From EEG signal to classification in Alzheimer disease: A mini review

Ghina H. Shaaban, Mira M. Antar, Mira O. Al-Yaman, Nataly I. Mousharafieh
Electrical and Computer Engineering Department
Rafik Hariri University
Mechref, Lebanon
(ShaabanGH, AntarMM, ALYamanMO, MousharafiehNI)@students.rhu.edu.lb

Maher M. Sabbah
Electrical and Computer Engineering Department
Rafik Hariri University
Mechref, Lebanon
SabbahMM@rhu.edu.lb

Mahmoud Hassan
MINDIG, 35000 Rennes, France
School of Science and Engineering
Reykjavik University
Reykjavik, Iceland.
mahmoud.hassan@mindig.io

Mohamad O. Diab
Electrical and Computer Engineering Department
Rafik Hariri University
Mechref, Lebanon
Diabmo@rhu.edu.lb

2023 Seventh International Conference on Advances in Biomedical Engineering (ICABME) | 979-8-3503-2585-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICABME59496.2023.10293112

***Abstract*—Alzheimer's disease (AD) is a progressive neurodegenerative disorder that is characterized by cognitive decline and memory loss. Early and accurate diagnosis of this disease is still very challenging. In order to identify relevant biomarkers to assess a subject's cognitive impairment, it has been proved that the electroencephalography (EEG) signals can help in the understanding of various brain dysfunctions. In this mini-review paper, our aim was to present the potential of EEG in the context of Alzheimer disease and the challenges to promote EEG as a diagnostic/prognostic tool for detecting AD. We focus on the integration of EEG data with machine learning algorithms to help in assessing AD. The approach involves a systematic process of EEG data collection, signal preprocessing, feature extraction, feature selection and classification.**

Keywords*—Alzheimer Disease, EEG Signals, SVM, KNN, DT

I. Introduction

Neurodegenerative disorders (NND), such as dementia, affect millions of people worldwide. They are long-term diseases that gradually deteriorate and destroy various sections of the nervous system, mainly the brain [1]. Alzheimer disease (AD) is the most common form of dementia and can contribute to 60–70% of the cases [2].

AD is considered the most prevalent neurological brain condition affecting adults aged 65 and older. The percentage of people with Alzheimer's dementia increases dramatically with age. Five percent of people age 65 to 74, 13.1% of people age 75 to 84, and 33.3% of people age 85 or older have Alzheimer's dementia [2]. It is described as a gradual condition that begins with mild memory loss and may proceed to loss of capacity to converse and adjust to the surroundings. Alzheimer's disease affects brain regions that govern cognition, memory, and language [2]. This disease could be diagnosed using a range of traditional clinical approaches; for instance, through the imaging methods like positron emission tomography (PET), magnetic resonance imaging (MRI) or computed tomography (CT) scan, nuclear magnetic resonance spectroscopy (NMRS) that can be used to measure biomarkers [3]. Nevertheless, these diagnostic tools are considered to be expensive and not always available. The electroencephalogram (EEG), on the other hand, can be a very promising in detecting Alzheimer's disease since it is non-invasive, safe and painless, has a high temporal resolution, cost-effective, portable, and accessible [6].

A large number of studies have been done and used EEG in AD for understanding physiopathology and classifying AD patients vs. healthy controls or vs. patients with mild cognitive impairment (MCI) [6].

EEG has been shown to be very important in improving our understanding of AD. Yet, EEG is not used as a diagnostic tool [6]. This is due to many factors including the focus of previous studies on group-level analysis, the lack of large sample size and the high variability in the EEG analysis pipeline [16]. Combining EEG with machine learning (ML) algorithms has been proposed as an approach to go beyond the average brain into precision medicine in AD [6].

This paper’s sections include a description of the methodology used to classify AD using EEG signals, potential applications of EEG, limitations of EEG, future research directions, and a conclusion of the paper.

II. Methodology

In this section, we provided a description of the methodology used to classify AD using EEG signals. At the beginning, we described the EEG data collection method. Following that, the EEG signal pre-processing method, which involved removing artifacts from the EEG signals, sampling, and filtering. Following that, the feature extraction method using different representations in which the EEG can be analyzed. Then, the feature selection method using different statistical tests. Lastly, the classification method using ML models.

A. EEG Data Collection

EEG is used to analyze the behavior of the brain through measuring the electrical activity using electrodes placed on the scalp (Fig. 1). The positioning of the electrodes follows an international 10–20 system which is a well-accepted technique for describing and applying the placement of scalp electrodes [6]. Electrical signals are detected by the EEG electrodes typically having a frequency range between 0.5 and 70 Hz.

