



**HAL**  
open science

# Analyse automatique de l'état de sommeil sur données EEG par utilisation de Transformers et de matrices de covariance

Mathieu Seraphim, Paul Dequidt, Alexis Lechervy, Florian Yger, Luc Brun,  
Olivier Etard

## ► To cite this version:

Mathieu Seraphim, Paul Dequidt, Alexis Lechervy, Florian Yger, Luc Brun, et al.. Analyse automatique de l'état de sommeil sur données EEG par utilisation de Transformers et de matrices de covariance. 19ème Colloque ORASIS (ORASIS 2023) :, journées francophones des jeunes chercheurs en vision par ordinateur,, Laboratoire d'Informatique et Systèmes UMR7020, May 2023, Carqueiranne, France. hal-04055874v2

**HAL Id: hal-04055874**

**<https://hal.science/hal-04055874v2>**

Submitted on 12 Apr 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Analyse automatique de l'état de sommeil sur données EEG par utilisation de Transformers et de matrices de covariance

Mathieu Seraphim<sup>1</sup> Paul Dequidt<sup>1</sup> Alexis Lechervy<sup>1</sup> Florian Yger<sup>2,1</sup> Luc Brun<sup>1</sup> Olivier Etard<sup>3</sup>

<sup>1</sup> Laboratoire GREYC, UMR CNRS 6072, Normandie Univ, CNRS, UNICAEN, ENSICAEN, France

<sup>2</sup> Laboratoire LAMSADE, UMR CNRS 7243, Université Paris-Dauphine, PSL, France

<sup>3</sup> Université de Caen Normandie, INSERM, COMETE U1075, CYCERON, CHU de Caen, Normandie Univ, F-14000 Caen, France  
mathieu.seraphim@unicaen.fr

## Résumé

Les données électroencéphalographiques (EEG) sont communément utilisées en médecine du sommeil. Il s'agit d'un ensemble de signaux électriques cérébraux issus de différents capteurs, subdivisés en segments devant être annotés manuellement pour quantifier les différents stades de sommeil. Ces dernières années, une littérature croissante s'est accumulée sur l'automatisation de ce processus d'annotation, offrant des résultats prometteurs, mais insuffisants pour une utilisation en milieu clinique. Nous proposons d'explorer une approche alternative afin d'améliorer la classification, basée sur l'étude de l'information portée par les covariations entre plusieurs signaux EEG représentatifs de différentes régions cérébrales. Ces covariations prennent la forme de séquences temporelles de matrices de covariance, exploitées au travers de mécanismes d'attention à l'échelle intra-époque et inter-époque. Nous validons nos résultats sur un jeu de données standard de l'État de l'Art.

## Mots Clef

Analyse du sommeil, EEG, apprentissage profond, attention, corrélations, matrices symétriques définies positives.

## Abstract

Electroencephalographic data (EEG) is commonly used in sleep medicine. It consists of a number of cerebral electrical signals measured from various brain locations, subdivided into segments that must be manually scored to reflect their sleep stage. These past few years, multiple implementations of an automatization of this scoring process have been attempted, with promising results, although they are not yet accurate enough to see clinical use. We propose a novel approach, that relies on the information contained within the covariations between multiple EEG signals, each signal representative of a different cerebral region. This is done through temporal sequences of covariance matrices, analyzed through attention mechanisms at both the intra- and inter-epoch levels. Evaluation is performed on a standard dataset, for comparison with the State of the Art.

## Keywords

Sleep analysis, EEG, Deep Learning, attention, correlations, symmetric positive definite matrices

## 1 Introduction

En médecine du sommeil, l'analyse du sommeil d'un sujet passe souvent par un examen de polysomnographie (PSG). Lors de cet examen, des électrodes sont placées sur le corps du sujet pour relever l'activité de divers organes sous forme de signaux électriques - on parle d'électroencéphalogramme (EEG), d'électrooculogramme (EOG), d'électromyogramme (EMG) et d'électrocardiogramme (ECG) respectivement pour l'activité du cerveau, des yeux, des muscles et du coeur. Ces signaux correspondent à la différence de potentiel entre deux électrodes au cours du temps, souvent avec une électrode servant de référence. Dans cet article, le terme "signal" sera utilisé exclusivement dans ce contexte.

Il peut exister un certain recouvrement entre ces signaux ; par exemple, un signal EEG peut être parasité par les signaux EOG. Les conventions les plus utilisées dans le cadre de l'analyse du sommeil par polysomnographie sont celles définies par l'Académie Américaine de la Médecine du Sommeil (AASM) [4]. Elles préconisent de subdiviser les signaux électriques en "époques" de 30 secondes (que nous appellerons époques de sommeil), et reconnaissent cinq états distincts : l'éveil, le sommeil paradoxal (REM), et trois états de sommeil lent (N1 à N3). Elles préconisent également l'acquisition de signaux non seulement EEG, mais aussi EOG et EMG, pour annoter au mieux l'état de sommeil. Cette annotation est guidée par la présence de divers événements ponctuels dans le signal EEG. Ces événements sont limités à des intervalles fréquentiels distincts, et leur analyse passe typiquement par une subdivision du domaine fréquentiel en bandes de fréquences. Les bandes que nous utilisons sont présentées en Tableau 1.

L'annotation de chaque époque de sommeil se fait manuellement par un expert. Ce processus d'annotation est laborieux, surtout lorsqu'il s'agit de nuits complètes. C'est

	Delta	Thêta	Alpha	Bêta <sub>bas</sub>	Bêta <sub>haut</sub>	Gamma
Hz	[0,5; 4[	[4; 8[	[8; 12[	[12; 22[	[22; 30[	[30; 45[

TABLEAU 1 – Bandes de fréquence considérées pour l’analyse de données EEG

pourquoi de nombreuses approches ont été développées pour l’automatiser.

