicales, le monitoring de la profondeur de l'anesthésie est un facteur déterminant pour : *i*) différencier une narcose excessive, adéquate ou éventuellement insuffisante ; *ii*) guider l'administration et l'ajustement des agents anesthésiques volatils ou intraveineux. Un dosage correct des agents hypnotiques permet de limiter la morbidité hémodynamique et d'améliorer la qualité et la rapidité du réveil. Une mauvaise appréciation de la profondeur d'anesthésie peut s'avérer dangereuse lors d'une intervention chirurgicale (en particulier chez les enfants, les patients fragiles ou les personnes âgées). Les principaux risques associés au sous-dosage de l'anesthésie sont la conscience peropératoire et les traumatismes psychologiques tels que le syndrome de stress post-traumatique. Inversement, un surdosage d'anesthésiques produit un état de narcose plus profond dont les risques associés sont principalement l'hypoperfusion des tissus, qui peut conduire à une défaillance des organes et à d'éventuelles lésions cérébrales, dysfonctionnements cognitifs ou délirium postopératoire. En dernier lieu cela peut représenter une menace vitale à des concentrations très élevées (Chan et al., 2017). En outre, la prévalence croissante de maladies chroniques telles que le diabète,

Détection d'anesthésie générale par analyse de l'électrocardiogramme en chirurgie

les maladies cardiovasculaires, le cancer, les accidents vasculaires cérébraux et les maladies rénales devrait entraîner une augmentation du nombre d'interventions chirurgicales. Dans ce contexte, il est essentiel de mesurer précisément la profondeur d'anesthésie, en particulier chez les patients à risque, afin de réduire les complications lors des interventions chirurgicales.

L'évaluation clinique de la profondeur de l'anesthésie (classification de Guedel) étant souvent insuffisante, des dispositifs de surveillance de l'activité cérébrale sont nécessaires pour accompagner le praticien. Les moniteurs basés sur l'électroencéphalogramme (EEG) sont aujourd'hui *de facto* la norme. Ils transforment le signal EEG brut en indices de profondeur d'anesthésie en utilisant des algorithmes propriétaires (Johansen, 2006) tels que l'indice bispectral (BIS)(Covidien), la E-entropie (GE Healthcare), l'indice Narcotrend (Monitor Technik) ou le *Patient State Index* (PSI)(Masimo). Dans le cas particulier du BIS, dont l'utilisation est la plus répandue et le seul à avoir obtenu un agrément de la FDA américaine, plusieurs études ont montré qu'il contribue à un meilleur dosage des agents anesthésiques, à une réduction des cas de narcose courte et à une amélioration de la phase de réveil (Gan et al., 1997). On notera toutefois que le monitoring à l'aide d'EEG ne peut pas être utilisé dans certaines situations cliniques, par exemple lors de chirurgies céphaliques ou en présence de lésions cutanées (brûlures) sur le front.

D'autre part, le suivi systématique de l'activité cardiaque des patients, par la pose d'un électrocardiogramme (ECG), est obligatoire pour des raisons légales et révèle certains effets des agents hypnotiques (Kanaya et al., 2003). Il convient d'observer que des mesures dérivées de l'ECG, telles que le signal RR (intervalle de temps entre deux ondes R) ou la variabilité de la fréquence cardiaque (VFC), ont depuis longtemps été décrites dans la littérature pour détecter les différentes phases de la conscience (Latson et al., 1992). Ainsi la VFC, qui est liée à la régulation du système autonome, est particulièrement affectée par l'anesthésie générale et varie en fonction des agents anesthésiques utilisés (Huhle et al., 2012). Par ailleurs, il a été montré que l'intervalle QT est aussi altéré par certains agents anesthésiques, notamment le propofol (Oji et al., 2013).

Pour autant l'estimation de l'état (éveillé vs anesthésié) ou de la profondeur d'anesthésie (indices qualitatifs ou quantitatifs) reste un défi pour les chercheurs concernés. La difficulté à estimer précisément la profondeur d'anesthésie est liée à l'existence de nombreux cofacteurs tel que l'usage de différentes drogues (*e.g.* analgésiques), l'existence de certaines pathologies secondaires (*e.g.* cardiovasculaire), la variabilité inter patients. Par conséquent les signaux électrophysiologiques à analyser sont non-stationnaires et présentent de nombreux artefacts, en raison aussi de l'environnement électro-magnétique présent dans les blocs opératoires. Néanmoins, des travaux utilisant les signaux EEG avec différentes approches méthodologiques sont décrits dans la littérature, par exemple : l'analyse dans les domaines fréquentiels et/ou temporels (Sigl et Chamoun, 1994); les réseaux de neurones artificiels (Ortolani et al., 2002); les mesures d'entropie ou de complexité (Castro et al., 2007); les méthodes multi-échelles (Shou-Zen Fan et al., 2012); l'analyses de Poincaré (Hayase et al., 2020); les réseaux de convolution et l'apprentissage profond (Afshar et al., 2021).