B. Signal Preprocessing

Preprocessing is a sequence of signal processing procedures (Fig. 1) performed on the EEG data and it usually refers to minimizing noise to acquire a clean signal of interest. It is a necessary step to extract the relevant information from noise and obtain a real neural signal of the brain since artifacts and noise might taint the original signal. Eye movements, blinks, electrode channel drift, magnetic fields of the electronic devices, blood pressure, breathing, limb movements, power line interference or other human movements are just among these noises and distortions that can contaminate EEG signals and prevent a clean representation of brain activity from being captured in the data. To avoid further incorrect interpretation, it is critical to eliminate noise and artifacts from the EEG data using filtering, segmentation, and artifact identification [3]. Different and important frequency bands are usually analyzed, such as delta (1-4 Hz), theta (4–8 Hz), alpha (8–12 Hz), mu (12–16 Hz), (16-20 Hz) beta, and gamma (25–40 Hz) [4]. Hence, filtering the EEG signal at this frequency range can assist in isolating particular frequency bands of interest and enhance the analysis' level of specificity and sensitivity.

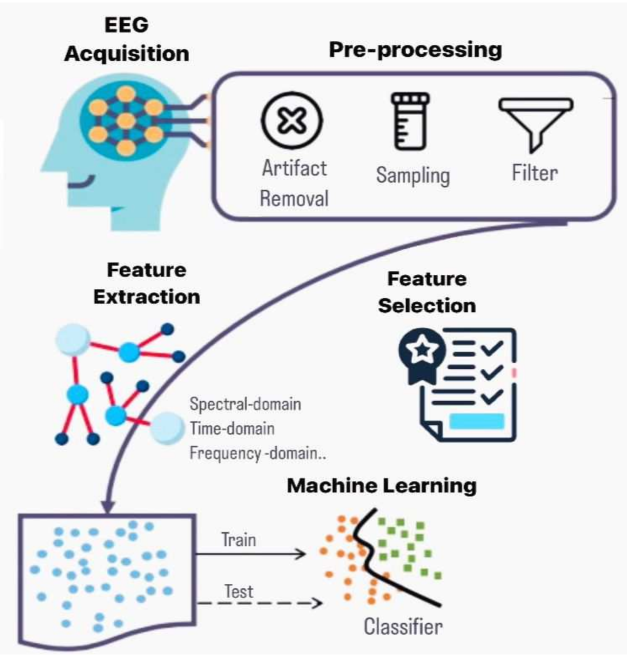

Fig .1: ML-based AD Classification Framework

C. Feature Extraction

Feature extraction is the process of converting data into numerical feature. It's an essential step for reducing the resources needed to process a large dataset without losing (when possible) crucial or pertinent information, removing redundant data from a dataset to reduce its dimensionality, building the ML model with less computational effort, speeding up learning and training, improving the accuracy of the models, lowering the risk of overfitting, and enhancing visualization of data [7]. Different parameters can be extracted from the EEG signal, such as amplitude, frequency, spectral power, or temporal pattern. Commonly used features include power spectral density, wavelet coefficients, statistical measures (mean, variance, correlation, etc.), and connectivity measures [7].

The spectral features of EEG data are meant to show the properties of the EEG frequency sub-bands at distinct scalp regions [5]. These features have been widely used due to their relationship with brain functions. Additionally, functional connectivity features have been used: In recent years, multiple studies have looked at brain connections to identify how data is processed, delivered to, received by, or exchanged between distinct brain areas during health and disease [5]. Functional connectivity is an area of neuroscience that aims to quantify the statistical correlations between the dynamics of simultaneously recorded signals in order to assess brain connectivity [5]. In general, synchronization likely estimates the synchronization between two signals by analyzing their nonlinear and linear relationships, which may be complicated and considerably different. Several other features also exist and can be extracted from the EEG signals such as nonlinearity, complexity, entropy, etc.

D. Feature Selection

Feature selection method is an optional step that can be applied on the features extracted from the EEG signals since only a limited number of features in the EEG dataset can be used for feeding machine learning algorithms, and the rest can be considered either redundant or unnecessary [7]. These redundant and irrelevant features in the dataset can cause a negative impact on the ML model's overall accuracy and performance. As a result, this step aims to reduce the dimensionality of the feature space while preserving the relevant information, so it is critical to discover and choose the most relevant features from the data while eliminating irrelevant ones which can be achieved through feature selection method [7]. Feature selection can be performed with the statistical tests t-test, Analysis of Variance (ANOVA) test, Chi-Square test, Least Absolute Shrinkage and Selection Operator (LASSO) test [7], etc.

E. Classification

For the classification step, the data will be split into a training set and a testing set as illustrated in Fig. 2. A model is trained to categorize the EEG signals into multiple categories or classes, such as healthy vs. diseased (AD) [6], after the relevant features are selected. Different machine learning methods, such as decision trees, support vector machines, K-nearest neighbors, random forests, neural networks, or deep learning models, can be used. The classification model learns the patterns and correlations between the characteristics and the related classes by training on labeled data where the class labels are known [8].

