

Détection d'épilepsie dans les signaux EEG par graphe de visibilité et un noyau de SVM adapté

Tristan AVERTY, Delphine DARÉ-EMZIVAT, Abdel-Ouahab BOUDRAA

IRENav, École Navale/Arts-Metiers ParisTech, CC 600, 29240 Brest Cedex 9, France.
[tristan.averty, delphine.dare, abdel.boudraa]@ecole-navale.fr

Résumé – Dans cet article, nous présentons une stratégie de détection d'épilepsie à partir de signaux EEG (issus d'un seul capteur) basée sur l'algorithme de visibilité, qui consiste à transformer une série temporelle en un graphe dit de visibilité (GV). Nous montrons la pertinence de la distribution des degrés des sommets du GV, qui, combinée à une fonction noyau type Jensen-Shannon d'une machine à vecteurs de support (SVM) permet une classification binaire performante : présence ou absence d'une crise d'épilepsie. La méthode proposée est illustrée sur des données réelles et les résultats comparés à des méthodes de la littérature. Les résultats obtenus en termes de précision, de sensibilité et de spécificité sont similaires à ceux de la littérature voire meilleurs, et ce, en exploitant un seul attribut, à savoir la distribution des degrés du graphe.

Abstract – In this paper, we present a strategy for epilepsy detection in EEG signals (from a single sensor) based on the visibility algorithm, which consists in transforming a time series into a so-called visibility graph (VG). We show the relevance of the degree distribution of the vertices of the VG, which, combined with a Jensen-Shannon kernel function of a support vector machine (SVM), allows a powerful binary classification: presence or absence of an epileptic seizure. The proposed method is illustrated on real data and the results are compared to literature methods. The results obtained in terms of accuracy, sensitivity and specificity are similar to those of the literature or even better, and this, by exploiting only one attribute, namely the degree distribution of the graph.

1 Introduction

L'électroencéphalogramme (EEG) constitue l'enregistrement de l'activité électrique du cerveau en mesurant les courants délivrés par les connexions nerveuses. C'est le moyen le plus utilisé pour détecter par exemple les crises d'épilepsie, ces dernières étant issues d'un dysfonctionnement passager d'un groupe de neurones. Ces dernières années, des études ont montré l'apport de l'apprentissage automatique pour la détection et la classification de signaux EEG chez les sujets épileptiques, soit en passant d'abord par l'extraction d'attributs pertinents du signal EEG, comme les caractéristiques temporelles et fréquentielles [1] ou les coefficients d'ondelettes [2], soit en manipulant directement ces signaux dans des réseaux de neurones dit récurrents [3].

Dans le but d'analyser et de traiter des signaux de plus en plus nombreux et provenant de plusieurs capteurs, la théorie des graphes est arrivée au renfort du traitement du signal et de l'image. Les objets appelés graphes, composés de sommets et d'arêtes (pondérés ou non), ont initialement été créés pour modéliser des systèmes complexes dans la génétique ou encore les réseaux sociaux et les réseaux informatiques. Cette théorie prend désormais une place prépondérante dans le traitement du signal. Une manière d'allier ces deux domaines est de considérer un graphe issu d'une transformation du signal telle que le graphe de visibilité (GV) [4], dont les propriétés topologiques sont liés à la dynamique du système physique ou biologique sous-jacent. Ce dernier est bien adapté pour l'analyse

de signaux liés à la biologie médicale tels que les ECG [5] ou les EEG [6]. En particulier, les distributions des degrés de ce graphe apportent des informations pertinentes pour la classification des signaux d'entrée.

Dans cet article, nous présentons une stratégie de détection de crise d'épilepsie exploitant la distribution des degrés de graphes de visibilité calculés à partir de signaux EEG. Pour comparer ces distributions, nous proposons une adaptation du traditionnel noyau gaussien de SVM avec la distance de Jensen-Shannon (JSD).

La notion de graphe de visibilité fait l'objet de la section 2. Une troisième partie expose le noyau de SVM basé sur la JSD. Après la présentation de la base de données en section 4, les résultats obtenus sont présentés et commentés dans la section 5.

