Université du Québec
Institut national de la recherche scientifique
Centre énergie matériaux télécommunications

CLASSIFICATION DES DONNÉES EEG PAR APPRENTISSAGE PROFOND POUR LA PRÉDICTION DES CRISES D'ÉPILEPSIE

Par

Imene Jemal

Thèse présentée pour l'obtention du grade de
Doctorat en philosophie, Ph.D.
en télécommunications

Jury d'évaluation

Examinateur externe	Mustapha, Kardouchi Université de Moncton
Examinateur externe	Sid Ahmed, Selouani Université de Moncton
Examinateur interne et président de jury	Serioja Ovidiu, Tatu INRS-EMT
Directeur de recherche	Amar Mitiche INRS-EMT
Codirectrice de recherche	Neila Mezghani TELUQ

©Imene Jemal, 2023

Remerciements

C'est avec grande reconnaissance que je réserve cette page pour tous ceux qui, directement ou indirectement, ont contribué à la réalisation de cette thèse.

Je tiens à exprimer mes sincères remerciements à mon directeur de thèse, Amar MITICHE, pour son dévouement, son écoute attentive, ses conseils précieux et son soutien tout au long de ces quatre années de recherches.

Je remercie également ma co-directrice de thèse, Neila MEZGHANI, pour son soutien, sa disponibilité, sa patience et ses encouragements. Sa participation constante a grandement contribué au succès de ce travail.

Je tiens également à remercier les membres du jury pour avoir accepté d'évaluer cette thèse, ainsi que pour leur intérêt porté à mes travaux.

Je remercie mes collègues, Lina ABOU-ABBAS et Khadidja HENNI, pour les discussions sur ma recherche et la collaboration que nous avons eue au fil des ans. Je leur suis également reconnaissant pour leur soutien, leur appuis et leur encouragement continu. Ce fut une belle une expérience enrichissante à la fois personnelle et professionnelle.

Je tiens à remercier tout particulièrement mes parents pour leur soutien, leur confiance et leur amour inconditionnel pendant les moments difficiles de ma thèse, ainsi que pour leurs sacrifices au fil des années d'études.

Je remercie également mes sœurs, mon frère et mes amis pour leur motivation, leur réconfort et leurs discussions passionnées, malgré les distances et les absences prolongées.

Je souhaite que tous ceux qui m'ont soutenue puissent trouver dans ce travail le fruit de leurs efforts et l'expression de ma profonde gratitude.

Résumé

L'épilepsie est une maladie chronique qui se caractérise par des crises imprévisibles et récurrentes. Bien que l'approche principale pour traiter cette maladie soit à travers une médication à long terme, environ un tiers des patients sont résistants à cette thérapie. De plus, les options chirurgicales pour l'épilepsie sont limitées en raison des perceptions négatives liées à la chirurgie, des préoccupations concernant les complications subséquentes et des taux de réussite modérés. Actuellement, les recherches se concentrent de plus en plus sur l'exploration du potentiel de la prédiction des crises à l'aide d'enregistrements électro-encéphalogrammes (EEG), ce qui pourrait ouvrir de nouvelles voies d'intervention et de traitement.

Bien que de nombreux modèles de prédiction des crises épileptiques aient été proposés dans la littérature scientifique, ils ne sont pas encore largement utilisés dans des contextes cliniques en raison de plusieurs limitations. Ces limitations incluent l'efficacité des caractéristiques extraites à partir des signaux EEG, les défis d'interprétation des modèles de prédiction et la difficulté de généraliser les modèles à de nouveaux patients. L'objectif de cette thèse est de développer et d'améliorer des modèles pour la prédiction des crises d'épilepsie en utilisant des données EEG.

Dans un premier temps, nous avons mené une étude pour évaluer la complexité des caractéristiques extraites à partir des données EEG qui sont utilisées pour la prédiction des crises d'épilepsie. Nous avons déterminé que des méthodes d'apprentissage profond complexes étaient nécessaires pour extraire des caractéristiques plus précises et robustes que celles obtenues manuellement. Ensuite, nous avons proposé une architecture de réseau de neurones interprétable conçue pour la prédiction de crises spécifiques au sujet. Cette architecture a non seulement mieux performé que les méthodes actuelles de l'état de l'art, mais elle a également amélioré et simplifié l'interprétabilité du modèle.

Pour résoudre le problème de la généralisation à de nouveaux patients, nous avons étudié les modèles multi-sujets et inter-sujets, qui offrent un niveau supérieur de généralisation et qui sont plus utiles à implémenter dans des situations réelles. L'évaluation de notre modèle multi-sujets a montré des résultats supérieurs à ceux des travaux antérieurs évalués sur les mêmes bases de données. Bien que les performances aient diminué avec le modèle inter-sujets, nous avons intégré des méthodes d'adaptation au domaine qui ont considérablement amélioré les performances du modèle.

En conclusion, cette thèse a apporté des contributions significatives pour la prédiction des crises d'épilepsie basées sur l'apprentissage profond à l'aide d'enregistrements EEG. Cependant, plus de travail est nécessaire pour améliorer la généralisation du modèle de prédiction avant la traduction de ces approches en dispositifs commerciaux.

Mots-clés : Prédiction des crises d'épilepsie, Électroencéphalographie; Apprentissage profond; Réseaux de neurone; Interprétation des réseaux de neurone; IA explicable; Adaptation du domaine; Généralisation du modèle.

Abstract

Epilepsy is a chronic illness characterized by the occurrence of repeated unpredictable seizures. Although the primary approach to treating it is through long-term medication, around one-third of patients are resistant to this therapy. Moreover, surgical options for epilepsy are limited due to negative perceptions associated with surgery, concerns regarding complications, and moderate success rates. Research is currently focused on exploring the potential of seizure prediction using electroencephalogram (EEG) recordings, which could open up new intervention avenues.

Although numerous models for predicting epileptic seizures have been proposed in literature, they are not yet widely used in clinical settings due to several limitations. These limitations include the effectiveness of features extracted from EEG signals, the challenges of interpreting prediction models, and the difficulty of generalizing models to new patients. The aim of this thesis is to develop and improve accurate models for predicting epilepsy seizures using EEG data.

Firstly, we conducted a study to evaluate the complexity of EEG features in predicting epilepsy seizures. We determined that complex deep learning methods were necessary to extract more precise and robust features than those obtained manually. Next, we proposed an interpretable neural network architecture designed for patient-specific seizure prediction. This architecture not only outperformed current state-of-the-art methods but also improved and simplified model interpretability.

To address the problem of generalizing to new patients, we studied multi-subject and inter-subject models, which offer a higher level of generalization and are more applicable in real-life situations. The evaluation of our multi-patient model showed superior results compared to previous works evaluated on the same datasets. Although performance decreased with the inter-subject model, we integrated domain adaptation methods that significantly improved the model's performance.

In conclusion, this thesis has made significant contributions to predicting epilepsy seizures based on deep learning using EEG recordings. However, more work is needed to improve the generalization of the prediction model before translating these approaches into commercial devices.

Keywords : Seizure prediction; Electroencephalography; Deep learning; Neural networks; Neural network interpretability; Explainable AI; Domain adaptation; Model generalization.

Table des matières

Liste des figures

Liste des tableaux

Liste des abréviations

AI	Artificial intelligence
AUC	Area Under the ROC Curve
CDAN	Domain Adversarial Conditional Adaptation
CNN	Convolutional neural network
DANN	Discriminative Adversarial Neural Network
DL	Deep Learning
EEG	Electroencephalography
FBCSP	Filter Bank Common Spatial Pattern
FFT	Fast Fourier Transform
FPR	False Positive Rate
LOPO	Leave-One-Patient-Out
LRP	Layer-wise Relevance Propagation
LSTM	Long short-term memory
NN	Neural Network
OMS	Organisation Mondiale de la Santé
RF	Random Forest
SVM	Support Vector Machine

Chapitre 1

Introduction

1.1 Contexte de la thèse

L'épilepsie est une maladie neurologique fréquente qui touche des millions de personnes dans le monde entier. Elle se manifeste par des crises récurrentes et imprévues causées par des décharges neuronales anormales et excessives dans le cerveau. Au Canada, environ 300 000 personnes en sont atteintes, tandis qu'au niveau mondial, l'organisation mondiale de la santé estime que 70 millions de personnes souffrent d'épilepsie who (2019).

Les crises d'épilepsie sont le résultat d'un dysfonctionnement cérébral focal ou généralisé, dont les symptômes varient considérablement d'une personne à l'autre, allant d'un court arrêt des activités de la personne à une altération de la conscience et à des crises violentes et incontrôlées. Pour diagnostiquer l'épilepsie, les neurologues utilisent généralement l'électroencéphalographie (EEG). L'EEG enregistre l'activité électrique du cerveau à l'aide des électrodes placées sur le crâne, tel qu'illustré dans la Figure 1.1 Klem (1999). L'électroencéphalographie offre une bonne résolution temporelle et est non invasive et peu coûteuse. Les enregistrements EEG sont ensuite analysés par des spécialistes pour détecter les crises, connaître leurs types et localiser la région épileptique.

L'inspection visuelle des heures ou des jours de données EEG est un processus long, épuisant qui nécessite la présence d'un expert. C'est pourquoi de nombreuses recherches ont été menées pour développer un système automatique de détection des crises de l'épilepsie à partir des signaux EEG. Les systèmes de détection d'épilepsie permettent l'identification des patterns EEG typiques liés à la

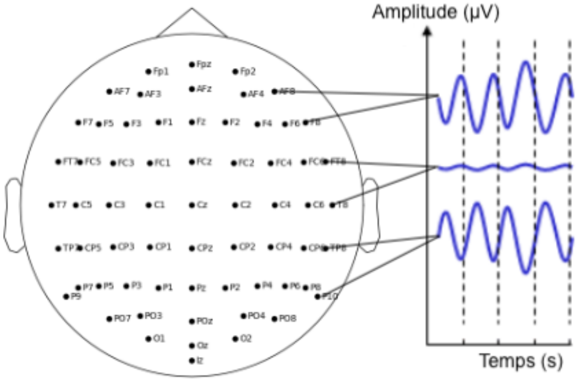


Figure 1.1 – Placement des électrodes pour l'enregistrement des signaux EEG.

crise et fournissent aux cliniciens des données détaillées utiles pour le traitement de l'épilepsie. Cependant, étant donné l'impact physique et psychologique des crises sur les patients, l'identification précoce d'une crise, permettant de prendre des mesures préventives, est cruciale. Alors que la détection consiste à identifier les crises en cours, la prédiction permet d'avertir suffisamment à l'avance de l'apparition d'une crise. Dans cette thèse, le principal objectif sera la prédiction des crises. Selon plusieurs études, il existe une période pré-ictale (pré-critique) de l'ordre d'une dizaine de minutes qui précède la survenue d'une crise Mormann *et al.* (2005); Gadhoumi *et al.* (2016). Pendant cette période, les enregistrements EEG présentent des motifs différents des patterns de la crise et aussi des périodes précédentes, appelé inter-ictale (inter-critique) (Figure 1.2). Par conséquent, prédire une crise d'épilepsie revient à identifier cette phase pré-ictale. Bien que différents paradigmes tels que la théorie des modèles autorégressifs et la méthode de variation des caractéristiques ont été utilisés pour prédire l'épilepsie, les principales méthodes de la littérature sont basées sur la reconnaissance des formes Duda *et al.* (2012); Bishop (2006). Ces méthodes consiste à extraire d'un signal EEG, en général des fragments de signal de quelques secondes, un certain nombre de caractéristiques pertinentes permettant de déterminer à quelle catégorie, inter-ictale ou pré-ictale, appartient une mesure observée. Ainsi, ces méthodes sont basées sur la classification des états inter-ictal et pré-ictal à partir des signaux EEG pour la prédiction de l'épilepsie. Dans cette thèse, l'objectif est de prédire les crises d'épilepsie par la classification des états inter-ictaux et pré-ictaux en utilisant l'apprentissage profond, un sous-domaine de la reconnaissance de formes.

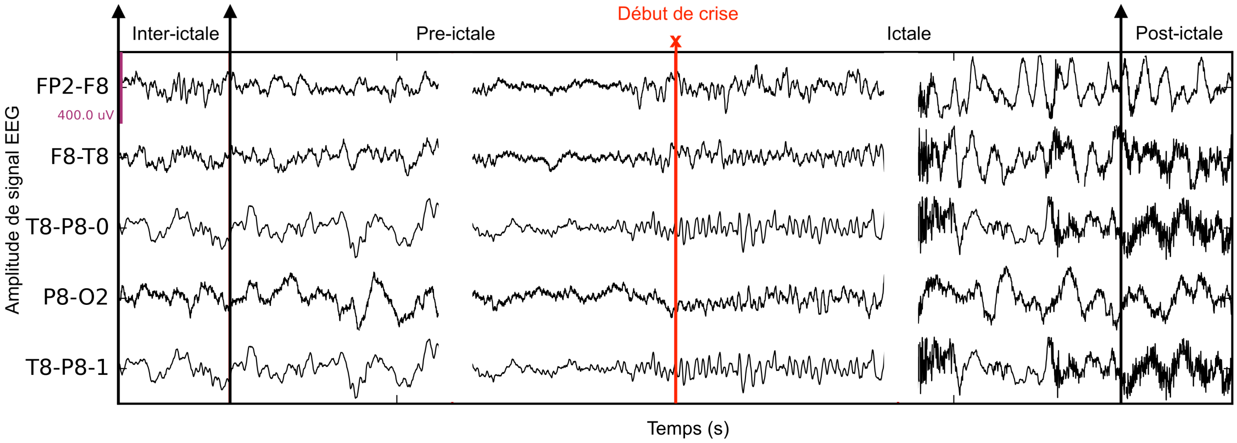


Figure 1.2 – Cinq canaux d'enregistrement EEG illustrant les différents états cérébraux au cours de la transition de passage d'un état normal à un état de crise.

1.2 Problématiques soulevées dans la littérature

Malgré les avancées réalisées dans la prédiction de l'épilepsie au cours des dernières années, il reste encore des défis à relever pour mettre en place des méthodes de prédiction automatisées dans des applications réelles.

— Problématique 1 : Efficacité des caractéristiques extraites des signaux EEG
  Les approches traditionnelles utilisées pour la prédiction des crises d'épilepsie se basent sur l'extraction de caractéristiques à partir des signaux EEG pour différencier les états inter-ictal et pré-ictal , afin de prédire une crise imminente. Dans la littérature, de nombreuses caractéristiques ont été étudiées pour l'analyse de signaux, telles que les caractéristiques statistiques, les caractéristiques extraites du domaine temporel ou fréquentiel, ainsi que celles issues de la théorie des systèmes dynamiques (voir Annexe A). Cependant, en raison de la non-linéarité et de la non-stationnarité des données EEG, ces caractéristiques sont peu efficace face à la variation des patterns de crises qui dépendent de la sévérité de la manifestation de la crise, de la zone cérébrale impliquée et du type d'épilepsie. Des méthodes plus complexes, telles que l'apprentissage profonds, peuvent être utilisées pour extraire des caractéristiques plus robustes et plus discriminantes à partir des signaux EEG.

— Problématique 2 : L'effet "boîte noire" des réseaux de neurones
  Les méthodes d'apprentissage profond telles que les réseaux de neurones profonds, permettent d'extraire automatiquement des caractéristiques pertinentes à partir de données brutes. Ces

caractéristiques sont souvent plus discriminantes et robustes que celles extraites manuellement. Toutefois, l'utilisation de ces réseaux de neurones, souvent considérés comme des boîtes noires, dans le domaine biomédical est encore limitée en raison de la complexité de leurs architectures internes et de la difficulté à comprendre leurs prédictions. Cette opacité des réseaux de neurones pose un défi pour les applications biomédicales, où les résultats doivent être interprétables pour les cliniciens. En effet, les cliniciens ont besoin de comprendre comment les prédictions sont faites et quelles caractéristiques des données contribuent le plus à la prédiction. Cela est particulièrement crucial pour l'épilepsie, où la fiabilité et l'interprétabilité des résultats sont essentielles pour une utilisation clinique.

— Problématique 3 : Généralisation des modèles de prédiction de l'épilepsie à de nouveaux patients

  Un défi majeur dans la prédiction des crises d'épilepsie est la généralisation des modèles à de nouveaux patients. Il existe trois types de modèles de prédiction : (1) les modèles spécifiques au sujet, qui sont conçus pour chaque patient en utilisant une partie de ses données pour l'apprentissage et le reste pour l'évaluation, mais leur utilisation est limitée par la quantité de données disponibles ; (2) les modèles multi-sujets , également appelés modèles indépendants du patient, qui sont appliqués à un ensemble spécifique de patients et ne sont pas limités par la quantité de données, mais posent le défi de l'adaptation à de nouveaux patients ; et (3) les modèles inter-sujets , également appelés modèles généralisés, qui utilisent des données de plusieurs patients et peuvent être appliqués à de nouveaux patients, mais qui sont beaucoup plus complexes. Cependant, la plupart des algorithmes de prédiction de la littérature sont spécifiques à un sujet en raison de la grande variabilité des données EEG entre les patients. Ces modèles sont généralement configurés et paramétrés sur de petits ensembles de données provenant d'un seul patient, ce qui limite leur capacité à se généraliser à de nouveaux patients. En outre, ces modèles nécessitent souvent la collecte de données de crises afin de prédire d'autres crises futures, ce qui peut être difficile pour les patients qui ont des crises peu fréquentes. Les modèles multi-sujets ont été peu explorés dans la littérature, tandis que, à notre connaissance, les modèles inter-sujets n'ont pas encore été étudiés. Cela souligne l'importance de relever le défi majeur du développement de modèles de prédiction de l'épilepsie permettant une meilleure généralisation à de nouveaux patients.

1.3 Objectifs de la thèse

L'objectif de cette thèse est de développer et améliorer des modèles de prédiction des crises d'épilepsie basés sur des enregistrements EEG. Pour atteindre cet objectif, plusieurs objectifs spécifiques ont été identifiés en réponse aux défis relevés dans la littérature.

1. Le premier objectif est d'évaluer l'efficacité des caractéristiques extraites des signaux EEG pour distinguer les états inter-ictal et pre-ictal, en vue de prédire les crises d'épilepsie. De plus, nous visons à évaluer la variabilité de ces caractéristiques entre les patients. L'analyse de ces aspects servira de base pour orienter les choix de conception des modèles de prédiction futurs.
2. Le deuxième objectif de cette thèse est d'aborder le problème de la "boîte noire" des réseaux de neurones utilisés pour la prédiction des crises d'épilepsie. Nous cherchons à concevoir une architecture de réseau de neurones interprétable pour la prédiction des crises basé sur les signaux EEG. Cela permettra d'une part de mieux comprendre les mécanismes sous-jacents de l'épilepsie, ce qui pourrait être très utile pour la mise au point clinique et l'amélioration des traitements. D'autre part, cela permettra également d'identifier les caractéristiques pertinentes qui contribuent à la prédiction des crises, offrant ainsi des résultats plus fiables.
3. Le troisième objectif consiste à étudier les modèles de prédiction de crises d'épilepsie multi-sujets et inter-sujets. Contrairement aux modèles spécifiques au sujet, ces modèles permettront une meilleure généralisation à de nouveaux patients. Cela permettra ainsi de maximiser l'utilisation des données disponibles pour obtenir des résultats fiables et précis pour un plus grand nombre de patients, tout en évitant la nécessité de recueillir des données spécifiques pour chaque patient individuellement.

1.4 Méthodologie de recherche

Afin d'atteindre les trois objectifs énoncés, nous avons adopté la méthodologie décrite dans la suite. D'abord, pour atteindre notre premier objectif, la base de données publique CHB-MIT, contenant des enregistrements EEG de 23 patients épileptiques, a été utilisée. Des caractéristiques linéaires et non linéaires des signaux EEG, soit unimodales (calculées à partir d'un seul canal) soit multimodales (extraites à partir de plusieurs canaux), généralement utilisées pour prédire une crise d'épilepsie ont été extraites. Ces caractéristiques comprenaient des caractéristiques statistiques, temporelles et fréquentielles, ainsi que des caractéristiques issues de la théorie des systèmes dynamiques. La complexité de ces caractéristiques a été étudiée en utilisant plusieurs mesures, y compris le ratio discriminant de Fisher, le volume de chevauchement et l'efficacité individuelle de chaque caractéristique. Enfin, ces mesures ont été comparées à celles obtenues à partir d'autres ensembles de données synthétiques et réelles.

Pour atteindre le deuxième objectif, une architecture de réseau de neurones interprétable pour la prédiction de l'épilepsie a été proposée, dont les couches ont été expliquées en termes de traitements classiques du signal, tels que les filtres passe-bande de fréquence et les filtres spatiaux. Cette architecture a été inspirée par l'algorithme Filter Bank Common Spatial Pattern (FBCSP). En outre, des méthodes d'interprétation des réseaux de neurones et d'explication des décisions de réseaux ont été explorées, notamment la visualisation des filtres appris et la technique de propagation de pertinence couche par couche (Layer-wise Relevance Propagation).

Enfin, pour atteindre le troisième objectif, des modèles de prédiction multi-sujets et inter-sujets ont été développés pour obtenir des prédictions précises et fiables pour un groupe plus large de patients. Afin d'améliorer la généralisation du modèle de prédiction à de nouveaux patients, des techniques d'adaptation de domaine ont été utilisées pour maximiser l'utilisation des données disponibles et éviter la nécessité de collecter des données spécifiques pour chaque patient individuellement.

En conclusion, la méthodologie employée dans cette thèse a permis d'atteindre les objectifs de recherche et d'obtenir des résultats significatifs dans le domaine de la prédiction des crises d'épilepsie.

1.5 Contributions de la thèse

Dans cette section, nous présentons un résumé des principales contributions de cette thèse, organisées par chapitre.

— **Chapitre 3 - Article 1 : A study of EEG feature complexity in epileptic seizure prediction [P1]** Cette étude a pour objectif d'évaluer la complexité des caractéristiques extraites des signaux EEG couramment utilisés pour la prédiction des crises épileptiques, ainsi que la variabilité inter-sujets de ces caractéristiques. En outre, elle met en évidence le défi de la variabilité importante entre les données EEG de différents patients, ce qui rend la prédiction des crises épileptiques plus complexe. En conséquence, la grande complexité des caractéristiques extraites des signaux EEG nous a orienté vers l'utilisation d'algorithmes avancés, tels que l'apprentissage profond, pour prédire efficacement les crises épileptiques à partir de données EEG.

— **Chapitre 4 - Article 2 : An interpretable deep learning classifier for epileptic seizure prediction using EEG data** [**P2**] Cette étude propose une architecture de réseau de neurones interprétable pour la prédiction des crises épileptiques dont les couches ont été expliquées en termes de traitements classiques du signal, tels que les filtres passe-bande de fréquence et les filtres spatiaux. Des techniques d'interprétation supplémentaires ont également été utilisées pour révéler les caractéristiques pertinentes qui peuvent expliquer les décisions de prédiction. Cette approche a permis d'améliorer les résultats par rapport à l'état de l'art tout en simplifiant l'interprétation do modèle et l'explication de ses décisions.

— **Chapitre 5 - Article 3 : Domain adaptation for cross-subject EEG-based seizure prediction** [**P6**] Dans cette étude, nous avons exploré différents modèles de prédiction de crises épileptiques, notamment les modèles multi-sujets, également appelés modèles indépendants du patient, et les modèles inter-sujets. Nous avons proposé un modèle de prédiction multi-sujets en utilisant une architecture d'apprentissage profond développée précédemment par notre équipe. Ce modèle nous a permis d'obtenir de meilleurs résultats que les travaux existants évalués sur les mêmes bases de données. Nous avons également proposé une modélisation inter-sujets pour permettre la généralisation à de nouveaux patients. Enfin, nous avons étudié trois méthodes d'adaptation au domaine pour améliorer les performances du modèle inter-sujets et garantir sa généralisation à de nouveaux patients.

— **Annexe B - Article 4 : An effective deep neural network architecture for cross-subject epileptic seizure detection in EEG data** [**P4**] Dans cette étude, l'intérêt est porté sur la détection plutôt que la prédiction, comme cela a été étudié dans d'autres chapitres de la thèse. Bien que cette tâche soit considérée comme plus facile que la prédiction dans un certain sens, elle n'en demeure pas moins importante. Cette étude propose une architecture modifiée de réseau de neurones convolutionnel (CNN) basée sur une convolution séparable en profondeur (separable depth-wise convolution) pour détecter efficacement et automatiquement les crises d'épilepsie. L'architecture a été conçue avec un nombre limité de paramètres pour réduire la complexité du modèle et les exigences de stockage, ce qui la rend facilement déployable dans un dispositif connecté pour une détection en temps réel des crises d'épilepsie.

1.6 Liste de publications

1.6.1 Articles dans une revue internationale avec comité de lecture - Papiers réguliers

[P1] **Imene Jemal**, Neila Mezghani, Lina Abou-Abbas, and Amar Mitiche. "An interpretable deep learning classifier for epileptic seizure prediction using EEG data." IEEE Access (2022).

[P2] **Imene Jemal**, Amar Mitiche, and Neila Mezghani. "A study of eeg feature complexity in epileptic seizure prediction." Applied Sciences 11, no. 4 (2021): 1579.

[P3] Lina Abou-Abbas, **Imene Jemal**, Khadidja Henni, Youssef Ouakrim, Amar Mitiche, and Neila Mezghani. "EEG Oscillatory Power and Complexity for Epileptic Seizure Detection." Applied Sciences 12, no. 9 (2022): 4181.

1.6.2 Communications internationales avec actes - Conférences

[P4] **Imene Jemal**, Amar Mitiche, Lina Abou-Abbas, Khadidja Henni, and Neila Mezghani. "An Effective Deep Neural Network Architecture for Cross-Subject Epileptic Seizure Detection in EEG Data." In Proceedings of CECNet 2021, pp. 54-62. IOS Press, 2021.

[P5] Lina Abou-Abbas, **Imene Jemal**, Khadidja Henni, Amar Mitiche, and Neila Mezghani. "Focal and Generalized Seizures Distinction by Rebalancing Class Data and Random Forest Classification." In International Conference on Bioengineering and Biomedical Signal and Image Processing, pp. 63-70. Springer, Cham, 2021.

1.6.3 Articles en finalisation

[P6] **Imene Jemal**, Amar Mitiche, Lina Abou-Abbas, Khadidja Henni, and Neila Mezghani. "Domain adaptation for cross-subject EEG-based seizure prediction" Neural Computing and Applications, 2023.

[P7] Lina Abou-Abbas, Khadidja Henni, **Imene Jemal**, Amar Mitiche and Neila Mezghani, "Patient-Independent epileptic seizure detection by stable feature selection", Expert Systems With Applications, 2023.

[P8] Khadidja Henni, Lina Abou-Abbas, **Imene Jemal**, Amar Mitiche and Neila Mezghani, "Imbalance-aware Machine Learning for Epileptic Seizure Detection", ICCSEA 2023.

1.7 Organisation de la thèse

Cette thèse est une thèse par articles. Elle est structurée comme suit: Le chapitre 1 décrit le contexte et les problématiques de la recherche ainsi que les différentes contributions. Le chapitre 2 présente une revue de la littérature des concepts abordés dans cette thèse. Le chapitre 3 propose une analyse approfondie de la complexité des données EEG utilisées pour la prédiction de l'épilepsie, ainsi qu'une évaluation de la variabilité de ces données entre les différents sujets. Ces résultats serviront de base pour orienter les choix de conception des modèles de prédiction ultérieurs. Le chapitre 4 présente une architecture interprétable proposée pour la prédiction de l'épilepsie pour des modèles spécifiques au sujet. Cette étude vise à résoudre la difficulté d'interpréter les réseaux de neurones. Des interprétations des filtres appris et des explications de la décision du réseau de neurones pour certains exemples ont également été ajoutées. Par la suite, dans le chapitre 5, nous avons exploré la question de la généralisation des modèles de prédiction en utilisant la même architecture pour une modélisation multi-sujets permettant ainsi une généralisation du modèle à l'ensemble des patients de la base de données. Nous avons également utilisé l'adaptation au domaine pour développer un modèle inter-sujets, permettant ainsi une généralisation du modèle à de nouveaux patients. Enfin, le chapitre 6 résume les conclusions générales de cette thèse, les limites des contributions décrites et présente quelques idées pour les travaux futurs liés à chaque contribution.

Chapitre 2

Revue de littérature

2.1 Introduction

Ce chapitre présente une revue de la littérature des concepts abordés dans cette thèse. Nous commençons par une revue des méthodes qui ont été développées pour prédire les crises d'épilepsie en utilisant l'apprentissage profond sur des signaux EEG. Ensuite, nous présentons les concepts d'interprétation et d'explication des réseaux de neurones, en décrivant quelques méthodes d'interprétation et d'explication et leurs applications sur des données EEG. Enfin nous expliquons également le contexte de l'adaptation de domaine et son utilisation dans différentes applications en utilisant les données EEG.

2.2 Système automatique de prédiction des crises d'épilepsie

Bien que différentes approches, telles que la théorie des modèles autorégressifs et la méthode de variation des caractéristiques, ont été utilisées pour prédire automatiquement l'épilepsie, les principales méthodes de la littérature sont basées sur la reconnaissance des formes Duda *et al.* (2012); Bishop (2006). Cette méthode consiste à extraire d'un signal EEG, généralement des fragments de quelques secondes, un certain nombre de caractéristiques pertinentes permettant de déterminer à quelle catégorie, inter-ictale ou pré-ictale, appartient une mesure observée. Ainsi, ces méthodes se basent sur la classification des états inter-ictal et pré-ictal à partir des signaux EEG pour la prédiction de l'épilepsie. Il existe deux approches principales de reconnaissance de formes : les approches traditionnelles et les approches d'apprentissage profond, comme indiqué dans la figure 2.1.

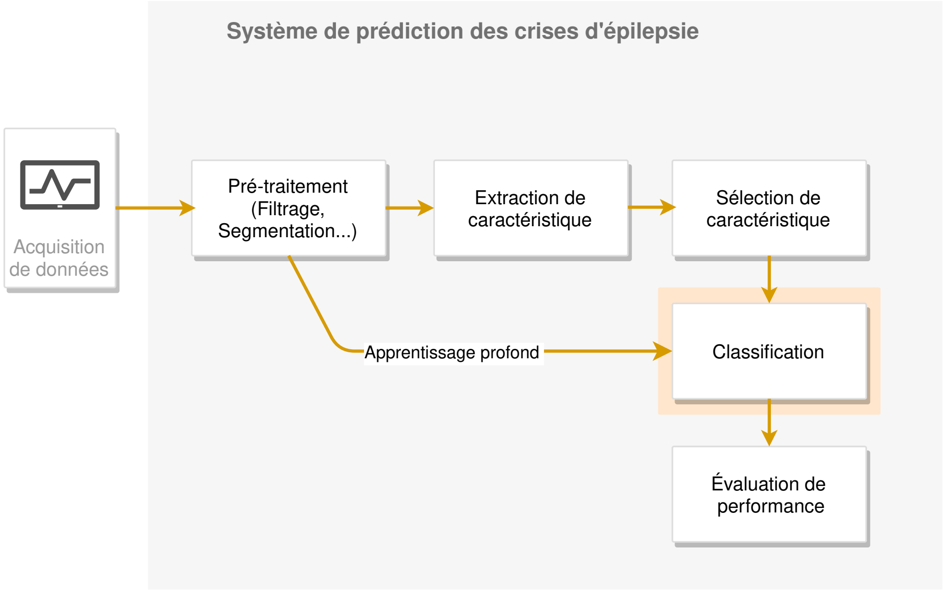

Système de prédiction des crises d'épilepsie
Acquisition de données
Pré-traitement (Filtrage, Segmentation...)
Extraction de caractéristique
Sélection de caractéristique
Apprentissage profond
Classification
Évaluation de performance

Figure 2.1 – Schéma typique d'un système automatique de prédiction des crises d'épilepsie.

