4I-RV1 - Projet Annuel - IA pour la Médecine

Yanis Aitaouit, Charles Batchaev, Elodie Pan, Yassine Guelaa

7 juin 2024

Détection des états épileptiques

Sommaire

I. Introduction

Ce rapport a été rédigé en utilisant LaTeX pour assurer un contrôle maximal sur la mise en page. L'objectif de ce document est de présenter les travaux réalisés dans le cadre du projet annuel "Détection des états épileptiques d'un sujet basée sur les algorithmes de machine learning en utilisant les techniques de « Ensemble feature extractor » et « Ensemble selection »".

L'électroencéphalographie (EEG) est une technique incontournable pour l'observation de l'activité cérébrale, notamment dans le domaine de la détection des états épileptiques. Grâce à sa capacité à fournir des informations détaillées sur l'activité électrique du cerveau, l'EEG joue un rôle crucial dans le diagnostic et la gestion des troubles épileptiques. Cependant, la classification précise des états épileptiques à partir des signaux EEG demeure un défi complexe en raison de la nature sophistiquée des données à analyser.

Les états épileptiques, tels que les phases interictale, préictale et ictale, se manifestent par des patterns spécifiques dans les signaux EEG. Pour détecter et prédire ces états avec précision, des approches avancées basées sur les algorithmes de machine learning et deep learning sont nécessaires. Traditionnellement, l'extraction et la classification des caractéristiques des signaux EEG étaient effectuées manuellement par des experts, mais cette méthode est non seulement laborieuse mais aussi sujette à des erreurs humaines. Avec les progrès technologiques récents, il est désormais possible d'automatiser ces processus en utilisant des techniques sophistiquées, offrant ainsi une nouvelle dimension d'efficacité et de précision.

La problématique principale de ce projet réside dans l'extraction et la classification efficaces des caractéristiques des signaux EEG. Jusqu'à présent, de nombreuses techniques ont été explorées, chacune présentant ses avantages et ses limites. Cependant, aucune méthode n'a encore permis d'atteindre des performances optimales de manière systématique. L'objectif de notre projet est de développer une approche innovante qui combine l'extraction et la sélection des caractéristiques de manière à maximiser la performance de la classification des états épileptiques.

Notre projet se concentre spécifiquement sur la mise en œuvre de deux techniques avancées : l'Ensemble Feature Extractor et l'Ensemble Selection. L'Ensemble Feature Extractor vise à associer plusieurs vecteurs caractéristiques obtenus à partir de différents extracteurs, afin d'exploiter pleinement les informations disponibles dans les signaux EEG. Parallèlement, l'Ensemble Selection sélectionnera les classifieurs les plus pertinents parmi ceux générés par diverses méthodes de machine learning, optimisant ainsi la performance globale de la classification.

Lors de nos travaux, nous avons identifié plusieurs défis majeurs à surmonter. Premièrement, il est crucial de choisir les méthodes d'extraction de caractéristiques les plus appropriées, telles que l'Analyse en Composantes Principales (PCA) ou les réseaux de neurones convolutionnels (CNN). Ensuite, il faut développer des techniques efficaces pour combiner ces caractéristiques de manière à maximiser l'information exploitable. Enfin, la sélection des algorithmes de machine learning les plus performants nécessite une évaluation rigoureuse basée sur des métriques telles que l'accuracy et le recall.

Pour atteindre nos objectifs, notre méthodologie se déroule en plusieurs étapes distinctes. Dans un premier temps, nous mettrons en œuvre et testerons diverses méthodes d'extraction

de caractéristiques. Ensuite, nous utiliserons ces caractéristiques pour alimenter des modèles de machine learning, en appliquant des techniques d'Ensemble Selection pour identifier les classifieurs les plus performants. Les performances des algorithmes seront évaluées de manière systématique, et les résultats obtenus serviront à affiner notre approche.

1.1 Définition de l'Épilepsie

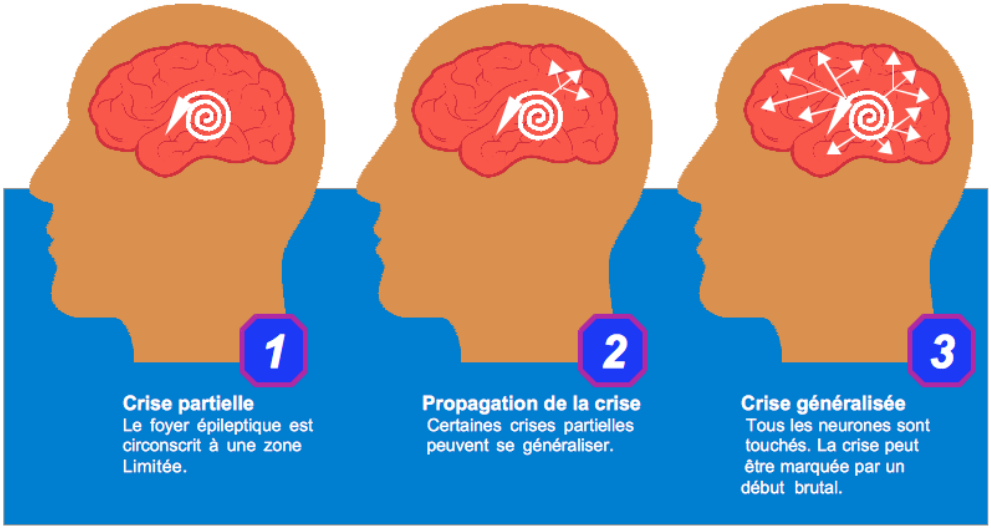

Figure 1: Etendue des crises.