2 Graphe de visibilité

L'analyse des séries temporelles en les transformant en graphes est une idée introduite par Zhang et Small [7], mais reste limitée aux séries pseudo-périodiques. On peut citer également comme méthodes de représentation de séries temporelles les réseaux de récurrence [8] ou les réseaux de transition [9]. Lacasa *et al.* [4] ont introduit l'algorithme de visibilité qui convertit une série temporelle en un graphe, dit de visibilité naturelle (GVN), où chaque échantillon de la série est transformé en un nœud et les arêtes entre les nœuds sont définis par un critère géométrique. L'intérêt du GVN est sa capacité d'analyser, avec une faible complexité algorithmique, une large classe de

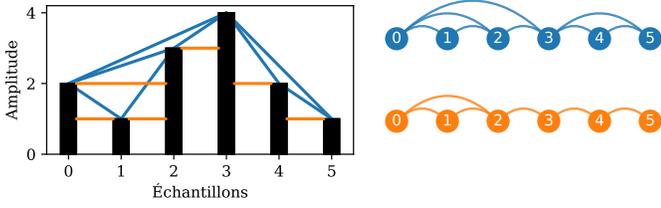


FIGURE 1 – À gauche : Séries temporelles (6 échantillons) et critères de visibilité / À droite : Graphes de visibilité associés (6 sommets) : GVN en haut (8 arêtes) / GVH en bas (6 arêtes).

séries temporelles, en particulier celles issues de systèmes physiques ou biologiques très complexes [10],[11]. Soit une série temporelle $\mathbf{x} = (x_i)_{1 \leq i \leq n}$ ayant n points, le GVN possédant n nœuds est construit en utilisant le critère géométrique suivant : une arête existe entre les sommets i et j si et seulement si les deux points (i, x_i) et (j, x_j) peuvent se voir mutuellement c'est-à-dire si tout autre échantillon (k, x_k) placé entre les deux satisfait :

$$x_k < x_i + (x_j - x_i) \frac{k - i}{j - i}, \quad \forall k, i < k < j \quad (1)$$

Une variante introduite simplifiant le GVN, appelée graphe de visibilité horizontale (GVH), a été introduite par Luque *et al.* [12]. Elle est basée sur le simple fait que deux sommets i et j seront connectés par une arête dans le graphe correspondant si et seulement si une ligne horizontale peut être tracée entre x_i et x_j sans intersecter de valeurs intermédiaires c'est-à-dire si et seulement si tout autre échantillon (k, x_k) placé entre les deux satisfait $x_k < \min(x_i, x_j), \forall k, i < k < j$. La figure 1 présente une illustration de ces deux types de critères de visibilité et leurs graphes associés.

Un élément pertinent dans l'étude d'un GV est la distribution probabiliste des degrés de chaque sommet de ce graphe (le degré d'un sommet étant égal au nombre d'arêtes liant ce sommet à d'autres sommets), car elle peut mettre en avant certaines caractéristiques du signal. Ainsi, l'analyse des distributions des degrés, et notamment ceux issus du GVH, permet de caractériser le type d'une série aléatoire [4], d'estimer l'exposant de Hurst d'un mouvement Brownien fractionnaire [10] ou de distinguer le bruit du chaos [11]. Afin d'exploiter ces distributions des degrés à une fin de classification, nous utilisons une fonction noyau de SVM basée sur une distance statistique appelée distance de Jensen-Shannon permettant de les comparer quantitativement.

3 Noyau adapté d'une SVM

Même si la *deep learning* prend une place de plus en plus importante dans les tâches de classification, la machine à vecteurs de support (ou SVM pour *Support Vector Machine*) reste utilisée car elle est robuste avec une grande capacité de généralisation et facile à implémenter. Soit l'ensemble d'apprentissage labellisé $\{(\mathbf{x}_i, y_i) \in \mathbb{R}^n \times \{-1, 1\}\}_{1 \leq i \leq p}$. Le principe de base de la SVM est de trouver l'hyperplan qui sépare linéairement les données en deux classes tout en maximisant les marges.