Les approches traditionnelles impliquent généralement quatre étapes distinctes : le pré-traitement des signaux EEG pour éliminer les bruits et les artefacts, l'extraction et la sélection des caractéristiques pertinentes, la classification des états inter-ictaux et pré-ictaux à l'aide d'algorithmes d'apprentissage automatique, et l'évaluation des performances pour tester l'efficacité de la méthode. Dans certaines études, une phase de régularisation est également incluse pour réduire le nombre de faux alarmes Assi *et al.* (2017).

Les approches d'apprentissage profond, en revanche, combinent toutes les étapes entre les données d'entrée et les résultats de sortie en un seul processus. Ces techniques automatisent l'extraction et la sélection des caractéristiques à partir des données brutes. Les différentes techniques d'apprentissage profond telles que les réseaux de neurones profonds, apprennent différents niveaux de représentation des données qui codent des caractéristiques abstraites représentatives Goodfellow *et al.* (2016).

2.3 Revue des méthodes de prédiction des crises d'épilepsie utilisant l'apprentissage profond

On peut distinguer deux grandes catégories de modèles existants de prédiction de crises d'épilepsie : les modèles de prédiction spécifiques au sujet, qui sont conçus pour s'adapter à chaque patient individuellement, et les modèles de prédiction multi-sujets, qui sont conçus pour s'appliquer à l'ensemble des patients. Les modèles spécifiques au sujet ont l'avantage de prendre en compte les caractéristiques individuelles du patient, mais ils nécessitent une collecte de données plus importante pour chaque patient et sont, par conséquent, plus coûteux à mettre en place. Les modèles multi-sujets, en revanche, peuvent être plus faciles à déployer et nécessitent moins de données par patient, mais ils peuvent manquer de précision en raison de la variabilité des données entre les patients.

2.3.1 Modèle de prédiction des crises d'épilepsie spécifiques au sujet

Au cours des dernières années, l'application de l'apprentissage profonds pour prédire les crises épileptiques a suscité un intérêt croissant. Par exemple, Tsiouris *et al.* (2018) a mené une étude utilisant un réseau Long Short-Term Memory (LSTM) profond qui a été entraîné avec diverses caractéristiques extraites des données EEG telles que des caractéristiques statistiques, des caractéristiques basées sur la puissance spectrale et la théorie des graphes. Cette méthode a été évaluée sur l'ensemble de données CHB-MIT et a rapporté des résultats impressionnants, avec une sensibilité et une spécificité de 99,84% et 99,86%, respectivement.

D'autres études ont utilisé des réseaux de neurones à convolution (CNN) pour prédire les crises, telles que Truong *et al.* (2018), qui a proposé une architecture CNN à trois couches entraînée sur des représentations de spectrogrammes de données EEG. Leur modèle a atteint une sensibilité de 81,4% et un taux de faux positifs de 0,16/h lorsqu'il a été testé sur 13 patients de l'ensemble de données CHB-MIT.

Dans Zhang *et al.* (2019b), les auteurs ont utilisé l'algorithme Filter Bank Common Spatial Pattern (FBCSP) pour extraire des caractéristiques pertinentes des données EEG, qui ont ensuite été alimentées dans un classificateur basé sur un CNN pour prédire les états de crises. Leur approche a obtenu des scores de précision, de spécificité et de sensibilité de 90%, 92% et 92%, respectivement, avec un taux de fausses alarmes relativement faible de 0,12/h, tel qu'évalué sur la base de données CHB-MIT. Pour résoudre le problème de données limitées, les auteurs ont utilisé une méthode de cutting-splicing d'augmentation de données ansi qu'un réseau antagoniste génératif (generative adversarial network) pour synthétiser de nouvelles données. Cependant, ils ont noté que de telles méthodes d'augmentation de données peuvent augmenter la complexité et le temps d'entraînement de réseau de neurones. Leur classificateur a atteint une précision de 92,2% avec une fenêtre de prédiction anticipée de 30 minutes pour 23 sujets de l'ensemble de données EEG CHB-MIT.

Une autre étude récente, menée par Zhao *et al.* (2020), a proposé une nouvelle architecture CNN unidimensionnelle qui a été directement entraînée sur des données EEG brutes pour prédire les crises. L'étude a obtenu des performances impressionnantes sur les ensembles de données de la Société américaine d'épilepsie (AES) et CHB-MIT, rapportant une aire sous la courbe (AUC) de 0,915, une sensibilité de 89,26% et un taux de fausses prédictions (FPR) de 0,117/h et 0,970, 94,69%, 0,095/h, respectivement.

2.3.2 Modèle de prédiction des crises d'épilepsie multi-sujets

Il y a eu peu de travaux récents sur les de prédiction muti-sujets. L'étude Tsiouris *et al.* (2017) a comparé différents algorithmes de classification pour la prédiction des crises, y compris l'algorithme RIPPER (Repeated Incremental Pruning to Produce Error Reduction), les machines à vecteurs de support et les réseaux de neurones, afin de distinguer les segments EEG pré-ictaux et inter-ictaux pour la prédiction des crises d'épilepsie. En utilisant un nombre équilibré d'enregistrements pré-ictaux et inter-ictaux sélectionnés de chaque patient dans la base de données CHB-MIT, le SVM a donné les meilleurs résultats avec une précision de 68,5%. Des études plus récentes, telles que Khan *et al.* (2017), ont montré des résultats améliorés en utilisant un réseau de neurones convolutif (CNN) sur la transformée en ondelettes des signaux EEG, pour atteindre une sensibilité de 87,8% avec un faible taux de fausses prédictions de 0,142 FP/h. Dans l'étude Dissanayake *et al.* (2021a), une approche d'apprentissage profond multi-tâches a été utilisée pour la prédiction des crises et la classification des patients en utilisant une architecture de réseau Siamese sur l'ensemble de données EEG de la base de données CHB-MIT, ce qui a donné une précision moyenne de 91,54%. Une autre étude, Wu *et al.* (2022), a utilisé la technique de la distillation de connaissances pour transférer des connaissances d'un modèle multi-sujets entraîné sur les données de plusieurs patients à un modèle spécifique au patient entraîné sur les données de ce patient en particulier. Cette approche a conduit à des résultats de prédiction spécifiques au sujet améliorés par rapport à quatre autres méthodes existantes, avec une amélioration moyenne de 3,37% en précision, 2,33% en sensibilité et une réduction des fausses prédictions de 0,044/h en moyenne lorsqu'elle a été testée sur 11 patients de l'ensemble de données CHB-MIT.

2.4 Interprétation et explication des réseaux de neurones

L'intelligence artificielle explicable (IAE) est un sujet de recherche actif depuis plus de cinq décennies et a récemment regagné en popularité. Des années 1970 aux années 1990, les chercheurs ont étudié des systèmes de raisonnement symbolique tels que MYCIN Buchanan & Shortliffe (1984), GUIDON Clancey (1987) et SOPHIE Brown (1982), qui visaient à expliquer leurs processus de raisonnement. Les développeurs du système d'identification antimicrobien MYCIN ont argumenté que l'IA ne pourrait être acceptée par la médecine que si elle fournissait des explications convaincantes. Dans les années 1990, les chercheurs ont commencé à explorer des techniques d'IA opaques, telles que les réseaux de neurones, et ont développé des cartes de saillance pour comprendre et visualiser les non-linéarités inhérentes aux réseaux de neurones. Ces techniques ont également été appliquées à l'imagerie médicale Morch *et al.* (1995). En 2001, Breiman a souligné l'importance de l'interprétabilité, même si le théorème de la "Occam's razor" stipule que les modèles les plus simples sont les meilleurs. Le conflit entre la simplicité (l'interprétabilité) et la précision est évident, car la régression linéaire est un modèle assez interprétable mais moins précis que les réseaux de neurones, surtout pour des données non linéairement séparables Breiman (2001b).

2.4.1 Interprétation des réseaux de neurones

Dans la littérature, l'intelligence artificielle explicable est associée à divers termes souvent utilisés de manière interchangeable, tels que "interpréter" ou "expliquer" un modèle. Selon Montavon *et al.* (2018), l'interprétation consiste à mapper un domaine abstrait (comme un espace vectoriel) vers un domaine interprétable compréhensible par les humains (comme une image). Ainsi, l'interprétation d'un modèle d'apprentissage profond nécessite de comprendre les concepts appris, tels que les activations ou les poids de réseau de neurones. Les techniques d'interprétation des réseaux de neurones comprennent la méthode de maximisation d'activité, qui recherche des motifs activant un neurone spécifique. Une autre technique se concentre sur l'interprétation des poids appris du réseau de neurones, également appelés noyaux ou filtres. Cette approche a été couronnée de succès en vision par ordinateur, produisant des interprétations intéressantes. Par exemple, Hinton *et al.* (2006) a constaté que les filtres de première couche appris par le réseau de neurone profond (Deep Belief Network) pour la reconnaissance de patchs d'images naturelles étaient similaires aux filtres de localisation, d'orientation et de fréquence spatiale, tels que les filtres de Gabor utilisés pour la détection de contours. De plus, Larochelle *et al.* (2009) a observé que le réseau de neurones utilisé sur l'ensemble de données MNIST apprenait des caractéristiques de bas niveau similaires aux détecteurs de traits généralement utilisés pour la localisation de texte. Une autre approche de l'interprétabilité Gilpin *et al.* (2018) consiste à créer un modèle produisant des explications avec une architecture conçue pour simplifier l'interprétation de son traitement et de ses représentations internes.

2.4.2 Explication des décisions d'un réseau de neurones

D'un autre côté, l'explication vise à répondre à la question de pourquoi une entrée spécifique conduit à une décision particulière. Les techniques d'explication identifient l'ensemble de caractéristiques (données d'entrée) qui ont contribué à la décision de classification pour une observation donnée. Il existe deux principales catégories d'explication: l'analyse de sensibilité et les méthodes de décomposition de la prédiction. L'analyse de sensibilité évalue l'impact qu'aurait le changement de la valeur d'une caractéristique (par exemple, un pixel) sur le score de prédiction. Par exemple, les cartes de chaleur obtenues en calculant la probabilité donnée par le réseau pour chaque pixel lors du masquage d'une partie de l'image centrée sur ce pixel. Les méthodes de décomposition attribuent un score de pertinence à chaque caractéristique reflétant sa contribution à la décision de classification. Par exemple, la technique de propagation de la pertinence couche par couche (layer-wise relevance propagation (LRP)) Bach *et al.* (2015) décompose la prédiction en scores de pertinence pour chaque neurone, en commençant par la dernière couche et en la propageant progressivement vers l'entrée.

LRP a été appliqué dans divers domaines. Par exemple, Becker *et al.* (2018) a utilisé cette technique d'explication pour examiner les résultats de la reconnaissance de chiffres et de la classification de genre sur l'ensemble de données AudioMNIST contenant 30 000 enregistrements de chiffres prononcés en utilisant à la fois la représentation spectrogramme et les données brutes. Pour la classification de chiffres avec la représentation spectrogramme, les auteurs ont observé que différentes régions de l'entrée sont essentielles pour chaque classe. Pour la classification de genre, la plage de basses fréquences s'est avérée pertinente, car elle est déjà connue pour être discriminante pour cette tâche. De plus, en utilisant les données brutes, ils ont découvert que les échantillons de grande amplitude sont importants. En comparant différentes manipulations de l'entrée, celles qui modifient l'ensemble d'entrées pertinentes ont causé la plus grande détérioration du modèle.

Une autre étude Lawhern *et al.* (2018) a utilisé PRL pour expliquer la classification des enregistrements EEG de sujets effectuant des mouvements imaginaires de la main gauche et droite. Le score de pertinence a été calculé pour chaque canal à chaque instant. Pour un instant spécifique, la pertinence du canal affiche un schéma d'activation moteur latéralisé typique, qui lorsqu'il est moyenné sur toutes les époques donne un schéma similaire.

2.5 Adaptation au domaine

Au cours de la dernière décennie, l'apprentissage profond est devenu une approche très efficace pour diverses applications. Cependant, il est souvent irréaliste en pratique de supposer que les données d'entraînement étiquetées et les données de test proviennent de la même distribution de données, car ces dernières peuvent être décalées ou même totalement différentes. Cela peut entraîner un biais potentiel ou un désalignement des données, couramment appelé problème de "décalage de domaine" Ben-David *et al.* (2010), qui est courant dans les applications pratiques et peut entraîner une dégradation significative des performances Ponce *et al.* (2006); Torralba & Efros (2011). Par exemple, dans la prédiction de crises, les données d'un nouveau patient peuvent différer considérablement de celles des patients utilisés pour entraîner le modèle de prédiction.

Pour atténuer le biais entre les données source et cible, l'adaptation de domaine peut être envisagée. Dans ce contexte, un domaine fait référence théoriquement à la distribution de probabilité à partir de laquelle les données du problème sont tirées. L'ensemble de données d'entraînement est appfelé données de domaine source et l'ensemble de données de test est appelé données de domaine cible. L'adaptation de domaine utilise des données sources étiquetées et des données cibles non étiquetées pour apprendre un modèle qui fonctionne bien dans les domaines cible et source. Cela est généralement réalisé en rééquilibrant les échantillons source pour minimiser le décalage de distribution, de sorte que les échantillons source les plus proches du domaine cible soient donnés plus d'importance Huang *et al.* (2006); Sugiyama *et al.* (2007).

Alternativement, on peut utiliser un modèle pré-entraîné du domaine source sur le nouveau domaine cible Oquab *et al.* (2014). Cette approche, connue sous le nom d'approche basée sur les paramètres, ne peut être appliquée que si des données étiquetées sont disponibles dans le domaine cible. Une autre approche sur laquelle nous nous concentrons dans cette thèse, appelée approche basée sur les caractéristiques Daumé III (2009); Ganin *et al.* (2016); Long *et al.* (2018), utilise les données source et cible pour apprendre des caractéristiques qui présentent un comportement similaire en classification sur les données des domaines source et cible. L'utilisation de l'adaptation de domaine pour classifier les données EEG a été fructueuse dans des applications telles que la reconnaissance d'émotions Li *et al.* (2019); Ma *et al.* (2019); Zhang *et al.* (2019a), la classification de l'imagerie motrice Tang & Zhang (2020); Wu *et al.* (2019), et l'évaluation de la qualité du sommeil Zhang *et al.* (2017).

2.6 Conclusion

En conclusion, ce chapitre a présenté un aperçu des méthodes de prédiction de l'épilepsie en mettant en évidence les plus récentes utilisées pour améliorer la précision des prédictions. Nous avons également exploré les différents outils et méthodes disponibles pour interpréter et expliquer les modèles de réseaux de neurones, qui seront utilisés pour améliorer la compréhension et la confiance dans les prédictions produites par les modèles de prédiction de crises d'épilepsie. Enfin, nous avons examiné les techniques d'adaptation au domaine, qui permettent d'améliorer les performances des prédictions de l'épilepsie pour de nouveaux patients.

Chapitre 3

Article 1 : A study of EEG feature complexity in epileptic seizure prediction

Imene Jemal[1,2], Amar Mitiche[1] and Neila Mezghani [2,3]

[1] Centre ÉMT, Institut National de la Recherche Scientifique, Montréal, Canada

[2] Centre de Recherche LICEF, Université TÉLUQ, Montréal, Canada

[3] Laboratoire LIO, Centre de Recherche du CHUM, Montréal, Canada

Applied Sciences, 2021, 11(4), 1579; https://doi.org/10.3390/app11041579

Received: 31 December 2020 / Revised: 2 February 2021 / Accepted: 5 February 2021 / Published: 9 February 2021

Résumé : Au cours des dernières années, la prédiction automatique de crises d'épilepsie à l'aide de données EEG a connu une croissance rapide. Les méthodes couramment utilisées se basent sur l'extraction de caractéristiques des données EEG, censées être des indicateurs de la présence d'une crise épileptique, pour différencier un état pré-ictal d'un état inter-ictal. Bien que ces méthodes utilisent souvent des résultats expérimentaux pour évaluer l'efficacité des différentes caractéristiques

et classificateurs proposés, aucune étude approfondie n'a été réalisée pour évaluer la complexité de la classification des données EEG basée sur ces caractéristiques. Une telle étude est importante car elle peut aider à déterminer les caractéristiques les plus pertinentes et à optimiser la conception des classificateurs, ce qui peut avoir un impact significatif sur la prédiction de l'épilepsie à partir de données EEG.

L'objectif de cette étude est (1) d'évaluer la complexité de la classification des données EEG en utilisant des caractéristiques extraites des signaux EEG et (2) d'évaluer la variabilité de ces caractéristiques entre les sujets. Ainsi, nous examinons une liste exhaustive de 88 caractéristiques (voir Annexe A), issues des domaines temporel, fréquentiel et de la théorie des systèmes dynamiques, en utilisant différentes mesures de complexité telles que le rapport discriminant de Fisher F1, le volume de chevauchement F2 et l'efficacité individuelle de la caractéristique F3 afin d'évaluer leurs capacités à distinguer les classes indépendamment du classificateur. Notre analyse se divise en deux étapes. Tout d'abord, nous utilisons ces mesures de complexité pour un problème à deux classes afin de déterminer la capacité de chaque caractéristique à distinguer les périodes inter-ictales et pré-ictales et d'évaluer la difficulté de la tâche de classification. Ensuite, nous analysons la complexité des caractéristiques extraites des données pré-ictales de tous les patients, où la tâche de classification consiste à catégoriser les instances pré-ictales par patient, dans l'hypothèse que la simplicité de la tâche de classification implique une grande variabilité des caractéristiques entre les patients.

Pour analyser la complexité des deux problèmes de classification, nous suivons la même procédure. Nous évaluons d'abord les mesures de complexité pour chaque caractéristique extraite pour déterminer leur capacité à distinguer les classes. Ensuite, nous comparons différentes distributions pour estimer la distribution de chaque mesure de complexité évaluée pour les 88 caractéristiques extraites. Nous utilisons un test de Kolmogorov-Smirnov pour évaluer la qualité de l'estimation. Les seuils de complexité sont déterminés en utilisant les quantiles $5^{ème}$ et $95^{ème}$ de la distribution qui convient le mieux aux données. Les caractéristiques qui dépassent ces seuils sont considérées comme potentiellement discriminantes. Après l'évaluation de la complexité, un *test-t* est effectué pour valider les résultats. Ce *test-t* est réalisé pour chaque caractéristique retenue de l'analyse de complexité des classes inter-ictale et pré-ictale pour déterminer la signification de la différence entre elles. Pour le deuxième problème de classification, qui concerne la distinction des échantillons pré-ictaux par patient, un test ANOVA unidirectionnel est utilisé pour vérifier si les caractéristiques prometteuses sont significativement différentes entre les sujets. Nous avons mené l'analyse de complexité sur la base de données publique CHB-MIT, qui comprend les enregistrements EEG de 23 patients. Les résultats de l'analyse de complexité de la classification des états pré-ictaux et inter-ictaux montrent que certaines caractéristiques peuvent différencier les états inter-ictal et pré-ictal, mais que le problème de classification est très complexe par rapport à d'autres problèmes de classification de référence. En effet, sur les 88 attributs analysés, seulement 10 ont été considérés comme peu complexes. Nous avons également comparé la complexité du problème de classification des états inter-ictaux et pré-ictaux en utilisant le rapport discriminant maximal de Fisher et l'efficacité maximale des caractéristiques, avec des problèmes de classification de référence et des données synthétiques étiquetées aléatoirement. Les problèmes comparés incluaient la classification d'Iris, la reconnaissance des lettres, la reconnaissance d'activité humaine à partir d'accéléromètre et de gyroscope, et deux ensembles de bruit aléatoire. Les résultats indiquent que les données contiennent des motifs exploitables, mais que la classification des caractéristiques extraites des enregistrements EEG reste très complexe.

L'analyse de la complexité de la classification des patients en utilisant les caractéristiques extraites des données pré-ictales révèle une variabilité importante entre les patients. Seulement 14 sur 88 caractéristiques peuvent efficacement distinguer les observations pré-ictales pour chaque patient. Cela signifie que les caractéristiques dérivées de l'état pré-ictal varient considérablement d'un patient à l'autre, ce qui rend incertaine la généralisation des modèles inter-sujets à de nouveaux patients dans des applications réelles. Par conséquent, il est important de poursuivre la recherche de nouvelles caractéristiques invariantes entre les patients, telles que celles extraites à l'aide de méthodes plus avancées, comme les réseaux de neurones profonds.

3.1 Abstract

The purpose of this study is (1) to provide EEG feature complexity analysis in seizure prediction by inter-ictal and pre-ital data classification and (2) to assess the between-subject variability of the considered features. Indeed, in the past several decades, there has been a sustained interest in predicting epilepsy seizure using EEG data. Most methods classify features extracted from EEG, which they assume are characteristic of the presence of an epilepsy episode, for instance by distinguishing a pre-ictal interval of data (which is in a given window just before the onset of a seizure) from inter-ictal (which is in preceding windows away from the seizure). To evaluate the difficulty of this classification problem independently of the classification model, we investigate the complexity of an exhaustive list of 88 features using various complexity metrics i.e., the Fisher discriminant ratio, the volume of overlap and the individual feature efficiency. Complexity measurements on real and synthetic data testbeds reveal that that seizure prediction by pre-ictal/inter-ictal feature distinction is a problem of significant complexity. It shows that several features are clearly useful, without decidedly identifying an optimal set.

Keywords : Data complexity measures; Epileptic seizure; Pre-ictal period; Hand-engineered features; Epilepsy prediction.

3.2 Introduction

Epilepsy is a chronic disorder of unprovoked recurrent seizures. It affects approximately 50 million people of all ages, which makes it the second most common neurological disease who (2019). Episodes of epilepsy seizure can have a significant psychological effect on patients. Also, sudden death, called sudden unexpected death in epilepsy, can occur during or following a seizure, although this is uncommon (about 1 in 1000 patients) Coll *et al.* (2016); Partemi *et al.* (2015). Therefore, early seizure prediction is crucial. Several studies have shown that the onset of a seizure generally follows a characteristic *pre-ictal* period, where the EEG pattern is different from the patterns of the seizure and also of periods preceding, called *inter-ictal* (Figure 3.1). Therefore, being able to distinguish pre-ictal and inter-ictal data patterns affords a way to predict a seizure. This can be done according to a standard pattern classification paradigm Duda *et al.* (2012); Bishop (2006): represent the sensed data by characteristic measurements, called features, and determine to which pattern class, pre-ictal of inter-ictal in this instance, an observed measurement belongs to. Research in seizure computer prediction has mainly followed this vein of thought. Although signal processing paradigms, such as time series discontinuity detection, are conceivable, feature-based pattern classification is a well understood and effective framework to study seizure prediction.

To be successful, a feature-based pattern classification scheme must use efficient features, which are features that well separate the classes of patterns in the problem. Features are generally chosen following practice, and basic analysis and processing. The traditional paradigm of pattern recognition Duda *et al.* (2012); Bishop (2006) uses a set of pattern-descriptive features to drive a particular classifier. The performance of the classifier-and-features combination is then evaluated on some pertinent test data. The purpose of the evaluation is not to study the discriminant potency of the features independently of the classifier, although such a study is essential to inform on the features complexity, i.e., the features classifier-independent discriminant potency. In contrast, our study is concerned explicitly with classifier-independent relative effectiveness of features. The purpose is to provide a feature complexity analysis in seizure prediction by inter-ictal/pre-ital data classification, which evaluates features using classifier-independent and statistically validated complexity metrics, such as Fisher discriminant ratio and class overlap volume, to gain some understanding of the features relative potency to inform on the complexity of epilepsy seizure prediction and the level of classification performance one can expect. This can also inform on between-patients data variability and its potential impact on epileptic seizure prediction as a pattern classifier problem. Before presenting this analysis in detail, we briefly review methods that have addressed EEG feature-based epileptic seizure prediction. The statistical study in Mormann *et al.* (2005) compared 30 features in terms of their ability to distinguish between the pre-ictal and pre-seizure periods, concluding that only a few features of synchronization showed discriminant potency. The method had the merit of not basing its analysis on a particular classifier. Rather, it measured a feature in pre-ictal and inter-ictal segments and evaluated the difference by statistical indicators such as the ROC curve. This difference is subsequently mapped onto the ability of the feature to separate pre-ictal from inter-ictal segments. In general, however, epilepsy seizure prediction methods, such as Assi *et al.* (2017); Gadhoumi *et al.* (2016), are classifier-based because they measure a feature classification potency by its performance using a particular classification algorithm. Although a justification to use the algorithm is general given, albeit informally, feature potency interpretation can change if a

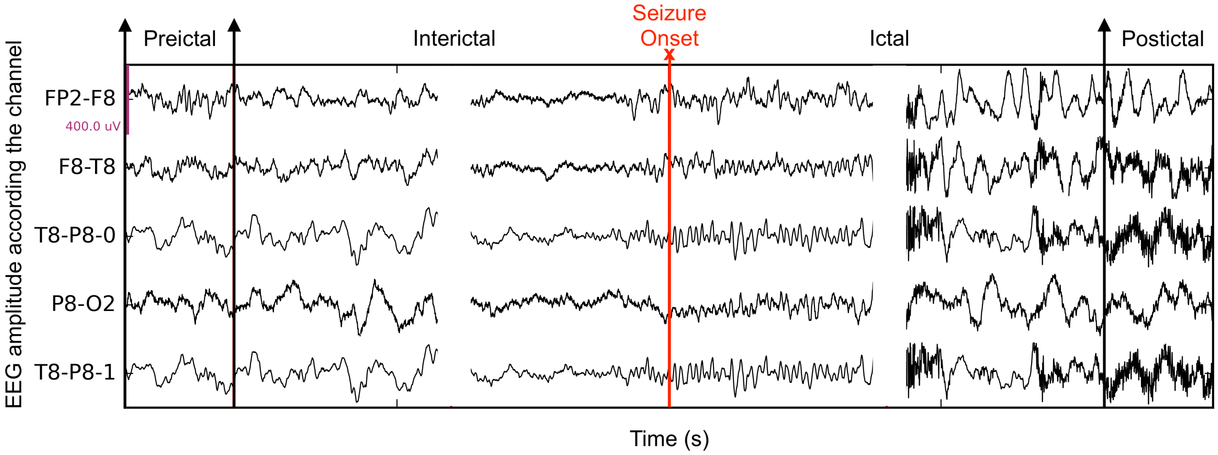


Figure 3.1 – The seizure phases in 5 EEG channels, including interictal, preictal, ictal and postictal.

different classifier is used. Observed feature interpretation discrepancies can often be explained by the absence of statistical validation in classifier-dependent methods. The importance of statistical validation has been emphasized in Teixeira *et al.* (2014), which used recordings of 278 patients to investigate the performance of a subject-specific classifier learned using 22 features. The study reports low sensitivity and high false alarm rates compared to studies that do not use statistical validation and which, instead, report generally optimistic results. Other methods Andrzejak *et al.* (2003, 2009); Kreuz *et al.* (2004) have resorted to simulated pre-ictal data, referred to as surrogate data, generated by a Monte Carlo scheme, for instance, for a classifier-independent means of evaluating a classifier running on a set of given features: If its performance is better on the training data than on the surrogate data, the method is taken to be sound, or is worthy of further investigation. However, no general conclusions are drawn regarding the investigated algorithm seizure prediction ability. The study of Cook *et al.* (2013) investigated a patient-specific monitoring system trained from long-term EEG records, and in which seizure prediction combines decision trees and nearest-neighbors classification. Along this vein, Moghim & Corne (2014) used a support vector machine to classify patient-dependent, hand-crafted features. Experiments reported indicate the methods decisions have high sensitivity and false alarm rates. Deep learning networks have also served seizure prediction, and studies mention that they can achieve a good compromise between sensitivity and false alarm rates Tsiouris *et al.* (2018); Daoud & Bayoumi (2019)

None of the studies we have reviewed has inquired into EEG feature classification complexity, in spite of the importance of this inquiry. As we have indicated earlier, a study of feature classification complexity is essential because it can inform on important properties of the data representation features, such as their mutual discriminant capability and extent, and their relative discriminant efficiency. This, in turn, can benefit feature selection and classifier design. Complexity is generally mentioned by clinicians, who acknowledge, often informally, that common subject-specific EEG features can be highly variable, and that cross-subject features are generally significantly more so. This explains in part why research has so far concentrated on subject-specific data features analysis, rather than cross-subject. Complexity analysis has the added advantage of applying equally to both subject-specific and cross-subject features, and thus offers an opportunity to draw beforehand some insight on cross-subject data.

The purpose of this study is to investigate the complexity of pre-ictal and inter-ictal classification using features extracted from EEG records, to explore the predictive potential of cross-subject classifiers for epileptic seizure prediction and evaluate the between-subject variability of the considered features. Using complexity metrics which correlate linearly to classification error Ho (2002); Bernadó-Mansilla & Ho (2005); Ho & Bernadó-Mansilla (2006); Mansilla & Ho (2004), this study provides an algorithm-independent cross-subject classification complexity analysis of a set of 88 prevailing features, collected from EEG data of 24 patients. We employed such complexity measures for two-class problems to examine the individual feature ability to distinguish the inter-ictal and pre-ictal periods and to inspect the difficulty of the classification problem. The same complexity metrics generalized to multiple classes are also utilized in order to highlight the variability between patients considering the extracted features.

The remainder of this chapter gives the details of the data, its processing, and analysis. It is organized as follows: Section 3.3 describes the materials and methods: the database, the features, the complexity metrics, and the statistical analysis. Section 3.4 presents experimental results and section 3.5 concludes the chapter.

3.3 Materials and Methods

3.3.1 Database

We performed the complexity analysis on the openly available database collected at the Children's Hospital Boston Shoeb (2009) which contains intracranial EEG records from 24 monitored patients. The EEG raws sampled at 256 Hz were filtered using notch and band-pass filters to remove some degree of artifacts and focus on relevant brain activity. After, we segmented the data to five seconds non-overlapping windows allowing capturing relevant patterns and satisfying the condition of stationarity Rasekhi *et al.* (2013); Assi *et al.* (2015); Bandarabadi *et al.* (2015b). We set the pre-ictal period to be 30 minutes before the onset of the seizure as suggested in Teixeira *et al.* (2014); Bandarabadi *et al.* (2015a) and eliminated 30 minutes after the beginning of the seizure to exclude effects from the post-ictal period, inducing a total of almost 828 remaining hours divided into 529,415 samples for the inter-ictal state and 66,782 for the pre-ictal interval.

3.3.2 Extracted features

We queried relevant papers and reviews Mormann *et al.* (2007); Teixeira *et al.* (2011); Assi *et al.* (2017) which addressed epileptic seizure prediction, to collect a superset of 88 features, univariate as well as bivariate, commonly used in epilepsy prediction. We focused on algorithm-based seizure prediction studies. Only studies which used features from EEG records were included in the study. Image-based representation studies, for instance, were excluded.