Dans cet article, nous étudions la pertinence de l’analyse de la connectivité fonctionnelle cérébrale pour déterminer l’état de sommeil d’un individu, au travers de l’étude des covariations entre signaux EEG. Nous utilisons pour cela des séquences temporelles de matrices de covariance, calculées pour différentes bandes de fréquence. Nous nous appuyons sur une architecture existante à base d’encodeurs de *Transformer* pour effectuer notre classification. Après une vue d’ensemble de l’État de l’Art en Section 2, nous expliquerons notre traitement des données en Section 3.1, avant de présenter l’architecture de notre modèle en 3.2. Enfin, la Section 4 est dédiée à la présentation de nos résultats, comprenant une présentation de la base de données utilisée et une comparaison aux méthodes de l’État de l’Art.

## 2 État de l’Art

### 2.1 Modèles prédictifs et états de sommeil

Certaines approches considèrent qu’un unique signal contient assez d’information pour prédire l’état de sommeil associé à chaque époque [27, 19, 18]. Une approche commune consiste à combiner un signal EEG et un signal EOG par soustraction [20, 7, 22, 23], afin d’obtenir ce signal d’entrée. D’autres approches peuvent prendre une multitude de signaux en entrée. Dans ce cas-là, elles intègrent généralement d’autres signaux, par exemple EOG ou EMG, à l’entrée du modèle. Phan et al. [15, 16, 17] intègrent au plus un signal de chaque origine (EEG, EOG, EMG), tandis que d’autres [6, 21, 10, 9] peuvent en utiliser plusieurs. Appliquées au même jeu de données, ces dernières approches semblent donner empiriquement de meilleurs résultats.

Un prétraitement commun des données EEG consiste à en extraire les composantes fréquentielles d’intérêt. Ainsi, Sun et al. [21] et Dong et al. [7] extraient la densité spectrale de puissance (DSP) de leurs signaux, tandis que Phan et al. [15, 16, 17, 18] calculent une image temps-fréquence. On peut justifier facilement une extraction explicite de composantes fréquentielles des signaux EEG, vu que les événements caractérisant les états de sommeil ont des signatures fréquentielles spécifiques (Section 1).

Les conventions d’annotation manuelle de l’état de sommeil correspondant à une époque de sommeil incluent la prise en compte de son contexte - en d’autres termes, les époques entourant l’époque de sommeil à annoter. L’architecture des réseaux reflète ce besoin, et peut souvent être découpée en deux sections : intra-époque et inter-époque. La première étape consiste en l’extraction de des-

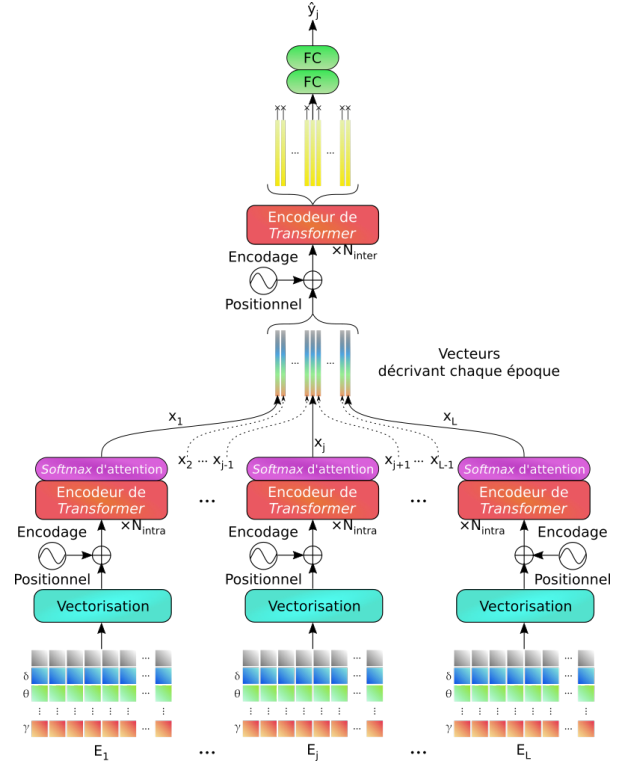


FIGURE 1 – L’architecture de notre réseau.  $(E_1, \dots, E_L)$  est la séquence d’époques en entrée, avec  $E_j$  l’époque centrale, et  $\hat{y}_j$  la sortie du réseau.  $N_{intra}$  et  $N_{inter}$  correspondent au nombre de couches de nos encodeurs intra- et inter-époque respectivement.

cripteurs spécifiques à chaque époque. Ces descripteurs sont ensuite combinés avec ceux des époques voisines, afin d’intégrer leur information contextuelle dans la classification. Une approche commune est d’utiliser des réseaux convolutifs (CNN) au stade intra-époque [19]. Cela peut être suivi par une simple combinaison de valeurs [15], ou par un réseau récurrent (RNN) uni- [7, 23] ou bi-directionnel [20, 22, 16, 21]. L’utilisation de réseaux bi-directionnels peut être justifiée, car dans la pratique, l’annotation manuelle des époques se fait en considérant non seulement les époques précédant l’époque considérée, mais aussi les époques suivantes. Phan et al. ont étendu cette approche, en utilisant un RNN bi-directionnel aux échelles intra- et inter-époque dans [16, 17].

Une autre approche intéressante consiste à non seulement considérer les relations entre époques voisines, mais aussi entre de multiples signaux EEG d’une même époque. Ainsi, Jia et al. [10, 9] construisent un graphe explicitant les relations entre différents signaux au sein d’une même époque (ce que l’on appelle la connectivité fonctionnelle, voir Section 2.2). Plus précisément, le réseau intra-époque consiste en un apprentissage sur graphes où le signal de chaque électrode est codé par un noeud du graphe. Ces graphes sont ensuite convolués spatialement et

temporellement pour obtenir une classification. Toutefois, un avantage des approches par graphes est leur invariance aux permutations, et dans une application où l'ensemble des sommets est fixe, cette propriété a peu d'intérêt. Elle peut même être contre-productive, vu que la spécificité des noeuds ne peut être prise en compte.