Supposons que les données soient « presque » linéairement séparables. En introduisant un paramètre dit de « régularisation » C , le problème dual à résoudre est :

$$\max_{\alpha} \sum_{i=1}^p \alpha_i - \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \quad (2)$$

$$\text{avec } \forall i, 0 \leq \alpha_i \leq C \text{ et } \sum_{i=1}^p \alpha_i y_i = 0$$

Dans les faits, l'hyperplan séparateur est souvent recherché dans un espace \mathcal{H} de plus grande dimension et muni d'un produit scalaire. Pour cela, les observations \mathbf{x}_i sont projetées à l'aide d'une fonction non-linéaire $\phi : \mathbb{R}^n \rightarrow \mathcal{H}$. La formulation du problème dual (2) ne dépendant de \mathbf{x}_i et \mathbf{x}_j qu'à travers un produit scalaire, plutôt que de choisir ϕ , il est préférable de définir une fonction symétrique dite « noyau » $k : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ représentant un produit scalaire dans \mathcal{H} : $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle_{\mathcal{H}}$. La matrice $(k(\mathbf{x}_i, \mathbf{x}_j))_{i,j}$ est appelée matrice de Gram du noyau k . L'un des noyaux les plus utilisés est le noyau gaussien défini à partir de la distance euclidienne :

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|_2) \quad (3)$$

Le paramètre γ strictement positif est appelé « coefficient de noyau » et est lié à l'écart-type de la gaussienne considérée.

Au regard des travaux récents qui étudient la classification de signaux EEG à l'aide de GV, les entrées des différents algorithmes de classification sont les degrés combinés à des grandeurs comme le poids moyen ou la modularité [13] ou encore la complexité [6]. L'objectif de ce travail est d'étudier si les degrés seuls sont pertinents pour la séparation des signaux EEG en deux classes.

Afin d'adapter le noyau (3) à cette problématique, la distance de Jensen-Shannon (JSD) [14] mesurant la similarité entre deux distributions de probabilité \mathbf{p} et \mathbf{q} est utilisée. Cette distance, basée sur une symétrisation de la divergence de Kullback-Leibler, peut être définie grâce à l'entropie de Shannon $\mathcal{S}(\cdot)$ comme :

$$\text{JSD}(\mathbf{p}, \mathbf{q}) = \mathcal{S}(\mathbf{m}) - \frac{\mathcal{S}(\mathbf{p}) + \mathcal{S}(\mathbf{q})}{2} \text{ où } \mathbf{m} = \frac{\mathbf{p} + \mathbf{q}}{2} \quad (4)$$

Dans notre cas, les distributions de probabilité sont les distributions des degrés. Ainsi, en notant $H^{(i)}$ (resp. $H^{(j)}$) la distribution des degrés du GV associé au $i^{\text{ème}}$ (resp. $j^{\text{ème}}$) signal, le noyau s'écrit :

$$k(H^{(i)}, H^{(j)}) = \exp(-\gamma \text{JSD}(H^{(i)}, H^{(j)})) \quad (5)$$

Le noyau (5) est plus adapté à la comparaison des distributions que le noyau (3). En effet, ce dernier les compare degré-à-degré tandis que la JSD est calculée sur l'ensemble de la distribution.

Le principe de la méthode est présenté en figure 2. Après avoir transformé les signaux EEG en GVH, les distributions des degrés sont calculées et la matrice de Gram constituée des JSD est construite. La SVM est alors entraînée grâce à cette matrice.

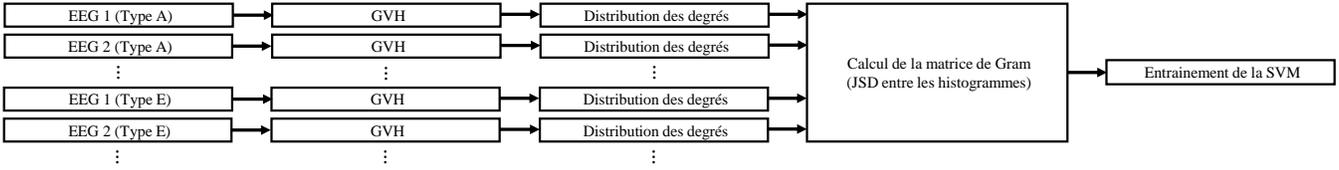


FIGURE 2 – Principe de la méthode (cas A vs. E)

4 Base de données

La détection est illustrée sur la base de données de l’Université de Bonn [15] composée de 5 ensembles d’EEG :

- A : Personnes saines avec les yeux ouverts ;
- B : Personnes saines avec les yeux fermés ;
- C : Personnes malades sans crise hippocampique ;
- D : Personnes malades sans crise d’épilepsie ;
- E : Personnes malades durant une crise d’épilepsie.