We extracted a total of 21 univariate linear features including statistical measures such as variance, σ^2, skewness, χ, and kurtosis, κ, temporal features such as Hjorth parameters, HM and HC Damaševičius *et al.* (2018), the de-correlation time, τ_0, and the prediction error of auto-regressive modeling, ϵ_{err}, as well as spectral attributes, for instance, the spectral band power of the delta, δ_r, theta ,θ_r, alpha, α_r, beta, β_r, and gamma, γ_r, bands, the spectral edge frequency, f_{50}, the wavelet energy, E_w, entropy, S_w, signal energy, E, and accumulated energy, AE. Eleven additional univariate non-linear features from the theory of dynamical systems Schuster & Just (2006); Ott (2002); Kantz & Schreiber (2004) have been used, characterizing the behavior of complex dynamical system, such the brain by using observable data (EEG records) Andrzejak *et al.* (2001b); Iasemidis *et al.* (1990). Time-delay embedding reconstruction of the state space trajectory from the raw data was used to calculate the non-linear features Damasevicius *et al.* (2014). The time delay, τ, and the embedding dimension, m, were chosen according to previous studies Iasemidis *et al.* (1990, 2000). Within this framework, we determine the correlation dimension, D_2 Lehnertz *et al.* (2001), and correlation density, D_ϵ Lerner (1996). We used also the largest Lyapunov exponent, L_{max} Iasemidis *et al.* (1990), and the local flow, Λ, to assess determinism, the algorithmic complexity, AC, and the loss of recurrence, LR, to evaluate non-stationarity and, finally, the marginal predictability, δ_m Savit & Green (1991). Moreover, we used a surrogate-corrected version of the correlation dimension, D_2^*, largest Lyapunov exponent, L_{max}^*, local flow, Λ^*, and algorithmic complexity, AC^*. We investigated also 48 linear bivariate attributes, including 45 bivariate spectral power features, b_k Bandarabadi *et al.* (2015b), cross-correlation, C_{max}, linear coherence, Γ and mutual information, MI. As to bivariate non-linear measures, we used 6 different characteristics for phase synchronization: the mean phase coherence, R, and the indexes based on conditional probability, λ_{cp}, and Shanon entropy, ρ_{se}, evaluated on both the Hilbert and wavelet transform. Finally, we also retained two measures of non-linear interdependence, S and H.

3.3.3 Complexity metrics for pre-ictal and inter-ictal feature classification

Data complexity analysis often has the goal to get some insight into the level of discrimination performance that can be achieved by classifiers taking into consideration intrinsic difficulties in the data. Ho & Bernadó-Mansilla (2006) observed that the difficulty of a classification problem arises from the presence of different sources of complexity: (1) the class ambiguity which describes the issue of non-distinguishable classes due to an intrinsic ambiguity or insufficient discriminant features Ho & Baird (1997), (2) the sample sparsity and feature space dimensionality expressing the impact of the number and representativeness of training set on the model's generalization capacity Bernadó-Mansilla & Ho (2005) and (3) the boundary complexity defined by the Kolmogorov complexity Kolmogorov (1965); Li *et al.* (2008) of the class decision boundary minimizing the Bayes error. Since the Kolmogorov complexity measuring the length of the shortest program describing the class boundary is claimed to be uncountable Maciejowski (1979), various geometrical complexity measures have been deployed to outline the decision region Basu & Ho (2006). Among the several types of complexity measures, the geometrical complexity is the most explored and used for complexity assessment Mezghani *et al.* (2018); Morán-Fernández *et al.* (2017); Sun *et al.* (2019). Thus, for each individual extracted feature we evaluate the geometrical complexity measures of overlaps in feature values from different classes inspecting the efficiency of a single feature to distinguish between the inter-ictal and the pre-ictal states. We considered the following feature complexity metrics:

— **Fisher discriminant ratio, F1**: quantifying the separability capability between the classes. It is computed as:

$$F1_i = \frac{(\mu_{i1} - \mu_{i2})^2}{\sigma_{i1}^2 + \sigma_{i2}^2} \tag{3.1}$$

where μ_{i1}, μ_{i2} and σ_{i1}^2, σ_{i2}^2 are the means and variances of the attribute i for each of the two classes : inter-ictal and pre-ictal respectively. The larger the value of F1 is, the wider the margin between classes and smaller variance within classes are, such that a high value presents a low complexity problem.

— **Volume of overlap region, F2**: measuring the width of the entire interval encompassing the two classes. It is denoted by:

$$F2_i = \frac{min(max_{i1}, max_{i2}) - max(min_{i1}, min_{i2})}{max(max_{i1}, max_{i2}) - min(min_{i1}, min_{i2})} \tag{3.2}$$

where max_{i1}, max_{i2}, min_{i1}, min_{i2} are the maximums and minimums values of the feature i for the two classes respectively. F2 is zero if the two classes are disjoint. A low value of F2 would correspond to small amount of the overlap among the classes indicating a simple classification problem.

— **Individual feature efficiency, F3**: describing how much an attribute contribute to distinguish between the two classes. It is defined by:

$$F3_i = \frac{|fi \in [min(max_{i1}, max_{i2}), max(min_{i1}, min_{i2})]|}{n} \tag{3.3}$$

where max_{i1}, max_{i2}, min_{i1}, min_{i2} are the maximums and minimums of the attribute i for each of the two classes respectively and n is the total number of samples in both classes. A high value of F3 refer to a good separability between the classes.

3.3.4 Complexity metrics for cross-subject variability assessement

Alternatively, we resorted to the extension of complexity measures designed for a binary problem to multiple-class classification in order to study the variability of the features of the pre-ictal state between patients where the classification task is to identify to which patient a pre-ictal instance belongs. The complexity of the patient's classification problem points out to the level of variability within patients. By converting the multi-class problem to many two-class sub-problems. The complexity metrics become:

— **The Fisher discriminant ratio, F1**, for C classes extended from equation 3.1 as:

$$F1_i = \frac{\sum_{j=1,k=1,j\neq k}^{C} p_{ij} p_{ik} (\mu_{ij} - \mu_{ik})^2}{\sum_{j=1}^{C} p_{ij} \sigma_{ij}^2} \tag{3.4}$$

where μ_{ij}, μ_{ik}, p_{ij}, p_{ik} and $sigma_{ij}^2$ are the means, the proportions and the variance of the feature i for the two classes j and k respectively.

— **The volume of overlap region, F2**, for a multi-class problem is computed as:

$$F2_i = \sum_{j=1,k=1,j\neq k}^{C} \frac{min(max_{ij}, max_{ik}) - max(min_{ij}, min_{ik})}{max(max_{ij}, max_{ik}) - min(min_{ij}, min_{ik})} \tag{3.5}$$

where max_{ij}, max_{ik}, min_{ij}, min_{ik} are the maximums and minimums values of the feature i for the two classes j and k respectively.

— **The individual feature efficiency** for multiple classes can be written as:

$$F3_i = \sum_{j=1,k=1,j\neq k}^{C} \frac{|fi \in [min(max_{ij}, max_{ik}), max(min_{ij}, min_{ik})]|}{n_{j,k}} \tag{3.6}$$

where max_{ij}, max_{ik}, min_{ij}, min_{ik} are the maximums and minimums of the attribute i for the classes i and j respectively and $n_{j,k}$ is the total number of samples.

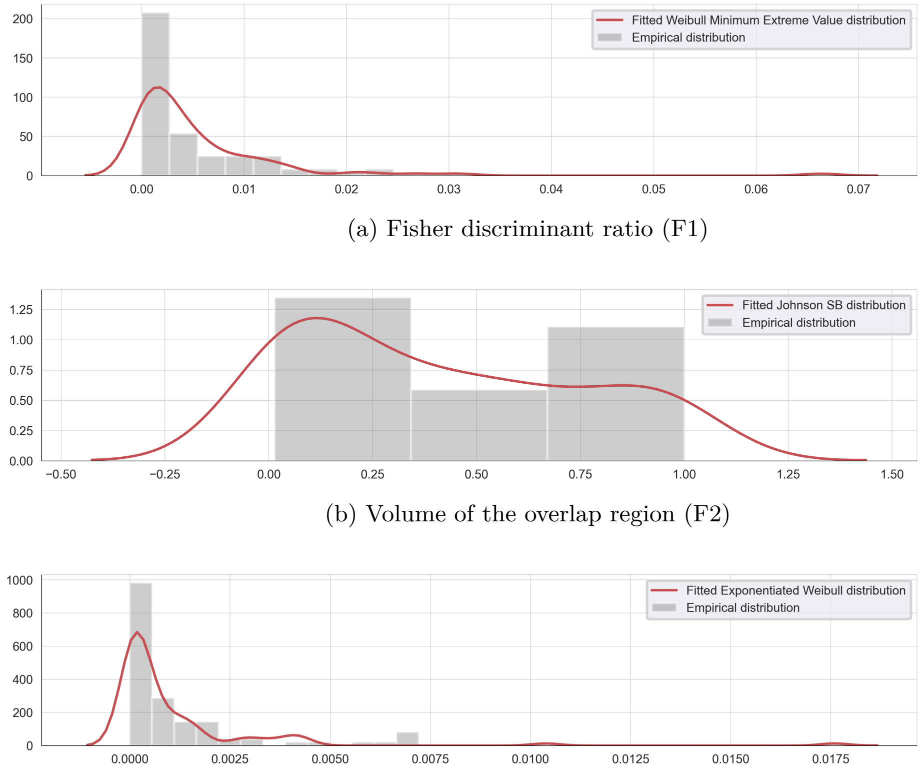
Furthermore, for each complexity measure evaluated for all 88 extracted features, we estimate the data distribution by fitting the data with 82 distribution functions available in the SciPy 0.12.0 Package. To test the goodness of fit, we perform a Kolmogorov-Smirnov test, with a significance level of 0.05. Lower and upper thresholds, on which the decision whether the feature is complex or not relies, are set according to the rule of thumbs to the 5^{th} and 95^{th} quantiles of the probability distribution which fits the best the data. The feature that exceeds the lower or higher threshold, depending on the complexity metric, is considered a potential discriminant feature.

3.3.5 Statistical analysis

Following the complexity metrics assessment, we conduct a statistical test to certify that the analysis results are significant. We performed the t-test for each retained feature, distinguishing the inter-ictal and pre-ictal classes, to assess how significant the difference between the categories.For the between-patient variability study, since classifying the pre-ictal samples by patients is a multi-class problem, we applied the one-way ANOVA test to verify that the promising discriminant features are significantly different between subjects. The classes are said to differ significantly if the p-value of the statistical test is smaller than a two-sided significance level of 0.05.

3.4 Results and discussion

As described before, we conducted our experiments on the public CHB-MIT dataset described in Section 3.3.1. A total of 88 univariate and bivariate features commonly used in epilepsy prediction have been extracted from the EEG records (as presented in section 3.3.2). The pre-processing of the dataset and the feature extraction were done using Matlab R2020a software.



(a) Fisher discriminant ratio (F1)

(b) Volume of the overlap region (F2)

(c) Feature efficiency (F3)

Figure 3.2 – Empirical and fitted distribution for the complexity measures (a) Fisher discriminant ratio F1, (b) the volume of the overlap interval F2, and (c) the individual feature efficiency F3.

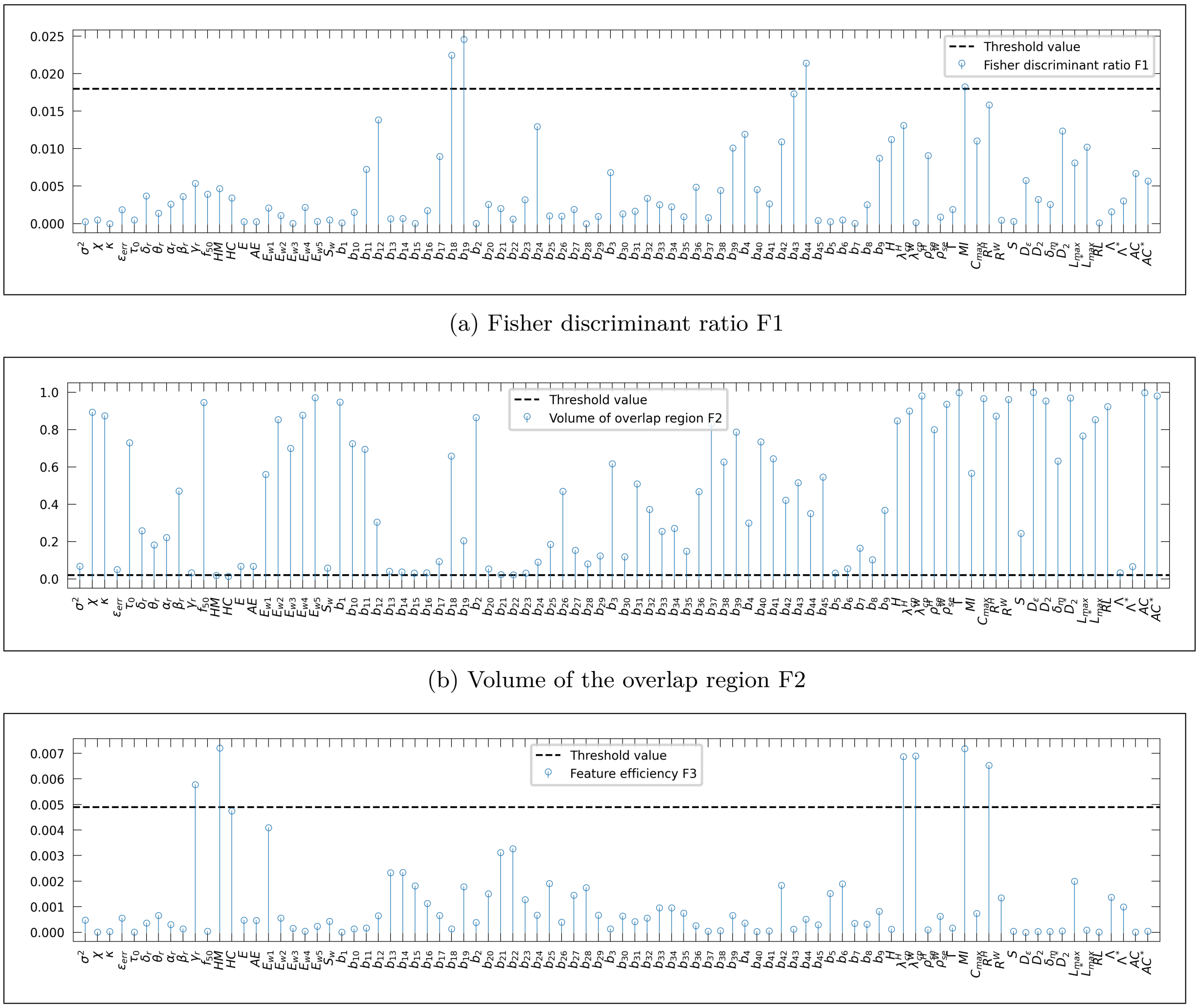
3.4.1 Analysis of the complexity of the pre-ictal and inter-ictal features

To illustrate the feature complexity analysis, three metrics are evaluated on the extensive list of the extracted features: F1, F2 and F3. The lower and higher threshold values have been identified, for each complexity metric, using a probability density which best fit each complexity data. Figure 3.2 illustrates the empirical and fitted distribution for each complexity measure.

The Fisher discriminant ratio F1, can be modeled by a Weilbull minimum extreme value distribution (Figure 3.2a), the volume of overlap region F2, follows a Johnson SB distribution (Figure 3.2b) and the feature efficiency values F3, can be approximated by an exponentiated Weibull distribution (Figure 3.2c). Following this modeling, the lower and upper threshold values are determined using the 5th and the 95th percentiles of each complexity measure empirical density as shown in Table 3.1.

Tableau 3.1 – Thresholds for the complexity metrics.

Complexity metrics	Lower and upper thresholds
Fisher discriminant ratio F1	$\{< 0.001, \mathbf{0.017}\}$
Volume of overlap region F2	$\{\mathbf{0.020}, 0.987\}$
Feature efficiency F3	$\{< 0.001, \mathbf{0.0048}\}$



(a) Fisher discriminant ratio F1

(b) Volume of the overlap region F2

(c) Feature efficiency F3

Figure 3.3 – Evaluation of the complexity metrics for each extracted value, (a) Fisher discriminant ratio, F1: a high value of F1 shows that the feature is discriminant (b) the volume of the overlap interval, F2: a low value of F2 indicates small amount of overlap among the classes and (c) the individual feature efficiency, F3: a high value of F3 implies a good separability between the classes. The horizontal dotted blue represents the threshold values.

The results of evaluating the Fisher discriminant ratio, F1, on various features are shown in Figure 3.3a. The threshold was set to 0.017 (Table 3.1 line 1). As shown in Figure 3.3a, three bivariate spectral power attributes, b_{18}, b_{19} and b_{44} and the mutual information, MI, surpass the threshold value. Because higher values indicate better class separation, only these features are retained.

Likewise, Figure 3.3b presents the results obtained by the assessment of the volume of overlap region, F2. Given the threshold 0.02 (Table 3.1, line 2), only the two Hjorth parameters, HM and HC, are retained since lower values indicate a small amount of overlap among the classes.

In figure 3.3c, the individual feature efficiency, F3, values for each characteristic are shown. For this complexity metric, the threshold is set to 0.0048 (Table 3.1, line 3). The features having a higher value of F3 than the threshold value are the relative gamma band spectral power, γ_r, the Hjorth parameter, HM, the mean phase coherence using the Hilbert transform, R^H, and the indexes measures for phase synchronization based on conditional probability, λ_{cp}^H and λ_{cp}^W. Thus, the features are retained because high values of F3 claim a good feature separability between the categories.

In summary, the analysis of the extracted features complexity reveals that only ten out of 88 attributes has been picked as not complex. We observed also an overlap between the different feature obtained by each complexity metrics such as the Hjorth parameter, HM, which has a low value of volume of overlap, F2, and high value of the feature efficiency, F3. According to Table 3.2 summarizing the list of the retained features, All the ten features have a p-value corresponding to the statistical t-test lower than the critical value for 5% significance level, which approves that the features are statistically significantly different and thus, confirm their ability to discriminate the two classes.

Tableau 3.2 – Retained attributes from the feature complexity analysis, their types, and the corresponding p-value of the t-test.

Feature	Type	p-value
Hjorth parameter, HM	linear, univariate	< 0.001
Hjorth parameter, HC	linear, univariate	< 0.001
Relative gamma band power spectral, γ_r	linear, univariate	< 0.001
Bivariate spectral power characteristics, b_{18}	linear, bivariate	< 0.001
Bivariate spectral power characteristics, b_{19}	linear, bivariate	< 0.001
Bivariate spectral power characteristics, b_{44}	linear, bivariate	< 0.001
Mutual information, MI	linear, bivariate	< 0.001
Phase synchronization index based on conditional probability using the wavelet transform, λ_{cp}^W	non-linear, bivariate	< 0.001
Phase synchronization index based on conditional probability using the Hilbert transform, λ_{cp}^H	non-linear, bivariate	< 0.001
Mean phase coherence, R	non-linear, bivariate	< 0.001

We observe that only three non-linear features, the mean phase coherence using the Hilbert transform and the two measures for phase synchronization, the indexes based on conditional probability, λ_{cp}^{W} and λ_{cp}^{H} , all bivariate, are retained as discriminant, declaring they are the best discriminant non-linear bivariate features and that most of the non-linear other features especially univariate characteristics are incapable to distinguish between the inter-ictal and pre-seizure times, as also observed by Mormann *et al.* (2005); Harrison *et al.* (2005); McSharry *et al.* (2003). The Hjorth parameters, HM, and HC, are also shown to be the most uni-variate linear discriminant features of the two states as substantiated in Mormann *et al.* (2005). The other linear univariate recalled feature as discriminant is the relative gamma-band spectral power, γ_r, which supports the studies Park *et al.* (2011)Assi *et al.* (2015) saying that the characteristics from the gamma band are more relevant than other bands for the epilepsy prediction.

Despite of extracting many linear and non-linear both univariate and bivariate features from EEG records, we found that a restricted number of attributes shown to be promising to distinguish the inter-inctal and pre-ictal classes.

Comparison with reference databases: To have some insight into the complexity of the classification problem of inter-ictal and pre-ictal states, we compared our results against known relatively simple classification problems from the UC-Irvine Machine Learning Depository Dua & Graff (2017) and other randomly labelled synthetic data. We estimate the complexity measures of three binary classification problems from the *Iris* data set and a linearly non-separable problem using the *Letter* database of 20000 samples and 16 attributes . We used also a more complex dataset for the Human activity recognition, HAR, using sensor signals (accelerometer and gyroscope) recorded from a waist-mounted smartphone. The dataset contains 10299 samples with 561 features. Moreover, we evaluate the complexity metrics on two artificial classification problems, obtained from randomly labeling uniformly distributed data points, containing, respectively, 10,000 samples with a single feature, and 600,000 with 88 dimensions.

The results are summarized in Table 3.3 shown that the maximum Fisher discriminant ratio of the epilepsy data set is lower than simpler problems for instance classification tasks from the *iris* and *letter* sets and higher than results from the random noise sets. Indeed, the maximum Fisher discriminant ratio is 0.0245 for epilepsy data, 31.19, 49.94 and 4.27 for the Versicolor-Virginica, Setosa-Virginica and Setosa-Versicolor, respectively, from the Iris data, 0.9 for the Letter data-set, 2.66 for the human activity recognition (HAR) data-set, $1.7e^{-5}$ and $1.66e^{-5}$ for the two random data-sets. Similarly, the maximum feature efficiency of easy classification problems from Iris and Letter data, and for a relatively complex classification problem such as human activity recognition, given the HAR data-set, is higher than 0.25, unlike the epilepsy data having a low value of 0.003 and the two random data having, respectively, a maximum feature efficiency of 0.007 and $1.66e^{-5}$. Hence, the maximum feature efficiency of the epilepsy prediction problem is not as high as the comparatively easy problem nor as low as the intrinsically complex random noise problem. Therefore, this shows evidence that the epilepsy data does contain learnable structures, yet the classification problem using the extracted features from the EEG records is highly complex.

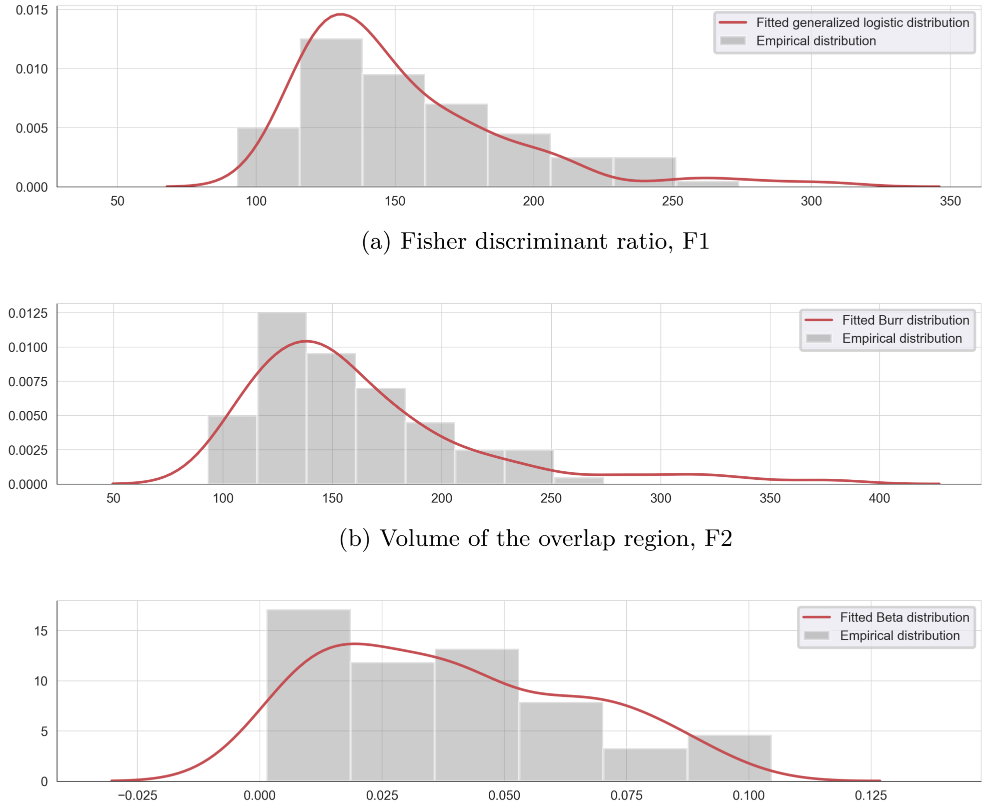
Tableau 3.3 – Comparison of the maximum Fisher discriminant ratio and the maximum feature efficiency for various classification problems: (1) Iris: Versicolor-Virginica, (2) Iris: Setosa-Virginica, (3) Iris: Setosa-Versicolor, (4) Letter recognition, (5) Epilepsy prediction, (6) Human activity recognition, HAR, and couple random noise sets, (7) Random data 1 and, (8) Random data 2.

Data-sets	Maximum Fisher discriminant ratio	Maximum feature efficiency
Iris: Setosa-Versicolor	31.19	1.0
Iris: Setosa-Virginica	49.94	1.0
Iris: Versicolor-Virginica	4.27	0.63
HAR	2.66	0.61
Letter	0.9	0.25
Epilepsy	0.024	0.003
Random data 1	$5.3e^{-5}$	0.007
Random data 2	$1.7e^{-5}$	$1.6e^{-5}$

3.4.2 Cross-patient variability assessement

To get a deeper insight into the epilepsy prediction problem formulated as a classification of inter-ictal and pre-ictal intervals using extracted features from EEG raws, it is necessary to check the variability of the information characterizing the pre-ictal state across patients. Therefore, Similar to the suggested strategy to evaluate the complexity of the pre-ictal and inter-ictal classification, we analyze the data complexity measures for the extracted features from the pre-ictal state of all patients, where the classification task is to categorize the pre-ictal instances by patients, under the hypothesis that the simplicity of the classification task implies a high variability of the features between patients.

The assessment of the complexity measures: the Fisher discriminant ratio, F1, the volume of overlap region, F2, and the feature efficiency, F3 are shown in Figure 3.5.



(a) Fisher discriminant ratio, F1

(b) Volume of the overlap region, F2

(c) Feature efficiency F3

Figure 3.4 – Empirical and fitted distribution for the complexity measures (a) Fisher discriminant ratio F1, (b) the volume of the overlap interval F2, and (c) the individual feature efficiency F3.

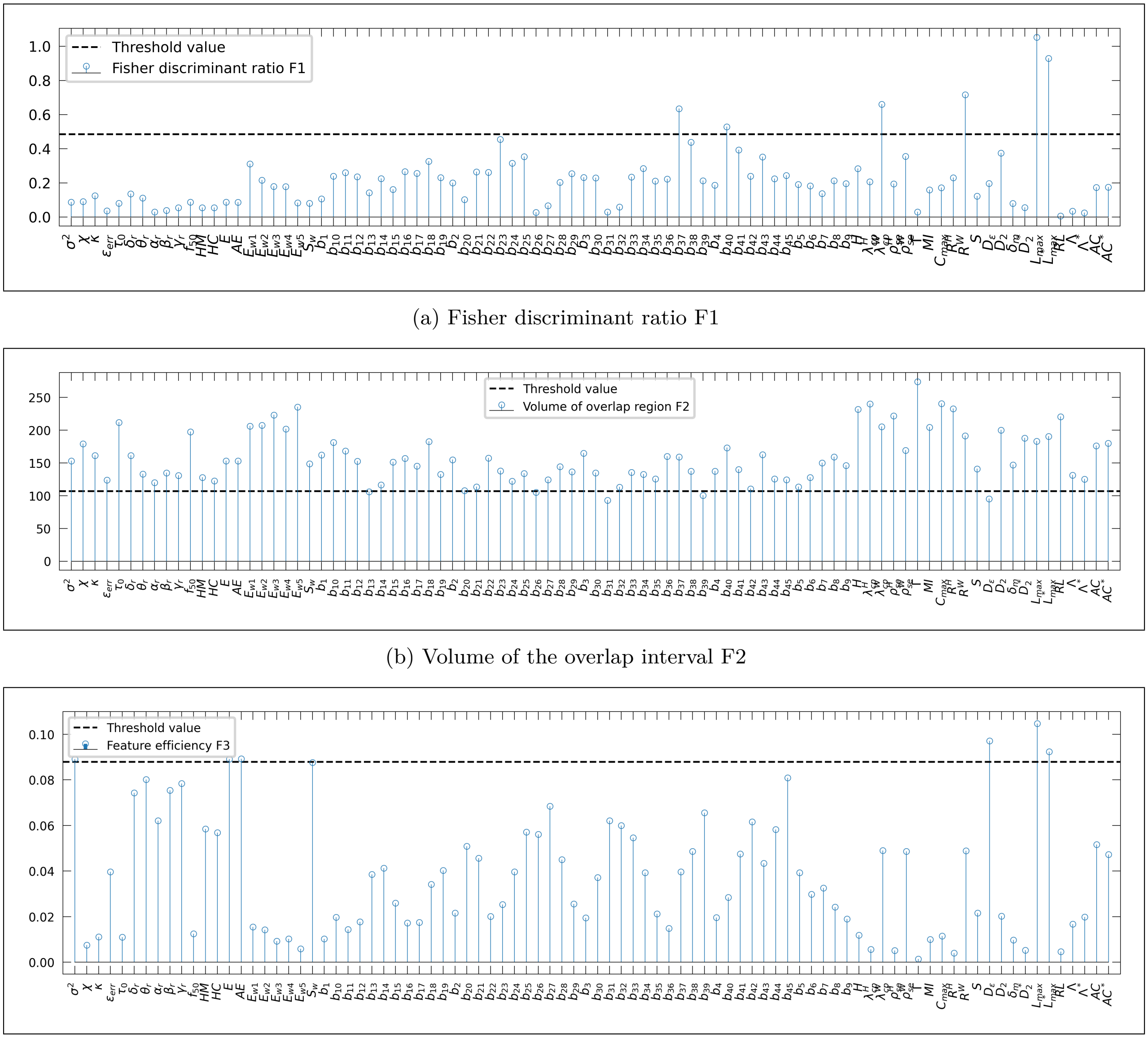
Threshold values were defined as the 5th and 95th quantiles of the best-fitted distribution of each complexity measure data shown in Figure 3.4 are resumed in Table 3.4. For the metrics F1, and F3, the thresholds were set to 0.48 (Table 3.4, line 1) and 0.087 (Table 3.4, line 3). While for F2, the threshold was set to 160 (Table 3.4 line 2).

Tableau 3.4 – Thresholds for the complexity metrics.

Complexity metrics	Lower and upper thresholds
Fisher discriminant ratio F1	$\{0.02, \mathbf{0.485}\}$
Volume of overlap region F2	$\{\mathbf{160}, 240\}$
Feature efficiency F3	$\{0.003, \mathbf{0.087}\}$

Figure 3.5 presents the evaluation with different complexity metrics. Figure 3.3a reveals that two bivariate spectral power attributes, b_{37} and b_{40}, the mean phase coherence, R, and the indexes based on conditional probability, λ_{cp}^{W}, and Shanon entropy, ρ_{se}^{W}, evaluated both using the wavelet transform, the largest Lyapunov exponent, L_{max}, and the surrogate-corrected version of the largest

Lyapunov exponent, L^{*}_{max}, exceed the Fisher discriminant ratio threshold value of 0.48. For the volume of the overlap region interval, only four bivariate spectral power attributes, b_{13}, b_{26}, b_{31}, and b_{39}, and the correlation density, D_{ϵ}, exceed the threshold value as shown in Figure 3.3b. Finally, Figure 3.3c displays the results of the assessment of the feature efficiency, showing that the variance, σ^2, signal energy, E, accumulated energy, AE, correlation density, D_{ϵ}, largest Lyapunov exponent, L_{max}, and the surrogate-corrected version of the largest Lyapunov exponent, L^{*}_{max}, have higher values than the threshold 0.087. In conclusion, a total of



(a) Fisher discriminant ratio F1

(b) Volume of the overlap interval F2

(c) Feature efficiency F3

Figure 3.5 – Evaluation of the complexity metrics: (a) Fisher discriminant ratio, F1 (b) the volume of the overlap interval, F2 (c) the individual feature efficiency, F3. The horizontal dotted blue indicates the threshold values.