L'épilepsie se définit comme une condition neurologique chronique caractérisée par la survenue de crises épileptiques récurrentes. Ces crises résultent d'une décharge excessive et synchronisée d'un groupe de neurones dans le cerveau, perturbant temporairement les fonctions cérébrales normales. La définition formelle inclut la présence de deux crises non provoquées survenant à plus de 24 heures d'intervalle.

1.2 Causes de l'Épilepsie

Les causes de l'épilepsie sont variées et peuvent être classifiées en plusieurs catégories :

- **Causes génétiques** : Certaines formes d'épilepsie sont héritées et peuvent être attribuées à des mutations spécifiques dans les gènes impliqués dans le fonctionnement neuronal.
- **Causes structurelles** : Des anomalies dans la structure du cerveau, telles que des malformations congénitales, des traumatismes crâniens, ou des tumeurs cérébrales, peuvent provoquer des crises.
- **Causes métaboliques** : Des déséquilibres métaboliques, tels que des troubles du métabolisme des glucides ou des lipides, peuvent aussi être à l'origine de l'épilepsie.
- **Causes infectieuses** : Les infections du système nerveux central, comme la méningite ou l'encéphalite, peuvent déclencher des crises épileptiques.
- **Causes immunitaires** : Certaines épilepsies peuvent être liées à des dysfonctionnements du système immunitaire, ou des anticorps attaquant les neurones.

1.3 Symptômes de l'Épilepsie

Les symptômes de l'épilepsie sont divers et dépendent du type de crise et de la région du cerveau affectée. Les principaux symptômes incluent :

- **Crises généralisées** : Affectent l'ensemble du cerveau et peuvent entraîner une perte de conscience, des convulsions, et des secousses musculaires incontrôlées.
- **Crises focales** : Affectent une partie spécifique du cerveau, provoquant des symptômes tels que des mouvements anormaux, des sensations étranges, ou des altérations de la perception.
- **Crises d'absence** : Caractérisées par des périodes brèves de perte de conscience, souvent sans mouvements anormaux notables, fréquentes chez les enfants.
- **Crises myocloniques** : Impliquent des secousses musculaires rapides et soudaines, souvent sans perte de conscience.

1.4 Conséquences de l'Épilepsie

L'épilepsie a des conséquences significatives sur la vie des individus affectés, impactant divers aspects :

- **Conséquences physiques** : Risque accru de blessures pendant les crises, ainsi que des effets secondaires liés aux traitements médicamenteux.
- **Conséquences psychologiques** : Dépression, anxiété et troubles de l'humeur sont fréquemment observés chez les patients épileptiques.
- **Conséquences sociales** : Stigmatisation, discrimination et isolement social sont courants, affectant la qualité de vie et les relations interpersonnelles.
- **Conséquences économiques** : Les coûts associés aux soins de santé, à la perte de productivité et à l'incapacité de travailler représentent une charge économique importante.

1.5 Pistes de Guérison et Traitements

Bien que l'épilepsie soit une maladie chronique, plusieurs pistes de guérison et traitements permettent de gérer et de réduire la fréquence des crises :

- **Traitements médicamenteux** : Les antiépileptiques sont les principaux médicaments utilisés pour contrôler les crises. Leur efficacité varie selon le type d'épilepsie et la réponse individuelle du patient.
- **Chirurgie** : Dans les cas réfractaires aux médicaments, une intervention chirurgicale peut être envisagée pour retirer la zone cérébrale à l'origine des crises.

- **Stimulation nerveuse** : La stimulation du nerf vague et la stimulation cérébrale profonde sont des techniques utilisées pour réduire la fréquence des crises chez certains patients.
- **Régimes alimentaires** : Le régime cétogène, riche en graisses et pauvre en glucides, a montré une efficacité chez certains patients, notamment les enfants.
- **Thérapies comportementales et psychologiques** : La prise en charge psychologique et le soutien psychothérapeutique sont essentiels pour améliorer la qualité de vie des patients épileptiques.

II. Feature Extraction

La complexité et l'hétérogénéité de l'épilepsie rendent nécessaire l'utilisation d'outils avancés pour analyser et interpréter les vastes quantités de données cliniques et biomédicales disponibles. Afin de mieux comprendre les mécanismes de la maladie, de faciliter le diagnostic et de personnaliser les traitements, il est essentiel de recourir à des techniques d'extraction de données et d'analyse statistique sophistiquées. Dans ce contexte, des méthodes d'extraction des caractéristiques sur des données tabulaires ou des données statiques sont important. Après une recherche et études des méthodes, nous avons retenu les méthodes d'extraction telles que l'Analyse en Composantes Principales (PCA), l'Analyse Discriminante Linéaire (LDA), Feature Agglomeration, Regression Ridge et Lasso et Convolutional Neural Network (CNN), une méthode qui n'est, à la base, pas conçu pour les données tabulaires ou données statiques. Les meilleures performances ont été montré par la méthode PCA (accuracy 78%) et CNN (accuracy 89%), c'est pourquoi nous les allons exposer en détail.

