

Etape 1 - Etude des données

Alexia Fournier

30 avril 2023

1 Introduction

Le projet de fin d'études que nous abordons a un objectif clair et ambitieux : développer une méthode pour prédire les crises d'épilepsie à partir de données EEG/iEEG. Pour atteindre cet objectif, nous avons choisi d'utiliser la base de données CHBMIT, qui est une source de données de qualité et reconnue dans le domaine de la recherche sur l'épilepsie.

Cependant, avant de pouvoir élaborer une nouvelle méthode, il est crucial de bien comprendre les subtilités de l'ensemble de données CHBMIT. C'est pourquoi la première étape de notre projet consiste à visualiser et analyser les données. Cette étape permettra aux chercheurs de mieux comprendre les caractéristiques de l'ensemble de données, telles que la qualité des signaux EEG/iEEG et la fréquence des crises d'épilepsie enregistrées.

Une analyse minutieuse des données nous permettra également d'identifier les caractéristiques les plus significatives à prendre en compte pour la prédiction des crises d'épilepsie. Cette première étape est donc essentielle pour le succès de notre projet, car elle nous permettra de développer une méthode de prédiction précise et fiable.

En résumé, notre projet de fin d'études est une recherche passionnante visant à développer une méthode novatrice pour prédire les crises d'épilepsie. Grâce à l'utilisation de la base de données CHBMIT et à une analyse minutieuse des données, notre équipe de recherche pourra avancer vers une meilleure compréhension de l'épilepsie et proposer une méthode prometteuse pour améliorer la qualité de vie des patients atteints de cette maladie.

2 Description de la base de données CHBMIT

La base de données CHBMIT est une ressource médicale précieuse, accessible pour la recherche dans les domaines de la neurologie et de l'apprentissage automatique (Base de données CHBMIT). Elle contient des enregistrements EEG de patients pédiatriques atteints d'épilepsie, et est hébergée sur le site PhysioNet. Les enregistrements EEG ont été effectués sur 23 patients, avec plus de 68 heures de données enregistrées pour chaque patient, à travers plusieurs sessions d'une heure chacune. Ces sessions ont été enregistrées à l'aide de 128 canaux EEG.

Il est important de noter que les données sont anonymisées afin de protéger la vie privée des patients. Cela signifie qu'aucune information personnelle identifiable n'est incluse dans les enregistrements. Cette anonymisation permet aux chercheurs d'accéder aux données sans violer la confidentialité des patients. Cependant, l'utilisation de ces données est limitée à des fins de recherche médicale.

La base de données CHBMIT est largement utilisée dans la recherche en neurologie et en apprentissage automatique. Elle est notamment utilisée pour le développement d'algorithmes de détection de l'épilepsie et de classification des états de sommeil. Les données EEG contenues dans cette base de données sont très utiles pour étudier différents aspects de l'épilepsie chez les enfants, ce qui peut aider les médecins à mieux comprendre et à mieux traiter cette maladie.

3 Analyse des données

L'analyse des crises observées chez différents patients révèle une grande variation dans le nombre et la durée de ces crises. Le nombre de crises varie considérablement d'un patient à l'autre, allant de 3 à 40, avec un total de 182 crises pour l'ensemble des patients étudiés.

3.1 Durée des crises

La durée des crises présente également une grande disparité, comme le montre la figure 1.

La durée minimale des crises varie de 6 à 134 secondes, tandis que la durée maximale s'étend de 8 à 752 secondes. Il convient de noter qu'une crise particulièrement longue de 752 secondes a été enregistrée chez le patient chb11. Cela souligne l'importance de prendre en compte les valeurs aberrantes lors de l'analyse des données.

La durée moyenne des crises, en prenant en compte l'ensemble des crises des patients, est d'environ 60,88 secondes. Cependant, cette mesure peut être influencée par les valeurs aberrantes, comme la crise