14 out of 88 features are shown to distinguish well the pre-ictal observations by patient. Therefore, performing the one-way ANOVA test for each of the fourteen recalled features exhibits significant differences between patients as null p-values are obtained for all the features, which validates that each individual characteristic has large between-class distances. As a result, it is safe to conclude that the extracted features from the pre-ictal state vary significantly between patients, which raises a concern when using cross-patient models in real-world applications. Moreover, the high variability of the extracted features motivates research on searching new invariant features between patients.

3.5 Conclusion

This study investigated the complexity of a superset of EEG-based features commonly practiced to distinguish an inter-ictal period from a pre-ictal in epileptic seizure prediction. The investigation is based on a classifier-independent complexity analysis which used complexity measures, such as the Fisher discriminant ratio and the volume of class overlap in feature space, to evaluate the discriminant potency of each feature. Implemented using the publicly available Boston Children's Hospital database of EEG records, the analysis supports the conclusion that the features and, thereof, feature-based distinction of the pre-ictal and inter-ictal periods in EEG records, are highly complex.

This study can be strengthened along three majors veins. Along one vein, larger amounts of data, using different other EEG databases, can confirm and strengthen its conclusions on feature complexity and inter-subject variability. Along a second vein, features other than those generally practiced, which are those used in this study, can be investigated. To this end, feature computation by deep machine learning is exceptionally promising, as it has had recently a remarkable performance with similar and more difficult data. Finally, it can be of significant benefit to investigate domain adaptation for EEG data, whereby the dependence on large amounts of data for accurate EEG-based decision can be alleviated.

Chapitre 4

Article 2 : An interpretable deep learning classifier for epileptic seizure prediction using EEG Data

Imene Jemal[1,2], Neila Mezghani[2,3], Lina Abou-Abbas[2,3] and Amar Mitiche[1]

[1] Centre ÉMT, Institut National de la Recherche Scientifique, Montréal, Canada

[2] Centre de Recherche LICEF, Université TÉLUQ, Montréal, Canada

[3] Laboratoire LIO, Centre de Recherche du CHUM, Montréal, Canada

IEEE Access, 2022, volume 10, 60141 - 60150; 10.1109/ACCESS.2022.3176367
Received: 15 March 2022 / Accepted: 4 May 2022 / Published: 18 May 2022

Résumé : L'apprentissage profond a été utilisé pour la classification des formes dans de nombreuses applications, avec des performances qui dépassent souvent considérablement celles d'autres approches d'apprentissage automatique. Cependant, cette approche implique des architectures complexes et non interprétables pour extraire des caractéristiques abstraites et effectuer la classification, ce qui rend les décisions difficilement compréhensibles en termes de connaissances pertinentes pour

l'application. Ce phénomène appelé "l'effet de boîte noire" peut être un obstacle majeur dans certaines applications critiques, telles que la prédiction des crises d'épilepsie.

L'objectif de cette étude est de proposer un réseau de neurones profond interprétable et explicable pour la prédiction de crises épileptiques à l'aide des données EEG. Le réseau de neurones est considéré comme interprétable pour les raisons suivantes : 1) son architecture s'inspire sur l'algorithme Filter Bank Common Spatial Pattern (FBCSP) dont les couches correspondent à des calculs de traitement du signal tels que les filtres passe-bande et les filtres spatiaux. Par conséquent, les caractéristiques extraites ne sont plus abstraites, car elles correspondent à des caractéristiques généralement utilisées pour le décodage des données EEG, 2) ses poids, correspondant aux filtres appris, sont visualisés pour une interprétation plus approfondie du codage abstrait des caractéristiques, et 3) la technique de propagation de la pertinence par couche (LRP) a été utilisée pour révéler les éléments d'entrée pertinents qui peuvent expliquer davantage les calculs menant à ses décisions.

D'abord, nous avons conçu une architecture de réseau de neurones simple à interpréter en termes de couches. Le réseau comprend dans sa première couche une couche de convolution 2D qui apprend des filtres de passe-bande. Cette couche est inspirée par la première étape de l'algorithme FBCSP qui utilise un banc de filtres passe-bande pour séparer les signaux EEG en plusieurs bandes de fréquence. La deuxième couche est une couche de convolution en profondeur qui utilise des filtres pour chaque carte de caractéristiques, sortie de la couche précédente, indépendamment. Cela permet d'apprendre des filtres spatiaux spécifiques pour une bande de fréquences. Dans le contexte de l'EEG, ces filtres sont des transformations linéaires qui projettent les données vers un espace de source pour séparer les activités. Ces filtres sont généralement estimés par la deuxième étape de FBCSP. Pour extraire les caractéristiques des signaux, nous utilisons une couche de convolution suivie d'une activation non linéaire et d'un sous-échantillonnage (average-pooling). Enfin, les caractéristiques sont transmises à une couche entièrement connectée et utilisées pour une classification Softmax. L'étude se concentre sur des modèles de prédiction d'épilepsie spécifiques au sujet, qui sont configurés et paramétrés avec peu de données d'un seul sujet. Par conséquent, la même architecture a été utilisé pour l'apprentissage d'un modèle séparé pour chaque sujet. En utilisant la base de données CHB-MIT, nous avons évalué l'architecture sur 23 patients pour prédire les crises d'épilepsie en classifiant les états préictal et interictal. La précision moyenne de tous les modèles est de 90,9%, avec une sensibilité de 96,1% et une spécificité de 84,6%. Le taux moyen de fausses alarmes par heure est de 0,041, ce qui indique une bonne capacité de prédiction. La comparaison avec d'autres méthodes proposées dans la littérature et évaluées sur la même base de données CHB-MIT montre que l'architecture proposée améliore considérablement les performances des modèles de prédiction.

Après l'apprentissage des modèles, nous avons examiné l'interprétation du réseau de neurones en visualisant les filtres appris. Les filtres appris sont les poids des noyaux de convolution du réseau. La visualisation des filtres permet de voir comment chaque couche extrait des caractéristiques de l'entrée. Nous étions particulièrement intéressés de savoir si la première couche de l'architecture proposée apprenait des filtres passe-bande. Par conséquent, nous avons visualisé les filtres appris pour différents modèles spécifiques aux sujets, ainsi que leur représentation dans le domaine fréquentiel dérivé de la transformée de Fourier rapide. Nous avons constaté que les filtres de la première couche étaient similaires à des filtres passe-bande et que, de plus, chaque modèle spécifique au sujet possède son propre ensemble de filtres. De plus, des filtres similaires, tels que ceux des bandes 0-5, 5-20 et 40-45, apparaissent, dans presque tous les modèles pour différents sujets. Nous avons également constaté que les modèles extraient des caractéristiques en apprenant des filtres dont la fréquence est supérieure à 25 Hz, ce qui est cohérent avec le fait que la prédiction de l'épilepsie est basée sur les caractéristiques de la bande gamma (30-140 Hz) qui sont plus pertinentes que les autres bandes.

Enfin, pour expliquer les décisions du réseau pour de nombreux échantillons, nous avons utilisé la technique de propagation de la pertinence Layer-wise relevant propagation (LRP). LRP calcule des scores de pertinence pour chaque caractéristique de l'entrée en fonction de leur impact sur la décision de classification finale. Nous avons évalué les scores de pertinence pour les échantillons EEG pré-ictaux, collectés auprès de sept patients atteints de crises frontales focales, correctement et incorrectement classifiés. Pour afficher la représentation topographique, les scores de pertinence ont été moyennés dans le temps. Les résultats ont montré que les caractéristiques extraites des canaux de la région source de la crise étaient les plus pertinents pour classer les segments pré-ictaux. En effet, les échantillons correctement classifiés avec une forte valeur de prédiction ont des scores de pertinence élevés dans les régions frontales où la crise se produira, alors que les échantillons incorrectement classifiés ont une distribution plus large des scores de pertinence dans toutes les régions du cerveau. En conclusion, l'architecture proposée améliore non seulement les performances du réseau de neurones, mais aussi facilite et clarifie son interprétation.

4.1 Abstract

Deep learning has served pattern classification in many applications, with a performance which often well exceeds that of other machine learning paradigms. Yet, in general, deep learning has used computational architectures built, albeit partially, by ad hoc means, and its classification decisions are not necessarily interpretable in terms of knowledge relevant to the application it serves. This is often referred to as the black box problem, which in certain applications, such as epileptic seizure prediction, can be a serious impediment. The purpose of this study is to investigate an interpretable deep learning classifier for epileptic EEG-driven seizure prediction. This neural network is interpretable because its layers can be visualized and interpreted as a result of a novel architecture where the learned weights follow from signal processing computations such as frequency sub-band and spatial filters. Consequently, the extracted features are no longer abstract as they correspond to the features commonly used for decoding EEG data. In addition, the network uses layer-wise relevance propagation to reveal pertinent features which can further explain the computations leading to the decisions. In seizure prediction experiments using the CHB-MIT data set, the method produced classification results which improved on the state-of-the art, with first network layer filters corresponding to clinically relevant frequency bands, and the input channels in the brain location in which the seizure originates contributing most significantly to the network predictions.

Keywords : Epileptic seizure prediction; Deep neural networks; Interpretable decisions; EEG signal.

4.2 Introduction

Deep neural networks have extended considerably the ability of common neural networks to learn and classify patterns, with striking, unprecedented results in long standing applications, and in challenging new ones as well LeCun *et al.* (2015). However, in many other important applications, such as EEG signal classification for epileptic seizure prediction, which is the subject of this study, pattern feature learning in deep neural networks, or deep learning (DL) Goodfellow *et al.* (2016), suffers from what is often referred to as the *black box* problem, where, in general, some prevalent network architecture is used, without explicit justification, to have its parameters learned from data by often adhoc trial and error experimentation. As a result, better classification is often missed.

Moreover, even when the classification is accurate, the results come essentially with no interpretation of how the network reached its classification decisions. The ability to interpret these decisions may not be an issue in simple classification tasks where a wrong outcome is of little consequence. However, in general, interpretation aids in learning better network parameters, for instance by biasing the network structure to favor learning of application-relevant features. In domains such as healthcare, it may be essential to develop efficient applications relevant in clinical settings. In this study, we consider a neural network to be *interpretable* at two levels: First, by designing layers that bias learned filters toward common signal processing computations, such as frequency sub-band and spatial filtering, which are relevant to the application that the network serves and, second by explaining the influence of the various input variables, or of the network learned features, in reaching the classification decisions. For instance, the deep neural network we investigate in this study for epileptic seizure prediction is interpretable in that it uses a convolution layer similar to a filter bank to extract characteristic filters corresponding to clinically relevant frequency bands, and the input channels in the brain location in which the seizure originates contribute most significantly to the network predictions.

The ability to interpret classifiers has been of general interest in Artificial intelligence and has first appeared in symbolic reasoning which supports decision making in expert systems, such as MYCIN Buchanan & Shortliffe (1984) which sets to diagnose patients on the basis of reported symptoms and medical tests results, and also GUIDON Clancey (1987) for knowledge based tutoring, and SOPHIE Brown (1982) for hazard identification. In machine learning as well, being able to explain the relationship between the input and output of predictive models, and interpret the outcomes of this relationship, has been a concern Breiman (2001b). Notable contributions in this regard are decision trees and related classifying structures, considered interpretable from the flow of their successive decision making stages Breiman (2001a).

Interpretable neural networks are of relatively recent interest Montavon *et al.* (2018), although the need for these is quite plain because, as mentioned earlier, neural networks, particularly deep neural networks, are currently used mostly in a black box manner, to process a multitude of classification applications which would be more accurate and of more flexible and general usage were their internal architecture are interpretable. Noteworthy investigations include studies and applications of saliency maps to visualize and understand non-linearity in neural networks Morch *et al.* (1995), and computer vision studies which related neural activity to image filtering Hinton *et al.* (2006).

The investigation in Hinton *et al.* (2006) was able to determine that the first-layer filters learned by a deep belief network for natural image patch recognition is analogous to location, orientation, and spatial frequency filters like Gabor filters used for edge detection. For digits recognition using the MNIST dataset, Larochelle *et al.* (2009) observed that the network learned low-level features similar to those found in stroke detectors typically used for text localization. For accrued interpretation abiliy Gilpin *et al.* (2018), an explanation producing model can be construed using an architecture designed so as to simplify interpretations of internal representations and corresponding processing.

A means to explain a deep neural network computations is the layer-wise relevance propagation (LRP) scheme Bach *et al.* (2015), by which a network decision is decomposed into relevance scores for each neuron, starting from last layer and propagating back towards the input. LRP has been a useful explanation tool in many applications. For instance, Becker *et al.* (2018) used LRP to explain digit recognition and gender classification using the AudioMNIST dataset, which contained spoken digit records. The spectrogram representation showed that different areas of the input were critical to each class for digit classification, and the low frequency range was a determinant for gender classification. Moreover, based on waveform data, large magnitude data were determined also important. In another study Lawhern *et al.* (2018) LRP was used to explain the classification of subjects EEG recordings while imagining left and right hand movements. The relevance score was calculated for each channel at each time point. The channel relevance of a time point reveals a typical lateralized motor activation pattern, which, when averaged over all epochs, yields a similar pattern.

In spite of the evident progress in building interpretable pattern classifiers for several important image-based applications, the subject remains actual and challenging for waveform data. Little work has been done in EEG-based applications, and none in epilepsy seizure prediction, the subject of this study. This study takes up the problem of developing an interpretable deep learning neural network applicable to epileptic seizure prediction in EEG recordings. Epilepsy seizure prediction is a subject worthy of investigation because epilepsy affects about 2.4 million people of all ages worldwide each year who (2019) and involve seizures with the risk of periodic disruptions in cognitive and behavioral functions. Predicting seizures would obviously benefits patients significantly, and also lighten physicians workload. Electroencephalography (EEG), which involves recording brain activity with electrodes placed on the scalp, has proven to be a reliable non-invasive clinical approach for epilepsy diagnosis. Predicting the eventual occurrence of seizures relies on identifying the pre-ictal period prior to the onset of a seizure Mormann *et al.* (2005), during which EEG recordings show different patterns from the patterns of the seizure and also from the earlier periods, so-called inter-ictal periods. As a result, the classification of inter-ictal and pre-ictal states simplifies the prediction of seizures. This study provides three contributions to the field: 1. Design of an architecture following the filter bank common spatial pattern (FBCSP) paradigm and build an explanation-producing model that biases learned filters toward relevant common sub-band frequency and spatial filters, 2. interpretation of the network abstract features encoding, by learned filters visualization and, 3. explanation of the network model decisions by layer relevance propagation. We tested the model on the CHB-MIT dataset for epilepsy prediction and its results outperformed those of the current state of the art. The model architecture showed a fair interpretability. Indeed, we found that the first layer trained filters gather data from specific frequency bands . Explanation of the model's decisions for several trials of patients with focal seizures reveals that the input channels in the brain location from which the seizure originates contribute most to the model's prediction.

The remainder of this chapiter is organized as follows: Section 4.3 describes the data set and the proposed architecture; Section B.5 details the experimental results, and Section 4.5 contains a discussion.

4.3 Materials and methods

The functional diagram of the seizure prediction task is illustrated in Figure 4.1. The proposed framework consists of three main steps: the first step consists of pre-processing and segmentation of the data(Section II-A). This is followed by training and evaluation of the neural network (Section II-B). The resulting models are interpreted by visualizing the learned filters and explaining the model decision for several trials (Section III).

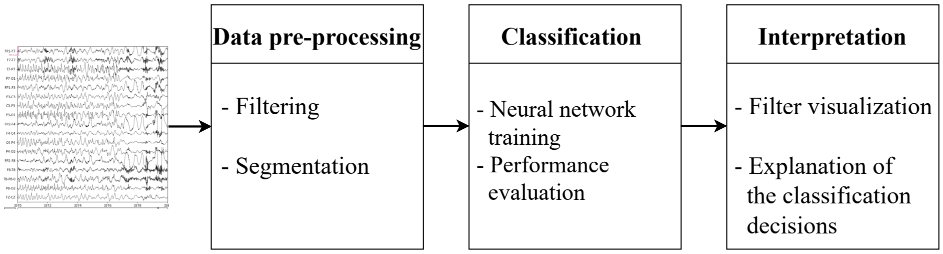


Figure 4.1 – The functional diagram of the seizure prediction task.

4.3.1 Dataset description and pre-processing

The dataset used for this study is the publicly available CHB-MIT dataset collected at the Boston Children's Hospital Shoeb (2009). CHB-MIT contains 940 hours of long-term continuous multi-channel scalp EEG recordings collected from 23 pediatric patients aged 1.5 to 19 years as shown in Table 4.1. A minimum of 17 electrodes was used in all trials distributed according to the international standard 10/20 system. The sampling rate was set to 256Hz. Using a notch filter with an upper cutoff frequency of 50Hz and a band-pass filter with a bandwidth of 0.5-70Hz, we eliminated noise and artifacts and focused on relevant frequencies. Based on published literature, we set the pre-ictal period to be 30 minutes before the onset of the seizure, as outlined in Teixeira *et al.* (2014); Bandarabadi *et al.* (2015b), and eliminated 30 minutes after the end of the seizure to exclude effects of the post-ictal periods. Subsequently, we divided the recordings into non-overlapping 5-second-windows yielding 529,415 and 66,782 samples of inter-ictal and pre-ictal activity respectively.

Tableau 4.1 – An overview of the CHB-MIT dataset.

Patient	Sex	Age	Seizure type*	Origin	# Seizure
1	F	11	SP, CP, GTC	Temporal	6
2	M	11	SP, CP	Frontal	3
3	F	14	SP, CP, GTC	Temporal	7
4	M	22	SP, CP, GTC	Temporal, Occipital	4
5	F	7	SP, CP	Frontal	4
6	F	1.5	SP, CP, GTC	Temporal	7
7	F	14.5	SP, CP, GTC	Temporal	3
8	M	3.5	SP, CP, GTC	Temporal	5
9	F	10	SP, CP	Frontal	4
10	M	3	SP, CP, GTC	Temporal	7
11	F	12	SP, CP	Frontal	3
12	F	2	SP, CP	Frontal	24
13	F	3	CP, GTC	Temporal, Occipital	12
14	F	9	SP, CP, GTC	Temporal	8
15	F	16	CP, GTC	Frontal, Temporal	20
16	M	7	SP, CP, GTC	Temporal	10
17	F	12	SP, CP, GTC	Temporal	3
18	F	18	CP, GTC	Temporal, Occipital	5
19	F	19	SP, CP	Frontal	3
20	F	6	SP, CP, GTC	Temporal	8
21	F	13	SP, CP, GTC	Temporal	4
22	F	9	CP, GTC	Temporal, Occipital	5
23	F	6	SP, CP	Frontal	8

*SP: simple partial, CP: complex partial, GTC:generalized tonic-clonic.

4.3.2 Neural network architecture

The deep neural network architecture uses the Filter Bank Common Spatial Pattern (FBCSP) algorithm Ang *et al.* (2008) as follows.

FBCSP aims at finding spatial filters that map the raw data into additive components that are capable of discriminating between the sources more efficiently. FBSCP is widely used for decoding EEG data in different applications, such as brain-computer interfaces experiments Schirrmeister *et al.* (2017), mental workload estimation Arvaneh *et al.* (2015), major depression detection Liao *et al.* (2017), and epilepsy prediction Alotaiby *et al.* (2017); Zhang *et al.* (2019b). The algorithm is composed of two main components: (1) A filter bank and (2) Spatial filtering using the Common Spatial Pattern (CSP) algorithm.

- The Filter bank consists of a set of band-pass filters that separates the input signal into multiple signals, each corresponding to a unique frequency sub-band of the original input.
- The spatial filters are linear transformations which project raw channel data into a spatial space known as "source space" to separate sources of activity Koles *et al.* (1990). The CSP algorithm computes a transformation matrix which maximize the variance of the output signal for one class and minimize it for the other.

The outputs of the algorithm are generally used to extract features such as the log-variance of each sub-component in all sub-frequency bands. These features are then used for the classification task.

In this study, we followed the steps of this algorithm to design a neural network architecture that simplifies the interpretation of its layers. The use of FBSCP-inspired architectures for different applications has been very little studied before. The study by Schirrmeister *et al.* (2017) suggested a convolutional neural network (CNN) with convolutional layers similar to the bandpass and spatial filters used to decode and visualize task-related information from EEG recordings. An alternative similar compact architecture Lawhern *et al.* (2018) has been used to classify EEG signals from different brain-computer interface paradigms. Neither of these studies considered long, continuous EEG data. Instead, they used the simpler data of event-related potentials (ERPs), which are brain responses to a specific sensory, cognitive, or motor events. Chambon *et al.* (2018) have investigated a convolutional neural network architecture whose first layer function is equivalent to spatial filtering for sleep-stage classification using multivariate, multimodal continuous time-series data. However, these studies did not include classification decisions explanations or interpretations. In this work, we present an architecture that is designed to take into account the type of long continuous EEG data. Moreover, a number of interpretations and explanations have been provided to delve further into the architecture.

Figure 4.2 and Table 4.2 show the overall diagram and the full detailed description of the proposed architecture.

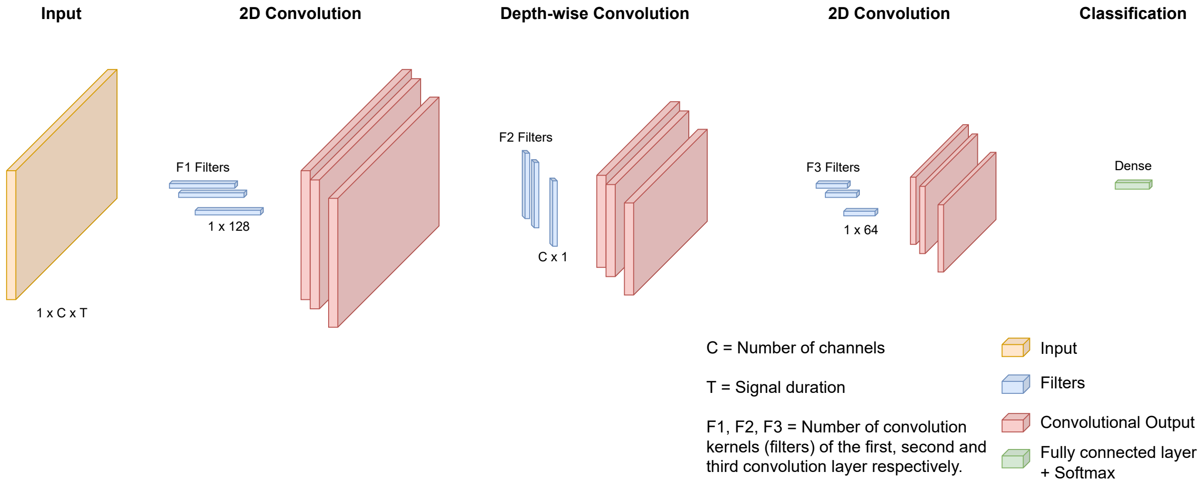


Figure 4.2 – Diagram of the proposed architecture. The network processes EEG inputs with standard convolution with filters of shape (1,128) allowing learning frequency filters. It uses depth-wise convolution to learn spatial filters for each feature-map output of the previous layer separately. Finally 2D convolution is used to extract features. The outputs of the pipeline are finally fed to a fully connected layer.

— The network first layer performs a standard temporal 2D convolution that learns a set of band-pass filters to output multiple components, each representing a frequency band within the original signal. This step is equivalent to the filter bank stage of the FBCSP algorithm. With a temporal kernel that is half the sampling frequency, it is possible to capture frequency information starting at 2 Hz. Following this operation, the batch normalization is used to stabilize the training.

— The subsequent component of the architecture starts with a depth-wise convolution, which is the application of convolution filters to each feature map (output from the previous layer) independently from the other maps. To learn spatial filters, this type of convolution is implemented using kernels of shape (C, 1) where C is the number of channels. Subsequently,

Tableau 4.2 – The detailed architecture of the network, where C = number of channels, T = signal duration, F1 = number of convolution kernels filters to learn frequency filters, F2 = number of convolution kernels to learn spatial filters, F3 = number of convolution kernels for feature extraction, N = number of classes, respectively.

Layer	#Filter	Filter size	#Parameters	Output	Activation
Input				(1,C,T)	
2D Convolution	F1	(1,128)	128 * F1	(F1,C,T)	Linear
Batch normalization	2*F1			(F1,C,T)	
Depth-wise convolution	F2*F1	(C,1)	C*F2*F1	(F2*F1,1,T)	
Batch normalization			2*F2*F1	(F2*F1,1,T)	
Activation				(F2*F1,1,T)	Relu
Average-pooling		(1,16)		(F2*F1,1,T//16)	
Dropout				(F2*F1,1,T//16)	
2D Convolution	F3	(1,64)	64*F3	(F3,F2*F1,T')	Linear
Batch normalization			2*F3	(F3,F2*F1,T')	
Activation				(F3,F2*F1,T')	Relu
Average-pooling		(1,16)		(F3,F2*F1,T'//16)	
Dropout				(F3,F2*F1,T'//16)	
Linear (flatten)				(F3*F2*F1*(T'//16))	
Dense				N = 2	Softmax

batch normalization, non-linear activation and average pooling were consecutively applied. A dropout layer is added to regularize the model.

- For the feature extraction from the activity source signals learned in the previous layers, we applied a combination of 2D convolutional layers, a nonlinear activation layer and an averaging layer. The outputs are finally passed through a dropout layer.
- The last stage of the architecture is a fully connected layer that flattens the features into a one-dimensional vector that is fed to a Softmax classifier.

4.3.3 Neural network interpretation

Filter visualization

In neural network terminology, the learned filters are simply the weights of the convolutional kernels of the network. Visualizing the learned filters allows us to see how each layer extracts information from the input. Generally, for standard convolution layers, interpreting the filters proves challenging since it performs both an in-channel and in-space computation at the same time. Using the 2D convolutional filter of size (1, 128) and the depth-wise convolution to learn filters for each input channel separately, it is possible to interpret time convolution as band-pass frequency filters and depth convolution as spatial filters.

Layer-wise relevance propagation

The RLP technique explains neural network decisions by assigning a score to each input point (e.g., pixels) related to its relevance to the classification decision. RLP is based on back-propagating the prediction score $f(x)$ according to specific propagation rules (see Figure 4.3). The prediction score is redistributed from the output layer down to the neurons of the lower layer and so forth until it reaches the input layer. For each point in the input layer, the relevance score corresponds to its contribution to the decision. A high relevance score indicates a relevant pattern. On the other hand, parts of input with a low relevance score are considered irrelevant. Let $f(x)$ and $R_j^{(l)}$ be the prediction score and relevance score of the neuron j in layer l respectively. Any propagation rule satisfying the following properties could be used for PRL. First, the relevance must be preserved between layers, so that the following equation is verified.

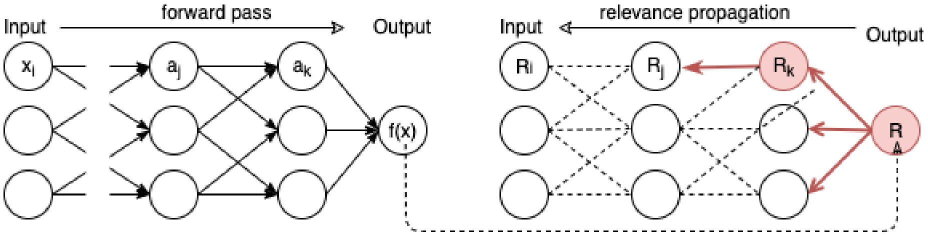


Figure 4.3 – Diagram of the LRP technique. The prediction score $f(x)$ is first computed through the forward pass. Then, it is back-propagated from the output layer to the input layer according to specific rules. The scores obtained on the input layer indicate the contribution of each feature in the classification decision.

$$f(x) = ... = \sum_{d \in l+1} R_d^{(l+1)} = \sum_{d \in l} R_d^{(l)} \tag{4.1}$$

Moreover, the relevance score of a node should be equal to the sum of the relevance score coming from nodes in upper layers and redistributed in same amount to nodes in lower layers as indicated in Equation 4.2 .

$$R_j^{(l)} = \sum_k R_{k \leftarrow j}^{(l,l+1)} \text{ and } R_k^{(l+1)} = \sum_j R_{j \leftarrow k}^{(l,l+1)} \tag{4.2}$$

where $R_{j \leftarrow k}^{(l,l+1)}$ is the relevance score sent from the neuron k in layer l to the neuron i in the next layer $l+1$.

Finally, the propagation rule must ensure that the relevance scores are related to the neuron activation or inhibition; a positive score corresponds to the existence of a feature whereas a negative or null score indicates to the absence of a pattern. There are several propagation rules that have proven effective in practice that satisfy the constraints listed above.

Basic rule $(LRP-0)$: This intuitive rule redistributes the relevance score in proportion to the contribution of each input to the neuron's pre-activation.

$$R_j = \sum_k \frac{a_j w_{jk}}{\sum_j a_j w_{jk}} R_k. \tag{4.3}$$

where a_j is the neuron activation from the previous layer and w_{jk} is the weight of the connection from unit j to unit k.

Epsilon rule $(LRP-\epsilon)$: To ensure that R_j does not take unbounded values for small or null values of neuron activation, a positive term ϵ is added to the denominator.

$$R_j = \sum_k \frac{a_j w_{jk}}{\epsilon + \sum_j a_j w_{jk}} R_k. \tag{4.4}$$

where a_j is the neuron activation from the previous layer and w_{jk} is the weight of the connection from unit j to unit k. **Gamma rule $(LRP-\gamma)$** : The $LRP-\gamma$ rule denoted by the equation 4.5 is used to highlight the positive contributions over the negative contributions. The parameter γ controls the importance of positive evidence.

$$R_j = \sum_k \frac{a_j (w_{jk} + \gamma w_{jk}^{+})}{\sum_j a_j (w_{jk} + \gamma w_{jk}^{+})} R_k. \tag{4.5}$$

where a_j the neuron activation from the previous layer, w_{jk} is the weight of the connection from unit j to unit k and w_{jk}^{+} is the positive part of the weight.

As suggested in Montavon *et al.* (2019) we used $LRP-0$ rule for the upper layers, the Epsilon rule for the middle layer and the Gamma rule for the lower layers.

4.3.4 Classification and implementation details

As mentioned earlier, an inter-ictal and pre-ictal segments classification could simplify seizure prediction. However, epileptic patterns vary widely from seizure to seizure as well as from patient to patient, which makes binary classification challenging. Seizure prediction can be performed through general cross-subject models applicable to all patients or by patient-specific modeling applicable to each patient individually. Models that are patient-specific are generally impractical since it requires recording a sufficient number of seizures for each patient. Cross-subject modeling does not require treating each patient separately, but it faces the major challenge of adapting the prediction algorithm to unseen data from new patients, mainly due to the high variability of cross-subject EEG patterns Jemal *et al.* (2021).

In this study, we focused mostly on the patient-specific modelling to evaluate the proposed architecture. Accordingly, a single architecture was designed and trained for each subject separately. Pytorch Paszke *et al.* (2017) was used to implement the proposed architecture. For data pre-processing The MNE-Python package Gramfort *et al.* (2013) was utilized. To ensure reliable generalization performance, a 5-fold stratified cross-validation test setup was used . A holdout validation is nested within a cross-validation procedure in order to further divide the training set of each fold into a validation set and a training set so that the early stopping criteria can be enforced to prevent over-fitting. In fact, the training runs up to 500 epochs, or until the validation loss remains constant for at least 20 epochs. Across all tasks, we use the gradient-based ADAM optimizer with coefficients β_1, and β_2 of 0.9 and 0.999 respectively because it is fast and reliable for reaching a global minimum. We use a learning rate of 0.005 and a dropout regularization value of 0.25.