2.1 Analyse en Composantes Principales (PCA)

L'analyse en composantes principales (PCA) est une technique d'extraction de features largement utilisée en statistiques et en apprentissage automatique. Son objectif principal est de réduire la dimensionnalité des données tout en conservant le maximum d'informations possible. PCA est basée sur la transformation linéaire des données dans un nouvel espace de dimensions réduites appelé espace des composantes principales.

2.1.1 Principes Fondamentaux

PCA utilise une transformation linéaire orthogonale pour convertir un ensemble de variables potentiellement corrélées en un ensemble de variables linéairement non corrélées appelées composantes principales. Les composantes principales sont ordonnées de telle sorte que la première composante principale capture la plus grande variance dans les données, la deuxième capture la deuxième plus grande variance, et ainsi de suite.

2.1.2 Méthodologie

- **Centrage des Données** : Avant d'appliquer PCA, il est essentiel de centrer les données en soustrayant la moyenne de chaque variable. Cela garantit que les composantes captureront la variance des données plutôt que leur niveau moyen.

- **Calcul des Composantes Principales** : Les composantes principales sont calculées en diagonalisant la matrice de covariance des données ou en utilisant la décomposition en valeurs singulières (SVD). La première composante principale est obtenue en maximisant la variance, tandis que les composantes suivantes sont obtenues de manière à maximiser la variance restante sous la contrainte d'orthogonalité.

- **Réduction de Dimension** : Une fois les composantes principales calculées, la dimension des données peut être réduite en ne conservant que les premières composantes

principales qui capturent la variance la plus significative. Cela permet de simplifier la représentation des données tout en conservant l'essentiel de l'information.

2.1.3 Interprétation des Résultats

La variance expliquée par chaque composante principale peut être utilisée pour évaluer l'importance de chaque composante dans la représentation des données. En traçant le ratio de variance expliquée par rapport au nombre de composantes principales, il est possible de déterminer le nombre optimal de composantes à conserver pour une représentation adéquate des données.

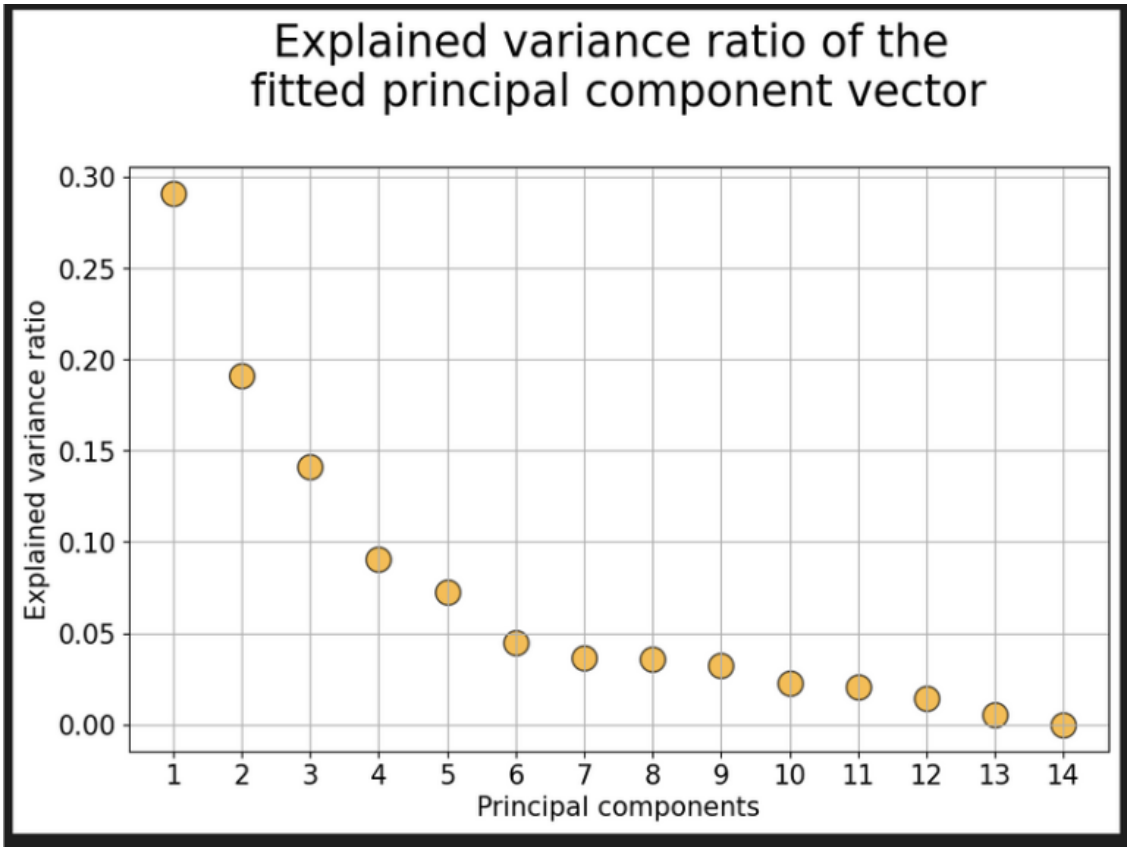

Figure 2: Visualisation des ratios de variance de chaque composant.