L'utilisation de l'ECG pour quantifier la profondeur d'anesthésie est également présent dans la littérature. Dans ces travaux, les auteurs utilisent des techniques de traitement du signal afin de filtrer et débruiter le signal ECG puis d'identifier le complexe d'ondes PQRST (Martinez et al., 2004). Cette étape de délimitation du signal brut permet de calculer différents descripteurs, tel que la VFC. Plusieurs méthodes utilisant des mesures d'entropie ou de com-

plexité, potentiellement à différentes échelles, pour une revue consulter Bravi et al. (2011), permettent de différencier les états d'anesthésie : par des analyses fréquentiels (Yoo et Yi, 2004) ou bien conjointement avec des algorithmes d'apprentissage automatique ou statistique. Ainsi dans Polk S. *et al.* (Polk et al., 2019) les auteurs ont utilisé un ensemble des descripteurs calculés à partir de l'analyse du spectre fréquentiel et l'analyse temporelle. Ils ont ensuite évalué les performances de trois algorithmes d'apprentissage (*K-nearest-neighbor*, la régression logistique et les arbres de décision) afin de prédire l'état de conscience : éveillé ou sous l'effet d'agents anesthésiques. Dans leur article, les auteurs Zhan et al. (2021) ont utilisé une approche similaire concernant la sélection des descripteurs : puissances haute fréquence (HF) et basses fréquences (LF); ratio de puissance HF/LF; complexité du signal par la méthode *sample entropy* (Richman et al., 2004). Ils ont ensuite mis en œuvre un réseau de neurones afin de discriminer trois états distincts de l'anesthésie : l'induction (transition entre l'état éveillé et sous anesthésie); la maintenance (état sous anesthésie stable); la récupération (transition entre maintenance et éveil).

Dans cet article nous décrivons une méthode de prédiction de l'état d'anesthésie (éveillé ou sous anesthésie) de patients ayant subi une intervention chirurgicale à partir d'extraits de 5 minutes d'un ECG. Ces descripteurs sont complétés par les informations physiologiques des patients (poids, taille, âge) dont il a été montré qu'ils influent le rythme cardiaque (Quer et al., 2020). Les différents algorithmes d'apprentissage utilisés ont été entraînés sur une cohorte constituée au CHRU de Tours (*c.f.* Matériel et méthodes). Afin d'évaluer la capacité de généralisation de ces algorithmes nous avons utilisé des bases de données de référence dans le domaine, Fantasia pour l'état éveillé (Iyengar et al., 1996) et VitalDB pour l'état anesthésié (Lee et al., 2022).

2 Matériel et méthodes

2.1 Description de la cohorte

L'étude a été validée par la CNIL¹ en amont des inclusions. La cohorte est constituée de patients opérés sous anesthésie générale induite par le propofol au CHU de Tours entre novembre 2021 et novembre 2022. Ont été exclus de l'étude les patients âgés de moins de 18 ans, n'ayant pas donné leur consentement, sous mesure de protection juridique à type de tutelle, présentant une fibrillation auriculaire permanente ou un stimulateur cardiaque, opérés en urgence ou en chirurgie céphalique. De plus, ont été exclus les patients dont la durée des enregistrements n'est pas suffisante pour cette étude : moins de 10 minutes en pré-anesthésie et moins de 35 minutes après l'induction. La description statistique des patients inclus dans cette cohorte est présentée dans le tableau 1.

2.2 Acquisition du signal

Les enregistrements de signaux physiologiques ont été effectués au moyen du logiciel VitalRecorder (Lee et Jung, 2018) version 1.10.3.0. Ce logiciel nous a permis d'extraire les données d'un moniteur portable CARESCAPE™ ONE et à un moniteur fixe CARESCAPE™ B650. Pour chaque patient, on a procédé à l'acquisition de deux enregistrements distincts. Le

1. Protocole enregistré au n° 2020_110 du registre des traitements informatiques du C.H.R.U. de Tours

Participants	48
Femme / Homme	18/30
Age	50.60 (± 16.19)
Taille (cm)	169.04 (± 8.85)
Poids (kg)	76.79 (± 13.05)