4.4 Results

In the following, we applied the proposed architecture to the classification of inter-ictal and pre-ictal brain states for seizure prediction. After the model training, the next step is to visualize the learning filters and explain the decisions made by the neural network using LRP-based technique.

4.4.1 Patient-specific seizure prediction

Using the CHB-MIT data, we evaluated the proposed architecture for specific-subject seizure prediction on 23 patients. Table 4.3 shows some performance measures such as prediction accuracy, sensitivity, specificity, precision, F1-score, Area under the ROC Curve(AUC) and false alarm per hour for each patient model. Across all patients, the overall averaged accuracy, sensitivity, specificity and F1-score across all patients are 90.9%, 96.1%, 84.6%, and 91.9%, respectively. An averaged area under the ROC curve of 0.918% was achieved. The models have a reasonably low averaged false prediction rate per hour(FPR/h) of 0.041 indicating good predictive power.

Tableau 4.3 – The performance of the proposed architecture on the 23 patients of the CHB-MIT dataset.

Patient ID	F1	Accuracy	Sensitivity	Specificity	Precision	FPR(h^{-1})	AUC
1	97.4	97.1	99.9	93.6	95.0	0.010	0.974
2	95.9	95.7	99.3	91.8	92.8	0.013	0.960
3	98.8	98.8	99.7	97.9	98.0	0.003	0.988
4	92.0	91.5	96.4	86.3	88.0	0.005	0.920
5	86.2	85.2	90.0	80.0	82.7	0.032	0.855
6	88.8	86.2	95.9	73.6	82.7	0.109	0.879
7	94.1	93.7	98.0	89.2	90.6	0.009	0.941
8	86.9	84.4	91.8	75.0	82.5	0.089	0.851
9	89.6	88.2	96.0	79.5	84.1	0.018	0.894
10	91.7	90.9	96.0	85.4	87.8	0.017	0.915
11	99.3	99.3	100.0	98.6	98.7	0.002	0.993
12	79.9	78.8	87.0	65.5	82.3	0.181	0.855
13	94.7	94.0	98.1	89.3	91.4	0.020	0.945
14	77.9	75.2	82.0	67.4	74.3	0.086	0.755
15	82.9	79.9	90.1	68.2	76.9	0.078	0.813
16	88.8	86.2	95.9	73.6	82.7	0.109	0.879
17	97.3	97.1	99.2	94.9	95.4	0.013	0.973
18	96.4	96.2	98.7	93.5	94.2	0.011	0.964
19	99.5	99.5	100.0	99.0	99.1	0.002	0.995
20	99.0	98.9	99.8	98.0	98.1	0.005	0.990
21	90.4	89.3	98.4	79.8	83.7	0.035	0.908
22	90.5	89.0	98.5	78.2	83.7	0.038	0.908
23	95.6	95.1	99.8	89.7	91.7	0.024	0.957
Average	91.9	90.9	96.1	84.7	88.5	0.040	0.918

For further evaluation of the proposed architecture, we compared our results to earlier publications that employed the same dataset, as given in Table 4.4. In the previous works considered, CNN architectures with different numbers of layers were used, such as the single-layer architecture as in Zhou *et al.* (2018), the three-layer architecture as in Truong *et al.* (2018) and the five-layers architecture as in Zhao *et al.* (2020). The authors of Zhang *et al.* (2019b) used the FBCSP algorithm followed by a CNN classifier. The proposed architecture achieved the highest sensitivity with the lowest false alarm rate.

Tableau 4.4 – Comparison to prior works on epileptic seizure prediction using the CHB-MIT dataset.

Article	Data	Method	Accuracy	Sensitivity	Specificity	FPR(h^{-1})	AUC
Truong *et al.* (2018)	CHB-MIT	CNN	-	81.2		0.16	-
Zhou *et al.* (2018)	CHB-MIT	CNN	95.6	94.2	96.9	-	-
Zhang *et al.* (2019b)	CHB-MIT	CSP+ CNN	90	92	92	0.12	0.90
Zhao *et al.* (2020)	CHB-MIT	Bi-CNN	-	94.69	-	0.095	0.97
This work	CHB-MIT	CNN	90.9	96.1	84.7	0.040	0.918

4.4.2 Filter visualization

Following model training, the interpretation of the learned data representation was conducted. As previously stated, the proposed architecture's first layer is supposed to be equivalent to a filter-bank. As a result, we are especially interested in seeing if the model was able to learn band-pass frequency filters. Therefore, for all subjects we visualized the filters learned on the first layer of the various patient-specific models. The convolution filter for the first layer of patient 20's model, as well as the frequency domain representation derived using the Fast Fourier Transform (FFT) are shown in Figure 4.4. The frequency bandwidths were calculated using the FFT. Figure 4.5 shows the frequency bandwidths of the seven learnt first layer filters in each of the 23 patients' subject-specific models.

4.4.3 Explaining model decision

The LRP technique is used to explain the model decision for many samples, which is the third level of interpretability explored in this work. As outlined in Section 4.3, LRP computes, on a sample basis, the relevance scores for individual features related to their contribution to the ultimate classification decision. Positive relevance values suggest features that support the classification

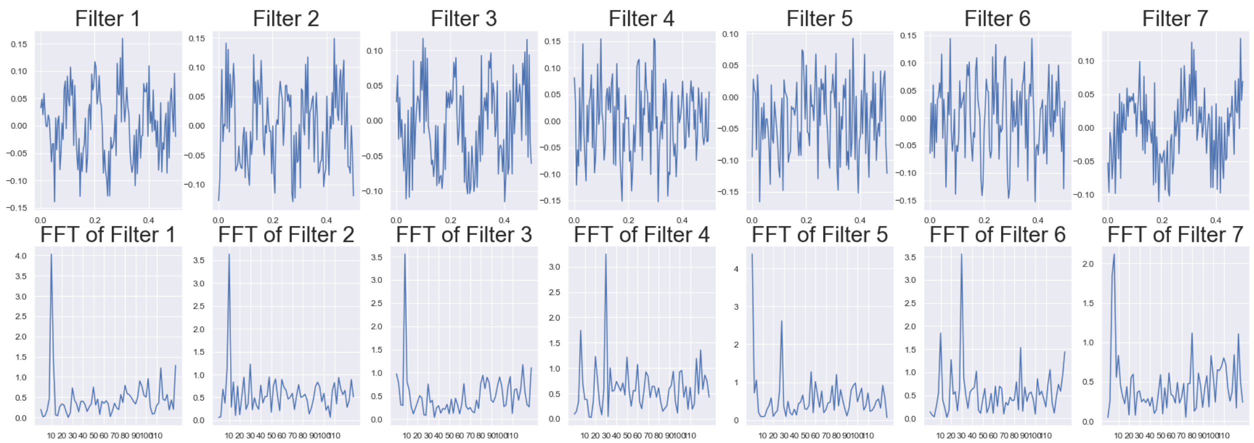


Figure 4.4 – Visualization of the learned convolution filters of the first layer of the patient 20's model. Top row shows the temporal kernels of shape (1,128) for a 0.5 window. Bottom row display the FFT calculated for each filter to determine the frequency bandwidths.

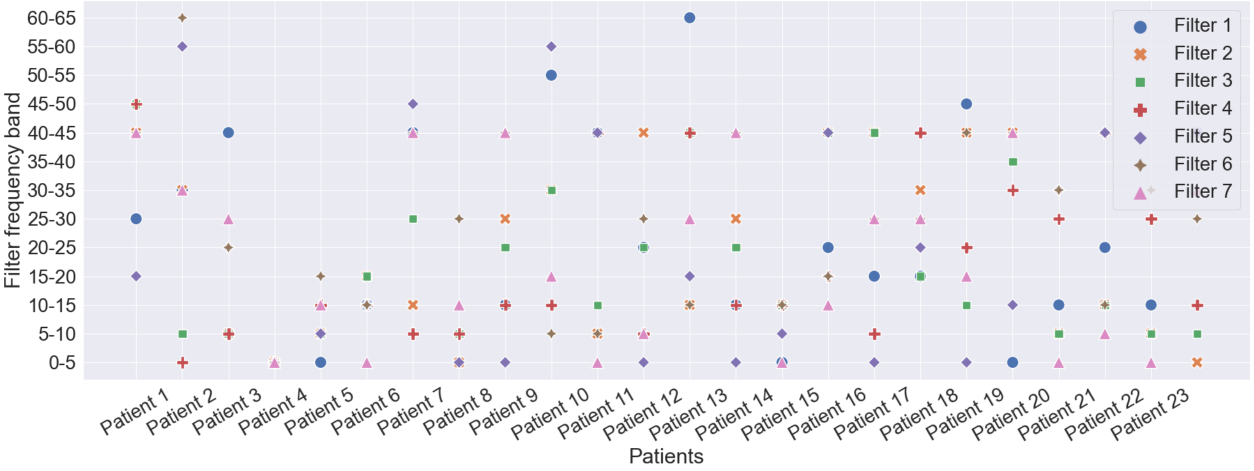


Figure 4.5 – The frequency bandwidths of the seven learned filters of the first layer in all subject-specific models of the 23 patients.

decision, whilst negative values indicate features that are irrelevant to the prediction. The relevance scores of individual features for successfully and inaccurately detected pre-ictal EEG samples were determined in this study. To display the topographic map, the relevance scores were averaged across time. Figure 4.6 shows the topographic representation of the relevance scores for various pre-ictal samples from seven patients with focal frontal seizures.

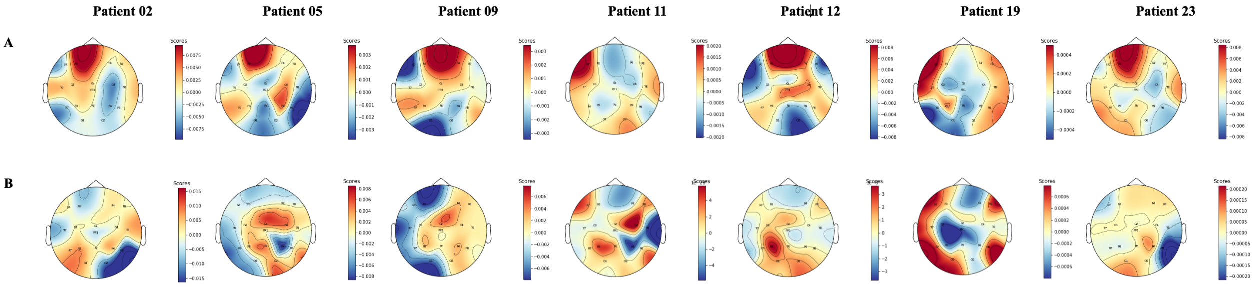


Figure 4.6 – Topographical representation of relevance scores for various pre-ictal samples from seven patients with focal frontal seizures. A high relevance score implies a relevant feature, whereas a low score indicates an irrelevant input. The top row A shows the relevance scores of the correctly classified samples. The bottom B row shows the relevance scores of the miss-classified segments.

4.4.4 Cross-subject seizure prediction

To evaluate the patient-independent model we used all of the data from the 23 patients at CHB-MIT. Thus, we divided the dataset into three stratified sets with the same proportions of classes : the training set, the validation set, and the test set. The proposed architecture yields satisfactory results. With a false prediction rate of 0.6/h, we were able to achieve a sensitivity of 67.17%. An F1 score of 65.84% was achieved.

4.5 Discussion and conclusion

This study investigated an interpretable deep learning model for seizure prediction using EEG signals. Its evaluation was conducted in three steps.

As a first step, we created an interpretable deep learning architecture whose earlier layers act according to the FBSCP scheme. The architecture was tested with a patient-specific seizure prediction task using the CHB-MIT dataset. The proposed architecture achieved a reasonably high level of prediction accuracy. Table 4.4 shows the benchmark of recent seizure prediction methods. Because these methods have been evaluated according to different metrics, the proposed classifier has been evaluated using several metrics commonly used in seizure prediction. From a clinical perspective, it is desirable to have a high sensitivity and a low false alarm rate. Authors of Truong *et al.* (2018) proposed a three-layer CNN architecture that yielded a sensitivity of 81.% and FPR of 0.16/h as tested with 13 patients from the CHB-MIT dataset. The study in Zhou *et al.* (2018) adopted a more compact single-layer CNN which performs much better, giving a better sensitivity of 94.2% a high accuracy of 95.6%, and a specificity of 96.9%. However, the evaluation is incomplete because no false alarm rate was reported. Another patient-specific CNN classifier has been described in Zhang *et al.* (2019b). The FBCSP algorithm was applied prior to the feature extraction step. The authors reported accuracy, specificity and sensitivity values of 90%, 92%, and 92%, respectively, with a relatively low false alarm rate of 0.12/h. A more advanced approach **?** used a five-layer one-dimensional binary convolutional neural network. They tested the model on only 5 patients from the CHB-MIT database and their seizure prediction sensitivity averaged 94.96% with an FPR of 0.096/h. Based on the results of evaluation on all 23 patients of the CHB-mit dataset, our model reaches the highest sensitivity of 96.1 with the lowest false alarm rate of 0.041. The overall average of the areas under the receiver operating curve was 0.91. The model not only improves the results of the recent others but also enhances and simplifies its interpretation.

Our next focus was on the interpretation of the learned filters. We found that the first layer's filters were found to be similar to band-pass frequency filters and moreover, each patient-specific model has its own set of filters. Additionally, similar filters, such as those in ranges 0-5, 5-20, and 40-45 appear frequently in almost all subjects as shown in Figure 4.5. The model learns low- and high-frequency filters in the range 0 to 60 Hz range for each subject, which is critical for the epilepsy prediction task, since abnormal seizure discharge is primarily observed in the 5 to 50 Hz frequency range. Likewise, we found that, the models recover essential features of each patient by learning filters with a frequency of 25 Hz or higher, which is consistent with the fact that epilepsy prediction relies more on characteristics in the gamma band (30-140 Hz) that are more relevant than other bands for epilepsy predictionPark *et al.* (2011).

Finally, LRP enabled us to interpret several of the classification decisions. We found that features extracted from the channels in the region of the seizure origin were shown to be the most relevant features for pre-ictal segments classification. Hence, we determined that well classified samples with a high prediction value (Figure 4.6 top row) possess high relevance scores in the frontal regions where the seizure will occur, while misclassified samples (4.6 bottom row) displayed a distribution of relevance scores more broadly distributed throughout the scalp.

Since EEG data vary greatly between subjects and only a few patients are available, developing patient-independent models is a complex task. Therefore, most researchers simplified the problem to develop models that are patient-specific. To our knowledge, this is the first study to examine between-subjects modeling. The proposed architecture yields satisfactory results when tested on the entire dataset of the 23 patients of the CHB-MIT dataset. However, due to the substantial variability of EEG data between patients, cross-subject seizure prediction performed somewhat worse than patient-specific modelling, but the model appears to have potential applicability to data from unknown subjects.

In summary, we introduced a novel interpretable neural network architecture to simplify its opaque representation of data. The proposed architecture is based on the common FBCSP paradigm where its layers where correspond to known signal processing calculations, such as sub-frequency band and spatial filtering. The architecture performance was evaluated using the CHB-MIT dataset for the patient-specific prediction task. The proposed architecture has achieved a reasonably high predictive accuracy compared to other deep learning methods. Next, the model was interpreted by visualizing the learned filters, showing that the first-layer filters are similar to the band-pass filters. Finally, using the LRP, we were able to explain several model decisions. We observed that for the pre-ictal segments, the channels in the seizure origin region were the most relevant characteristics for classification.

The study could be strengthened by using larger amounts of data and using different EEG databases. Furthermore, it is highly useful to study how to transfer learning for cross-patient modelling, which could help to learn new representations shared between-data subjects that would transfer knowledge gained from multiple patients to new unseen patients. Finally, the proposed architecture and explanation scheme can be applied to other EEG-based classification tasks, such as seizure diagnosis and seizure type categorization, as well as autism and Alzheimer’s disease detection.

Chapitre 5

Article 3 : Domain adaptation for deep learning of EEG-based, cross-subject epileptic seizure prediction

Imene Jemal[1,2], Neila Mezghani[2,3], Lina Abou-Abbas[2,3], Khadidja Henni[2,3] and Amar Mitiche[1]

[1] Centre ÉMT, Institut National de la Recherche Scientifique, Montréal, Canada

[2] Centre de Recherche LICEF, Université TÉLUQ, Montréal, Canada

[3] Laboratoire LIO, Centre de Recherche du CHUM, Montréal, Canada

Submitted to : Neural Computing and Applications

Résumé : La capacité de prédire une crise d'épilepsie est une protection contre les blessures et les complications de santé des patients. La principale difficulté de la prédiction de la crise d'épilepsie provient de la variation considérable des données entre les patients. Il est donc difficile de développer des modèles qui peuvent prendre en compte cette variabilité pour fonctionner correctement lorsqu'elles sont développées pour certains patients mais appliquées à d'autres. Il existe trois types de modèles: (1) le modèle spécifique au sujet qui est conçu pour chaque patient et utilise une partie

de ses données pour l'apprentissage de modèle et le reste pour l'évaluation ; (2) le modèle multi-sujets, également appelé modèle indépendant du patient, qui est appliqué à un ensemble dédié de patients, mais pose le défi de l'adaptation du modèle à de nouveau patient ; (3) le modèle inter-sujets, qui utilise des données de plusieurs patients et peut être appliqué à de nouveaux patients. Les méthodes courantes sont spécifiques aux sujets et n'ont pas la propriété de généralisation à des nouveau patients.

Le but de cette étude est d'examiner les modèles multi-sujets et inter-sujets qui permettent une meilleure généralisation aux nouveaux patients que les modèles spécifiques aux sujets. Nous avons commencé par appliquer notre architecture de réseau de neurones pur un modèle de prédiction multi-patients. L'architecture de réseau de neurones est ensuite adaptée pour la prédiction inter-sujets en utilisant la stratégie Leave-One-Subject-Out, fournissant ainsi un contexte d'application plus large et plus réaliste. La modélisation inter-sujets de la prédiction des crises est nécessaire en raison de la significative variabilité de données entre différents patients. Les données d'un nouveau patient peuvent différer significativement des données utilisées pour l'apprentissage du modèle inter-sujets, entraînant un problème connu sous le nom du décalage de domaine, ce qui entraîne souvent une baisse des performances de classification. Pour remédier à cela, l'adaptation de domaine est employée. Dans cette étude, on a examiné trois algorithmes d'adaptation de domaine basés sur les caractéristiques: Discriminative Adversarial Neural Network (DANN), Domain Adversarial Conditional Adaptation (CDAN) et CDAN+E, la variante d'entropie de conditionnement de CDAN pour améliorer les performances du modèle inter-sujets.

L'évaluation des modèles multi-sujets et inter-sujets pour la prédiction de crises d'épilepsie a été effectuée sur les ensembles de données CHB-MIT et SIENA. Le modèle multi-sujets a obtenu d'excellents résultats, avec une précision de 96,01% sur l'ensemble de données SIENA et de 97,36% sur l'ensemble de données CHB-MIT. Ces résultats surpassent les modèles actuels évalués sur les mêmes ensembles de données. Le modèle inter-sujets a obtenu une performance moyenne de 48,69% de précision sur l'ensemble de données SIENA pour tous les patients. Les résultats étaient légèrement meilleurs pour l'ensemble de données CHB-MIT, qui avait un plus grand nombre de patients, avec une précision de 63,5%. La réduction de performance peut être attribuée au décalage entre les données des nouveaux patients et celles d'entraînement, un problème fréquent dans la prédiction de crises d'épilepsie en raison de la variabilité élevée des données des patients.

Afin d'atténuer le potentiel de décalage des données, nous avons exploré trois méthodes d'adaptation de domaine : DANN, CDAN et CDAN+E. Les résultats de l'évaluation de ces trois méthodes ont montré qu'elles ont toutes amélioré les performances sur les deux ensembles de données SIENA et CHB-MIT. La méthode CDAN s'est révélée particulièrement efficace avec une précision de 60,27% sur l'ensemble de données SIENA, tandis qu'une amélioration significative des performances a été observée sur l'ensemble de données CHB-MIT avec une précision de 70,90% en utilisant la méthode CDAN+E.

5.1 Abstract

The ability to predict the occurrence of an epileptic seizure is a safeguard against patient injury and health complications. The purpose of this study is to investigate EEG-based, deep learning of epileptic seizure prediction. In general, the main difficulty in epileptic seizure prediction stems from the considerable variation known to occur in the data of different patients, and the challenge is to develop methods that can account for this variability so as to perform well when these are developed for some patients but applied to others. Common, prevalent, patient-specific methods, which apply to each patient independently, do not have such generalization property. The property is dependent on the ability to process simultaneously the data of several different patients. This study addresses the problem first by investigating a new method of multiple-subject prediction by deep learning. In general, multiple-subject modeling broadens the scope of patient-specific modeling to account for the data from a dedicated ensemble of patients, thereby providing some useful, though relatively modest, level of generalization. The basic neural network architecture of this method is then adapted to cross-subject prediction using the leave-one-out strategy, thereby providing a broader, more realistic, context of application. For accrued performance, and generalization ability, cross-subject modeling is enhanced by domain adaptation so as to better account for data unseen during training. Experimental evaluation using the publicly available CHB-MIT and SIENA data datasets shows that the multiple-subject method of this study performs better than others, and provides a useful epileptic prediction application. Cross-subject processing experiments with and without domain adaptation, using the same datasets, expose the effect of data variability, and highlight the role and importance of domain adaptation.

Keywords : Epileptic seizure prediction; Deep learning; Domain adaptation; EEG.

5.2 Introduction

Epilepsy is a neurological disorder which causes recurrent seizures resulting from brain dysfunction. Symptoms of seizures vary greatly among patients and can range from brief disruptions in activity to loss of consciousness and severe convulsions. To diagnose epilepsy, physicians use electroencephalography (EEG), which records the electrical activity of the brain using electrodes placed on the skull. Studies have shown that there is a pre-ictal period, lasting several minutes before the onset of a seizure, during which EEG recordings display patterns that are different from those of the seizures and also from normal periods, called inter-ictalMormann *et al.* (2005). Thus, by distinguishing pre-ictal from inter-ictal states, it is possible to predict seizures.

Computer-aided models for seizure prediction can be grouped into three categories: (1) patient-specific modeling, which is tailored to each individual patient, (2) multiple-subject modeling, also called patient-independent modeling, which is applied to a dedicated set of patients, and (3) cross-patient modeling, a generalized model that uses data from multiple patients and can be applied to new, unseen patients.

1. **Patient-specific modeling** involves using a portion of a single patient's data for model training and the remaining data for performance evaluation. However, this type of model is not practical as it is limited by the amount of data available and the need to record a sufficient number of seizures for each individual patient.
2. **multiple-subject modeling**, also called patient-independent modeling, is a complex task that predicts seizures for subjects in a dedicated ensemble of subjects. This modeling does not have the limitation of lack of data as it utilizes all patient data grouped into at least two sets to train and test the model. However, this approach faces the major challenge of adapting the prediction model to new data from unseen patients.
3. **Cross-subject modeling**, also known as generalized modeling, is the most complex type for seizure prediction. This approach allows generalization to other patients and does not require labeled data for new patients in the training phase.

Research has mostly focused on multiple-subject modeling for seizure prediction Tsiouris *et al.* (2017); Khan *et al.* (2017); Dissanayake *et al.* (2021a). As far as we know, cross-subject seizure prediction has not been investigated, although cross-subject modeling has been successfully applied to other tasks, such as seizure detection Zhang *et al.* (2020), emotion recognition Li *et al.* (2018), and mental load assessment Albuquerque *et al.* (2019). Cross-subject modeling of seizure prediction is justified by the high variability between the data of different patients Jemal *et al.* (2022), because the data from a new patient may differ sensibly from that of patients whose data served to train the cross-subject model. This is often referred to as a domain shift Ben-David *et al.* (2010). It is common in real data applications, and can result in a significant drop in classification performance Ponce *et al.* (2006). To address this issue, domain adaptation can be used. In this context, a domain refers theoretically to the probability distribution from which the problenm data are drawn. The training dataset is called the source domain data and the test dataset is called the target domain data. Domain adaptation uses labeled source data and unlabeled target data to learn a model that performs well in both the target and source domains. This is generally achieved by re-weighting the source samples to minimize the distribution shift, so that the source samples closest to the target domain are given more importance Huang *et al.* (2006); Sugiyama *et al.* (2007). Alternatively, one can use a pre-trained model from the source domain on the new target domain Oquab *et al.* (2014). This approach, known as a parameter-based approach, can only be applied if some labeled data are available in the target domain. Another approach on which we focus in this study, called feature-based Daumé III (2009); Ganin *et al.* (2016); Long *et al.* (2018) uses the source and target data to learn features that display similar behavior in classification on both the source and the target domains data. The use of domain adaptation to classify EEG data has been successful in applications such as emotion recognition Li *et al.* (2019); Ma *et al.* (2019); Zhang *et al.* (2019a), motor imagery classification Tang & Zhang (2020); Wu *et al.* (2019), and evaluation of sleep quality Zhang *et al.* (2017).

In this paper, we study epileptic seizure prediction. We begin by developing and investigating a new method of multiple-subject prediction by deep learning, and compare it to the state-of-the-art of such methods. multiple-subject modeling broadens the scope of patient-specific modeling to account for the data from a fixed set of patients, thereby providing some useful level of generalization. For seizure prediction of accrued flexibility and generalization ability to data from new patients, unseen during model training, we develop and investigate cross-subject processing, with and without domain adaptation to expose the effect of data variability, and highlight the role and importance of domain adaptation. Results show a significant improvement in accuracy, F1-score, and Area under the curve, on both CHB-MIT and SIENA datasets.

The remainder of this paper is organized as follows: Section 5.3 provides a summary of previous research on seizure prediction. Section 5.4 describes the EEG databases, the deep neural network architecture, as well as the domain adaptation methods used. Section B.5 presents the experimental setup and results. Section 5.6 contains a conclusion and alludes to future directions of research.

5.3 Related work

Existing seizure prediction models fall into two major categories: the general multiple-subject modeling that applies to all patients, and the patient-specific modeling that addresses each patient individually. There has been little recent work on multiple-subject prediction modeling. The study Tsiouris *et al.* (2017) compared different classification algorithms for seizure prediction, including the repeated incremental pruning to produce error reduction (RIPPER) algorithm, support vector machines (SVM), and neural networks (NN), in order to distinguish between pre-ictal and inter-ictal EEG segments in data from multiple patients. Using a balanced number of selected pre-ictal and inter-ictal records from each patient in the CHB-MIT database, the SVM was found to have the best results with an accuracy of 68.5%. More recent studies, such as Khan *et al.* (2017), have demonstrated improved results using a convolutional neural network (CNN) on the wavelet transform of EEG signals, to achieve a sensitivity of 87.8% with a low false prediction rate of 0.142 FP/h. In Dissanayake *et al.* (2021a), a multi-task deep learning approach was used for both seizure classification and patient prediction using a Siamese network architecture on the CHB-MIT-EEG dataset, resulting in an average accuracy of 91.54%. Another study, Wu *et al.* (2022), utilized knowledge distillation to transfer information from a multiple-subject model trained on data from N-1 patients to a patient-specific model trained on the remaining patient's data. This approach led to improved patient-specific prediction results compared to four other existing methods, with an average improvement of 3.37% in accuracy, 2.33% in sensitivity, and a reduction in false predictions by an average of 0.044/h when tested on 11 patients from the CHB-MIT dataset.

Recently, deep learning has gained attention for its application in seizure prediction. For instance, a study by Tsiouris *et al.* (2018) employed a deep Long Short-Term Memory (LSTM) network to predict seizures using EEG segments with a pre-ictal duration of 120 minutes. The study reported high sensitivity and specificity of 99.84% and 99.86%, respectively, using the CHB-MIT dataset. Other studies have applied CNNs, such as Truong *et al.* (2018) which used spectrogram representations of EEG data, and Zhang *et al.* (2019b) which applied a common spatial algorithm model prior to the CNN. Another recent study Zhao *et al.* (2020) proposed a one-dimensional CNN trained on raw EEG data trained with raw EEG data to predict seizure occurrence. The study reported an area under the curve, sensitivity, and false prediction rate of 0.915, 89.26%, 0.117/h and 0.970, 94.69%, 0.095/h on american epilepsy society (AES) and CHB-MIT data, respectively. An alternative study proposed adversarial training for data augmentation to account for the limited amount of pre-ictal data, which improved the performance and robustness of the model. Recently, there has been a focus on developing interpretable models for patient-specific seizure prediction, as seen in studies such as Jemal *et al.* (2022) and Pinto *et al.* (2021), which utilize a genetic algorithm and a deep learning classifier, respectively.

5.4 Material and methods

5.4.1 Datasets and pre-processing

In this study, experiments were conducted using two open-access datasets, the SIENA EEG database and the CHB-MIT dataset. The datasets description is summarized in Table 5.1.

Tableau 5.1 – Overview of SIENA and CHB-MIT datasets used in this study for seizure prediction.

	SIENA	CHB-MIT
Number of subjects	14	23
Age of subjects	20-71	1.5-19
Number of seizures	47	198
Type of recordings	Scalp	Scalp
Total hours of EEG recordings (h)	128	940
Number of channel	29	23
Sampling frequency (Hz)	512	256

The SIENA dataset Detti *et al.* (2020), acquired at the unit of neurology and neurophysiology of the university of SIENA, contains recordings from 14 epileptic subjects aged 20 to 71 years. The subjects were monitored using video EEG. A total of 29 EEG channels sampled at 512 Hz were recorded following the standard 10-20 system. During 128 hours of EEG recording, 47 epileptic seizures were recorded. The time of the onset of a seizure and its duration were identified by experts. The CHB-MIT dataset Shoeb (2009), collected at Boston children's hospital, contains 940 hours of long-term continuous multichannel scalp EEG recordings from 23 epileptic subjects aged

1.5 to 19 years. A minimum of 19 EEG channels sampled at 256 Hz were recorded according to the international 10/20 standard. In total, these recordings included 198 seizures in which the onset and the end were precisely annotated by clinicians with expertise in neuroscience. In this study, we eliminated recordings with fewer than 23 electrodes.

The raw EEG channels from both datasets were filtered to focus on frequencies relevant to epilepsy analysis and to eliminate noise sources. This was done by using a notch filter with a cutoff frequency of 50Hz and a band-pass filter with a bandwidth of 0.5-70Hz. The pre-ictal period, which is the time before a seizure starts, was set to 1 hour based on published literature and the post-ictal period, which is the time after the seizure ends, was eliminated to exclude any effects Dissanayake *et al.* (2021a); Daoud & Bayoumi (2019). Non-overlapping windows of 10 seconds were extracted from inter-ictal and pre-ictal recordings. To address the limited number of pre-ictal samples, under-sampling was used to randomly select examples from the majority class. The windows were then normalized so that the channels had zero mean and unit standard deviation. This resulted in a total of 77,529 inter-ictal samples and 89,783 pre-ictal samples from the CHB-MIT dataset and 197,805 inter-ictal samples and 80,845 pre-ictal samples from the SIENA dataset, which were split between training, validation, and testing data.

5.4.2 Deep learning architecture

The architecture used for the multiple-subject and cross-subject models in this study was previously proposed by our team for the prediction of patient-specific seizures, as outlined in Jemal *et al.* (2022). As shown in Figure5.1, it consists of a three-layer convolutional neural network designed to be interpretable. The first layer uses standard 2D convolutions to extract relevant frequency components of the signal, the second layer uses depth-wise filters to learn spatial filters from the previous outputs. These two steps are similar to the Filter Bank Common Spatial Pattern (FBCSP) algorithm commonly used for EEG data encoding. The third convolutional layer is used for feature extraction, and the output is passed through a fully connected layer with a Softmax activation function.