De nombreux modèles incluent des mécanismes d'attention. Les mécanismes d'attention peuvent se comprendre comme une émulation de l'attention humaine - à partir d'une séquence de données d'entrée, le mécanisme va apprendre quels éléments de cette séquence sont les plus importants pour la tâche en cours, et augmenter leur impact sur la classification finale. En particulier, les *Transformers* présentés par Vaswani et al. [24] utilisent des mécanismes d'autoattention, où différents descripteurs des mêmes données sont comparés au sein d'un mécanisme attentionnel. Sun et al. [21] et Jia et al. [10, 9] utilisent des mécanismes d'attention, tandis que Qu et al. [19] implémentent un bloc d'attention *multi-head* comme réseau inter-époque. Le réseau SleepTransformer de Phan et al. [18] utilise des encodeurs de *Transformer* aux échelles intra- et inter-époque, émulant ainsi leurs travaux précédents utilisant des RNN bi-directionnels à ces deux échelles. Similairement, Zhu et al. [27] utilisent aux deux échelles des blocs attentionnels similaires à ces encodeurs, entrelacés de couches plus classiques (convolutions, *dropout*, etc.).

## 2.2 Connectivité fonctionnelle

Une approche sous-représentée dans la littérature concerne la relation entre différentes régions du cerveau, ou connectivité fonctionnelle. Il s'agit de l'étude de la connectivité entre les différentes régions du cerveau par les corrélations détectées entre lesdites régions, bien souvent indépendamment de la connectivité structurelle (physique) existant entre elles [8]. Dans le cadre de l'étude du sommeil, il a été prouvé que l'état de sommeil induit une réponse cérébrale caractéristique, descriptible en termes de connectivité fonctionnelle [5].

Il existe de nombreuses manières d'estimer la connectivité entre différentes zones du cerveau, mais dans le cadre de l'analyse de l'état de sommeil au travers de données EEG, nous nous sommes intéressés à l'étude des covariations entre de multiples signaux EEG. Plus précisément, nous étudions les matrices de covariance issues de ces signaux. De telles approches ont déjà été implémentées avec succès, notamment dans le contexte des interfaces cerveau-machine (BCI) [25, 11].

Toute matrice de covariance est théoriquement symétrique semi-définie positive, mais dans notre cas (données physiologiques non simulées), elles sont généralement symétriques définies positives (ou SDP). L'ensemble des matrices SDP de  $\mathbb{R}^{n \times n}$  forme l'intérieur d'un cône qui peut être muni d'une géométrie Riemannienne. En d'autres termes, il s'agit d'une variété de Riemann, ou espace métrique courbe, sur lequel des opérations d'interpolation et d'extrapolation sont possibles. Pour effectuer ces opéra-

tions, nous pouvons définir deux familles de métriques sur la variété.

Tout d'abord, les métriques invariantes par transformations affines, ou *affine invariant* [14] sont souvent privilégiées, mais présentent des inconvénients. Par exemple, il est impossible de calculer une moyenne algébriquement en utilisant ces métriques, bien qu'il existe des approximations algorithmiques. Une autre famille est celle des métriques dites LogEuclidiennes [2], qui ne présentent pas les mêmes propriétés d'invariance, mais offrent des facilités de calcul non négligeables. Ces métriques prennent avantage de la bijection existant entre l'ensemble des matrices SDP et l'ensemble des matrices symétriques, par les opérateurs de logarithme et d'exponentielle matricielle. Plus précisément, la distance LogEuclidienne classique entre deux matrices SDP  $A$  et  $B$  est :

$$\delta_{LE}(A, B) = \|\log(A) - \log(B)\|_F \quad (1)$$

avec  $\|\cdot\|_F$  la norme de Frobenius de  $\mathbb{R}^{n \times n}$ . Nous pouvons paramétrer cette équation par une matrice SDP  $P$  [26], nous donnant la formule générale suivante pour une distance LogEuclidienne :

$$\delta_{LE}^P(A, B) = \|\log(P^{-1/2}AP^{-1/2}) - \log(P^{-1/2}BP^{-1/2})\|_F \quad (2)$$

L'Équation 1 peut être vue comme un cas particulier de l'Équation 2, avec  $P = \mathbb{I}_n$ , et  $\mathbb{I}_n$  l'identité. Essentiellement, l'utilisation d'une métrique LogEuclidienne consiste à projeter les matrices SDP sur l'espace tangent de la variété (ensemble des matrices symétriques), centré en un point  $P$ .

En tant que données biologiques, les données EEG sont fortement marquées par les spécificités du sujet dont elles proviennent, et ces spécificités seront transmises à nos matrices de covariance. Afin de les retirer au mieux, nous effectuons une opération de *whitening* sur nos matrices [26] :

$$M' = G^{-1/2}MG^{-1/2} \quad (3)$$

avec  $M$  une matrice SDP calculée sur un segment d'enregistrement EEG, et  $G$  la matrice SDP moyenne de l'enregistrement dont  $M$  est tirée.  $G$  est obtenue en calculant la matrice de covariance du signal dans sa globalité. Utiliser cette opération, en conjonction avec la métrique LogEuclidienne non paramétrée (Équation 1), revient à utiliser une métrique paramétrée par  $G$  (cf. Équation 2). Cela peut être interprété comme un "transport" des données centrées en  $G$  pour qu'elles soient centrées en  $\mathbb{I}_n$ , permettant ainsi de comparer des données tirées de différents sujets en utilisant uniquement la métrique présentée en Équation 1. Cela est important, au vu du manque d'invariances de la métrique LogEuclidienne. Recentrer les données en  $\mathbb{I}_n$  permet de minimiser d'éventuelles déformations.