TAB. 1 – *Description de notre cohorte.*

premier a été effectué préalablement à l'anesthésie générale en salle préopératoire. Il constitue l'état basal, le patient étant éveillé en position allongée. Le second enregistrement est réalisé pendant toute la durée de l'opération chirurgicale. Il débute lorsque le patient est installé sur la table d'opération et se termine juste avant la sortie du bloc opératoire. L'horodatage du début de l'induction est ajouté aux annotations au cours de l'intervention, *via* le logiciel VitalRecorder. Pour la suite de cette étude on a conservé la période pendant laquelle le patient est sous l'effet d'agents anesthésiques. Par ailleurs, les 5 minutes suivant l'induction sont exclues, le rapport signal sur bruit est beaucoup trop faible et ne saurait être exploitable ou bien le signal est absent. Dans la majorité des cas ceci est la conséquence de la période d'installation du patient pour la chirurgie.

Afin de constituer un jeu de données d'entraînement équilibré, pour chaque patient nous avons choisi de ne conserver que les 30 minutes de signal les moins bruitées. Les 40 premiers patients ont été utilisés pour évaluer la performance des modèles par validation croisée et 8 enregistrements acquis ultérieurement ont servi d'ensemble de test.

2.3 Traitement du signal ECG

Le signal ECG est débruité avec un filtre passe-bande de type Butterworth dont les fréquences de coupure sont entre 0.5Hz et 100Hz. Chaque enregistrement est ensuite segmenté en fenêtres d'une durée de 5 minutes avec 50% de recouvrement. Les segments contenant des pertes de signal sont exclus. Pour chacun de ces segments on procède à l'identification des ondes PQRST caractéristiques de l'ECG par une méthode de délinéation. Nous avons implémenté l'algorithme décrit par les auteurs Martinez et al. (2004). L'implémentation a été réalisée avec Python 3.10.12 et NumPy 1.24.4. Cet algorithme calcule une transformée par ondelettes de l'ECG, puis détecte les pics par un processus seuillage. L'intérêt de cette méthode réside dans l'utilisation de plusieurs échelles de la transformée en ondelettes qui procure une plus importante robustesse au bruit.

2.4 Descripteurs

Le signal après délinéation est exploité pour calculer un ensemble de descripteurs à partir des segments d'ECG selon des méthodologies différentes : VFC (Electrophysiology, 1996); décomposition fréquentielle de la VFC (Electrophysiology, 1996); Indices de fragmentation (Costa et al., 2017); Indices de complexité; Durées inter-ondes. Les caractéristiques suivantes des patients sont également ajoutées aux descripteurs : l'âge, la taille et le poids. L'inclusion

de ces caractéristiques est importante car certaines influent sur les mesures de VFC. La table 2 contient la liste exhaustive des descripteurs considérés pour le reste de cette étude.

	Descripteurs
VFC Statistique (Electrophysiology, 1996)	SDNN, RMSSD, NN50, pNN50
VFC Fréquentielles (Electrophysiology, 1996)	Très basses fréquences (0,0033-0,04Hz) Basses fréquences (0,04-0,15Hz) Hautes fréquences (0,15-0,4Hz) Balance sympatho-vagal Puissance totale
Indices de fragmentations (Costa et al., 2017)	IALS, PIP, PSS, PAS
Indices de complexité	Permutation Entropy (Bandt et Pompe, 2002) Approximate Entropy (Pincus, 1991) Sample Entropy (Richman et al., 2004) Dispersion Entropy (Rostaghi et Azami, 2016)
Segments (moyenne et écart-type)	QT, PR, ST, QRS
Caractéristiques des patients	Age, Taille, Poids, IMC

TAB. 2 – caractéristiques extraites.

2.5 Modèles d'apprentissage automatique

Lors de cette étude nous avons évalué 6 modèles d'apprentissage différents : les arbres de décision, la régression logistique, les plus proches voisins, les forêts d'arbres décisionnels, les machines à vecteurs de support (SVM) et la méthode de *gradient boosting* (XGBoost). Pour chaque modèle, une validation croisée imbriquée a permis d'en mesurer les performances. À chaque étape de la validation croisée, le jeu de données d'entraînement est utilisé pour trouver les meilleurs hyperparamètres par une recherche aléatoire. Dans cette recherche aléatoire, la performance est évaluée par validation croisée. La séparation du jeu d'entraînement est effectuée sur les patients par la méthode *Leave One Out* dans laquelle un patient sert de données de test. Nous avons utilisé l'implémentation du package XGBoost version 1.7.4 pour le modèle de *gradient boosting*. Pour tous les autres modèles, nous avons utilisé les implémentations de Scikit-Learn version 1.3.0.