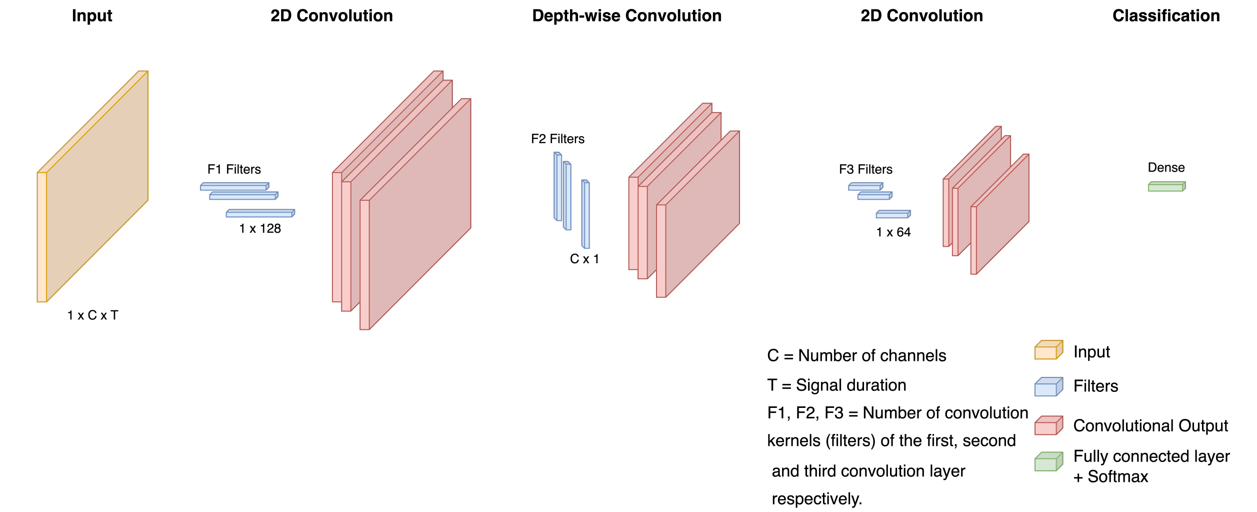
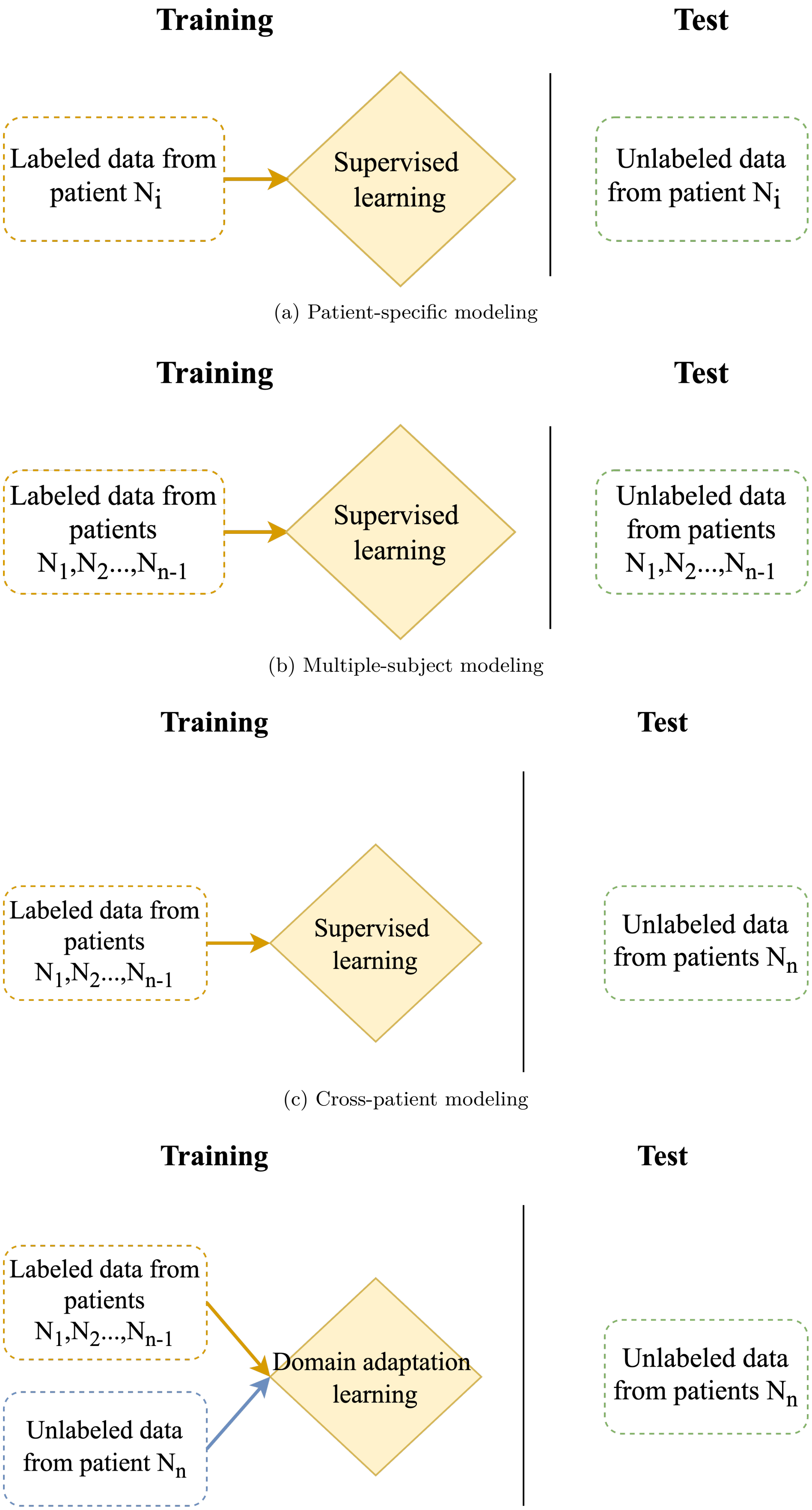


Figure 5.1 – Diagram of the deep-learning architecture. The network processes EEG inputs with standard 2D convolution to learn frequency filters. Next, it uses depth-wise convolution to learn spatial filters. Finally, a 2D convolution is used to extract features. The pipeline outputs are finally forwarded to a fully connected layer, followed by a Softmax activation.

5.4.3 Multiple-subject vs cross-patient modeling

The task of seizure prediction can be approached using several models as depicted in Figure 5.2. Patient-specific modeling involves using data from a single patient to train a unique model for that patient, as shown in Figure 5.2a. A more practical solution is multiple-subject modeling (Figure 5.2b), which uses data from multiple patients grouped into training and test sets to learn a single model that can be applied to all patients. However, this model may not generalize well to new patients. The focus of this work is cross-subject modeling (Figure 5.2c), which involves using labeled data from N-1 patients for training and data from the remaining patient for testing.

This study investigates domain adaptation methods to improve cross-subject modeling (Figure 5.2c). Domain adaptation involves using labeled data from N-1 patients (source domain) and unlabeled data from a new patient (target domain) to transfer the model. With this method, the model is able to perform well on both the N-1 patients used for training and the new patient. To accomplish this, we adopt a feature-based approach, which aims at learning features which yield good classification in both the source domain and the target domain. To do this, we investigate our architecture with three different domain adaptation algorithms: Discriminative Adversarial Neural Network (DANN), Domain Adversarial Conditional Adaptation (CDAN), and the Entropic conditioning variant of CDAN (CDAN+E).


Training
Test
Labeled data from patient N_i
Supervised learning
Unlabeled data from patient N_i
Training
Test
Labeled data from patients N_1,N_2...,N_{n-1}
Supervised learning
Unlabeled data from patients N_1,N_2...,N_{n-1}
Training
Test
Labeled data from patients N_1,N_2...,N_{n-1}
Supervised learning
Unlabeled data from patients N_n
Training
Test
Labeled data from patients N_1,N_2...,N_{n-1}
Unlabeled data from patient N_n
Domain adaptation learning
Unlabeled data from patients N_n

(a) Patient-specific modeling

(b) Multiple-subject modeling

(c) Cross-patient modeling

(d) Cross-patient modeling with domain adaptation

Figure 5.2 – Different modeling for seizure prediction.

5.4.4 Domain adaptation and cross-subject generalization

Supervised learning

Let X be the input space, the set of all possible examples or data points and Y be its corresponding label space. For example, Y would be $\{0, 1\}$ for a binary classification case. A domain D is defined as a distribution over X.

Moreover, let $F : X \to Y$ be a deterministic mapping function such as a neural network. In general, the quality of the predictor $F(x)$ can be measured using loss function $l(F(x), y)$. Supervised learning, can be defined as searching the optimal predictor F^* using the optimization problem of the following form

$$\min_F L_{task}(F(X), y) \tag{5.1}$$

The training data are used to find the optimal predictor F^*, and the test data for the evaluation. Generally, the training and test data are assumed drawn from the same distribution. However, this assumption often does not hold in practice, thus justifying domain adaptation.

Domain adaptation

Let X_S be the source domain data (training data) drawn from the distribution $P_S(X_S)$ and the target domain data (test data) denoted as X_T are drawn from the distribution $P_T(X_T)$.

We assume that there are sufficient labeled source domain data $D_S = \{(x_i^S, y_i^S)\}$ and unlabeled target domain data, $D_T = \{(x_i^T, y_i^T)\}$, in the training stage. The input spaces and label spaces between domains are assumed the same: if $x_S = x_T$, then $y_S = y_T$. However, due to the data shift, $P_S(X_S) \neq P_T(X_T)$ and $P_S(Y_S/X_S) \neq P_t(Y_S/X_T)$.

The task of domain adaptation involves adapting a model trained on a source domain, to perform well on a new target domain. The feature-based approach aims to learn features that minimize the difference between the source and target distributions. We will discuss three different methods for achieving this: the Discriminative Adversarial Neural Network (DANN), the Domain Adversarial Conditional Adaptation (CDAN), and the Entropic conditioning variant of CDAN (CDAN+E).

Discriminative Adversarial Neural Network (DANN)

The DANN method, as described by Ganin *et al.* (2016), aims at learning a feature representation that is discriminative for the classification task on the source domain and not so regarding the shift between domains. It is based on the assumption that unlabeled target domain data is available. As shown in Figure 5.3, it consists of three main components: a feature extractorϕ, a label predictor F and a domain discriminator D. The feature extractor ϕ also referred to as the generator, is a neural network that is trained using data from both the source and target domains to learn a feature representation that is not specific to any particular domain. The label predictor F is trained to minimize the classification error on the source domain data, while the domain discriminator D is trained to differentiate between the source domain and the target domain. The label predictor and the domain discriminator work adversarially, encouraging the feature extractor to learn domain-invariant representations. The parameters of all three components are optimized according to the following objective function:

$$\begin{aligned} &\min_{\phi,F} L_{task}(F(\phi(X_S)), y_S) - \lambda(log(1 - D(\phi(X_S))) + log(D(\phi(X_T)))), \\ &\max_{D} log(1 - D(\phi(X_S))) + log(D(\phi(X_T))) \end{aligned} \tag{5.2}$$

where λ is a trade-off parameter.

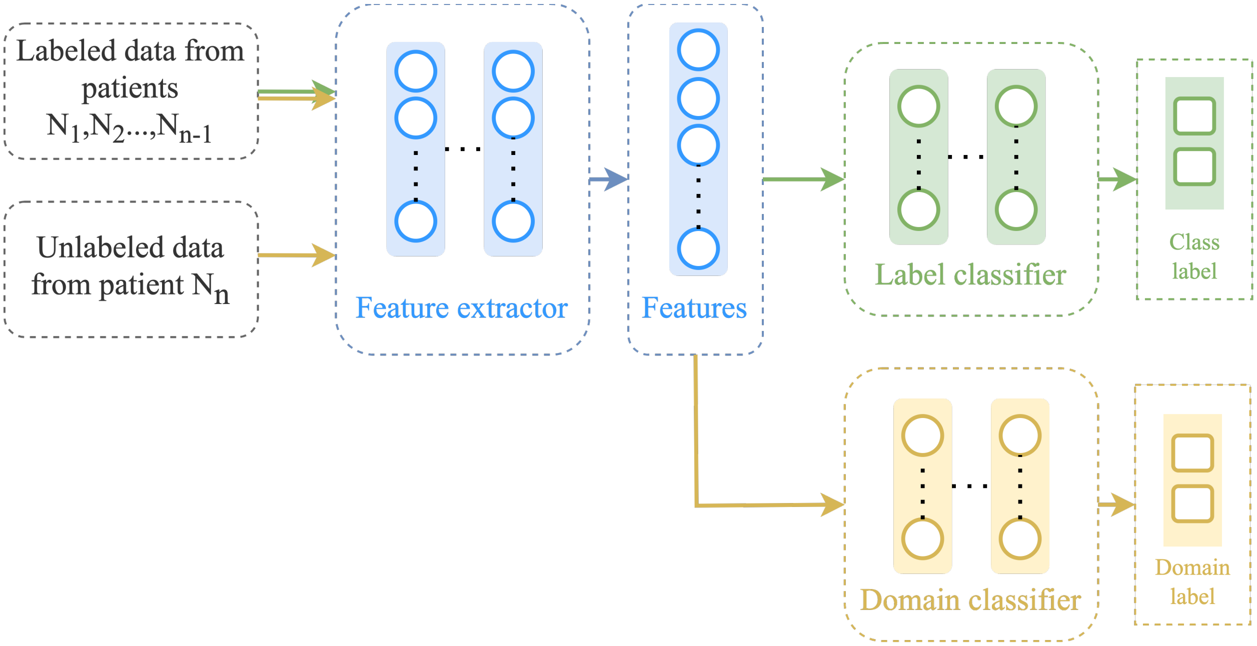

Labeled data from patients N$_1$,N$_2$...,N$_{n-1}$
Unlabeled data from patient N$_n$
Feature extractor
Features
Label classifier
Class label
Domain classifier
Domain label

Figure 5.3 – Discriminative adversarial neural network architecture.

Conditional Domain Adaptation Network (CDAN) and Entropic conditioning variant of CDAN (CDAN+E)

The CDAN approach, proposed by Long et al. Long *et al.* (2018), is similar to the DANN approach and also contains three main components: a feature extractor ϕ, a label predictor F, and a domain discriminator D. However, in CDAN, a conditional discriminator D is used through the joint variable $H = (\phi, F)$ in order to improve discriminability by capturing the cross-covariance between feature representations and classifier predictions. The method uses a multilinear conditioning strategy to combine the feature vector with the predicted label. The label predictor and domain discriminator are trained alternatively to minimize the label classification and domain classification losses, respectively. The optimization formulation for CDAN is as follows:

$$\begin{aligned}
&\min_{\phi,F} L_{task}(F(\phi(X_S)), y_S) - \lambda(log(1 - D(\phi(X_S) \otimes F(\phi(X_S)) + log(D(\phi(X_T) \otimes F(X_T))) \\
&\max_{D} log(1 - D(\phi(X_S) \otimes F(X_S)) + log(D(\phi(X_T) \otimes F(X_T))
\end{aligned} \tag{5.3}$$

where λ is a trade-off parameter, and, $\phi(X_S) \otimes F(X_S)$ is the multilinear map between the encoded sources and the task predictions.

In addition, an extension of the CDAN algorithm, known as CDAN+E, was also proposed by Long et al., in which an entropy conditioning strategy was introduced to improve transferability. This approach involves using a score that quantifies the uncertainty of the classifier predictions based on an entropy criterion to re-weight each example used by the conditional domain discriminator. This helps obtain better transferability.

5.5 Results

Performance evaluation of this study multiple-subject and cross-patient models for EEG based seizure prediction was conducted on CHB-MIT and SIENA datasets. Pytorch Paszke *et al.* (2017) was used to implement the proposed architecture. Data pre-processing was done using the MNE-Python package Gramfort *et al.* (2013). Across all models, we employed the gradient-based ADAM optimizer with coefficients β_1 β_2 set to 0.9 and 0.999 respectively for its efficiency and reliability in

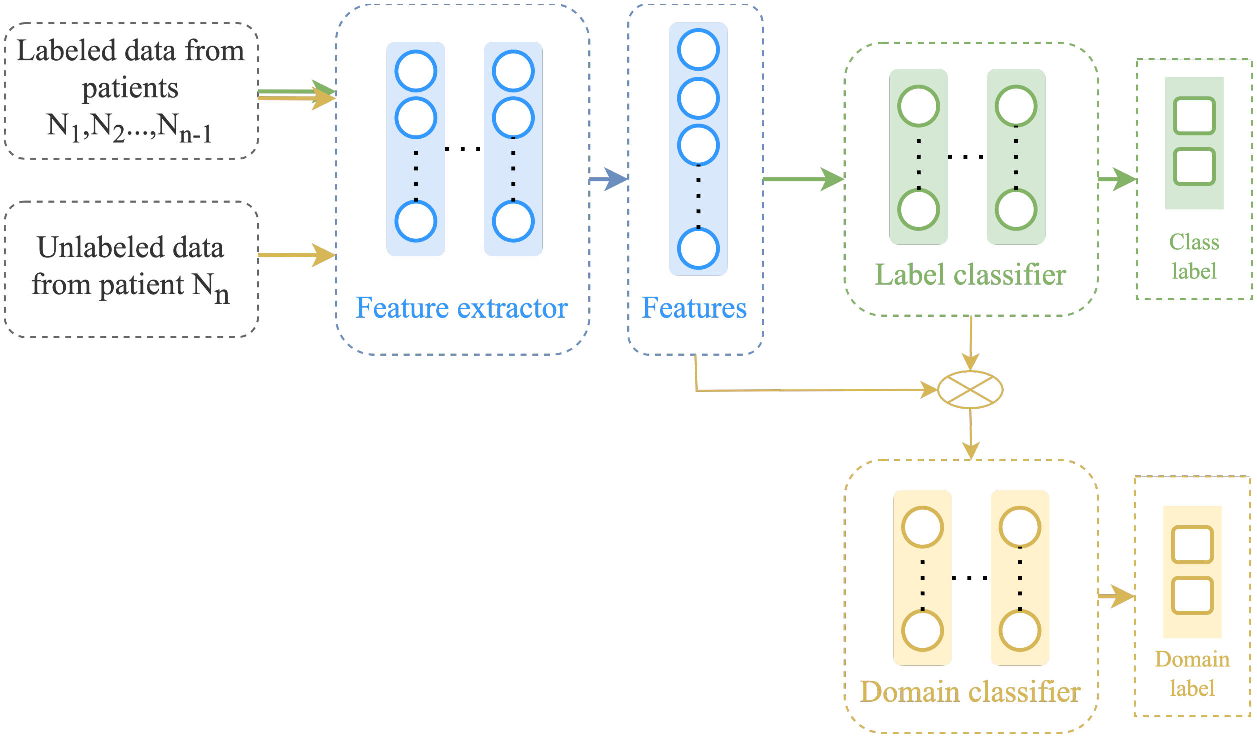


Figure 5.4 – Conditional domain adaptation network architecture.

reaching a global minimum. The learning rate was set to 0.005. To prevent over-fitting, we used a holdout validation method to divide the data into a validation set and a training set, and the training stopped after 500 epochs or when the validation loss remained constant for at least 20 epochs. To evaluate the models, we used various metrics including accuracy, precision, recall, F1-score, as well as the receiver operating characteristic (ROC) and the area under the curve (AUC).

5.5.1 Multiple-subject modeling

The multiple-subject model was evaluated using data from all patients in the SIENA dataset, which was divided into training, validation, and test sets. The model achieved a high accuracy of 96.01%, sensitivity of 97.24%, and specificity of 94.57%. The model also produced a high AUC value of 0.96. The training and validation loss curves indicate that the model does not suffer from over-fitting, as shown in Figure 5.5. The confusion matrix in Figure 5.6 demonstrates the model's classification performance. Comparison with current state-of-the-art seizure prediction models, as shown in Table 5.2, indicates that the model has comparable performance.

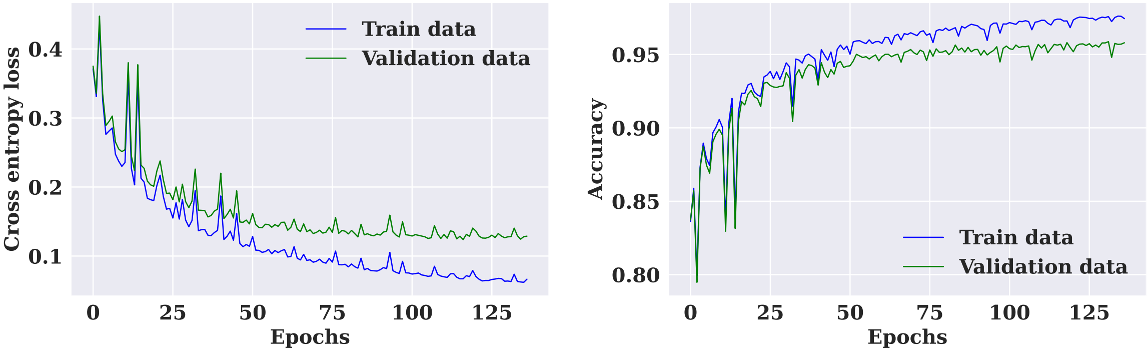


Figure 5.5 – Training and validation loss and accuracy curves of the multiple-subject seizure prediction model trained using SIENA dataset.

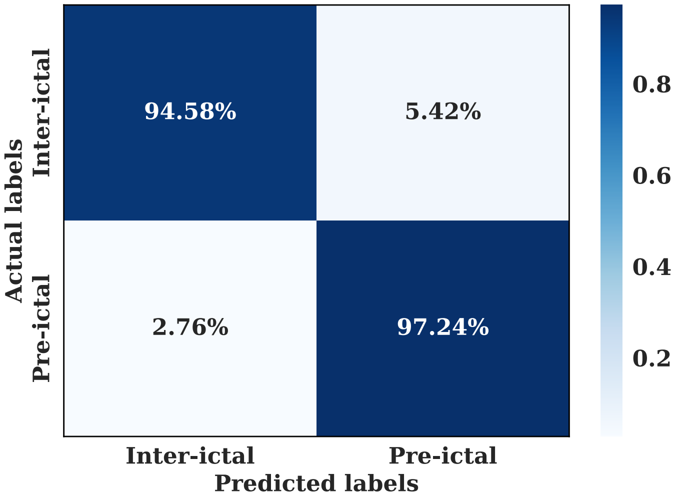


Figure 5.6 – Confusion matrix of multiple-subject seizure prediction model trained using SIENA dataset.

Tableau 5.2 – Comparisons of state-of-the-art methods on SIENA dataset using a multiple-subject modeling.

Article	Year	Method	Accuracy (%)	Sensitivity (%)	Specificity (%)
Dissanayake *et al.* (2021b)	2022	Geometric DL	95.56	95.33	95.11
Dissanayake *et al.* (2021b)	2022	Geometric DL	96.05	96.05	96.61
This work	**2022**	**CNN**	**96.01**	**97.24**	**94.57**

The multiple-subject model showed also high performance when evaluated using the CHB-MIT dataset, with an accuracy of 97.36%, a sensitivity of 98.31%, and a specificity of 96.97%. As shown in Figure 5.7, the training and validation loss decreased to a stable point, with a small gap between the training and validation curves, indicating that the model is well-fitting. The confusion matrix in Figure 5.8 shows that inter-ictal instances were confused with pre-ictal instances (3.03%), and pre-ictal instances were confused with inter-ictal instances (1.69%). According to Table 5.3, the model outperforms current state-of-the-art models evaluated on the same dataset.

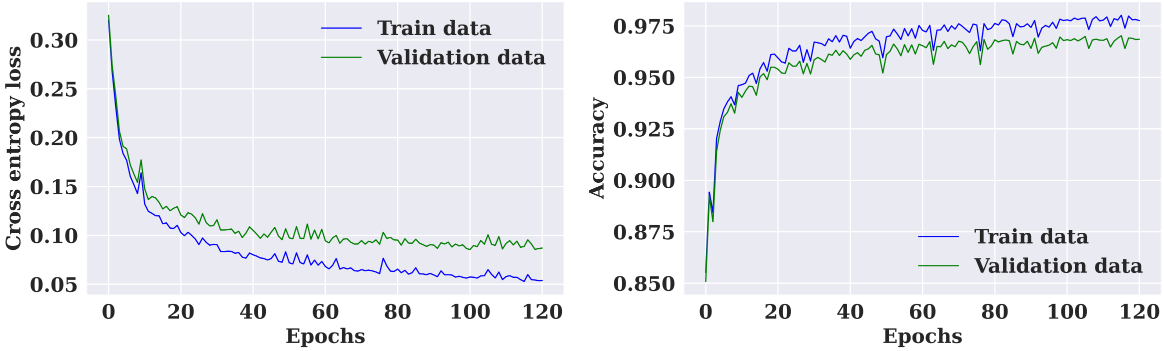


Figure 5.7 – Training and validation loss and accuracy curves of the multiple-subject seizure prediction model trained using CHB-MIT dataset.

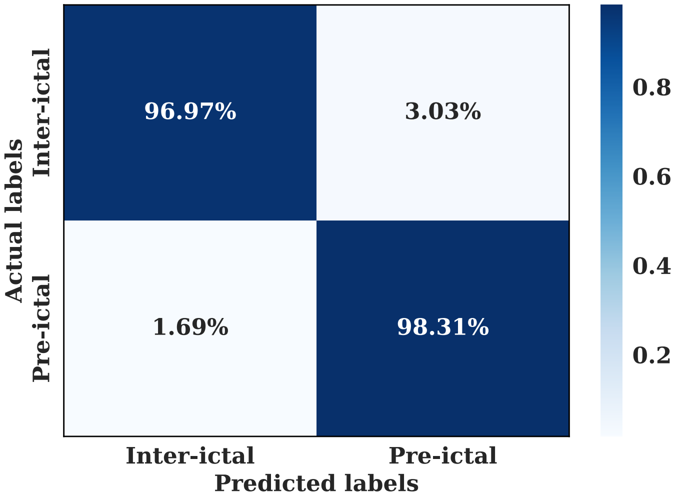


Figure 5.8 – Confusion matrix of multiple-subject seizure prediction model trained using the CHB-MIT dataset.

5.5.2 Cross-subject modeling

In contrast to the multiple-subject modeling approach, the cross-subject modeling employs a validation strategy called leave-one-patient-out. This strategy involves using data from each patient in the dataset, one at a time, for testing while training the classifier with data from the remaining

Tableau 5.3 – Comparison to the state-of-the-art methods on the CHB-MIT dataset using a multiple-subject modeling.

Article	Year	Method	Accuracy (%)	Sensitivity (%)	Specificity (%)
Tsiouris *et al.* (2017)	2017	SVM	68.50	81.20	-
Khan *et al.* (2017)	2017	CNN	-	87.80	-
Dissanayake *et al.* (2021a)	2021	Multitask DL	91.50	92.45	89.94
Dissanayake *et al.* (2021b)	2022	Geometric DL	95.38	94.47	94.26
Dissanayake *et al.* (2021b)	2022	Geometric DL	95.02	95.94	93.52
This work	2022	CNN	97.36	98.31	96.97

N-1 patients. This method is commonly used to assess the ability of the classifier to generalize to new patients.

We evaluated the proposed cross-subject method using both data sets by measuring performance in terms of F1-score, accuracy, and area under the curve (AUC). The results, presented in tables 5.4 and 5.5, show the evaluation results for each patient in the SIENA and CHB-MIT datasets respectively. The overall averages across all patients in the SIENA dataset were 39.81% for F1-score, 48.69% for accuracy, and 0.48 for AUC. Results were slightly better with the CHB-MIT dataset which uses more patients, with averaged F1-score, accuracy, and AUC, equal to, respectively, 55.34%, 63.5%, and 0.69. The performance degradation can be attributed to the mismatch between the distribution of the new patient data and the training distribution. This issue is particularly pronounced in seizure prediction, as most studies in this field focus on patient-specific models. Additionally, our previous research Jemal *et al.* (2021) on the complexity of EEG features in predicting epileptic seizures, highlighted the significant variability in EEG data between patients. To mitigate the potential data shift, we explored three domain adaptation methods: DANN, CDAN, and CDAN+E.

5.5.3 Domain adaptation for cross-subject seizure prediction

We assessed the effectiveness of three different domain adaptation methods against a baseline cross-subject modeling method using leave-one-patient-out. The results, displayed in Tables 5.6 and 5.7, indicate that all three methods (DANN, CDAN, and CDAN+E) enhanced performance on both SIENA and CHB-MIT datasets. Notably, the CDAN method performed exceptionally well with 60.27% accuracy, 59.77% F1 score, and 0.61 AUC on the SIENA dataset. Significant improvement in performance was also observed on the CHB-MIT dataset with 70.90% accuracy, 66.45% F1-score, and 0.75 AUC for CDAN+E adaptation.

Tableau 5.4 – Evaluation of cross-subject modeling by the leave-one-patient out strategy (SIENA dataset).

Patient	F1-score(%)	Accuracy(%)	AUC (%)
PN01	29.75	30.75	0.31
Pn03	53.78	56.27	0.56
PN05	54.76	53.23	0.53
PN06	14.05	40.32	0.34
PN07	28.40	41.04	0.40
PN09	61.39	49.03	0.48
PN11	54.93	64.18	0.67
PN12	30.33	52.49	0.54
PN13	8.58	50.10	0.50
PN14	24.45	50.81	0.51
PN16	56.20	45.67	0.44
PN17	61.14	50.40	0.50
Average	**39.81 ± 19.14**	**48.69 ± 8.53**	**0.48± 0.09**

The comparisons shown in Figure 5.9 reveal the importance of incorporating domain adaptation in cross-subject modeling compared to a traditional leave-one-patient-out approach. The CDAN method was found to enhance the accuracy, F1-score, and AUC by 11.58%, 19.59%, and 0.13, respectively. The results were even better when evaluated on the CHB-MIT dataset (Figure 5.10), with an average improvement in accuracy, F1-score, and AUC of +7.40%, +11.11%, and +0.06%, respectively, using the CDAN+E method. Additionally, it was noted that the model's performance improved as the number of patients in the dataset increased.

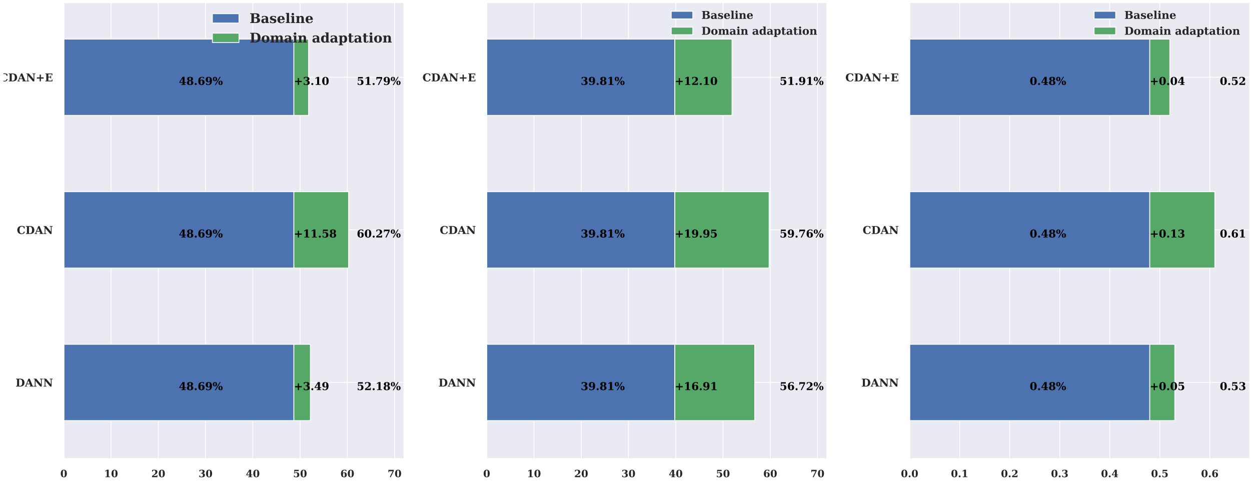


Figure 5.9 – Domain adaptation for cross-subject modeling Comparison of DANN, CDAN and CDAN+E adaptation to baseline cross-subject modeling by the leave-one-patient-out strategy (SIENA dataset). Improvements over the baseline are consistently significant in terms of accuracy, F1-score, and AUC.