Il existe une autre bijection entre l'ensemble des matrices symétriques de  $\mathbb{R}^{n \times n}$  et l'ensemble  $\mathbb{R}^{n(n+1)/2}$ . Étant donnée une matrice SDP, nous pouvons la vectoriser par combinaison du logarithme matriciel et de cette bijection. Cette

transformation opère une réduction de dimensionnalité, de manière réversible et sans perte d'information. Cela peut être interprété comme une extension des métriques LogEuclidiennes, nous permettant d'effectuer des opérations euclidiennes sur les vecteurs en conservant l'interprétabilité Riemannienne de nos données. Par conséquent, nous pouvons exprimer toute époque de sommeil en séquence temporelle de vecteurs représentatifs des covariations entre  $n$  canaux.

Notre approche repose sur le postulat qu'il est possible d'obtenir une meilleure classification de l'état de sommeil à partir des covariations entre signaux d'EEG représentatifs de différentes zones du cerveau. Afin d'explorer ce postulat, nous avons décidé d'analyser des séquences temporelles de matrices de covariance entre ces signaux EEG. Notre approche étant basée sur une analyse du cerveau uniquement, nous n'avons pas inclus de signaux d'une autre nature (EOG, EMG...). Le détail du choix des signaux EEG utilisés est expliqué en Section 4.1. Nous avons basé notre architecture sur le réseau SleepTransformer [18], au vu de son utilisation de mécanismes d'attention.

## 3 Notre analyse de l'état de sommeil

### 3.1 Données d'entrée

De même que Zhu et al. [27], nous normalisons les signaux (*z-score normalization*), afin qu'ils aient une moyenne de 0 et un écart-type de 1.

D'après l'AASM [4], les événements ponctuels permettant d'annoter le signal EEG sont environ de l'ordre de la seconde. Pour pouvoir relever ces signaux, nous avons décidé de subdiviser le signal brut en fenêtres d'une seconde chacune. Chaque époque de sommeil est donc décomposée en 30 sous-fenêtres (Section 1), sans recouvrement. Sur chacune de ces 30 fenêtres, une matrice de covariance est calculée entre les  $n$  signaux d'entrée. Afin de s'assurer que ces matrices de  $\mathbb{R}^{n \times n}$  soient bien conditionnées, nous ajoutons à toute matrice non-SDP la matrice  $\mathbb{I}_n \times 10^{-5}$ . Les événements indicatifs de l'état de sommeil en EEG ont des composantes fréquentielles distinctes (Section 1). Afin de permettre au réseau de les analyser plus efficacement, nous avons filtré les signaux EEG suivant les six bandes de fréquence présentées en Section 1. Ces filtrages ont été effectués par un filtre Butterworth passe-bande d'ordre 4. Des matrices de covariance ont été calculées entre les signaux filtrés, résultant en un total de 7 canaux en entrée du réseau (en comptant la matrice tirée du signal non filtré) par époque de sommeil.

Étant donné une matrice SDP de  $\mathbb{R}^{n \times n}$  et un vecteur de  $\mathbb{R}^n$  quelconque, il est possible de construire une matrice SDP de  $\mathbb{R}^{(n+1) \times (n+1)}$  par l'opération présentée en Équation 5. Afin d'inclure dans nos matrices une composante spécifique à chaque signal, nous tirons profit de cette propriété. Pour chaque matrice, nous définissons un vecteur  $V$  correspondant à une statistique calculée sur chacun des  $n$  signaux EEG la composant, et ce par fenêtre d'une seconde. Nous avons choisi plusieurs statistiques possibles :

la moyenne, le maximum, le maximum moins le minimum, et la densité spectrale de puissance (DSP). Afin de paramétrer l'importance relative du vecteur  $V$ , nous le multiplions par une constante  $\alpha$  :

$$V_\alpha = V \times \alpha \quad (4)$$

Le vecteur  $V_\alpha$  est ensuite combiné à la matrice de covariance  $C$  correspondante, résultant en une matrice finale  $M$  :

$$M = \left( \begin{array}{c|c} C + V_\alpha V_\alpha^T & V_\alpha \\ \hline V_\alpha^T & 1 \end{array} \right) \quad (5)$$

Cette opération peut aussi être appliquée aux signaux filtrés, nous donnant 7 matrices SPD dans  $\mathbb{R}^{(n+1) \times (n+1)}$  par seconde. La DSP est calculée à partir du spectrogramme du signal non filtré, moyenné sur chaque bande de fréquence. Dans le reste de cet article, nous appellerons ce vecteur  $V$  un vecteur de bord, et l'opération décrite en Équation 5 l'augmentation des matrices de covariance par le vecteur de bord associé. Par abus de langage, qualifions le choix de la statistique à utiliser (tel que décrit ci-avant) de "choix du vecteur de bord".

À noter que la DSP moyenne d'un signal sur une bande de fréquence est proportionnelle à sa variance sur cette même bande de fréquence, par application du théorème de Parseval. Or, la variance est déjà présente sur la diagonale de nos matrices de covariance. Néanmoins, si  $V$  code la DSP, les éléments de la matrice  $V_\alpha V_\alpha^T$  sont homogènes à une variance au carré, et introduisent donc des moments d'ordre supérieur à nos données d'entrée.

Afin d'éliminer au mieux dans nos matrices les spécificités inhérentes aux sujets, nous effectuons une opération de *whitening*, telle que définie en Section 2.2. Les matrices moyennes utilisées sont augmentées du vecteur moyen choisi. Cette opération est appliquée aux signaux filtrés.

Ainsi, chaque époque de sommeil en entrée de notre modèle est composée de 7 canaux de 30 matrices SDP, y compris 6 issues de signaux filtrés.