3 Résultats

3.1 Sélection des descripteurs

La sélection des variables pour la construction des modèles a été réalisée sur la base d'une analyse de variance et d'une analyse de corrélations. On calcule les coefficients de corrélation

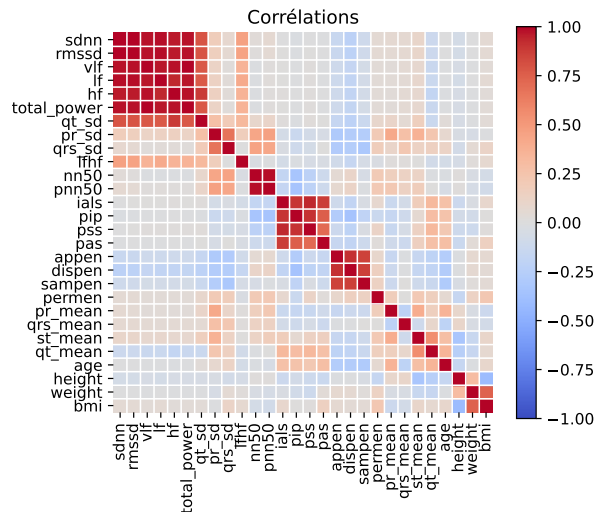


FIG. 1 – Représentation en heatmap des coefficients de corrélations de Pearson entre tous les descripteurs.

de Pearson entre toutes les paires de descripteurs, représenté sur figure 1, ce qui met en évidence des groupes de descripteurs fortement corrélés. Ainsi on observe que les indices de fragmentation sont très corrélés entre eux, de même que les indices de complexité (à l’exception de la mesure de *permutation entropy*) et de la plupart des indices statistiques et fréquentiels. Pour chaque groupe de descripteurs corrélés, une analyse de variance a permis d’identifier le descripteur le plus pertinent. Après avoir calculé le score de Fisher (F-score) de tous les descripteurs, les 10 descripteurs avec les F-score les plus élevés ont été sélectionnés. Ces descripteurs sont : le PIP, la *dispersion entropy*, la moyenne des QT, le NN50, le poids, la taille, la moyenne des PR, la moyenne des ST, l’âge et la *permutation entropy*. Cette analyse révèle que les indices de fragmentation et les indices de complexité ont les pouvoirs explicatifs les plus importants.

3.2 Mesures de performances des modèles

Algorithme	Exactitude	F1	Précision	Sensibilité	Spécificité
Decision Tree	0.87 (0.82-0.91)	0.90 (0.86-0.94)	0.92 (0.88-0.96)	0.91 (0.85-0.96)	0.75 (0.64-0.87)
Logistic Regression	0.85 (0.78-0.91)	0.87 (0.81-0.93)	0.92 (0.86-0.97)	0.88 (0.81-0.95)	0.80 (0.69-0.92)
Nearest Neighbors	0.82 (0.77-0.88)	0.86 (0.81-0.91)	0.92 (0.87-0.96)	0.85 (0.79-0.92)	0.74 (0.61-0.86)
Random Forest	0.88 (0.83-0.92)	0.91 (0.88-0.94)	0.90 (0.85-0.94)	0.95 (0.92-0.99)	0.70 (0.58-0.83)
SVM	0.86 (0.81-0.91)	0.89 (0.83-0.94)	0.92 (0.88-0.96)	0.91 (0.84-0.97)	0.71 (0.58-0.84)
XGBoost	0.88 (0.84-0.91)	0.91 (0.88-0.94)	0.91 (0.87-0.95)	0.93 (0.89-0.97)	0.75 (0.63-0.87)

TAB. 3 – Scores de validation croisée (avec intervalles de confiance à 95%)

La validation croisée réalisée lors de l'étape de sélection de variables (voir les résultats sur le tableau 3) a permis de sélectionner les meilleurs hyperparamètres pour chaque modèle. Ces modèles, avec les paramètres optimisés, ont été à nouveau entraînés avec la base d'entraînement avant d'être évalués avec la base de test. Pour cela nous avons utilisé les mesures de performances suivantes : l'exactitude, la précision, la sensibilité, la spécificité et le score F1. Les résultats sont indiqués dans le tableau 4.

Algorithme	Exactitude	F1	Précision	Sensibilité	Spécificité
Decision Tree	0.7600	0.8142	0.8679	0.7667	0.7455
Logistic Regression	0.8057	0.8482	0.9135	0.7917	0.8364
Nearest Neighbors	0.8514	0.8968	0.8561	0.9417	0.6545
Random Forest	0.8286	0.8760	0.8689	0.8833	0.7091
SVM	0.8629	0.9048	0.8636	0.9500	0.6727
XGBoost	0.8629	0.8983	0.9138	0.8833	0.8182

TAB. 4 – Performances des modèles sur les données de test.