Chaque ensemble est constitué de 100 signaux chacun contenant 4097 échantillons. Pour augmenter la taille de la base de données sans perte d’information, chaque signal EEG est segmenté en 4 tranches de 1024 échantillons [13]. La base de données est ainsi constituée de 5 ensembles de 400 signaux possédant 1024 échantillons chacun. Étant donné que la classification intuitive est de déterminer si une personne est dans une crise d’épilepsie ou pas, seuls les tests A vs. E, B vs. E, C vs. E, D vs. E et (ABCD) vs. E seront à l’étude dans cet article.

Un exemple de deux signaux EEG (l’un de type A et l’autre de type E) et leurs distributions des degrés associées est présenté en figure 3. Une observation directe des séries temporelles montre une plus forte activité électrique liée à la crise épileptique pour le signal appartenant à l’ensemble E. Le kurtosis du vecteur $\mathbf{d} = (d_i)_{1 \leq i \leq n}$ composé des degrés du graphe de visibilité, défini par $\text{Kurt}[\mathbf{d}] := \frac{1}{n\sigma_d^4} \sum_{i=1}^n (d_i - \mu_d)^4$ avec μ_d (resp. σ_d) la moyenne (resp. l’écart-type) de \mathbf{d} , peut également être calculé. La distribution des degrés de l’EEG de type E apparaît plus « pointue » ($\text{Kurt}[\mathbf{d}_E] = 17.16$) que celle de l’EEG de type A ($\text{Kurt}[\mathbf{d}_A] = 4.03$). Ainsi, la différence des distributions peut être quantifiée grâce à la distance JSD (4).

5 Analyse des résultats

Pour de meilleurs résultats de précision, une validation croisée à 10 couches (90 % des échantillons pour l’entraînement et 10 % pour le test) est utilisée. Par ailleurs, une sélection des hyper-paramètres est effectuée avec le paramètre de régularisation $C \in \{1, 10, 50, 100, 250, 500, 750, 1000\}$ et le coefficient de noyau $\gamma \in \{0.1, 0.5, 1, 5, 10, 50\}$. Cette expérience est répétée 5 fois et les résultats sont moyennés. La classification est validée par les grandeurs suivantes : la précision, la sensibilité et la spécificité.

Nous comparons les méthodes de classification uniquement en utilisant les GVH. Le tableau 1 montre que la méthodologie proposée atteint et dépasse dans certains cas les résultats de la littérature. Les résultats obtenus pour le cas (A vs. E) ainsi que la précision du test (B vs. E), bien que moins élevés, présentent

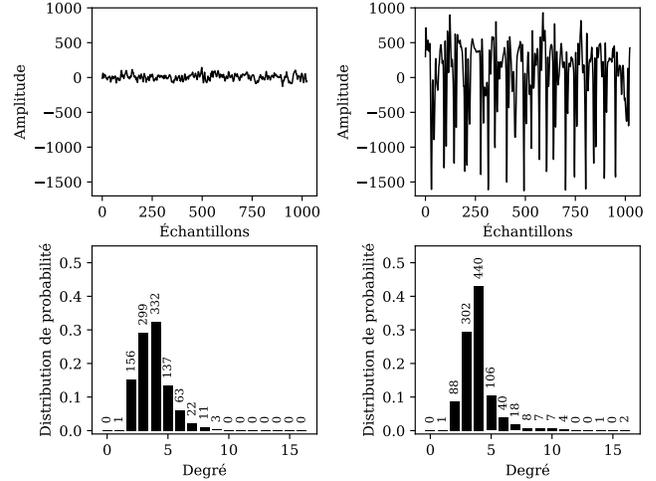


FIGURE 3 – Deux exemples de signaux EEG (1024 échantillons) et la distribution des degrés de leurs graphes de visibilité associés (1024 sommets) : Type A à gauche (1935 arêtes) / Type E à droite (2013 arêtes).

une différence maximale de 0.5 % avec les autres méthodes à l’inverse des sensibilités et des spécificités pour les tests (B vs. E), (C vs. E) et (D vs. E) qui sont améliorées d’au moins 0.5 %. Il est à noter que dans les cas (C vs. E) et (D vs. E), qui sont censés être les tests les plus ardues, la précision atteint 99.75 % dans le premier et 97 % dans le deuxième, ce qui surpasse de presque 2 % les autres méthodes utilisant les GVH.