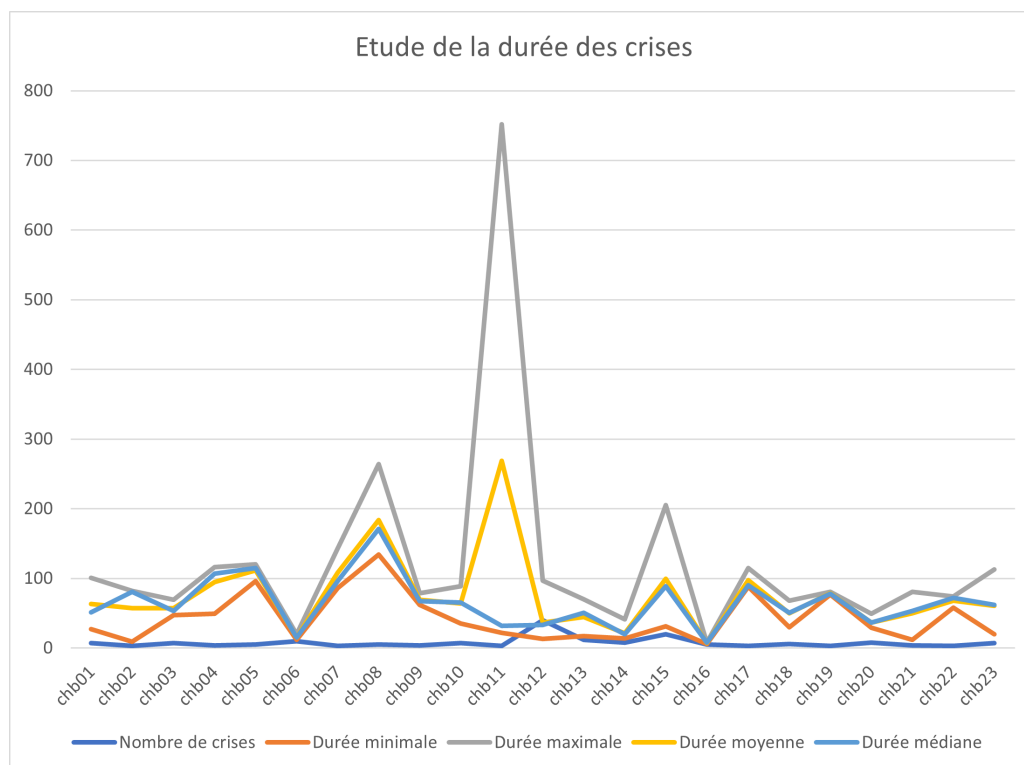


FIGURE 1 – Analyse de la durée des crises

exceptionnellement longue mentionnée précédemment. Par conséquent, il est essentiel de considérer également la durée médiane des crises, qui offre une mesure plus robuste. Dans ce cas, la durée médiane est de 47 secondes pour l'ensemble des patients.

L'analyse des données révèle des différences significatives dans le nombre et la durée des crises entre les patients. Certains patients présentent seulement 3 crises, tandis que d'autres en ont jusqu'à 40. De même, la durée des crises peut varier considérablement, allant de 6 à 752 secondes.

Il est important de noter que les caractéristiques des crises peuvent être influencées par divers facteurs tels que l'état de santé sous-jacent du patient, les médicaments pris et d'autres facteurs individuels. Ces éléments doivent être pris en considération pour interpréter les données de manière adéquate.

3.2 Durée D

Les crises d'épilepsie peuvent survenir en groupes appelés « orages de crises ». Afin de mieux comprendre les types de crises et de prédire les épisodes futurs, il est important de distinguer deux catégories de crises : les premières crises et les répliques. Cette distinction repose sur la durée minimale sans crise précédant une nouvelle crise. Dans ce contexte, la durée D est utilisée pour effectuer cette distinction. Les différents cas ont été étudiés pour chaque patient et dans leur ensemble, comme le montre la figure 2.

La figure 2 révèle que sur les 182 crises recensées, 48 d'entre elles (soit 26%) se sont produites après une durée D de 2 heures. Les durées D de 1 heure, 30 minutes et 0 minute représentent respectivement 12%, 15% et 41% des crises. Les crises survenues dans un intervalle inférieur à 30 minutes sont les plus fréquentes.

L'analyse de ces données permet de conclure que les crises épileptiques sont plus susceptibles de se produire dans les périodes les plus rapprochées depuis la dernière crise. Les durées plus longues entre les crises sont associées à une diminution du nombre de crises. Cela suggère qu'il existe une corrélation entre la fréquence des crises et la durée écoulée depuis la dernière crise.