Tableau 5.5 – Evaluation of cross-subject modeling by leave-one-patient out strategy (CHB-MIT dataset).

Patient	F1-score(%)	Accuracy(%)	AUC (%)
Chb01	15.30	32.93	0.29
Chb02	64.63	52.46	0.55
Chb03	32.74	49.18	0.49
Chb04	40.31	51.78	0.52
Chb05	40.08	59.59	0.67
Chb06	56.73	54.39	0.54
Chb07	67.44	56.58	0.61
Chb08	65.25	52.27	0.55
Chb 09	54.54	55.56	0.55
Chb 10	57.94	54.61	0.55
Chb 11	95.31	95.52	0.96
Chb 13	89.73	90.62	0.92
Chb 14	10.90	52.88	0.76
Chb 15	41.95	63.27	0.79
Chb 16	66.70	74.98	0.83
Chb 17	43.76	63.72	0.78
Chb 18	59.34	71.09	0.82
Chb 19	68.78	76.21	0.84
Chb 20	6.51	51.16	0.63
Chb 21	81.48	84.37	0.88
Chb 22	88.64	89.80	0.91
Chb 23	69.38	64.19	0.66
Average	**55.34 ± 24.57**	**63.5 ± 15.91**	**0.69 ± 0.17**

5.6 Conclusion

In this study, our goal was to assess the generalization capability of a seizure prediction model to new patients. Therefore, we used a deep-learning architecture previously developed for patient-specific modeling in both multiple-subject and cross-subject scenarios. Our deep-learning architecture for multiple-subject seizure prediction was compared to existing state-of-the-art models and demonstrated superior performance. However, despite the impressive accuracy of the model, its ability to generalize to new patients not in the dataset was uncertain. Thus, to better assess the model's performance and its ability to generalize to new patients, we employed a cross-subject modeling approach. But, this resulted in a noticeable decrease in performance when tested on open-access data. To overcome this issue, we investigated various domain adaptation methods to enhance the performance of cross-subject modeling. The results showed that all three methods (DANN, CDAN, and CDAN+E) significantly improved performance on both SIENA and CHB-MIT datasets.

Tableau 5.6 – Domain adaptation for cross-subject modeling. Results with DANN, CDAN and CDAN+E adaptation for cross-subject modeling by the leave-one-patient-out strategy (SIENA dataset).

Patient	DANN			CDAN			CDAN+E		
	F1(%)	ACC(%)	AUC	F1(%)	ACC(%)	AUC	F1(%)	ACC(%)	AUC
PN01	57.18	50.50	0.50	75.48	73.32	0.74	34.56	32.06	0.32
Pn03	67.84	57.66	0.62	62.49	65.90	0.66	58.92	59.61	0.60
PN05	54.86	55.64	0.56	57.89	56.91	0.56	59.07	56.86	0.57
PN06	32.586	38.53	0.38	45.67	49.16	0.49	20.02	37.70	0.5
PN07	67.19	51.63	0.66	76.72	73.16	0.75	65.13	57.48	0.59
PN09	66.10	50.54	0.54	65.38	50.11	0.51	65.87	51.62	0.56
PN11	60.94	57.80	0.58	66.17	71.10	0.73	65.87	63.71	0.64
PN12	61.62	55.58	0.56	55.96	56.83	0.57	60.35	60.27	0.60
PN13	24.84	51.56	0.53	11.84	51.46	0.57	11.71	50.88	0.54
PN14	61.56	60.11	0.60	69.67	64.50	0.66	62.73	57.97	0.58
PN16	65.16	48.73	0.39	63.34	52.73	0.54	61.85	45.42	0.32
PN17	60.78	47.86	0.46	66.66	58.10	0.61	56.79	47.86	0.47
Average	**56.72** ±13.74	**52.18** ±5.8	**0.53** ±0.08	**59.77** ±17.28	**60.27** ±9.00	**0.61** ±0.09	**51.91** ±18.85	**51.79** ±9.61	**0.52** ±0.10

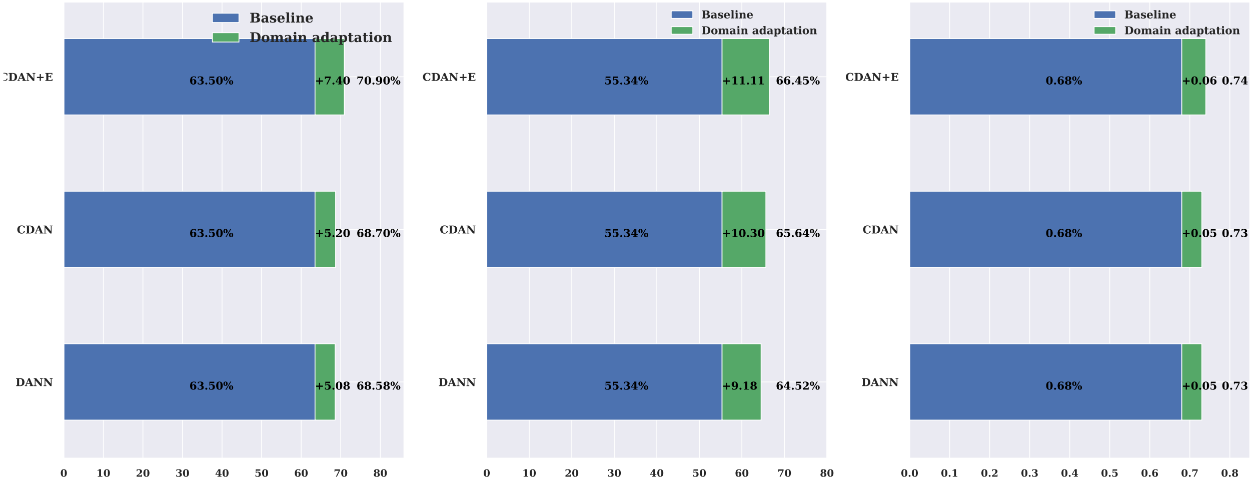


Figure 5.10 – Domain adaptation for cross-subject modeling. Comparison of DANN, CDAN, and CDAN+E adaptation to baseline cross-subject modeling by the leave-one-patient-out strategy (CHB-MIT dataset). Improvements over the baseline are consistently significant in terms of accuracy, F1-score, and AUC.

Although this study realized significant progress in epileptic seizure prediction, it currently has limitations which can be the subject of future research to resolve. One limitation is that data undersampling was resorted to in order to address the occurrence of data imbalance. This not only affects the method's computational cost of training, but may also remove data which possibly contains information that has information relevant to seizure prediction. Another limitation is related to the size of the datasets. Although the amount of data in this study allowed a principled, coherent

Tableau 5.7 – Domain adaptation for cross-subject modeling Results of DANN, CDAN and CDAN+E adaptation for cross-subject modeling by the leave-one-patient-out strategy (CHB-MIT dataset).

Patient	DANN			CDAN			CDAN+E		
	F1(%)	ACC(%)	AUC	F1(%)	ACC(%)	AUC	F1(%)	ACC(%)	AUC
Chb01	18.10	34.77	0.32	29.28	39.44	0.38	17.97	35.58	0.32
Chb02	73.47	64.86	0.79	68.00	56.76	0.67	75.00	67.57	0.80
Chb03	54.17	52.79	0.53	57.48	59.37	0.60	40.62	50.64	0.51
Chb04	55.79	61.44	0.62	71.15	67.98	0.69	44.01	52.64	0.53
Chb05	47.65	63.07	0.70	60.14	69.51	0.75	70.69	73.34	0.74
Chb06	59.64	54.38	0.55	56.89	54.56	0.55	56.02	54.77	0.55
Chb07	67.77	56.88	0.63	67.33	56.29	0.61	67.99	58.00	0.63
Chb08	66.67	52.69	0.61	67.16	52.69	0.65	74.34	68.82	0.74
Chb09	59.19	63.51	0.64	68.13	63.79	0.65	74.88	70.21	0.73
Chb10	55.31	52.18	0.52	63.08	57.48	0.58	68.88	58.59	0.65
Chb11	97.62	97.70	0.98	95.12	95.40	0.96	96.47	96.55	0.96
Chb13	98.83	98.85	0.99	94.89	95.15	0.95	99.31	99.3	0.99
Chb14	13.92	53.74	0.76	45.17	64.59	0.80	31.86	59.48	0.78
Chb15	72.13	78.11	0.84	70.00	76.82	0.83	45.33	64.81	0.79
Chb16	70.81	77.37	0.84	76.09	80.73	0.86	95.89	95.77	0.96
Chb17	63.20	73.06	0.82	46.08	64.71	0.78	82.07	84.77	0.88
Chb18	68.87	76.13	0.83	70.14	77.02	0.84	76.29	80.85	0.86
Chb19	82.19	85.06	0.88	75.36	80.46	0.86	75.36	80.46	0.86
Chb20	39.71	62.13	0.77	12.60	52.83	0.67	9.48	52.38	0.71
Chb21	88.69	89.84	0.91	86.72	88.28	0.90	90.60	91.41	0.93
Chb22	90.21	91.09	0.92	90.52	91.35	0.93	93.73	94.10	0.95
Chb23	75.45	69.25	0.76	72.68	66.55	0.71	75.03	69.84	0.74
Average	**64.52** ±21.92	**68.59** ±16.72	**0.73** ±0.16	**65.64** ±19.81	**68.72** ±15.21	**0.74** ±0.14	**66.45** ±25.27	**70.90** ±17.63	**0.75** ±0.17

investigation of the problem, more data, obtained from other EEG databases, will provide additional support to this study, and, therefore, further confirm its findings. Finally, a K-patients-leave-out strategy may be worthy of investigation to improve on this study leave-one-patient-out strategy. might be more convenient and would provide a better evaluation of the generalization of the model to more than one new patient.

Chapitre 6

Conclusion et travaux futurs

6.1 Contributions principales

Les travaux de recherche réalisés dans le cadre de cette thèse représentent des contributions significatives à l'amélioration des modèles de prédiction de l'épilepsie en utilisant des réseaux de neurones profonds.

Dans un premier temps, nous avons analysé la complexité d'un ensemble de caractéristiques extraites des données EEG généralement utilisées dans la prédiction des crises d'épilepsie. Cette analyse a révélé la complexité élevée des caractéristiques, ce qui justifie l'utilisation de méthodes plus avancées telles que les réseaux de neurones qui permettent d'extraire d'autres caractéristiques plus discriminantes. De plus, cette étude a montré la grande variabilité entre les données des patients, soulignant le besoin de modèles spécifiques au sujet et le défi de généralisation à de nouveaux patients.

Dans un second temps, nous avons proposé une nouvelle architecture interprétable de réseau de neurones et utilisé des techniques d'explication de décision des réseaux de neurones pour résoudre le problème de l'effet boîte noire des réseaux de neurones, qui limite leur utilisation dans les applications médicales et les rend difficiles à comprendre pour les cliniciens. Le modèle non seulement a atteint une précision prédictive élevée par rapport aux autres méthodes d'apprentissage profond, mais il permet également de simplifier son interprétation et d'identifier les caractéristiques les plus importantes des données EEG qui contribuent à la prédiction pour une meilleure compréhension des mécanismes sous-jacents de l'épilepsie.

En dernier lieu, nous avons abordé le défi de la généralisation des modèles de prédiction de l'épilepsie. Nous avons examiné ce problème en utilisant des techniques telles que l'adaptation au domaine. Les modèles de prédiction spécifiques au sujet ne peuvent pas être généralisés à d'autres patients, donc nous avons étudié les modèles multi-sujets et inter-sujets, qui offrent un meilleur niveau de généralisation et sont plus applicables dans des situations réelles. L'évaluation de notre modèle multi-sujets a montré des résultats supérieurs à des travaux antérieurs évalués sur les mêmes bases de données. Bien que les performances aient diminué avec le modèle inter-sujets, nous avons intégré des méthodes d'adaptation qui ont considérablement amélioré les performances du modèle.

En améliorant la précision de la prédiction de l'épilepsie, cette recherche a des implications importantes pour le diagnostic et le traitement de cette maladie. Les professionnels de la santé peuvent prendre des décisions plus éclairées sur les traitements et les interventions nécessaires pour prévenir les crises, ce qui peut améliorer considérablement la qualité de vie des patients atteints d'épilepsie.

Perspectives

Les résultats encourageants de cette recherche ouvrent des perspectives intéressantes pour améliorer les modèles de prédiction de crises d'épilepsie. Cependant, il est possible d'aller plus loin en explorant trois axes d'amélioration.

Tout d'abord, il pourrait être intéressant d'utiliser plus de données provenant de différentes bases de données EEG pour renforcer les conclusions sur la complexité des caractéristiques et la variabilité inter-sujets. En effet, en utilisant des données provenant de différentes sources, il est possible d'obtenir une vue plus complète des caractéristiques des signaux EEG et de réduire le biais lié à une seule source de données. De plus, l'utilisation de toutes les données disponibles plutôt que la technique de sous-échantillonnage pourrait améliorer les performances du modèle de prédiction, car cela permettrait de mieux exploiter les données disponibles.

Ensuite, bien que la stratégie de *Leave-one-patient-out* ait été utilisée pour évaluer la généralisation du modèle, la stratégie de *Leave-k-patient-out* pourrait être plus pratique et permettrait une meilleure évaluation de la généralisation du modèle à plusieurs nouveaux patients. En utilisant cette stratégie, il serait possible d'évaluer le modèle sur un plus grand nombre de patients tout en gardant une partie de l'ensemble de données pour l'entraînement. Cela permettrait également de mieux évaluer la capacité du modèle à généraliser à de nouveaux patients.

Enfin, pour améliorer la performance du modèle inter-sujets, d'autres algorithmes pour l'adaptation au domaine devront être déployés. Ces algorithmes permettent de réduire l'effet de la variabilité inter-sujets en adaptant le modèle à chaque sujet individuellement. Cela peut être particulièrement important dans le cas de l'épilepsie, où les caractéristiques des signaux EEG peuvent varier considérablement d'un patient à l'autre.

Des travaux futurs pourraient également se concentrer sur l'extension de l'utilisation de cette architecture et de la méthode d'interprétation proposées pour d'autres tâches de classification basées sur l'EEG, telles que le diagnostic et la catégorisation des types de crises d'épilepsie, ainsi que la détection de l'autisme et de la maladie d'Alzheimer. En étendant l'utilisation de cette méthode à d'autres tâches de classification, il est possible d'obtenir des résultats encore plus prometteurs pour améliorer le diagnostic et le traitement de ces maladies.

En conclusion, ces améliorations pourraient permettre d'améliorer encore davantage la robustesse et la généralisation des modèles de prédiction de crises d'épilepsie. Les implications potentielles pour le diagnostic et le traitement de cette maladie sont considérables, et il est important de poursuivre les recherches dans ce domaine. En utilisant ces améliorations, il sera possible d'obtenir des résultats plus précis et plus fiables pour prédire les crises d'épilepsie, ce qui pourrait avoir un impact significatif sur la qualité de vie des patients.

Références

(2019). *Epilepsy.* World health organization.

Albuquerque I, Monteiro J, Rosanne O, Tiwari A, Gagnon JF & Falk TH (2019). Cross-subject statistical shift estimation for generalized electroencephalography-based mental workload assessment. *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, IEEE, pages 3647–3653.

Alotaiby TN, Alshebeili SA, Alotaibi FM & Alrshoud SR (2017). Epileptic seizure prediction using csp and lda for scalp eeg signals. *Computational intelligence and neuroscience*, 2017.

Andrzejak RG, Chicharro D, Elger CE & Mormann F (2009). Seizure prediction: Any better than chance? *Clinical Neurophysiology*, 120(8):1465–1478.

Andrzejak RG, Lehnertz K, Mormann F, Rieke C, David P & Elger CE (2001a). 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.

Andrzejak RG, Mormann F, Kreuz T, Rieke C, Kraskov A, Elger CE & Lehnertz K (2003). Testing the null hypothesis of the nonexistence of a preseizure state. *Physical Review E*, 67(1):010901.

Andrzejak RG, Widman G, Lehnertz K, Rieke C, David P & Elger C (2001b). The epileptic process as nonlinear deterministic dynamics in a stochastic environment: an evaluation on mesial temporal lobe epilepsy. *Epilepsy research*, 44(2-3):129–140.

Ang KK, Chin ZY, Zhang H & Guan C (2008). Filter bank common spatial pattern (fbcsp) in brain-computer interface. *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, IEEE, pages 2390–2397.

Antoniades A, Spyrou L, Took CC & Sanei S (2016). Deep learning for epileptic intracranial eeg data. *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, IEEE, pages 1–6.

Arvaneh M, Umilta A & Robertson IH (2015). Filter bank common spatial patterns in mental workload estimation. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, pages 4749–4752.

Assi EB, Nguyen DK, Rihana S & Sawan M (2017). Towards accurate prediction of epileptic seizures: A review. *Biomedical Signal Processing and Control*, 34:144–157.

Assi EB, Sawan M, Nguyen D & Rihana S (2015). A hybrid mrmr-genetic based selection method for the prediction of epileptic seizures. *2015 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, IEEE, pages 1–4.

Bach S, Binder A, Montavon G, Klauschen F, Müller KR & Samek W (2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PloS one*, 10(7):e0130140.

Bandarabadi M, Rasekhi J, Teixeira CA, Karami MR & Dourado A (2015a). On the proper selection of preictal period for seizure prediction. *Epilepsy & Behavior*, 46:158–166.

Bandarabadi M, Teixeira CA, Rasekhi J & Dourado A (2015b). Epileptic seizure prediction using relative spectral power features. *Clinical Neurophysiology*, 126(2):237–248.

Basu M & Ho TK (2006). *Data complexity in pattern recognition.* Springer Science & Business Media.

Becker S, Ackermann M, Lapuschkin S, Müller KR & Samek W (2018). Interpreting and explaining deep neural networks for classification of audio signals. *arXiv preprint arXiv:1807.03418.*

Ben-David S, Blitzer J, Crammer K, Kulesza A, Pereira F & Vaughan JW (2010). A theory of learning from different domains. *Machine learning*, 79(1):151–175.

Bernadó-Mansilla E & Ho TK (2005). Domain of competence of xcs classifier system in complexity measurement space. *IEEE Transactions on Evolutionary Computation*, 9(1):82–104.

Bishop CM (2006). *Pattern recognition and machine learning.* springer.

Breiman L (2001a). Random forests. *Machine learning*, 45(1):5–32.

Breiman L (2001b). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, 16(3):199–231.

Brown JS (1982). Pedagogical, natural language, and knowlege engineering techniques in sophie i, ii, and iii. *Intelligent tutoring systems*, Academic Press.

Buchanan BG & Shortliffe EH (1984). Rule-based expert systems: the mycin experiments of the stanford heuristic programming project.

Chambon S, Galtier MN, Arnal PJ, Wainrib G & Gramfort A (2018). 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):758–769.

Chollet F (2016). Xception: deep learning with depthwise separable convolutions (2016). *arXiv preprint arXiv:1610.02357.*

Chollet F (2017). Xception: Deep learning with depthwise separable convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258.

Clancey WJ (1987). *Knowledge-based tutoring: The GUIDON program.* MIT press.

Coll M, Allegue C, Partemi S, Mates J, Del Olmo B, Campuzano O, Pascali V, Iglesias A, Striano P, Oliva A *et al.* (2016). Genetic investigation of sudden unexpected death in epilepsy cohort by panel target resequencing. *International journal of legal medicine*, 130(2):331–339.

Cook MJ, O'Brien TJ, Berkovic SF, Murphy M, Morokoff A, Fabinyi G, D'Souza W, Yerra R, Archer J, Litewka L *et al.* (2013). Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study. *The Lancet Neurology*, 12(6):563–571.

Damasevicius R, Martisius I, Jusas V & Birvinskas D (2014). Fractional delay time embedding of eeg signals into high dimensional phase space. *Elektronika ir Elektrotechnika*, 20(8):55–58.

Damaševičius R, Maskeliūnas R, Woźniak M & Połap D (2018). Visualization of physiologic signals based on hjorth parameters and gramian angular fields. *2018 IEEE 16th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, IEEE, pages 000091–000096.

Daoud H & Bayoumi MA (2019). Efficient epileptic seizure prediction based on deep learning. *IEEE transactions on biomedical circuits and systems*, 13(5):804–813.

Daumé III H (2009). Frustratingly easy domain adaptation. *arXiv preprint arXiv:0907.1815.*

Detti P, Vatti G & Zabalo Manrique de Lara G (2020). Eeg synchronization analysis for seizure prediction: A study on data of noninvasive recordings. *Processes*, 8(7):846.

Dissanayake T, Fernando T, Denman S, Sridharan S & Fookes C (2021a). Deep learning for patient-independent epileptic seizure prediction using scalp eeg signals. *IEEE Sensors Journal*, 21(7): 9377–9388.

Dissanayake T, Fernando T, Denman S, Sridharan S & Fookes C (2021b). Geometric deep learning for subject-independent epileptic seizure prediction using scalp eeg signals. *IEEE Journal of Biomedical and Health Informatics.*

Dua D & Graff C (2017). *UCI Machine Learning Repository.* http://archive.ics.uci.edu/ml.

Duda RO, Hart PE & Stork DG (2012). *Pattern classification.* John Wiley & Sons.

Gadhoumi K, Lina JM, Mormann F & Gotman J (2016). Seizure prediction for therapeutic devices: A review. *Journal of neuroscience methods*, 260:270–282.

Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, Marchand M & Lempitsky V (2016). Domain-adversarial training of neural networks. *The journal of machine learning research*, 17(1):2096–2030.

Gilpin LH, Bau D, Yuan BZ, Bajwa A, Specter M & Kagal L (2018). Explaining explanations: An overview of interpretability of machine learning. *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*, IEEE, pages 80–89.

Goodfellow I, Bengio Y & Courville A (2016). *Deep learning.* MIT press.

Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Goj R, Jas M, Brooks T, Parkkonen L *et al.* (2013). Meg and eeg data analysis with mne-python. *Frontiers in neuroscience*, 7:267.

Guo L, Rivero D, Dorado J, Rabunal JR & Pazos A (2010). Automatic epileptic seizure detection in eegs based on line length feature and artificial neural networks. *Journal of neuroscience methods*, 191(1):101–109.

Harrison MAF, Osorio I, Frei MG, Asuri S & Lai YC (2005). Correlation dimension and integral do not predict epileptic seizures. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 15(3):033106.

Hinton G, Osindero S, Welling M & Teh YW (2006). Unsupervised discovery of nonlinear structure using contrastive backpropagation. *Cognitive science*, 30(4):725–731.

Ho TK (2002). A data complexity analysis of comparative advantages of decision forest constructors. *Pattern Analysis & Applications*, 5(2):102–112.

Ho TK & Baird HS (1997). Large-scale simulation studies in image pattern recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(10):1067–1079.

Ho TK & Bernadó-Mansilla E (2006). Classifier domains of competence in data complexity space. *Data complexity in pattern recognition*, Springer, pages 135–152.

Huang J, Gretton A, Borgwardt K, Schölkopf B & Smola A (2006). Correcting sample selection bias by unlabeled data. *Advances in neural information processing systems*, 19.

Iasemidis L, Principe J & Sackellares J (2000). Measurement and quantification of spatiotemporal dynamics of human epileptic seizures. *Nonlinear biomedical signal processing*, 2:294–318.

Iasemidis LD, Sackellares JC, Zaveri HP & Williams WJ (1990). Phase space topography and the lyapunov exponent of electrocorticograms in partial seizures. *Brain topography*, 2(3):187–201.

Jemal I, Mezghani N, Abou-Abbas L & Mitiche A (2022). An interpretable deep learning classifier for epileptic seizure prediction using eeg data. *IEEE Access*.

Jemal I, Mitiche A & Mezghani N (2021). A study of eeg feature complexity in epileptic seizure prediction. *Applied Sciences*, 11(4):1579.

Kantz H & Schreiber T (2004). *Nonlinear time series analysis.* volume 7. Cambridge university press.

Khan H, Marcuse L, Fields M, Swann K & Yener B (2017). Focal onset seizure prediction using convolutional networks. *IEEE Transactions on Biomedical Engineering*, 65(9):2109–2118.

Klem GH (1999). The ten-twenty electrode system of the international federation. the international federation of clinical neurophysiology. *Electroencephalogr. Clin. Neurophysiol. Suppl.*, 52:3–6.

Koles ZJ, Lazar MS & Zhou SZ (1990). Spatial patterns underlying population differences in the background eeg. *Brain topography*, 2(4):275–284.

Kolmogorov AN (1965). Three approaches to the quantitative definition ofinformation'. *Problems of information transmission*, 1(1):1–7.

Kreuz T, Andrzejak RG, Mormann F, Kraskov A, Stögbauer H, Elger CE, Lehnertz K & Grassberger P (2004). Measure profile surrogates: a method to validate the performance of epileptic seizure prediction algorithms. *Physical Review E*, 69(6):061915.

Larochelle H, Bengio Y, Louradour J & Lamblin P (2009). Exploring strategies for training deep neural networks. *Journal of machine learning research*, 10(1).

Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP & Lance BJ (2018). Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of neural engineering*, 15(5):056013.

LeCun Y, Bengio Y *et al.* (1995a). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995.

LeCun Y, Bengio Y & Hinton G (2015). Deep learning. *nature*, 521(7553):436–444.

LeCun Y, Jackel LD, Bottou L, Cortes C, Denker JS, Drucker H, Guyon I, Muller UA, Sackinger E, Simard P *et al.* (1995b). Learning algorithms for classification: A comparison on handwritten digit recognition. *Neural networks: the statistical mechanics perspective*, 261(276):2.

Lehnertz K, Andrzejak RG, Arnhold J, Kreuz T, Mormann F, Rieke C, Widman G & Elger CE (2001). Nonlinear eeg analysis in epilepsy: Its possible use for interictal focus localization, seizure anticipation, and. *Journal of Clinical Neurophysiology*, 18(3):209–222.

Lerner DE (1996). Monitoring changing dynamics with correlation integrals: case study of an epileptic seizure. *Physica-Section D*, 97(4):563–576.

Li J, Qiu S, Du C, Wang Y & He H (2019). Domain adaptation for eeg emotion recognition based on latent representation similarity. *IEEE Transactions on Cognitive and Developmental Systems*, 12(2):344–353.

Li M, Vitányi P *et al.* (2008). *An introduction to Kolmogorov complexity and its applications.* volume 3. Springer.

Li X, Song D, Zhang P, Zhang Y, Hou Y & Hu B (2018). Exploring eeg features in cross-subject emotion recognition. *Frontiers in neuroscience*, 12:162.

Liao SC, Wu CT, Huang HC, Cheng WT & Liu YH (2017). Major depression detection from eeg signals using kernel eigen-filter-bank common spatial patterns. *Sensors*, 17(6):1385.

Long M, Cao Z, Wang J & Jordan MI (2018). Conditional adversarial domain adaptation. *Advances in neural information processing systems*, 31.

Ma BQ, Li H, Zheng WL & Lu BL (2019). Reducing the subject variability of eeg signals with adversarial domain generalization. *International Conference on Neural Information Processing*, Springer, pages 30–42.

Maciejowski JM (1979). Model discrimination using an algorithmic information criterion. *Automatica*, 15(5):579–593.

Mansilla EB & Ho TK (2004). On classifier domains of competence. *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, IEEE, volume 1, pages 136–139.

McSharry PE, Smith LA & Tarassenko L (2003). Prediction of epileptic seizures: are nonlinear methods relevant? *Nature medicine*, 9(3):241–242.

Mezghani N, Mechmeche I, Mitiche A, Ouakrim Y & De Guise JA (2018). An analysis of 3d knee kinematic data complexity in knee osteoarthritis and asymptomatic controls. *PloS one*, 13(10):e0202348.

Moghim N & Corne DW (2014). Predicting epileptic seizures in advance. *PloS one*, 9(6):e99334.

Montavon G, Binder A, Lapuschkin S, Samek W & Müller KR (2019). Layer-wise relevance propagation: an overview. *Explainable AI: interpreting, explaining and visualizing deep learning*, pages 193–209.

Montavon G, Samek W & Müller KR (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73:1–15.

Morán-Fernández L, Bolón-Canedo V & Alonso-Betanzos A (2017). Centralized vs. distributed feature selection methods based on data complexity measures. *Knowledge-Based Systems*, 117:27–45.

Morch NJ, Kjems U, Hansen LK, Svarer C, Law I, Lautrup B, Strother S & Rehm K (1995). Visualization of neural networks using saliency maps. *Proceedings of ICNN'95-International Conference on Neural Networks*, IEEE, volume 4, pages 2085–2090.

Mormann F, Andrzejak RG, Elger CE & Lehnertz K (2007). Seizure prediction: the long and winding road. *Brain*, 130(2):314–333.

Mormann F, Kreuz T, Rieke C, Andrzejak R, Kraskov A, David P, Elger C & Lehnertz K (2005). On the predictability of epileptic seizures. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 116:569–87. DOI:10.1016/j.clinph.2004.08.025.

Oquab M, Bottou L, Laptev I & Sivic J (2014). Learning and transferring mid-level image representations using convolutional neural networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1717–1724.

Ott E (2002). *Chaos in dynamical systems.* Cambridge university press.

Park Y, Luo L, Parhi KK & Netoff T (2011). Seizure prediction with spectral power of eeg using cost-sensitive support vector machines. *Epilepsia*, 52(10):1761–1770.

Partemi S, Vidal MC, Striano P, Campuzano O, Allegue C, Pezzella M, Elia M, Parisi P, Belcastro V, Casellato S *et al.* (2015). Genetic and forensic implications in epilepsy and cardiac arrhythmias: a case series. *International journal of legal medicine*, 129(3):495–504.

Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z, Desmaison A, Antiga L & Lerer A (2017). Automatic differentiation in pytorch.

Pinto M, Leal A, Lopes F, Dourado A, Martins P, Teixeira CA *et al.* (2021). A personalized and evolutionary algorithm for interpretable eeg epilepsy seizure prediction. *Scientific reports*, 11(1):1–12.

Ponce J, Berg TL, Everingham M, Forsyth DA, Hebert M, Lazebnik S, Marszalek M, Schmid C, Russell BC, Torralba A *et al.* (2006). Dataset issues in object recognition. *Toward category-level object recognition*, Springer, pages 29–48.

Rasekhi J, Mollaei MRK, Bandarabadi M, Teixeira CA & Dourado A (2013). Preprocessing effects of 22 linear univariate features on the performance of seizure prediction methods. *Journal of neuroscience methods*, 217(1-2):9–16.

Savit R & Green M (1991). Time series and dependent variables. *Physica D: Nonlinear Phenomena*, 50(1):95–116.

Schirrmeister RT, Springenberg JT, Fiederer LDJ, Glasstetter M, Eggensperger K, Tangermann M, Hutter F, Burgard W & Ball T (2017). Deep learning with convolutional neural networks for eeg decoding and visualization. *Human brain mapping*, 38(11):5391–5420.