Bien entendu, cette méthode est généralisable avec d’autres distances : la distance en variation totale $\sup |\mathbf{p} - \mathbf{q}|$, la distance euclidienne $\|\mathbf{p} - \mathbf{q}\|_2$, etc. Néanmoins, le tableau 2 montre que, dans notre étude, la JSD est meilleure que ces dernières pour tous les tests. La pertinence de la méthode basée sur un noyau adapté de SVM proposée dans cet article est supportée par les résultats reportés dans le tableau 2. En effet, si les degrés seuls sont si importants pour discriminer deux classes d’EEG, un modèle de type Random Forest prenant en entrée les distributions des degrés, entraîné suivant la même validation croisée que la SVM, devrait donner de très bons résultats. Cependant, ces derniers sont moins performants que la méthode SVM + JSD avec une différence minimale de précision égale à 0.5 % dans le cas (A vs. E) et (C vs. E) et une différence maximale égale à 3.125 % dans le cas (B vs. E).

6 Conclusion

Dans cet article, une nouvelle méthode de détection de crises d’épilepsie en classifiant des signaux EEG *via* la distribution des degrés des GVH associés à ces signaux est présentée. Pour

Tests	Méthodes	Précision (ACC)	Sensibilité (SEN)	Spécificité (SPE)
A vs. E	Zhu <i>et al.</i> [16]	99 %	100 %	99 %
	Supriya <i>et al.</i> [13]	100 %	100 %	100 %
	Rajadurai <i>et al.</i> [17]	100 %	-	-
	SVM (Jensen-Shannon)	99.5 %	99.75 %	99.5 %
B vs. E	Zhu <i>et al.</i> [16]	97 %	99 %	96 %
	Supriya <i>et al.</i> [13]	97.25 %	99.47 %	95.21 %
	Rajadurai <i>et al.</i> [17]	97.5 %	-	-
	SVM (Jensen-Shannon)	97 %	100 %	96.5 %
C vs. E	Zhu <i>et al.</i> [16]	98 %	98 %	99 %
	Supriya <i>et al.</i> [13]	98.5 %	98.5 %	98.5 %
	Rajadurai <i>et al.</i> [17]	97.75 %	-	-
	SVM (Jensen-Shannon)	99.75 %	100 %	99.75 %
D vs. E	Zhu <i>et al.</i> [16]	94 %	93 %	95 %
	Supriya <i>et al.</i> [13]	90.6 %	96.25 %	93.25 %
	Rajadurai <i>et al.</i> [17]	94.375 %	-	-
	SVM (Jensen-Shannon)	97 %	97 %	97.5 %
(ABCD) vs. E	Zhu <i>et al.</i> [16]	97 %	98 %	94 %
	SVM (Jensen-Shannon)	97.65 %	99.56 %	92.5 %

TABLEAU 1 – Comparaison des résultats avec des méthodes de la littérature à partir de graphes de visibilité.

Tests	Méthodes alternatives	Précision (ACC)
A vs. E	SVM (Jensen-Shannon)	99.5 %
	SVM (Variation totale)	99.125 %
	SVM (Euclidienne)	99.125 %
	Random Forest	99 %
B vs. E	SVM (Jensen-Shannon)	97 %
	SVM (Variation totale)	92.375 %
	SVM (Euclidienne)	95.375 %
	Random Forest	93.875 %
C vs. E	SVM (Jensen-Shannon)	99.75 %
	SVM (Variation totale)	98.625 %
	SVM (Euclidienne)	98.75 %
	Random Forest	99.25 %
D vs. E	SVM (Jensen-Shannon)	97 %
	SVM (Variation totale)	95.125 %
	SVM (Euclidienne)	96.125 %
	Random Forest	95.875 %
(ABCD) vs. E	SVM (Jensen-Shannon)	97.65 %
	SVM (Variation totale)	95.35 %
	SVM (Euclidienne)	96.8 %
	Random Forest	94.55 %

TABLEAU 2 – Comparaisons des résultats avec la méthode Random Forest et les distances variation totale et euclidienne.

cela, une adaptation du noyau gaussien pour une SVM a été proposée en introduisant une distance statistique permettant de comparer quantitativement les différentes distributions : la JSD. Cette méthode a été évaluée sur une base de données d'EEG réels. La précision et la spécificité sont améliorées dans 3 sur 5 cas et la sensibilité dans 4 sur 5 cas, avec notamment une précision bonifiée de plus de 2.5% égale à 96.75% dans le cas (D vs. E) : les résultats présentés sont donc satisfaisants et nous ont permis de montrer la pertinence des degrés en termes d'information pour la classification des signaux EEG.