Pour le modèle SVM nous avons aussi calculé les valeurs de SHAP (*SHapley Additive ex-Planations*), qui corroborent les résultats obtenus avec l'analyse de variance (voir figure 2). En particulier, on peut observer que les indices de fragmentations et de complexité ont le plus d'importance pour les prédictions. Il est intéressant de noter que les valeurs SHAP montrent que l'âge est le second descripteur le plus important, contrairement aux résultats de l'analyse de variance. En effet, l'importance de la contribution de l'âge, du poids et de la taille apparaît quand ils sont des cofacteurs d'autres descripteurs, notamment la VFC. Par exemple, il est établi que les indices de fragmentation augmentent avec l'âge chez les patients sains (Costa et al., 2017). Les valeurs SHAP permettent aussi d'analyser le niveau de contribution des descripteurs pour les prédictions d'un modèle. Nous avons réalisé cette analyse dont les résultats sont présentés sur la figure 2. On peut observer que la prédiction de l'état "anesthésié" est associée à des indices de fragmentations élevés et une complexité faible. De plus, l'augmentation de la durée de l'intervalle QT est cohérente avec l'effet clinique du propofol tel que décrit par Scalese et al. (2016).

3.3 Évaluation de la capacité de généralisation des modèles

Afin d'évaluer la capacité de généralisation des modèles nous avons constitué un jeu de test à partir de bases de données publiquement accessibles et correspondant aux états "éveillé" et "anesthésié". Pour les premiers il s'agit de la base Fantasia de Iyengar et al. (1996) qui regroupe les données ECG de 40 sujets jeunes et âgés en position allongée. Les données de patients anesthésiés proviennent de la base VitalDB (Lee et al., 2022), qui regroupe les enregistrements de 6,388 cas. Nous avons filtré cette base afin de ne conserver que les enregistrements similaires aux critères d'inclusion de cet article (anesthésie générale induite avec du propofol), puis nous avons aléatoirement échantillonné 294 patients afin d'obtenir un nombre de segments d'ECG semblable à celui de cette étude. La base Fantasia ne contenant pas les informations de poids et de taille des sujets, seul les modèles pouvant traiter les données manquantes ont été évalués, soit : les arbres de décision et la méthode du *gradient boosting* (XGBoost). Les mesures de

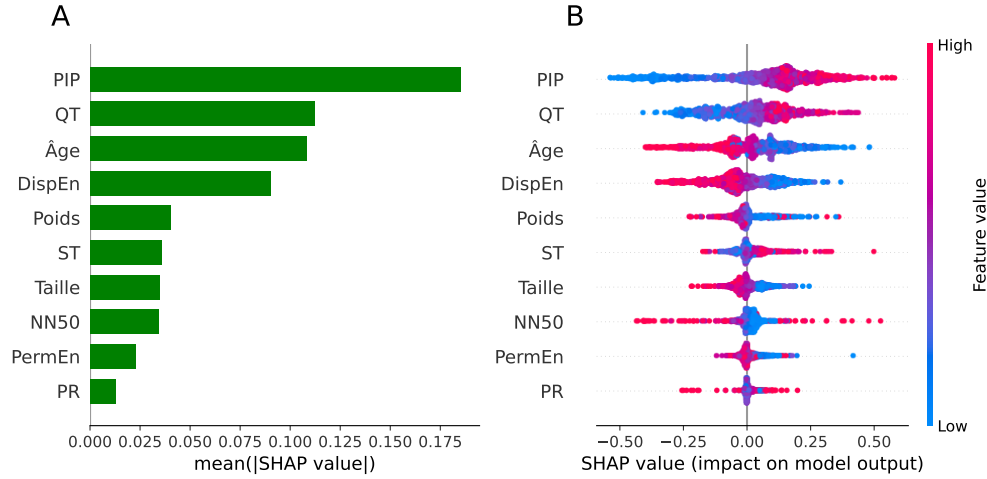


FIG. 2 – **A)** Importance des descripteurs dans le modèle SVM. **B)** Représentation de l'influence des descripteurs dans le modèle SVM. Chaque point correspond à un descripteur (en abscisse) et une valeur de ce descripteur (couleur). La couleur bleue indique que le descripteur a une valeur basse, le rouge indique une valeur élevée. La valeur SHAP (en ordonnée) correspond à l'influence sur la sortie du modèle (négative pour l'état éveillé, positive pour l'état anesthésié).

performances calculées sont identiques à celles précédemment utilisées (voir section 3.2) et les résultats présentés dans le tableau 5. On peut observer une baisse significative des performances globales pour les deux modèles, dans une moindre mesure concernant la sensibilité pour la méthode XGBoost (l'indice diminuant de 0,8833 à 0.8200).