### 3.2 Le modèle

Notre réseau prend en entrée une séquence temporelle d'époques de sommeil, avec une époque centrale augmentée des  $l$  époques précédentes et suivantes, pour un total de  $L = 2l + 1$ . Comme développé en Section 2, ces époques voisins fournissent une information contextuelle utile à la classification de l'époque centrale. Notre modèle extrait un descripteur par époque de la séquence, avant de les combiner pour obtenir cette classification.

La première couche du réseau est la couche de vectorisation. Elle opère tout d'abord l'augmentation des matrices de covariance par un vecteur de bord (Section 3.1). Le choix du vecteur de bord, et celui du paramètre  $\alpha$  (Équation 4), sont des hyperparamètres du réseau. Vient ensuite l'opération de *whitening*. Enfin, les matrices sont

	Méthode	MF1	Accuracy générale	Accuracy macro	Kappa
1	IITNet [20]	77,36 ± 4,76	85,08 ± 3,72	77,17 ± 4,08	0,774 ± 0,052
2	DeepSleepNet [22]	79,51 ± 4,29	<b>86,25</b> ± 2,69	79,75 ± 3,95	<b>0,795</b> ± 0,040
3	GraphSleepNet [10]	76,03 ± 4,01	83,48 ± 3,73	76,35 ± 4,53	0,751 ± 0,056
4	Notre méthode (DSP)	<b>79,78</b> ± 4,90	85,10 ± 4,17	<b>81,87</b> ± 4,83	0,778 ± 0,060
5	Notre méthode (DSP), sans covariances	61,97 ± 12,62	68,75 ± 13,77	68,66 ± 9,29	0,571 ± 0,164
6	Notre méthode, sans vecteurs de bord	78,96 ± 4,26	84,21 ± 3,88	80,61 ± 4,32	0,765 ± 0,057

TABLEAU 2 – *Ablation study* et comparaison à l'État de l'Art

vectorisées par la méthode présentée en Section 2.2. En sortie de cette couche, chaque époque est représentée par une séquence temporelle de vecteurs, toujours selon 7 canaux.

La couche suivante est un premier encodage positionnel, conformément à l'article de Vaswani et al. [24]. Ensuite, les différents canaux sont combinés par concaténation des vecteurs de chaque canal dans le sens de la longueur. Ces vecteurs sont alors entrés dans un encodeur de *Transformer*. Chaque couche d'encodeur peut être composée de multiples "sous-couches" agencées séquentiellement. Les couches fortement connectées présentes dans chaque sous-couche de l'encodeur permettent un brassage des vecteurs tirés de canaux fréquents différents. Pour combiner les vecteurs tirés d'une même époque en sortie de l'encodeur, nous utilisons une couche dite de *softmax* d'attention [18]. Il s'agit essentiellement d'une moyenne des vecteurs pondérée par des poids attentionnels. Le mécanisme est adapté de l'attention globale décrite par Luong et al. [12], avec le vecteur contextuel en paramètre appris du réseau plutôt que calculé. Les  $L$  vecteurs en sortie (descripteurs de chaque époque) passent par un autre encodeur positionnel, puis un encodeur de *Transformer* inter-époque. Le vecteur de sortie correspondant à l'époque de sommeil centrale passe enfin par deux couches fortement connectées (chacune suivie d'une activation ReLU et d'une couche de *dropout*), pour donner en sortie la classification voulue. Nous utilisons une fonction de coût par entropie croisée pour entraîner le réseau.

Nous utilisons l'algorithme Adam pour l'optimisation, avec ses paramètres  $\beta_1$ ,  $\beta_2$  et  $\epsilon$  fixés à 0,9; 0,999 et  $10^{-7}$  respectivement. Le paramètre de dégradation des pondérations (*weight decay*) est un hyperparamètre du réseau. Le taux d'apprentissage  $\lambda$  est aussi un hyperparamètre du réseau, de même que son paramètre de déclin exponentiel  $\gamma_\lambda$ .

Une vue d'ensemble de l'architecture de notre réseau est disponible en Figure 1. Les paramètres des encodeurs, y compris le nombre de sous-couches et le nombre de têtes d'attention par couche, et ceux de la couche de *softmax* d'attention et des couches fortement connectées, sont tous des hyperparamètres du réseau.

L'obtention des valeurs des hyperparamètres est décrite en Section 4.2. Les valeurs obtenues sont présentées en annexe.

## 4 Expérimentation

### 4.1 Base de données utilisée

Pour valider notre modèle, nous avons choisie la base de données Montreal Archive of Sleep Studies (MASS) [13], car elle présente le double avantage d'être très utilisée dans la littérature, tout en ayant un grand nombre d'électrodes disponibles pour notre analyse. La base de donnée MASS est composée de cinq sous-groupes, SS1 à SS5. Nous avons décidé de travailler sur la base de données SS3, la seule à être annotée selon les règles de l'AASM tout en étant composée uniquement de sujets sains. La base SS3 est composée de 62 sujets, avec un enregistrement par sujet, et 20 canaux EEG communs à tous les sujets (43 sujets en ont 21). Chaque signal EEG est filtré par un filtre notch à 60 Hz, ainsi que des filtres passe-haut et passe-bas aux fréquences de coupure de 0,30 Hz et 100 Hz respectivement [13]. Cette base de données correspond à des enregistrements sur des nuits complètes et est donc relativement déséquilibrée, avec 50,24% des époques dans l'état N2 et 8,16% dans l'état N1, le moins représenté.

Afin de récupérer un éventail significatif de signaux, et pour limiter la redondance entre électrodes voisines, nous avons décidé de travailler avec les électrodes F3, F4, C3, C4, T3, T4, O1 et O2 (respectivement frontales, centrales, temporales et occipitales). Cette sélection nous offre une distribution relativement homogène sur le crâne, avec symétrie inter-hémisphérique pour enregistrer toute variation pouvant exister entre les hémisphères. Sur SS3, ces signaux sont capturés avec le signal acquis derrière l'oreille gauche en référence.