2.1.4 Application Pratique

Une fois que les composantes principales ont été calculées, elles peuvent être utilisées pour visualiser les données dans un espace de dimension réduite. Cela permet souvent une meilleure compréhension de la structure des données et peut faciliter la détection de motifs ou de groupes.

2.1.5 Conclusion

En résumé, l'analyse en composantes principales est une technique puissante pour réduire la dimensionnalité des données tout en préservant l'information importante. Elle est largement utilisée dans de nombreux domaines, y compris l'apprentissage automatique, la vision par ordinateur et la bioinformatique, pour explorer et analyser les données de manière efficace et informative..

2.2 Analyse Discriminante Linéaire (LDA)

L'analyse discriminante linéaire (LDA) est une méthode d'extraction de features utilisée pour trouver une combinaison linéaire des caractéristiques qui sépare au mieux deux ou plusieurs classes de données. Contrairement à PCA, qui se concentre sur la maximisation de la variance, LDA cherche à maximiser la séparation entre les classes.

2.2.1 Principes Fondamentaux

LDA cherche à maximiser le rapport des variances entre les classes et à minimiser le rapport des variances à l'intérieur des classes. Pour ce faire, LDA projette les données sur un nouvel espace de features de dimension inférieure tout en préservant la séparation maximale entre les classes..

2.2.2 Méthodologie

- **Prétraitement des Données** : Comme pour toute technique d'apprentissage automatique, il est important de prétraiter les données en enlevant le bruit, en gérant les valeurs manquantes et en normalisant les caractéristiques si nécessaire.
- **Calcul des Transformations Linéaires** : LDA calcule les vecteurs propres et les valeurs propres de la matrice de dispersion entre les classes et de la matrice de dispersion intra-classes. Ces vecteurs propres sont utilisés pour projeter les données sur un nouvel espace de features.
- **Projection des Données** : Les données sont projetées sur le nouvel espace de features en utilisant les vecteurs propres calculés. Cette projection permet de réduire la dimensionnalité des données tout en conservant la structure discriminante.

2.2.3 Interprétation des Résultats

Les composantes résultantes de LDA peuvent être interprétées comme des axes qui maximisent la séparation entre les classes. En visualisant ces composantes, il est possible d'observer la distribution des données dans un espace de dimension réduite et de détecter les motifs ou les clusters qui séparent les différentes classes.

2.2.4 Application Pratique

Une fois que les données ont été projetées sur les composantes principales de LDA, elles peuvent être utilisées pour entraîner des modèles de classification tels que les classificateurs linéaires, les arbres de décision ou les réseaux de neurones. Ces modèles peuvent ensuite être utilisés pour prédire les classes des nouvelles données.

2.2.5 Conclusion

En résumé, l'analyse discriminante linéaire (LDA) est une technique puissante pour réduire la dimensionnalité des données tout en préservant la structure discriminante entre les classes. Elle est largement utilisée dans de nombreux domaines, y compris la reconnaissance de motifs, la classification et la vision par ordinateur, pour analyser et interpréter des ensembles de données complexes.

2.3 Réseau Neuronal convolutif (CNN)

Les réseaux de neurones convolutifs (CNN) sont principalement utilisés pour l'analyse d'images, mais ils peuvent également être adaptés pour l'extraction de caractéristiques à partir de données tabulaires. Cette section décrit le processus d'extraction des caractéristiques à l'aide d'un CNN.

2.3.1 Préparation des Données

```
# Vérification et reformattage de y_train_sampled pour la classification multiclasse
label_encoder = LabelEncoder()
y_train_encoded = label_encoder.fit_transform(y_train_sampled.ravel())
y_train_categorical = to_categorical(y_train_encoded)

# Mise à jour de la dernière couche Dense pour qu'elle corresponde au nombre de classes
#num_classes = y_train_categorical.shape[0]
num_classes = y_train_categorical.shape[1]
```

Figure 3: processus d'encodage des étiquettes et le remodelage des données pour les rendre compatibles avec CNN.

Les CNN sont des modèles de réseaux de neurones profonds qui utilisent des convolutions pour extraire des caractéristiques pertinentes des données. Ils sont particulièrement efficaces pour détecter des motifs spatiaux, ce qui les rend idéaux pour l'analyse d'images. Cependant, leurs principes peuvent être appliqués à d'autres types de données après une transformation appropriée.