Modèle	Exactitude	F1	Sensibilité	Précision	Spécificité
Decision Tree	0.6123	0.6499	0.7435	0.5772	0.4892
XGBoost	0.6617	0.7011	0.8200	0.6124	0.5134

TAB. 5 – Performances des capacités de généralisation des modèles.

4 Discussion

Dans cet article nous avons montré que l'ECG permet de différencier les états éveillé vs sous anesthésie générale. L'utilisation de descripteurs tels que les indices de fragmentation et de complexité du signal associés aux paramètres physiologiques des patients a permis d'obtenir des performances ouvrant des perspectives pour la définition d'indices quantitatifs de la profondeur d'anesthésie.

Le signal ECG est aussi utilisé pour le monitoring de la nociception. L'hypothèse clinique sous-jacente étant que la nociception altère la balance sympathique/parasympathique du système nerveux autonome et indirectement la VFC. On peut donc s'interroger si nos modèles discriminent l'effet des analgésiques plutôt que celui des anesthésiques. L'indice Analgésie/Nociception (ANI) de la société Mdoloris (Lille, France) basé sur la VFC procède par un filtrage du signal RR normalisé entre 0,15Hz et 0,4Hz pour ne conserver que les variations parasympathiques. *A contrario* nos modèles sont construits sans ce filtre. Cependant il serait intéressant de comparer ces deux modèles, notamment en utilisant les annotations sur les ajustements des doses d'agents analgésiques/anesthésiques présentes dans notre cohorte. Cette question de causalité entre les différentes drogues utilisées et l'activité cardiaque lors de l'anesthésie est un challenge pour la recherche concernée. D'autres facteurs peuvent aussi altérer le signal ECG. Ainsi l'enregistrement sous anesthésie, lors de la chirurgie, est soumis à plus de sources de bruit que celui en salle préopératoire. Il est sujet à plus d'interférences (*e.g.* bistouri électrique, arrachement d'électrodes). Étudier l'influence de ces interférences permettrait de vérifier la robustesse de nos modèles. Enfin, on notera que notre cohorte inclut uniquement des anesthésies générales induites par le propofol qui agit sur les récepteurs GABA. D'autres travaux seraient nécessaires pour vérifier que ces modèles restent valides avec d'autres agents anesthésiques comme la kétamine qui agit sur des récepteurs différents (NMDA).

Le lien entre la VFC et l'anesthésie a été décrits dans plusieurs études (Tarvainen et al., 2012; Huang et al., 1997). Cependant ces études se concentrent sur les VFC statistiques et fréquentielles. Nous avons montré que les indices de fragmentation sont plus pertinents pour caractériser l'état sous anesthésie. Les indices de complexité ont été utilisés dans Zhan et al. (2021) pour discriminer entre la phase de maintien et les phases d'induction et de réveil. Nous confirmons ici l'intérêt de ces descripteurs. Le choix de garder seulement 10 descripteurs est arbitraire mais l'analyse de corrélation semble indiquer que c'est un nombre suffisant, ce que tend à confirmer l'analyse des valeurs SHAP. On observe en effet que les descripteurs au-delà du 10^e ont une influence marginale sur la prédiction du modèle SVM.

Les performances de classification sont meilleures que Polk et al. (2019) qui constitue à notre connaissance l'état de l'art. Néanmoins, cette étude ne porte pas sur des anesthésies avec le propofol. Il est dès lors difficile de comparer ces résultats. Zhan et al. (2021) obtient une exactitude plus élevée. Néanmoins la tâche n'est pas exactement la même. La classification porte entre la phase de maintenance et les phases d'induction et de réveil là où nous discriminons entre la phase de maintenance et l'éveil. De plus, des différences dans la méthodologie d'évaluation rendent la comparaison difficile.

L'utilisation des bases de données VitalDB et Fantasia nous a permis d'évaluer la capacité de généralisation de nos modèles. Bien qu'utiles pour de l'évaluation, ces bases ne sont pas adaptées pour l'entraînement. En effet, aucune de ces deux bases ne permet d'avoir des enregistrements éveillé et anesthésié provenant des mêmes patients. On prendrait alors le risque d'un surapprentissage, les modèles apprenant à différencier les patients plutôt que leurs états. Ces craintes sont renforcées par les fortes variations interindividuelles des descripteurs de VFC et le faible volume de données disponible.