De toutes les matrices tirées de ces données, seules 11, calculées à partir du signal filtré en bande delta, ne sont pas SDP. Nous avons corrigé ceci par la méthode présentée en Section 3.1.

### 4.2 Validation du modèle

Afin de rechercher les hyperparamètres les mieux adaptés à notre réseau, nous avons décidé de découper notre base de données en trois ensembles : entraînement, validation et test. Nous utilisons une validation croisée à  $k$  blocs pour valider nos résultats ; à ce titre, nous avons décidé d'utiliser les 31 blocs de Seo et al. [20] afin de faciliter la comparaison. Chaque bloc comprend 50 sujets pour la base d'entraînement, 10 pour celle de validation, et 2 pour celle de test. Tous ensemble, les bases de test regroupent les 62 sujets

Méthode	F1 sur N3	F1 sur N2	F1 sur N1	F1 sur REM	F1 sur Éveil
IITNet [20]	80,60 ± 10,27	90,45 ± 2,14	51,73 ± 11,82	85,08 ± 8,34	78,92 ± 9,14
DeepSleepNet [22]	<b>81,73</b> ± 11,78	<b>90,63</b> ± 2,07	53,92 ± 11,22	<b>88,92</b> ± 4,49	82,35 ± 10,09
GraphSleepNet [10]	72,14 ± 15,05	87,83 ± 3,22	50,29 ± 7,40	84,88 ± 8,70	85,00 ± 7,12
Notre méthode (DSP)	77,56 ± 12,48	88,82 ± 3,38	<b>58,66</b> ± 7,68	87,10 ± 7,83	<b>86,76</b> ± 10,89

TABLEAU 3 – Scores F1 par classe.

sans recouvrement. Le nombre d'époques différant d'un enregistrement à l'autre, la taille des ensembles d'entraînement, de validation et de test varie d'un bloc à l'autre. Nous avons choisi de fixer le paramètre  $l$  du réseau à 10, résultant en  $L = 21$ , similairement à Seo et al. Par conséquent, les 10 premiers et derniers enregistrements de chaque enregistrement ne seront pas classifiés. Au vu du déséquilibre des classes et de la difficulté associée à la classification des époques en sommeil N1 (voir Section 4.3), nous avons rééquilibré l'ensemble d'entraînement de chaque bloc par répétition, afin que chaque classe ait autant d'éléments que la classe N2. Les ensembles de validation et de test ne sont en revanche jamais rééquilibrés.

Nous avons effectué tout d'abord une recherche d'hyperparamètres en utilisant l'algorithme dit de *Tree-structured Parzen Estimator* (TPE) [3], au travers de l'outil Optuna [1]. Cette recherche fut faite sur un bloc choisi au hasard, et nous avons sélectionné le bloc 11 (avec indexation commençant à 0). Les meilleurs hyperparamètres sont ensuite utilisés pour entraîner le réseau sur les 30 autres blocs. Nous avons sélectionné le score F1 moyenné par classe (communément appelé MF1) comme métrique principale du succès de notre réseau, car elle est indicatrice de la précision du réseau par classe, et permet de prendre en compte des effets tels que le déséquilibre des classes. Cette métrique est également fréquemment utilisée dans la littérature. En termes de statistiques globales, nous calculons également la précision (*accuracy*) globale, la précision moyennée par classe (ou *macro averaged*) et le kappa de Cohen (Tableau 2). Enfin, nous calculons le score F1 par classe, afin de déterminer plus en détail les performances de notre réseau (Tableau 3).

Afin d'agréger les résultats entre les 31 blocs, nous calculons la moyenne et l'écart type obtenu sur ces blocs.

### 4.3 Reproduction de l'État de l'Art

Afin de comparer nos résultats à l'État de l'Art, nous avons sélectionné trois méthodes : DeepSleepNet de Supratak et al. [22], souvent utilisé comme benchmark dans l'État de l'Art, IITNet de Seo et al. [20], dont nous avons récupéré les blocs de validation croisée, et GraphSleepNet de Jia et al. [10], qui est également basé sur une analyse de la connectivité fonctionnelle. Les trois ont leur code disponible sur Github, et ont été entraînés sur MASS SS3 originellement.

Bien que dans leur implémentation originelle les trois modèles sont validés par validation croisée à 31 blocs, les blocs de DeepSleepNet [22] et GraphSleepNet [10] sont

subdivisés en deux ensembles uniquement, avec 60 sujets pour la base d'entraînement et 2 sujets pour la base de test. Comme pour IITNet (Section 4.2), les bases de test regroupent les 62 sujets sans recouvrement. L'absence d'une base de validation peut être justifiée pour DeepSleepNet, qui utilise un nombre fixe d'époques d'apprentissage. Pour GraphSleepNet, néanmoins, le code semble récupérer le modèle présentant les meilleurs résultats sur la base de test. Ces résultats sont alors utilisés pour calculer les statistiques globales.

Les métriques en sortie de la validation croisée sont, pour les trois articles, obtenues en concaténant les tenseurs de prédictions sur les bases de test de chaque bloc en un unique tenseur. Les statistiques sont ainsi calculées sur les 62 sujets à la fois, sans omission ou répétition. Cette approche est néanmoins discutable. Tout d'abord, elle regroupe les résultats obtenus à partir de 31 jeux différents de paramètres (un par bloc). De plus, au vu de la nature de nos données, le nombre d'époques de sommeil varie d'un sujet à l'autre, et donc d'un ensemble de test à l'autre. Tout cela crée un effet de pondération relative, donnant plus de poids aux blocs dont la base de test est plus grande. Nous pouvons noter que la métrique privilégiée par les auteurs est la précision (*accuracy*) globale, négligeant ainsi le déséquilibre des classes et les variations de précision pouvant exister d'une classe à l'autre.