2.3.2 Construction du Modèle CNN

Tout d'abord, il est nécessaire de préparer les données pour qu'elles soient compatibles avec les attentes du modèle CNN. Cela implique d'encoder les étiquettes de classe pour qu'elles puissent être utilisées dans une classification multiclasses. Ensuite, les données d'entrée

```python
# Modèle CNN avec trois couches convolutionnelles
cnn_model = Sequential([
    # Augmente la taille de l'input si nécessaire ou ajuste les paramètres
    Conv1D(128, 3, activation='relu', padding='same', input_shape=(x_train_sampled.shape[1], 1)),
    MaxPooling1D(2),
    # Ajout d'une seconde couche convolutive avec padding pour conserver la dimension
    Conv1D(128, 3, activation='relu', padding='same'),
    MaxPooling1D(2),
    Conv1D(128, 3, activation='relu', padding='same'),
    MaxPooling1D(2),
    Flatten(name='flatten_layer'),
    Dense(100, activation='relu'),
    Dropout(0.5),
    Dense(num_classes, activation='softmax')
])

# Compilation du modèle avec la métrique correcte
cnn_model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])
```

Figure 4: Architecture du modèle CNN avec des couches convolutives, des couches de pooling et des couches denses.

doivent être remodelées pour correspondre aux attentes du CNN, ce qui peut nécessiter l'ajout de dimensions supplémentaires.

2.3.3 Entraînement du Modèle

Le modèle CNN est construit avec plusieurs couches convolutives suivies de couches de pooling et de couches denses. Les couches convolutives sont responsables de l'extraction des caractéristiques, tandis que les couches de pooling réduisent la dimensionnalité des données, rendant le modèle plus efficace et moins susceptible de surajuster. Le modèle est ensuite compilé avec un optimiseur (tel que Adam) et une fonction de perte appropriée (comme la cross-entropie catégorique) pour préparer le processus d'entraînement.

2.3.4 Extraction des Caractéristiques

L'entraînement du modèle implique l'ajustement du modèle sur les données avec un certain nombre d'époques et une taille de lot définie. Pendant l'entraînement, les performances du modèle, mesurées en termes de perte et de précision, sont surveillées à chaque époque. Les résultats de l'entraînement montrent une amélioration progressive de la précision, indiquant que le modèle apprend efficacement à partir des données.

2.3.5 Sauvegarde et Utilisation des Caractéristiques Extraites

Une fois le modèle entraîné, il peut être utilisé pour extraire des caractéristiques des ensembles de données d'entraînement et de test. Cela se fait en utilisant la couche flatten du CNN, qui convertit les données en une forme adaptée pour l'analyse ultérieure. Les caractéristiques extraites peuvent être sauvegardées pour une utilisation future dans d'autres modèles ou analyses, offrant une grande flexibilité pour des études ultérieures.

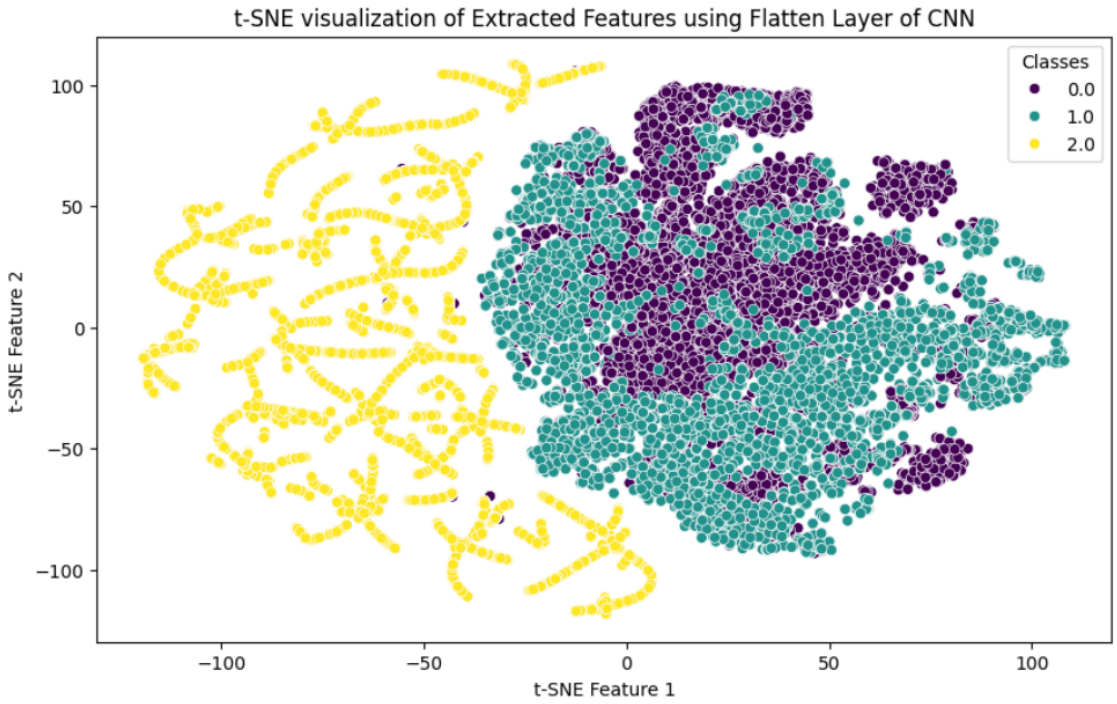

Figure 5: Extraction des Caractéristiques à partir de la Couche Flatten.