Cette compréhension des schémas de crises épileptiques en fonction de la durée D peut être utile pour la prédiction et le traitement des épisodes à l'avenir. Il est important de surveiller de près les

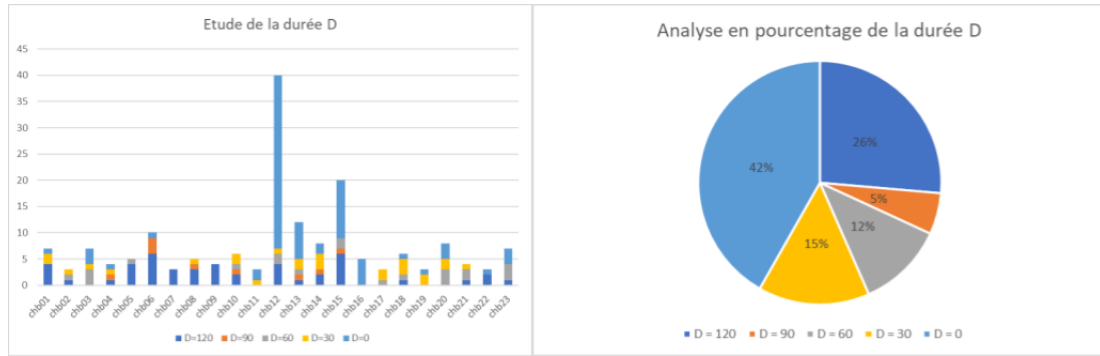


FIGURE 2 – Analyse des différentes durées D

patients ayant des durées D plus courtes, car ils sont plus susceptibles de subir des crises récurrentes. Les informations présentées dans ce tableau fournissent des indications précieuses pour la gestion de l'épilepsie et la mise en place de mesures préventives adaptées à chaque patient.

Dans le cadre de ce projet, les crises ayant une durée D de 2 heures sont considérées comme les premières crises, tandis que les autres crises sont considérées comme des répliques.

3.3 Durée S

Nous présentons ici une analyse approfondie des crises d'épilepsie en fonction de la durée de séparation (durée S) entre les épisodes. Les données fournies incluent le nombre de crises et le pourcentage correspondant pour chaque durée de séparation spécifiée. Les résultats de l'analyse des crises en fonction de la durée S sont présentés dans la figure 3.

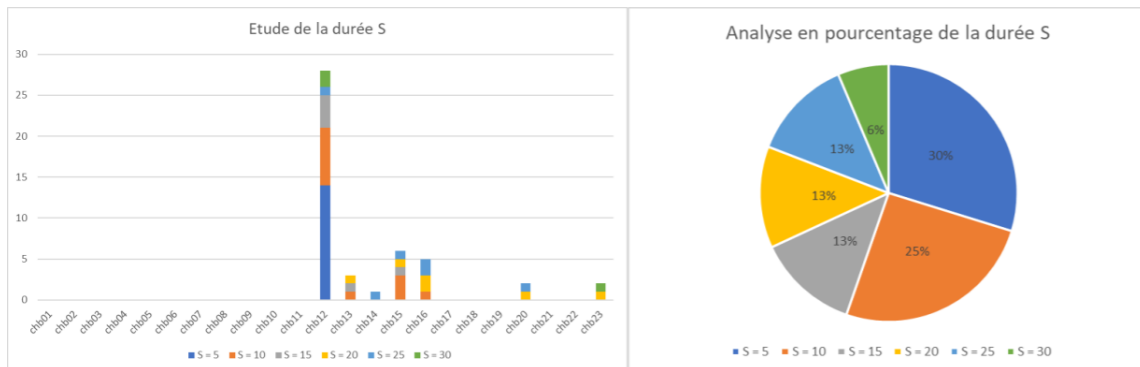


FIGURE 3 – Analyse des différentes durées S

La durée S représente l'intervalle de temps minimum entre deux crises pour qu'elles soient considérées comme distinctes. L'analyse de ce tableau permet de mieux comprendre la répartition des crises en fonction de cette durée de séparation.

En examinant les résultats, on constate que la durée de séparation la plus fréquemment observée est de S = 5 minutes, avec 14 crises représentant 30% de l'ensemble des crises. Cela suggère que de nombreux patients ont des crises qui surviennent dans des intervalles de temps très rapprochés.

Les durées de séparation de S = 10, 15, 20 et 25 minutes ont des proportions similaires, représentant chacune environ 13% des crises. Cela indique qu'un nombre considérable de patients connaissent des épisodes qui se produisent à des intervalles légèrement plus espacés, mais toujours relativement rapprochés.

La durée de séparation la plus longue répertoriée dans le tableau est de S = 30 minutes, avec seulement 3 crises, représentant 6% de l'ensemble des crises. Cela suggère que les intervalles de temps plus longs entre les crises sont moins fréquents, mais ils ne peuvent pas être négligés pour certains patients.