Par souci de comparaison à notre modèle, nous avons réentraîné les trois modèles sélectionnés, en utilisant nos métriques, nos blocs, et notre méthode d'agrégation de données (Tableaux 2 et 3). Ces résultats sont différents de ceux publiés par les auteurs originaux, ce qui peut être expliqué par les différences de méthodologie expliquées ci-avant.

### 4.4 Analyse des résultats

Notre recherche d'hyperparamètres décrite en Section 4.2 a révélé que le vecteur de bord le plus pertinent était celui issu de la DSP. Outre cela, le meilleur jeu d'hyperparamètres est présenté en annexe.

Les trois dernières lignes du Tableau 2 présentent les résultats que nous avons obtenus. La ligne 4 correspond à notre modèle, entraîné sur les meilleurs hyperparamètres, tel que décrit jusqu'à présent. Un hyperparamètre surprenant est un  $\alpha$  à 89,43 (cf. Équation 4). Cela sous-entend qu'une très large pondération du vecteur de bord offre les meilleurs résultats, en tout cas lorsqu'il s'agit de la DSP; en d'autres termes, que notre réseau privilégie une information obtenue indépendamment pour chaque signal. Afin de déterminer l'influence de ce paramètre, la ligne 5 du ta-

bleau correspond au même modèle, mais avec les données d'entrée dépourvues des informations de covariance (les éléments non diagonaux des matrices de covariance mis à 0 avant leur augmentation). Le jeu d'hyperparamètres utilisé est obtenu par une recherche d'hyperparamètres spécifique à cette configuration. Notons que comme développé en Section 3.1, la DSP étant proportionnelle à la variance, nous pouvons interpréter ce modèle comme entraînée uniquement au travers de la variance des signaux d'entrée sur des fenêtres d'une seconde, et ce suivant plusieurs bandes de fréquence. Comme nous pouvons le voir, cela induit des résultats bien moindres, et un grand écart type par rapport aux autres entrées du tableau, témoignant d'une certaine instabilité. Enfin, la ligne 6 correspond à une configuration pour laquelle les matrices ne sont pas augmentées d'un vecteur de bord. Le jeu d'hyperparamètres utilisé est également obtenu par une recherche d'hyperparamètres spécifique à cette configuration. Nous obtenons des résultats inférieurs, mais similaires, à ceux de la ligne 4. Nous pouvons en conclure que même lorsque l'information portée par le vecteur de bord est absente, l'information contenue dans les matrices de covariance est suffisante pour obtenir de bons résultats.

Les lignes 1 à 4 du Tableau 2, montrent que notre méthode est la plus performante en MF1 et en *accuracy* macro-moyennée, et la deuxième plus performante en termes d'*accuracy* générale et du kappa de Cohen.

De plus, le Tableau 3 montre qu'en termes de score F1 par classe, notre méthode est la meilleure pour la classification de l'éveil et du sommeil N1 et la seconde sur la classe REM. La qualité des prédictions varie grandement par classe, et ce autant pour l'État de l'Art que pour notre méthode. Tout particulièrement, les époques de sommeil de classe N1 semblent être difficiles à classifier, bien que notre approche possède une avance (4 points) dans ce domaine. Cette marge explique notre performance en termes de MF1 (Tableau 2).

Les Tableaux 2 et 3 montrent donc que notre approche fournit des résultats supérieurs ou équivalents à ceux de l'État de l'Art selon les statistiques considérées. Les écarts entre les méthodes restent toutefois inférieurs aux écarts-types.

## 5 Conclusion

Nous avons présenté une méthode d'annotation automatique des états de sommeil, basée sur les covariations entre signaux EEG. Nos résultats valident notre hypothèse concernant la pertinence d'une étude des covariations entre signaux EEG pour l'annotation automatique de l'état de sommeil, et par extension, sur la pertinence d'une analyse par la connectivité fonctionnelle.

Nous avons relevé un certain nombre d'axes d'amélioration de notre réseau. Une première piste consiste à introduire un mécanisme d'attention sur les coordonnées de nos vecteurs d'entrée. Une seconde piste consiste à modifier notre réseau pour mieux respecter la structure Riemmanienne des données en entrée. Enfin, il est possible que les covariances

entre les électrodes, bien que porteuses d'information, ne soient pas suffisantes pour classifier efficacement l'état de sommeil. Nous pourrions introduire de nouveaux canaux, chacun représentant une information issue d'un unique signal.

## Annexe

Les hyperparamètres obtenus, pour la meilleure version de notre modèle présentée en quatrième ligne du Tableau 2, sont :

Vecteur de bord : DSP ;  $\alpha$  : 89,43 ; encodeur intra-époque : 3 sous-couches, 5 têtes d'attention, composantes fortement connectées de taille 1024, *dropout* à  $3,8 \times 10^{-4}$  ; taille du *softmax* d'attention : 8 ; encodeur inter-époque : 5 sous-couches, 5 têtes d'attention, composantes fortement connectées de taille 1024, *dropout* à  $4,3 \times 10^{-4}$  ; couches fortement connectées finales : taille de 2048, *dropout* à  $1,5 \times 10^{-3}$  ; taux d'apprentissage ( $\lambda$ ) : valeur de  $6,6 \times 10^{-5}$ ,  $\gamma_\lambda$  à 0,92 ; *weight decay* à  $7,0 \times 10^{-6}$ .

## Remerciements

Ce travail a été cofinancé par la Région Normandie et l'Agence Nationale de la Recherche (ANR) au travers d'une bourse de thèse HAISCoDE, et s'inscrit dans le cadre du programme opérationnel FEDER/FSE 2014-2020 (RIN PredicAlert), cofinancé par la Région Normandie et l'Union Européenne.

Ce travail a bénéficié d'un accès aux moyens de calcul de l'IDRIS au travers de l'allocation de ressources 2022-AD010613618 attribuée par GENCI, et du mésocentre CRIANN (Centre Régional Informatique et d'Application Numériques de Normandie).