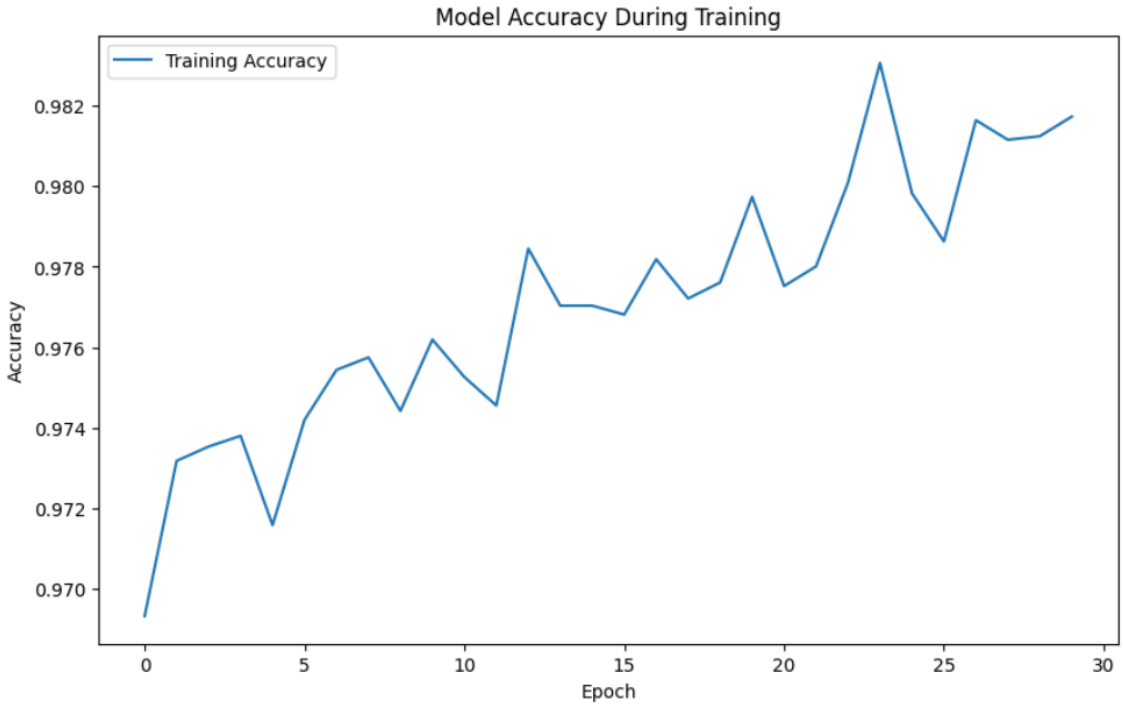

Figure 6: Précision du Modèle sur les Données d'Entraînement.

2.3.6 Résultats et Interprétation

Description : Graphique illustrant la précision élevée du modèle sur les ensembles de données d'entraînement, démontrant l'efficacité du CNN dans l'extraction des caractéristiques.

Les résultats obtenus montrent une précision élevée sur les ensembles de données d'entraînement, ce qui indique que le modèle CNN est capable de capturer des caractéristiques pertinentes pour la tâche de classification. Cette méthode a démontré des performances supérieures par rapport à d'autres techniques d'extraction de caractéristiques, justifiant son utilisation dans ce contexte.

En résumé, les réseaux de neurones convolutifs offrent une méthode puissante pour l'extraction de caractéristiques, même pour des données qui ne sont pas naturellement des images, grâce à leur capacité à détecter des motifs complexes et pertinents. Leurs applications vont bien au-delà de l'analyse d'images, rendant cette technique précieuse pour divers types de données et problèmes d'analyse.

2.4 Justification du Choix du Modèle CNN

Le modèle CNN a montré une précision impressionnante de 97.8% sur les données d'apprentissage, surpassant les autres techniques testées. Bien que principalement utilisés pour les images, les CNN peuvent être adaptés pour les données tabulaires avec un prétraitement approprié. Les résultats expérimentaux ont démontré une haute précision et une stabilité des performances, ce qui renforce la confiance dans cette méthode.

En raison de ces avantages, nous avons sélectionné le modèle CNN comme méthode principale d'extraction de caractéristiques. Cette approche robuste et efficace permet de capturer des informations essentielles pour améliorer la compréhension des mécanismes de l'épilepsie, faciliter le diagnostic et personnaliser les traitements.

III. Ensemble Selection

Les états épileptiques sont classés en trois catégories : Inter-ictal (pendant les crises), Pre-ictal (avant les crises) et Ictal (durant les crises). Une fois les données traitées et extraites via les méthodes Feature Extraction, l'étape suivante consiste à utiliser un modèle permettant de distinguer et détecter ces trois états. Pour cela, différents modèles existants tels que les modèles Random Forest (RF), Decision Tree (DT) ou encore K-Nearest Neighbours (KNN) peuvent être utilisés. Cependant, les résultats sont jugés insatisfaisants, avec une précision aux alentours de 60%. Améliorer ces performances est nécessaire pour détecter correctement et précisément les états épilettiques. Pour ce faire, une technique possible consiste à assembler plusieurs modèles de base et combiner leurs prédictions : c'est l'Ensemble Selection (Sélection d'Ensemble).