En résumé, cette analyse met en évidence une variabilité des durées de séparation entre les crises d'épilepsie chez les patients étudiés. La plupart des crises surviennent dans des intervalles de temps très

courts ($S = 5$ minutes), tandis que les intervalles légèrement plus longs ($S = 10, 15, 20$ et 25 minutes) sont également fréquents.

Dans le cadre de ce projet, les crises espacées de moins de 5 minutes sont considérées comme identiques.

4 Sélection des données

En fonction des critères établis précédemment, c'est-à-dire une durée D de 2 heures et une durée S de 5 minutes, nous avons sélectionné les données correspondantes. Nous avons commencé par travailler sur les premières crises, ce qui représente 48 crises. La proportion du jeu de données utilisées est illustrée en figure 4.

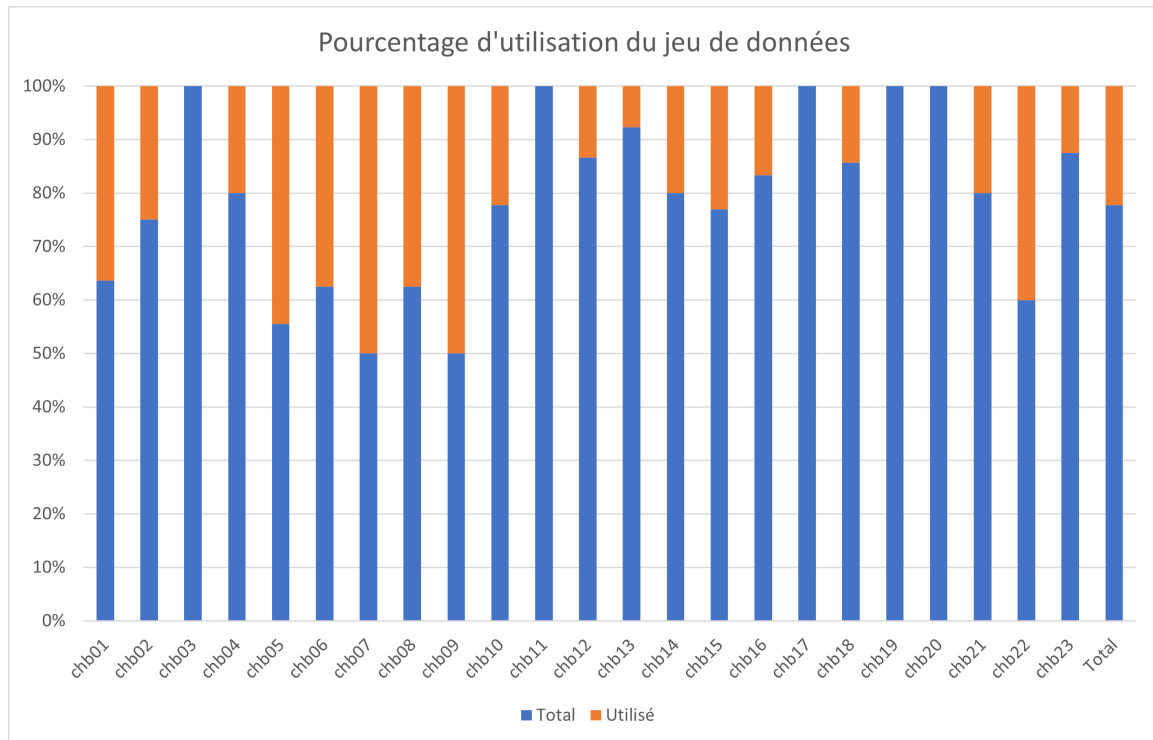


FIGURE 4 – Proportion des crises sélectionnées

Cette sélection de données nous permet d'obtenir un équilibre entre le nombre de crises prises en compte et la durée de séparation nécessaire pour éviter la fusion de deux crises en une seule.

Cette approche de sélection de données nous permettra de construire un modèle de prédiction plus précis et adapté au contexte des patients souffrant d'épilepsie. De plus, cette approche nous permettra d'optimiser les ressources utilisées en évitant de traiter des données inutiles ou non pertinentes pour le problème de prédiction.

Le code utilisé pour cette étape est disponible sur le GitHub de l'étape d'étude des données.

5 Conclusion

En conclusion, ce projet a pour objectif de développer un modèle de prédiction de crises d'épilepsie à partir des données EEG de la base de données CHBMIT. L'analyse des données a permis de déterminer deux paramètres importants pour la prédiction des crises : la durée D , qui permet de distinguer les premières crises des répliques, et la durée S , qui permet de déterminer si deux crises successives doivent être considérées comme une seule ou non.