Tout en sachant que les EEG sont des signaux très particuliers, il serait intéressant de voir les résultats que cette méthode peut apporter dans le cadre d'une classification de signaux d'origines plus diverses comme ceux de la base de données de l'Université de Californie¹. Il serait par ailleurs pertinent de savoir si la JSD est la meilleure quelque soit le type de signal considéré (audio, médical, électrique, etc.) et, si ce n'est pas le cas, d'étudier la variabilité des résultats suivant les types

de signaux et les distances statistiques considérées.

Enfin, la stratégie de détection présentée dans cet article a été appliquée sur des signaux EEG acquis à partir d'un seul capteur mais il peut être profitable de l'étendre à l'ensemble du réseau de connectivité fonctionnelle. De plus, c'est une stratégie de détection et non de prédiction, la base de données considérée ici ne nous permettant pas d'en faire. Il serait intéressant de voir si les GV peuvent être utilisés pour classifier les patients épileptiques hors crise.

Références

- [1] V. Srinivasan et al. Artificial neural network based epileptic detection using time-domain and frequency-domain features. *J. Med. Sys.*, 29(6) :647–660, 2005.
- [2] L. Guo et al. Automatic feature extraction using genetic programming : An application to epileptic EEG classification. *ESA*, 38(8) :10425–10436, 2011.
- [3] M. U Abbasi et al. Detection of epilepsy seizures in neo-natal EEG using LSTM architecture. *IEEE Access*, 7 :179074–179085, 2019.
- [4] L. Lacasa et al. From time series to complex networks : The visibility graph. *PNAS*, 105(13) :4972–4975, 2008.
- [5] N. Khodor et al. Evaluation de la maturation des bébés prématurés par graphe de visibilité. GRETSI, 2017.
- [6] M. Ahmadiou et al. New diagnostic EEG markers of the Alzheimer's disease using visibility graph. *Journal of neural transmission*, 117(9) :1099–1109, 2010.
- [7] M. Zhang, J. et Small. Complex network from pseudoperiodic time series : Topology versus dynamics. *Physical review letters*, 96(23) :238701, 2006.
- [8] N. Marwan et al. Generalised recurrence plot analysis for spatial data. *Physics Letters A*, 360(4-5) :545–551, 2007.
- [9] A. Shirazi et al. Mapping stochastic processes onto complex networks. *Journal of Statistical Mechanics : Theory and Experiment*, 2009(07) :P07046, 2009.
- [10] L. Lacasa et al. The visibility graph : A new method for estimating the Hurst exponent of fractional Brownian motion. *EPL*, 86(3) :30001, 2009.
- [11] M. Ravetti et al. Distinguishing noise from chaos : objective versus subjective criteria using horizontal visibility graph. *PLoS one*, 9(9) :e108004, 2014.
- [12] B. Luque et al. Horizontal visibility graphs : Exact results for random time series. *Physical Review E*, 80(4) :046103, 2009.
- [13] S. Supriya et al. Weighted visibility graph with complex network features in the detection of epilepsy. *IEEE access*, 4 :6554–6566, 2016.
- [14] J. Lin. Divergence measures based on the Shannon entropy. *IEEE TIT*, 37(1) :145–151, 1991.
- [15] R. Andrzejak et al. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity : Dependence on recording region and brain state. *Physical Review E*, 64(6) :061907, 2001.
- [16] G. Zhu et al. Epileptic seizure detection in EEGs signals using a fast weighted horizontal visibility algorithm. *Comput. Meth. Prog. Biomed.*, 115(2) :64–75, 2014.
- [17] T. Rajadurai et al. Epileptic Seizure Prediction Using Weighted Visibility Graph. In *Int. Conf. Soft Computing Systems*, pages 453–461. Springer, 2018.

1. www.timeseriesclassification.com/index.php