3.1 Performances individuelles des modèles de base

Pour entraîner les modèles de base, les données ont été préalablement traitées et sont passées par l'étape d'extraction des caractéristiques avec la méthode jugée la plus performante : le réseau CNN.

3.1.1 Decision Tree (DT)

La méthode de Decision Tree (Arbre de Décision en français) est une technique de Machine Learning supervisée qui est utilisée pour des problèmes de classification et de régression. Il s'agit d'une structure composée de noeuds. Chaque noeud correspond à une condition qui peut amener à plusieurs réponses, et se dirigeant vers le prochain noeud. L'arbre de décision sert souvent comme élément de base pour plusieurs méthodes.

En utilisant le dataset fourni concernant les états épileptiques, la méthode DT a permis d'atteindre une précision générale de 75%.

	precision	recall	f1-score	support
0.0	0.86	0.81	0.83	2500
1.0	0.44	0.51	0.47	692
2.0	0.05	0.09	0.07	11
accuracy			0.75	3203
macro avg	0.45	0.47	0.46	3203
weighted avg	0.76	0.75	0.75	3203

Figure 7: Rapport des résultats de la méthode DT sur l'ensemble de test issu du dataset sur les état épileptiques après extraction des caractéristiques avec la méthode CNN.

3.1.2 Random Forest (RF)

Une random Forest (Forêt d'arbres décision en français) est une méthode d'ensemble qui combine les résultats des dizaines, voire centaines d'arbres de décision qui la composent. Sur un dataset donné, chaque arbre est entraîné sur un sous-ensemble. Ainsi, les résultats de tous les arbres sont combinés pour obtenir une réponse finale.

En utilisant le dataset fourni concernant les états épileptiques, la méthode RF a permis d'atteindre une précision générale de 79%.

	precision	recall	f1-score	support
0.0	0.86	0.87	0.86	2500
1.0	0.52	0.51	0.51	692
2.0	0.14	0.18	0.16	11
accuracy			0.79	3203
macro avg	0.51	0.52	0.51	3203
weighted avg	0.79	0.79	0.79	3203

Figure 8: Rapport des résultats de la méthode RF sur l'ensemble de test issu du dataset sur les état épileptiques après extraction des caractéristiques avec la méthode CNN.

3.1.3 K-Nearest Neighbours (KNN)

La méthode des K-Nearest Neighbours (K plus proches voisins en français) est une méthode de classification supervisée qui consiste à estimer la sortie associée (la classe) à la valeur en entrée en fonction des k plus proches voisins (k étant un nombre positif défini). Cette technique évalue la distance entre la valeur x d'entrée et les points de données dont nous connaissons la classe. La classe affectée à la valeur x sera la classe des points les plus proches.

En utilisant le dataset fourni concernant les états épileptiques, la méthode KNN a permis d'atteindre une précision générale de 76%.

	precision	recall	f1-score	support
0.0	0.89	0.80	0.84	2500
1.0	0.48	0.65	0.55	692
2.0	0.15	0.27	0.19	11
accuracy			0.76	3203
macro avg	0.51	0.57	0.53	3203
weighted avg	0.80	0.76	0.78	3203

Figure 9: Rapport des résultats de la méthode KNN sur l'ensemble de test issu du dataset sur les état épileptiques après extraction des caractéristiques avec la méthode CNN.

3.1.4 Intérprétation des résultats

A partir des résultats obtenus suite à l'entraînement des trois modèles de base, nous en déduisons que la méthode Random Forest est la plus performante pour notre dataset sur les états épileptiques avec une précision générale de 79%.

Cependant, pour un domaine aussi sensible que le domaine médical, tout erreur de prédiction est difficilement pardonnable. Ainsi, il est nécessaire de trouver des moyens d'améliorer ces performances afin d'éviter tout mauvais diagnostic. Pour cela, nous utiliserons une méthode de Sélection d'Ensemble : la méthode Boosting.