Schuster HG & Just W (2006). *Deterministic chaos: an introduction.* John Wiley & Sons.

Shoeb AH (2009). *Application of machine learning to epileptic seizure onset detection and treatment.* Thèse de doctorat, Massachusetts Institute of Technology.

Shoeb AH & Guttag JV (2010). Application of machine learning to epileptic seizure detection. *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 975–982.

Sifre L & Mallat S (2014). Rigid-motion scattering for texture classification. *arXiv preprint arXiv:1403.1687*.

Sugiyama M, Nakajima S, Kashima H, Buenau P & Kawanabe M (2007). Direct importance estimation with model selection and its application to covariate shift adaptation. *Advances in neural information processing systems*, 20.

Sun M, Liu K, Wu Q, Hong Q, Wang B & Zhang H (2019). A novel ecoc algorithm for multiclass microarray data classification based on data complexity analysis. *Pattern Recognition*, 90:346–362.

Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A *et al.* (2014). Going deeper with convolutions. arxiv 2014. *arXiv preprint arXiv:1409.4842*, 1409.

Tang X & Zhang X (2020). Conditional adversarial domain adaptation neural network for motor imagery eeg decoding. *Entropy*, 22(1):96.

Teixeira C, Direito B, Feldwisch-Drentrup H, Valderrama M, Costa R, Alvarado-Rojas C, Nikolopoulos S, Le Van Quyen M, Timmer J, Schelter B *et al.* (2011). Epilab: A software package for studies on the prediction of epileptic seizures. *Journal of Neuroscience Methods*, 200(2):257–271.

Teixeira CA, Direito B, Bandarabadi M, Le Van Quyen M, Valderrama M, Schelter B, Schulze-Bonhage A, Navarro V, Sales F & Dourado A (2014). Epileptic seizure predictors based on computational intelligence techniques: A comparative study with 278 patients. *Computer methods and programs in biomedicine*, 114(3):324–336.

Thodoroff P, Pineau J & Lim A (2016). Learning robust features using deep learning for automatic seizure detection. *Machine learning for healthcare conference*, PMLR, pages 178–190.

Torralba A & Efros AA (2011). Unbiased look at dataset bias. *CVPR 2011*, IEEE, pages 1521–1528.

Truong ND, Nguyen AD, Kuhlmann L, Bonyadi MR, Yang J, Ippolito S & Kavehei O (2018). Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram. *Neural Networks*, 105:104–111.

Tsiouris KM, Pezoulas VC, Koutsouris DD, Zervakis M & Fotiadis DI (2017). Discrimination of preictal and interictal brain states from long-term eeg data. *2017 IEEE 30th international symposium on computer-based medical systems (CBMS)*, IEEE, pages 318–323.

Tsiouris KM, Pezoulas VC, Zervakis M, Konitsiotis S, Koutsouris DD & Fotiadis DI (2018). A long short-term memory deep learning network for the prediction of epileptic seizures using eeg signals. *Computers in biology and medicine*, 99:24–37.

Tzallas AT, Tsipouras MG & Fotiadis DI (2009). Epileptic seizure detection in eegs using time–frequency analysis. *IEEE transactions on information technology in biomedicine*, 13(5):703–710.

Ullah I, Hussain M, Aboalsamh H *et al.* (2018). An automated system for epilepsy detection using eeg brain signals based on deep learning approach. *Expert Systems with Applications*, 107:61–71.

Uyttenhove T, Maes A, Van Steenkiste T, Deschrijver D & Dhaene T (2020). Interpretable epilepsy detection in routine, interictal eeg data using deep learning. *Machine Learning for Health*, PMLR, pages 355–366.

Wu D, Yang J & Sawan M (2022). Bridging the gap between patient-specific and patient-independent seizure prediction via knowledge distillation. *arXiv preprint arXiv:2202.12598*.

Wu H, Niu Y, Li F, Li Y, Fu B, Shi G & Dong M (2019). A parallel multiscale filter bank convolutional neural networks for motor imagery eeg classification. *Frontiers in neuroscience*, 13:1275.

Zhang B, Wang W, Xiao Y, Xiao S, Chen S, Chen S, Xu G & Che W (2020). Cross-subject seizure detection in eegs using deep transfer learning. *Computational and Mathematical Methods in Medicine*, 2020.

Zhang W, Wang F, Jiang Y, Xu Z, Wu S & Zhang Y (2019a). Cross-subject eeg-based emotion recognition with deep domain confusion. *International conference on intelligent robotics and applications*, Springer, pages 558–570.

Zhang XZ, Zheng WL & Lu BL (2017). Eeg-based sleep quality evaluation with deep transfer learning. *International Conference on Neural Information Processing*, Springer, pages 543–552.

Zhang Y, Guo Y, Yang P, Chen W & Lo B (2019b). Epilepsy seizure prediction on eeg using common spatial pattern and convolutional neural network. *IEEE Journal of Biomedical and Health Informatics*, 24(2):465–474.

Zhao S, Yang J, Xu Y & Sawan M (2020). Binary single-dimensional convolutional neural network for seizure prediction. *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, IEEE, pages 1–5.

Zhou M, Tian C, Cao R, Wang B, Niu Y, Hu T, Guo H & Xiang J (2018). Epileptic seizure detection based on eeg signals and cnn. *Frontiers in neuroinformatics*, 12:95.

Annexe A

Article 1 : A Study of EEG Feature Complexity in Epileptic Seizure Prediction

Symbols	**Extracted features**
σ^2	variance
χ	skewness
κ	kurtosis
HM	Hjorth parameter (mobility)
HC	Hjorth parameter (complexity)
τ_0	decorrelation time
ϵ_{err}	error of the auto-regressive modeling
δ_r	relative power of the delta spectral band
θ_r	relative power of the theta spectral band
α_r	relative power of the alpha spectral band
β_r	relative power of the beta spectral band
γ_r	relative power of the gamma spectral band
f_{50}	spectral edge frequency
E_w	wavelet energy
S_w	wavelet entropy
E	signal energy
AE	signal accumulated energy
D_2	correlation dimension
D_ϵ	correlation density
L_{max}	largest Lyapunov exponent
Λ	local flow
AC	algorithmic complexity
LR	loss of recurrence
δ_m	marginal predictability
D_2^*	surrogate-corrected version of the correlation dimension
L_{max}^*	surrogate-corrected version of the largest Lyapunov exponent
Λ^*	surrogate-corrected version of the local flow

AC^*	surrogate-corrected version of the algorithmic complexity
b_k	bivariate spectral power features
C_{max}	cross correlation
Γ	linear coherence
MI	mutual information
R	mean phase coherence
λ_{cp}^{W}	index based on conditional probability using the wavelet transform
	index based on conditional probability using the Hilbert transform
ρ_{se}^{W}	index based on Shannon entropy using the wavelet transform
ρ_{se}^{H}	index based on Shannon entropy using the Hilbert transform
S	non-linear interdependence measure
H	non-linear interdependence normalized measure

Annexe B

Article 4 : An Effective Deep Neural Network Architecture for Cross-Subject Epileptic Seizure Detection in EEG Data

Imene Jemal[1,2], Amar Mitiche[1], Lina Abou-Abbas[2,3], Khadidja Henni[2,3] and Neila Mezghani[2,3]

[1] Centre ÉMT, Institut National de la Recherche Scientifique, Montréal, Canada
[2] Centre de Recherche LICEF, Université TÉLUQ, Montréal, Canada
[3] Laboratoire LIO, Centre de Recherche du CHUM, Montréal, Canada

Proceedings of CECNet 2021: The 11th International Conference on Electronics, Communications and Networks (CECNet), November 18-21, 2021, vol. 345, p. 54. IOS Press, 2022.

Résumé : La détection des crises d'épilepsie est un domaine de recherche actif depuis plusieurs décennies. Bien que les algorithmes spécifiques aux patients actuels aient une performance satisfaisante, leur application clinique peut encore être améliorée. La caractérisation et la détection des crises d'épilepsie par apprentissage automatique restent un défi en raison de la variabilité significative des motifs de données EEG entre les patients. Dans cette étude, nous proposons une architecture modifiée de réseau de neurones convolutionnel (CNN) basée sur les convolutions de profondeur séparable pour une détection automatique et efficace des crises d'épilepsie. Notre architecture a été conçue avec un nombre réduit de paramètres pour réduire la complexité du modèle et les exigences de stockage. Elle peut être facilement déployée sur un dispositif connecté pour permettre une détection en temps réel des crises. Nous avons évalué la performance de notre méthode sur deux jeux de données publics, l'un provenant de l'Hôpital pour enfants de Boston (CHB-MIT) et l'autre de l'Université de Bonn (Ubonn). Nous avons obtenu une sensibilité et un taux de faux positifs par heure de 91,93% - 0,005 et 100% - 0,057 pour les jeux de données CHB-MIT et Ubonn, respectivement.

B.1 Abstract

For several decades now, epileptic seizure detection has been an active research topic. The performance of current patient-specific algorithms is satisfactory, although clinical application can benefit from improvements. Due to significant EEG data pattern variability between patients, machine learning cross-subject seizure characterization and detection is still a challenging task. The purpose of this study is to propose and investigate a modified convolutional neural network (CNN) architecture based on separable depth-wise convolution for effective automatic cross-subject seizure detection. The architecture is conceived with a reduced number of trainable parameters in order to lower the complexity of the model and the storage requirements to deploy it easily in connected device for real-time seizure detection. Performance of the proposed method is evaluated on two public datasets of 5 and 23 subjects collected in the Children's Hospital Boston and the University of Bonn respectively. The method reaches the highest sensitivity-false positive rate/h of 91.93%–0.005, 100%–0.057 for the CHB-MIT and Ubonn datasets respectively.
Keywords : Epilepsy; Seizure prediction; Deep learning; Convolutional Neural Network; EEG.

B.2 Introduction

Epilepsy is a neuro-degenerative disorder of unprovoked, recurrent seizures. It is the second most frequent neurological disease **?**. Most often, EEG records are the basis for a diagnosis. The visual inspection of hours of EEG data is impractical because it is time-consuming and requires interpretations by experts. As a result, several studies have been conducted to develop computer-aided diagnosis systems which can detect seizures automaticallyTzallas *et al.* (2009); Shoeb & Guttag (2010); Guo *et al.* (2010). Several EEG-based epilepsy detection models have been proposed. However, epileptic patterns are highly variable across seizures and across patients, which makes real-time application of these models in clinical settings quite a challenging task. Models go along two veins: General cross-subject modeling which apply to patients at large, and patient-specific modeling which applies to patients individually. Patient-specific modeling is generally impracticable because it requires recording sufficient seizure onsets for each patient, separately from others. Cross-subject modeling does not have to treat each patient separately, but it faces the major problem of adapting the detection algorithm to unseen data of new patients, mainly due to the significant cross-subject EEG pattern variability Jemal *et al.* (2021).

Recent studies of deep learning (DL)Goodfellow *et al.* (2016) for epilepsy detection, which automatically encodes EEG features characteristic of epilepsy, have been much more potent than traditional feature selection and classification methods LeCun *et al.* (1995a). One of first the deep learning investigations of epilepsy detection Thodoroff *et al.* (2016) used a convolutional neural network (CNN) for feature extraction in an image-based representation of EEG signals, followed by Long Short Memory units (LSTM) for classification. The method was evaluated on the CHB-MIT dataset for subject-specific (sensitivity = 95-100%) and cross-subject (sensitivity = 85%) models. Along this vein, Antoniades *et al.* (2016) used CNNs to distinguish interictal epileptiform discharges from normal activity. The method achieved the higher accuracy of 87.51% on the CHB-MIT dataset. The study in Zhou *et al.* (2018) compared time domain and frequency domain subject-specific EEG representations in CNN feature coding for seizure prediction. On the CHB-MIT and the Freiburg datasets, frequency domain modeling gave sensibly better results (97.5% vs 95.4% accuracy). The investigation of Ullah *et al.* (2018) proposed a pyramidal one-dimensional convolution neural network architecture, obtaining higher detection sensitivity, specificity, and accuracy of 89%, 99% and 98.2% respectively. The experiments, however, were carried out on a relatively small amount of data from 5 patients in the Ubonn database. Although it did not address inter-ictal and ictal period classification, the study in Uyttenhove *et al.* (2020) used a relatively large EEG records dataset of 300 patients from the Temple University Hospital EEG database to compare CNN to conventional classification (support vector machine (SVM) and random forest (RF)) for distinguishing healthy from epileptic patients in cross-subject EEG data. Performance values were better in terms of area under the precision-recall curve (AUPR) with the tiny visual geometry group CNN architecture (AUPR = 0.9242), than SVM (AUPR = 0.8651) and RF (AUPR = 0.8578).

All of these studies highlight the importance of deep neural network processing of seizure detection, and the importance of dataset size in validation.

Convolutional neural networks Goodfellow *et al.* (2016) can learn effective input data nonlinear local features of increasing complexity as processing progresses from the input layer to the output layer. CNNs were first described by LeCun *et al.* (1995b) as neural networks composed of a sequence of convolution and pooling layers. The original CNN was subsequently upgraded to have a larger architecture, called AlexNet, which was followed by even more complex structures keeping the original basic ideas. The investigation of Szegedy *et al.* (2014) introduced the Inception-V1 architecture (GoogLeNet) with processing steps that express correlation between channels followed by spatial pattern learning. The architecture allowed richer pattern feature learning using less network parameters. Similar in concept, the Xception architechture Chollet (2016) starts with depth-wise convolution applied on channels to be followed by a point-wise convolution to combine the coded features. It has the particularity of not using non-linearity between layers. The architecture showed better performances than Inception-V3 in classification tasks on the ImageNet dataset and a larger image classification dataset comprising 350 million images and 17,000 classes.

The purpose of this study is to investigate epilepsy detection in EEG data by a new CNN architecture based on separable depth-wise layers. Unlike others, this CNN initial layer performs a convolution to learn a representation of the raw signal in terms of frequency components. This is in agreement with feature extraction by filter bank signal decomposition Schirrmeister *et al.* (2017). The architecture also includes separable depth-wise layers: this necessitates significantly fewer network parameters than the standard 2D convolution layer, which has the effect of: 1) lowering model complexity and subsequent execution, 2) decreasing storage requirements so as to allow execution on connected devices and, 3) allowing model training on either small or large datasets.

This architecture, pertinent to cross-subject modeling, is investigated here to distinguish ictal from inter-ictal periods in EEG data. The cross-subject modeling can increase significantly algorithm applicability because it allows processing data of unseen subjects, unlike models that are patient-specific.

This CNN architecture was evaluated using the publicly available CHB-MIT and Ubonn databases. As described in detail subsequently, it reached high performance, with 91.82% (5 patients in CHB-MIT) and 99.60% (23 patients in Ubonn).

The remainder of this paper is organized as follows. Section B.3 presents the method in detail. The experimental setup is presented in Section B.4. Finally, Section B.5 contains the results and a discussion.

B.3 Methods

In this section, we first explain the difference between standard and depth-wise separable convolution operations. Afterward, a detailed description of the deployed architecture is presented.

B.3.1 Standard vs Separable depth-wise convolution

A standard convolution layer simply applies a convolution operation between the input and learnable weighted filters to obtain a new data representation called a feature-map (See figure B.1a). The purpose of the learned filters is to capture spatial and cross-channel correlation simultaneously.

A separable convolution layer divides the convolution kernel into two smaller kernels, which has the effect of reducing the number of parameters. A classic example is the decomposition of the Sobel edge detector kernel into two smaller kernels as shown in Equation B.1

$$\begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} * \begin{pmatrix} -1 & 0 & 1 \end{pmatrix} \tag{B.1}$$

The depth-wise separable convolution Sifre & Mallat (2014) is similar to the separable convolution. However, the convolution operation is decomposed differently into two steps : (1) a depth-wise convolution to learn local patterns for each channel and (2) a point-wise convolution to find linear combination of the extracted feature capturing the between-channel correlations (See figure B.1b).

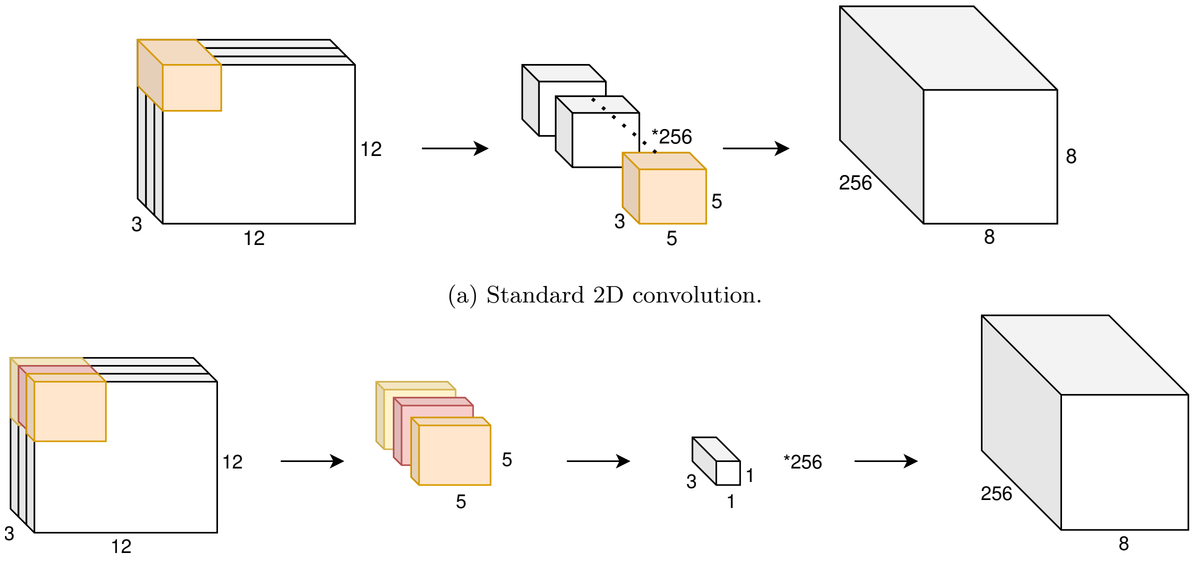
- Depth-wise convolution : This type of layer is so named because it takes on consideration the depth dimension (the number of channels) where the convolution of each channel with a different kernel is done separately as shown in Figure B.1b. This step allows learning filters for each channel.
- Point-wise convolution : This convolution uses a 1x1 kernel with depth equal to the number of channels to iterate through every point to learn a linear combination of the feature-maps (output from the depth convolution). This step allows capturing the correlation across the channels.

An example of standard and separable depth-wise convolution is shown in Figure B.1. The normal convolution transforms the input 256 times using Kernels of size $5 * 3 * 3$ leading to a total number of parameters of $5 * 3 * 3 * 256$. Contrarily, the separable depth-wise convolution apply a single transformation (kernel of size $5 * 3 * 3$) and simply elongate it to 256 channels using 256 kernels of size $3 * 1 * 1$. The number of parameters is reduced to $5 * 3 * 3 + 3 * 1 * 1 * 256$.

B.3.2 Architecture design

In this section, we introduce a convolution neural network architecture inspired by Chollet (2017); Schirrmeister *et al.* (2017). The architecture uses a reduced number of parameters allowing it to be trained with very limited data as well as with larger datasets. Full details of the network architecture are summarized in the table B.1.

- The network starts with a 2D convolution to learn F1 frequency filters. Indeed, this block is inspired by the concept of the filter bank, which is a set of band-pass filters that separate



(a) Standard 2D convolution.

(b) Separable depth-wise convolution.

Figure B.1 – Illustration of different convolution layer types.

the input signal into several components, each corresponding to a single frequency sub-band of the original signal. This technique is usually performed before the feature extraction step Schirrmeister *et al.* (2017).

- Subsequently, we alternate between separable depth-wise convolution layers and pooling layers. As mentioned above, the separable depth-wise convolution begin with the depth-wise convolution to learn specific filters for each channel (component of the signal in a specific frequency sub-band). This is followed by a point-wise convolution combining the learned features across channels. Pooling layers are used to reduce the dimensions. We applied batch normalization before the non-linear activation to stabilize the training. In order to regularize the model, we added a dropout layer.
- Finally, the features are flattened by the fully connected layer and fitted to a softmax classification.

B.4 Experimental setup

B.4.1 EEG data

We evaluated the proposed architecture on two publicly available datasets of EEG data (see Table B.2).

CHB-MIT database

The database collected at the Boston Children's HospitalShoeb (2009) contains scalp EEG data of 23 patients. The EEG data were recorded through 19 electrodes distributed over the scalp according to the international standard 10/20. The signals were amplified and sampled with a

Tableau B.1 – The detailed network architecture, where C = number of channels, T, T', T" = signal duration, F1, F2, F3 = number of convolution kernels to learn, N = number of classes, respectively.

Layer	#Filters	Size	#Output	Activation
Input			(1,C,T)	
Conv2D	F1	(1,128)	(F1,C,T)	Linear
BatchNorm2d			(F1,C,T)	
Reshape			(F1*C,1,T)	
DepthwiseConv	F2	(1,32)	(32*F2*F1*C,1,T')	Linear
PointwiseConv	2	(1,1)	(2,1,T')	Linear
BatchNorm2D			(2,1,T')	
Activation			(2,1,T')	Relu
AveragePool2D		(1,8)	(2,1,T'//8)	
Dropout			(2,1,T'//8)	
DepthwiseConv	F3	(1,16)	(16*F3*2,1,T")	Linear
PointwiseConv	2	(1,1)	(2,1,T")	Linear
BatchNorm2D			(2,1,T")	
Activation			(2,1,T")	Relu
AveragePool2D		(1,4)	(2,1,T"//4)	
Dropout			(2,1,T"//4)	
Linear			(2*T"//4)	
Dense			N=2	Softmax

frequency of 256 Hz. During 940 hours of EEG recording, 198 epileptic seizures were recorded. The time of onset of a seizure and its duration has been identified by experts.

UBonn university database

The data were collected from 5 monitored patients at Bonn University Andrzejak *et al.* (2001a). It consists of five sets (denoted A-E) each containing 100 single-channel EEG segments of 23.6 seconds. Sets A and B contain surface EEG recordings from healthy people. Sets C and D were recorded from epileptic patients in seizure-free intervals. Set E is the only set that contains activity recorded during seizures. The data was sampled at a rate of 173.61 Hz. The segments were selected after a visual inspection for artifacts like muscle activity or eye blinking. For classifying seizure-free and seizure EEG segments, set A-D are labeled as normal EEG records and set E is reserved for seizure events.

B.4.2 Pre-processing

For the two datasets, the raw EEG channels were filtered using notch filter with a cutoff frequency of 50Hz and band-pass filter with a bandwidth 0.5-75 Hz to remove artifacts and certain amount of noise. The data was segmented into five-second time intervals. Regarding long-term EEG recordings in the CHB-MIT database, 30 mintues were eliminated after the beginning of the seizure to exclude effects from the post-ictal period.

Tableau B.2 – Public databases for seizure detection

Dataset	CHB-MIT	UBonn
Number of subjects	23	5
Number of seizure	198	100*
Total duration(hour)	940	3.24
Recording type	Scalp	Scalp and Intracranial
Number of channels	17	1**
Sampling frequency(Hz)	256	173.73

*100 seizures file each of 23.36 s duration.

**Multi-channel data was converted to a single channel.

B.4.3 Performance metrics

The most commonly used metrics for the evaluation of epilepsy prediction are sensitivity, specificity and false alarm rate. The sensitivity (SS) which corresponds to the number of positives (seizure) that are correctly identified is the most critical metric. Indeed, it is better to identify non-seizure window as seizure than to miss a seizure. However, due to the significant implications for the patient, reducing the false positive rate (FPR) should be a priority. Thus, for epilepsy prediction, we aim to reach a good compromise between the sensitivity and the false alarm rate.

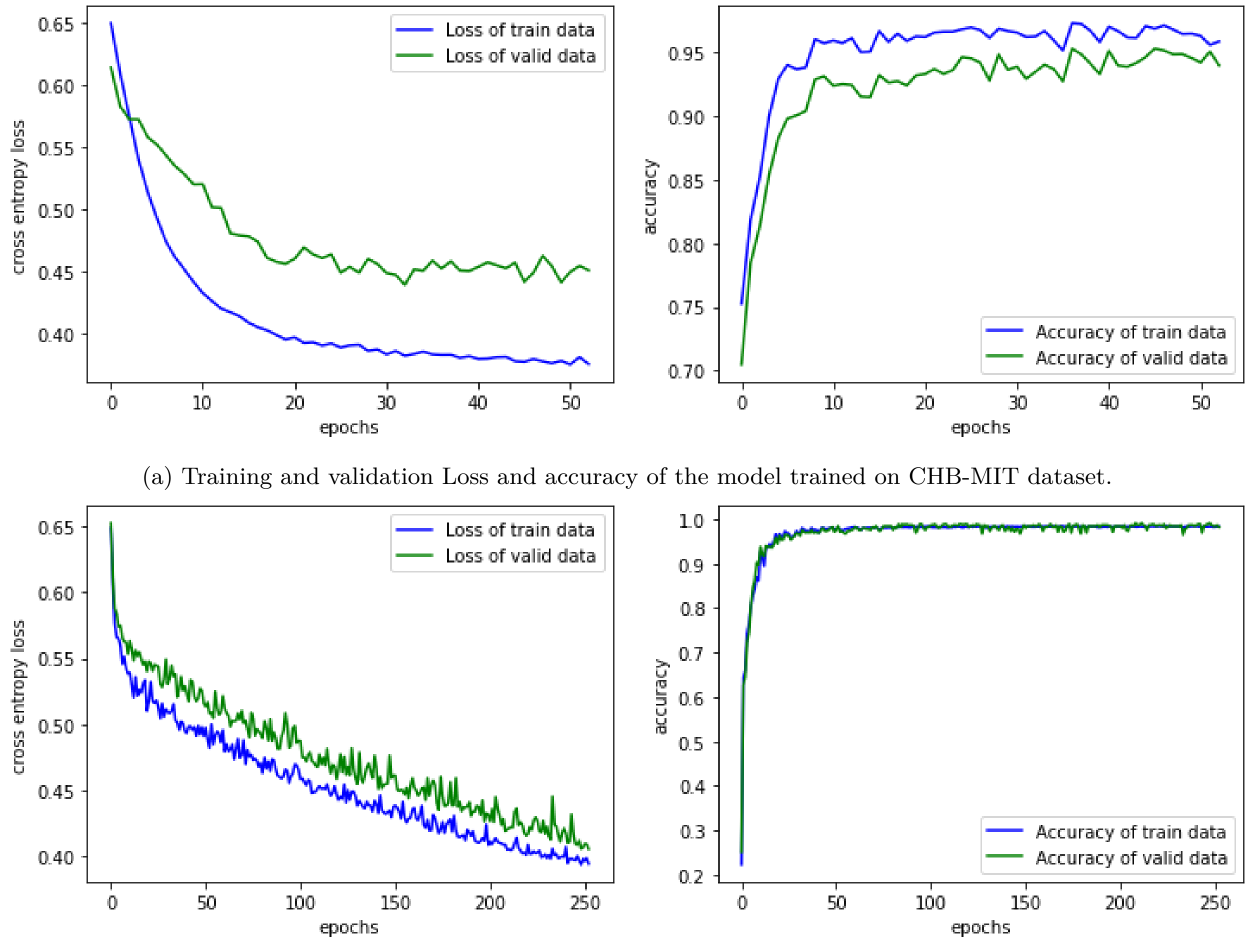
B.4.4 Model training

The model (Table B.1) was trained using each dataset (Table B.2) separately. Classifiers were implemented using Pytorch Paszke *et al.* (2017) , while data pre-processing was done using MNE-Python package Gramfort *et al.* (2013). The three-way holdout method was employed to optimise the performances in the hyper-parameter tuning step. In fact, we divided the data into stratified (having same classes proportions) three sets: training, validation and test set, each having data from different patients. To tackle the problem of imbalanced dataset, we used cost-sensitive cross-entropy loss. Adam optimizer was proposed as a gradient-based method with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and a learning rate of 0.002. We used the early stopping criteria to prevent over-fitting where training runs up to 500 epochs, or until the validation loss does not decrease anymore for at least 20 epochs.

B.5 Results and Discussion

We evaluated the performance of the proposed cross-subject method on the CHB-MIT and Ubonn university public databases. Figure B.2 shows the learning curves: the cross-entropy loss and the model accuracy curves, for both the train and the validation data, as a function of epochs. The training and validation loss decreases to a point of stability, with a small gap between the loss values indicating that the models are well-fitted.

Table B.3 presents the accuracy, sensitivity, F1-measure, and the false positive rate per hour, obtained when testing the model on unseen data for both databases. The results reveal that the



(a) Training and validation Loss and accuracy of the model trained on CHB-MIT dataset.

(b) Training and validation Loss and accuracy of the model trained on Ubonn.

Figure B.2 – Learning curves for all three datasets.

proposed architecture performs quite well. Indeed, using the smaller amount of data from in the CHB-MIT dataset, the accuracy was 91.82, sensitivity 91.93 %, and false alarm rate 0.005/hour. With the Ubonn university database, results were boosted to 99.60%, sensitivity of 100%, and false alarm rate of 0.057/hour.

Tableau B.3 – Model's performances for different datasets.

Database	Accuracy	Sensitivity	F-measure	FRP
CHM-MIT	91.82	91.93	95.73	0.005
UBonn	99.60	100	99.75	0.057

These results clearly show that CNNs are able to extract discriminative features in EEG data to allow cross-subject classification of inter-ictal and ictal data intervals. The hyper-parameters related to the network structure (number of layers, size of convolution filters), the activation and regularization function parameters, as well as the training parameters (learning rate and batch size, and optimization algorithm parameters), have been carefully chosen with the three-hold out method by observing the train and validation learning curves. Although it was not done in this study, performance can possibly benefit from extensive fine-tuning of hyperparameters. A comparison of our method to other CNN-based solutions is given in Table B.4. For a fair comparison, we focused on networks evaluated on the same datasets. However, unlike our cross-subject solution, all models are patient-specific except for Ullah *et al.* (2018). Regarding the CHB-MIT dataset, our method has a significantly better performance than the patient-specific CNN-based model in Antoniades *et al.* (2016). Although the method has a slightly lower accuracy than Zhou *et al.* (2018), it has the significant advantage of generalization ability across patients. In addition, the method had a better performance, although slightly, on the Ubonn university data than Ullah *et al.* (2018).

Tableau B.4 – Comparision of Benchmarking of recent seizure detction CNN-based studies and our work.

study	Dataset	Model	Method	Acc
Antoniades *et al.* (2016)	CHM-MIT	patient-specific	CNN on raw EEG signals	87.51
Zhou *et al.* (2018)	CHM-MIT	patient-specific	CNN on frequency domain signals	95.4
This work	CHM-MIT	cross-subject	CNN with separable depth -wise convolutions	91.82
Ullah *et al.* (2018)	Ubonn	cross-subject	pyramidal one-dimensional CNN	98.2
This work	Ubonn	cross-subject	CNN with separable depth -wise convolutions	99.60

Thanks to the specific choice of the type of layers in our study, the proposed architecture does not have more than 2,700 parameters in total. Such a low number of parameters has the advantage of allowing an implementation in connected devices for real-time seizure prediction. By comparison, other neural networks for epileptic seizure detection Ullah *et al.* (2018) Uyttenhove *et al.* (2020), used 8,326 and 16,401 parameters.

Overall, the classification results show that this study CNN architecture, which uses depth-wise convolution layers, performs well for epileptic seizure detection, using both a small and a fairly large database. Future work will focus on evaluating the architecture on even larger datasets, as well as application to other EEG-based classification tasks, such as epilepsy prediction, rather than detection, and seizure types categorization.