Remerciements à Huy Phan [15, 16, 17, 18] pour son aide lors de nos correspondances.

## Références

- [1] Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M. : Optuna : A next-generation hyperparameter optimization framework. In : Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 2623–2631 (2019)
- [2] Arsigny, V., Fillard, P., Pennec, X., Ayache, N. : Log-euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine* **56**(2), 411–421 (2006)
- [3] Bergstra, J., Bardenet, R., Bengio, Y., Kégl, B. : Algorithms for hyper-parameter optimization. *Advances in neural information processing systems* **24** (2011)
- [4] Berry, R.B., Brooks, R., Gamaldo, C., Harding, S.M., Lloyd, R.M., Quan, S.F., Troester, M.T., Vaughn, B.V. : Aasm scoring manual updates for 2017 (version 2.4) (2017)
- [5] Bouchard, M., Lina, J.M., Gaudreault, P.O., Dubé, J., Gosselin, N., Carrier, J. : EEG connectivity across sleep cycles and age. *Sleep* **43**(3) (11 2019)



- [6] Chambon, S., Galtier, M., Arnal, P.J., Wainrib, G., Gramfort, A. : A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **26**(4), 17683810 (Mar 2018)
- [7] Dong, H., Supratak, A., Pan, W., Wu, C., Matthews, P.M., Guo, Y. : Mixed neural network approach for temporal sleep stage classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **26**(2), 324–333 (2018)
- [8] Eickhoff, S., Müller, V. : Functional connectivity. In : Toga, A.W. (ed.) *Brain Mapping*, pp. 187–201. Academic Press, Waltham (2015)
- [9] Jia, Z., Lin, Y., Wang, J., Ning, X., He, Y., Zhou, R., Zhou, Y., Lehman, L.w.H. : Multi-view spatial-temporal graph convolutional networks with domain generalization for sleep stage classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **29**, 1977–1986 (2021)
- [10] Jia, Z., Lin, Y., Wang, J., Zhou, R., Ning, X., He, Y., Zhao, Y. : Graphsleepnet : Adaptive spatial-temporal graph convolutional networks for sleep stage classification. In : *IJCAI*. pp. 1324–1330 (2020)
- [11] Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., Arnaldi, B. : A review of classification algorithms for eeg-based brain–computer interfaces. *Journal of Neural Engineering* **4**(2), R1 (jan 2007)
- [12] Luong, T., Pham, H., Manning, C.D. : Effective approaches to attention-based neural machine translation. In : *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. pp. 1412–1421. Association for Computational Linguistics, Lisbon, Portugal (Sep 2015)
- [13] O’reilly, C., Gosselin, N., Carrier, J., Nielsen, T. : Montreal archive of sleep studies : an open-access resource for instrument benchmarking and exploratory research. *Journal of sleep research* **23**(6), 628–635 (2014)
- [14] Pennec, X., Fillard, P., Ayache, N. : A riemannian framework for tensor computing. *International Journal of Computer Vision* **66**(1), 41–66 (Jan 2006)
- [15] Phan, H., Andreotti, F., Cooray, N., Chén, O., de Vos, M. : Joint classification and prediction cnn framework for automatic sleep stage classification. *IEEE Transactions on Biomedical Engineering* **66**, 1285–1296 (05 2019)
- [16] Phan, H., Andreotti, F., Cooray, N., Chén, O.Y., De Vos, M. : Seqsleepnet : End-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **27**(3), 400–410 (2019)
- [17] Phan, H., Chén, O.Y., Tran, M.C., Koch, P., Mertins, A., De Vos, M. : Xsleepnet : Multi-view sequential model for automatic sleep staging. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **44**(9), 5903–5915 (2022)
- [18] Phan, H., Mikkelsen, K., Chén, O.Y., Koch, P., Mertins, A., De Vos, M. : Sleeptransformer : Automatic sleep staging with interpretability and uncertainty quantification. *IEEE Transactions on Biomedical Engineering* **69**(8), 2456–2467 (2022)
- [19] Qu, W., Wang, Z., Hong, H., Chi, Z., Feng, D.D., Grunstein, R., Gordon, C. : A residual based attention model for eeg based sleep staging. *IEEE journal of biomedical and health informatics* **24**(10), 2833–2843 (2020)
- [20] Seo, H., Back, S., Lee, S., Park, D., Kim, T., Lee, K. : Intra- and inter-epoch temporal context network (iitnet) using sub-epoch features for automatic sleep scoring on raw single-channel eeg. *Biomedical Signal Processing and Control* **61**, 102037 (2020)
- [21] Sun, C., Chen, C., Li, W., Fan, J., Chen, W. : A hierarchical neural network for sleep stage classification based on comprehensive feature learning and multi-flow sequence learning. *IEEE Journal of Biomedical and Health Informatics* **24**(5), 1351–1366 (2020)
- [22] Supratak, A., Dong, H., Wu, C., Guo, Y. : Deepsleepnet : a model for automatic sleep stage scoring based on raw single-channel eeg. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **25**(11), 1998–2008 (Nov 2017)
- [23] Supratak, A., Guo, Y. : Tinsleepnet : An efficient deep learning model for sleep stage scoring based on raw single-channel eeg. In : *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*. pp. 641–644 (2020)
- [24] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L.u., Polosukhin, I. : Attention is all you need. In : Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) *Advances in Neural Information Processing Systems*. vol. 30. Curran Associates, Inc. (2017)
- [25] Yger, F., Berar, M., Lotte, F. : Riemannian approaches in brain-computer interfaces : A review. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **25**(10), 1753–1762 (2017)
- [26] Yger, F., Sugiyama, M. : Supervised logeuclidean metric learning for symmetric positive definite matrices (2015)
- [27] Zhu, T., Luo, W., Yu, F. : Convolution-and Attention-Based Neural Network for Automated Sleep Stage Classification. *Int J Environ Res Public Health* **17**(11) (Jun 2020)