3.2 Boosting

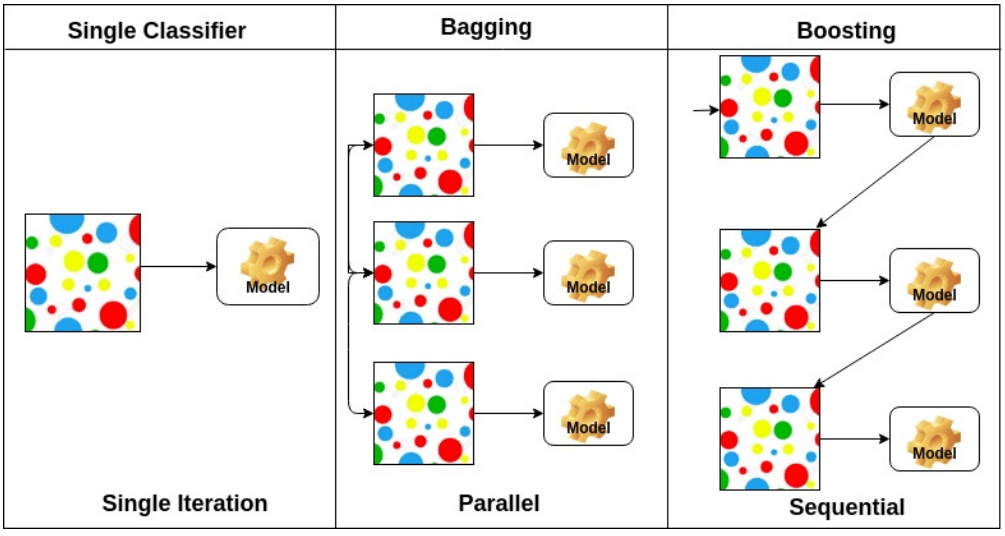

Figure 10: Comparaison des Types d'Ensemble : Classifieur Unique, Bagging et Boosting.

La méthode Boosting est une méthode d'apprentissage ensembliste qui combine un

ensemble d'algorithmes à faibles performances (appelés "weak learners") individuellement pour former un nouveau modèle plus efficace. Dans cette méthode, les "weak learners" sont entraînés à la suite, un par un. Ainsi, chaque "weak learner" tente de compenser les erreurs des algorithmes précédents pour obtenir le meilleur résultat au final.

Dans notre cas, nous utilisons l'algorithme Adaboost. Cet algorithme prend généralement les arbres décisionnels en tant que "weak learners" mais pour notre expérience, nous utilisons la forêt d'arbres car il s'agit de la méthode avec les meilleures performances individuelles.

3.2.1 Adaboost

L'algorithme Adaboost atrtibue dans un premier temps le même poids pour chaque ligne du dataset donné. Suite à l'entraînement du modèle, une note lui est attibué en fonction de ses performances. Cette note permettra de classer les "weak learners" par ordre d'importance lors du vote final au moment des prédictions. Pour permettre au "weak learner" suivant de compenser les erreurs de son prédécesseur, les poids sont augmentés pour les lignes du dataset présentant des erreurs. De cette façon, le "weak learner" sera entraîné de manière à maximiser les bonnes réponses sur les lignes à poids élevés. Et ce processus se répète pour chaque "weak learner".

Avec cette technique, la précision générale a pu atteindre la valeur de 80%.

	precision	recall	f1-score	support
0.0	0.87	0.86	0.87	2500
1.0	0.54	0.56	0.55	692
2.0	0.17	0.18	0.17	11
accuracy			0.80	3203
macro avg	0.53	0.54	0.53	3203
weighted avg	0.80	0.80	0.80	3203

Figure 11: Rapport des résultats de la méthode Adaboost sur l'ensemble de test issu du dataset sur les état épileptiques après extraction des caractéristiques avec la méthode CNN.

IV. Conclusion

La détection des états épileptiques représente un défi complexe en raison de la nature sophistiquée et hétérogène des signaux EEG. À travers notre état de l'art, nous avons exploré diverses approches avancées de machine learning et de deep learning pour surmonter ces défis. Les méthodes étudiées incluent l'Analyse en Composantes Principales (PCA), l'Analyse Discriminante Linéaire (LDA), et les Réseaux de Neurones Convolutifs (CNN), chacune offrant des avantages distincts en termes de réduction de la dimensionnalité, de séparation des classes et d'extraction de caractéristiques pertinentes.

L'analyse de PCA et de LDA a montré des résultats prometteurs en termes de réduction de la dimensionnalité et de maximisation de la variance et de la séparation entre les classes respectivement. Toutefois, c'est l'application des CNN qui a démontré la plus grande efficacité, atteignant une précision de 97.8% sur notre dataset, grâce à leur capacité à extraire des caractéristiques complexes et à détecter des motifs pertinents même dans des données tabulaires transformées.

Pour améliorer encore les performances, nous avons utilisé des techniques d'ensemble, telles que Random Forest et Boosting, qui ont permis d'augmenter la précision globale jusqu'à 80%. La méthode de Boosting, en particulier avec l'algorithme Adaboost, a montré une capacité significative à améliorer les performances en compensant les erreurs des classifieurs individuels.

La combinaison de ces approches a permis d'obtenir des résultats robustes, soulignant l'importance d'une extraction et d'une sélection des caractéristiques efficaces pour la classification des états épileptiques. Notre choix final de la méthode CNN, soutenue par des techniques d'ensemble, se justifie par sa précision et sa robustesse supérieures dans ce contexte médical sensible.

En conclusion, notre projet annuel a démontré la faisabilité et l'efficacité de l'application des techniques avancées de machine learning et deep learning pour la détection des états épileptiques. Les méthodes développées offrent des perspectives prometteuses pour améliorer le diagnostic et la gestion de l'épilepsie, avec des implications significatives pour la personnalisation des traitements et l'amélioration de la qualité de vie des patients. Les futures recherches devraient se concentrer sur l'intégration de ces approches dans des systèmes cliniques en temps réel pour maximiser leur impact pratique.