La forte baisse de performance lors de l'évaluation en généralisation peut s'expliquer en partie par l'absence de la taille et du poids dans Fantasia ainsi que l'indication de l'âge par tranches. Cette explication nous semble plausible étant donné que la chute la plus forte est observée pour la spécificité alors que la baisse de sensibilité est bien moins prononcée. La

baisse de la sensibilité prouve néanmoins que ce n'est pas la cause exclusive de la baisse de performance. Une différence dans la distribution des données nous semble le plus probable. D'autant plus que la cohorte VitalDB et celle du CHRU de Tours ont chacune été constituées dans un seul hôpital dans des continents différents.

5 Conclusion

Dans cet article nous avons montré que l'exploitation de l'électrocardiogramme permet de discriminer les états éveillé et anesthésié avec plus de 85% d'exactitude. Nous avons également montré que les indices de fragmentation sont parmi les plus pertinents pour détecter l'état anesthésique.

Références

- Afshar, S., R. Boostani, et S. Sanei (2021). A Combinatorial Deep Learning Structure for Precise Depth of Anesthesia Estimation From EEG Signals. *IEEE journal of biomedical and health informatics* 25(9), 3408–3415.
- Bandt, C. et B. Pompe (2002). Permutation Entropy : A Natural Complexity Measure for Time Series. *Physical Review Letters* 88(17), 174102.
- Bravi, A., A. Longtin, et A. J. Seely (2011). Review and classification of variability analysis techniques with clinical applications. *BioMedical Engineering OnLine* 10(1), 90.
- Castro, A., N. Bressan, L. Antunes, et C. S. Nunes (2007). EEG entropy monitoring of depth of anaesthesia : Pharmacokinetic and dynamic modelling. In *2007 European Control Conference (ECC)*, Kos, Greece, pp. 2535–2540. IEEE.
- Chan, M. T., M. H. Chu, C. K. Lam, B. Jia, S. Tsang, et W. K. Wu (2017). Deep anesthesia : too much of a good thing ? *Canadian Journal of Anesthesia* 64, 574–580.
- Costa, M. D., R. B. Davis, et A. L. Goldberger (2017). Heart Rate Fragmentation : A New Approach to the Analysis of Cardiac Interbeat Interval Dynamics. *Frontiers in Physiology* 8.
- Electrophysiology, T. F. o. t. E. S. o. C. t. N. A. S. o. P. (1996). Heart Rate Variability. *Circulation* 93(5), 1043–1065.
- Gan, T. J., P. S. Glass, A. Windsor, F. Payne, C. Rosow, P. Sebel, et P. Manberg (1997). Bispectral Index Monitoring Allows Faster Emergence and Improved Recovery from Propofol, Alfentanil, and Nitrous Oxide Anesthesia. *Anesthesiology* 87(4), 808–15. Publication Title : Anesthesiology.
- Hayase, K., K. Hayashi, et T. Sawa (2020). Hierarchical Poincaré analysis for anaesthesia monitoring. *Journal of Clinical Monitoring and Computing* 34(6), 1321–1330.
- Huang, H. H., H. L. Chan, P. L. Lin, C. P. Wu, et C. H. Huang (1997). Time-frequency spectral analysis of heart rate variability during induction of general anaesthesia. *British Journal of Anaesthesia* 79(6), 754–758.

- en
- Huhle, R., M. Burghardt, S. Zaunseder, N. Wessel, T. Koch, H. Malberg, et A. R. Heller (2012). Effects of awareness and nociception on heart rate variability during general anaesthesia. *Physiological Measurement* 33(2), 207–217.
- Iyengar, N., C. K. Peng, R. Morin, A. L. Goldberger, et L. A. Lipsitz (1996). Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 271(4), R1078–R1084.
- en
- Johansen, J. W. (2006). Update on Bispectral Index monitoring. *Best Practice & Research Clinical Anaesthesiology* 20(1), 81–99.
- en
- Kanaya, N., N. Hirata, S. Kurosawa, M. Nakayama, et A. Namiki (2003). Differential Effects of Propofol and Sevoflurane on Heart Rate Variability. *Anesthesiology* 98(1), 34–40.
- eng
- Latson, T. W., S. M. McCarroll, M. A. Mirhej, V. A. Hyndman, C. W. Whitten, et J. M. Lipton (1992). Effects of three anesthetic induction techniques on heart rate variability. *Journal of Clinical Anesthesia* 4(4), 265–276.
- Lee, H.-C. et C.-W. Jung (2018). Vital Recorder—a free research tool for automatic recording of high-resolution time-synchronised physiological data from multiple anaesthesia devices. *Scientific Reports* 8(1), 1527.
- Lee, H.-C., Y. Park, S. B. Yoon, S. M. Yang, D. Park, et C.-W. Jung (2022). VitalDB, a high-fidelity multi-parameter vital signs database in surgical patients. *Scientific Data* 9(1), 279.
- Martinez, J., R. Almeida, S. Olmos, A. Rocha, et P. Laguna (2004). A wavelet-based ECG delineator : Evaluation on standard databases. *IEEE Transactions on Biomedical Engineering* 51(4), 570–581.
- en
- Oji, M., Y. Terao, T. Toyoda, T. Kuriyama, K. Miura, M. Fukusaki, et K. Sumikawa (2013). Differential effects of propofol and sevoflurane on QT interval during anesthetic induction. *Journal of Clinical Monitoring and Computing* 27(3), 243–248.
- en
- Ortolani, O., A. Conti, A. Di Filippo, C. Adembri, E. Moraldi, A. Evangelisti, M. Maggini, et S. Roberts (2002). EEG signal processing in anaesthesia. Use of a neural network technique for monitoring depth of anaesthesia. *British Journal of Anaesthesia* 88(5), 644–648.
- Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences* 88(6), 2297–2301.
- en
- Polk, S. L., K. Kashkooli, S. B. Nagaraj, S. Chamadia, J. M. Murphy, H. Sun, M. B. Westover, R. Barbieri, et O. Akeju (2019). Automatic Detection of General Anesthetic-States using ECG-Derived Autonomic Nervous System Features. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, Germany, pp. 2019–2022. IEEE.
- en

- Quer, G., P. Gouda, M. Galarnyk, E. J. Topol, et S. R. Steinhubl (2020). Inter- and intraindividual variability in daily resting heart rate and its associations with age, sex, sleep, BMI, and time of year : Retrospective, longitudinal cohort study of 92,457 adults. *PLOS ONE* 15(2), e0227709.
- Richman, J. S., D. E. Lake, et J. R. Moorman (2004). Sample Entropy. In *Methods in Enzymology*, Volume 384 of *Numerical Computer Methods, Part E*, pp. 172–184. Academic Press.
- Rostaghi, M. et H. Azami (2016). Dispersion Entropy : A Measure for Time-Series Analysis. *IEEE Signal Processing Letters* 23(5), 610–614.
- Scalese, M. J., H. R. Herring, R. C. Rathbun, G. H. Skrepnek, et T. L. Ripley (2016). Propofol-associated QTc prolongation. *Therapeutic Advances in Drug Safety* 7(3), 68–78.
- Shou-Zen Fan, Cheng-Wei Lu, Jiann-Shing Shieh, Maysam F. Abbod, Tzu-Yu Lin, Qin Wei, et Quan Liu (2012). Adaptive Computation of Multiscale Entropy and Its Application in EEG Signals for Monitoring Depth of Anesthesia During Surgery. *Entropy* 14(6), 978–978–992.
- en
- Sigl, J. C. et N. G. Chamoun (1994). An introduction to bispectral analysis for the electroencephalogram. *Journal of Clinical Monitoring* 10(6), 392–404.
- Tarvainen, M. P., S. Georgiadis, T. Laitio, J. A. Lipponen, P. A. Karjalainen, K. Kaskinoro, et H. Scheinin (2012). Heart Rate Variability Dynamics During Low-Dose Propofol and Dexmedetomidine Anesthesia. *Annals of Biomedical Engineering* 40(8), 1802–1813.
- Yoo, C. S. et S. H. Yi (2004). Effects of Detrending for Analysis of Heart Rate Variability and Applications to the Estimation of Depth of Anesthesia. In *Journal of the Korean Physical Society*. ISSN : 03744884.
- en
- Zhan, J., Z.-x. Wu, Z.-x. Duan, G.-y. Yang, Z.-y. Du, X.-h. Bao, et H. Li (2021). Heart rate variability-derived features based on deep neural network for distinguishing different anaesthesia states. *BMC Anesthesiology* 21(1), 66.

Summary

Anesthesia monitoring tools are essentials for anesthesiologists. They help prevent underdosing and overdosing of anesthetic agents and their consequences. Current monitoring devices used during surgery rely upon electroencephalogram analysis. Therefore, they can't be used during head surgery or on burned patients. In this article, we propose the use of machine learning algorithms on ECG derived features to detect anesthesia state. We build an annotated dataset of patients undergoing surgery. With that dataset, we show that fragmentation indices are particularly important for our task. We also show that electrocardiogram could be an alternative to electroencephalogram to monitor depth of anesthesia.