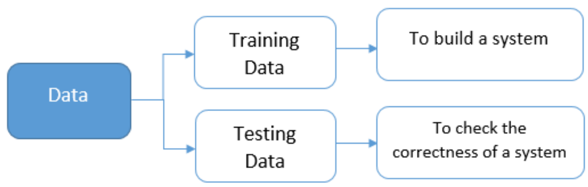

Fig. 2: Training and Testing Sets

Performance indicators like accuracy, sensitivity, specificity, precision, or F-score can be used to assess the trained classification model. This evaluation measures the model's accuracy in classifying EEG signals and offers an estimate of how well it generalizes [8].

The trained model is tested and validated by using it on EEG data in this stage. This data should be distinct from the training data. Besides, the model's performance on the testing data provides an indication of its robustness and generalizability [8].

Hyperparameter tuning is a crucial stage done to develop robust predictive models. Techniques like grid search, random search, or Bayesian optimization can be used to optimize the hyperparameters of the classification model in order to enhance its performance (improved accuracy, precision, recall, and other metrics) [9]. Optimization can also lead to better generalization, avoidance of overfitting, and performance.

Several algorithms exist for classification such as the Support Vector Machine (SVM), the K-Nearest Neighbor (KNN) and Decision Tree (DT).

SVM is a supervised learning classifier used to solve classification and regression problems. SVM classifier works by making a straight line called hyperplane (Fig.3) between two classes. All of the data points on one side of the line will represent a category and the data points on the other side of the line will be put into a different category [10].

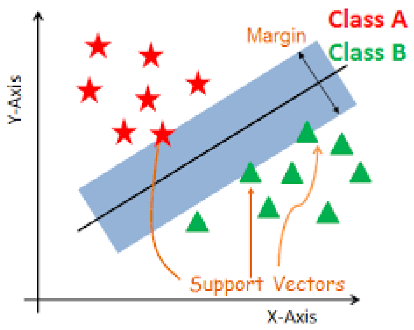

Fig. 3: Support Vector Machine Architecture

Moreover, KNN is a supervised learning classifier used to solve classification and regression problems. KNN works by calculating the distance between points on a graph (Fig. 4). The smaller the distance between two points, the more similar they are (belong to the same class or share the same label) [10].

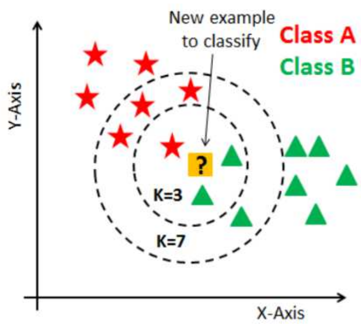

Fig. 4: K-Nearest Neighbor Architecture

The Decision Tree (DT) is a supervised learning classifier used to solve classification and regression problems. The decision tree algorithm works based on the decision on the conditions of the features. Nodes are the conditions or tests on an attribute (Fig. 5). Branches represent the outcome of the tests. Leaf nodes are the decisions based on the conditions. A DT simply asks a question, and based on the answer (Yes/No), it further splits the tree into subtrees [11].

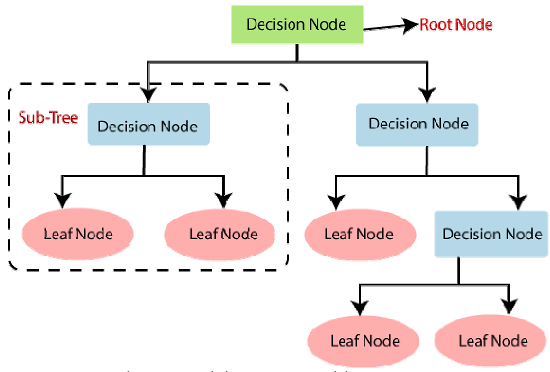

Fig. 5: Decision Tree Architecture

III. Potential Applications Of EEG

In this section, we discussed the potential applications of EEG. These applications include the diagnosis and prediction of AD, as well as the therapeutic follow-up.

A. Diagnosis and Prediction of Disease

EEG can help in several applications related to Alzheimer's disease.

1. EEG can be used to understand the physiopathology of the disease. Several studies have been done in this direction. Some of the most consistent results are the slowing of the EEG rhythms with the severity of the disease. Specifically, a decrease in the alpha and theta frequency bands was shown as an indication of a cognitive impairment related to AD.

2. EEG can be used to develop new biomarkers of AD. This can include diagnostic biomarkers (detecting AD), prognostic (predicting AD) and predictive of the therapy.

B. Therapeutic Follow-Up of Disease

EEG has potential therapeutic uses in AD research. For instance, by monitoring variations in brain activity before and after therapy, an EEG can determine how well pharmacological therapies work. Moreover, real-time feedback of an individual's brain activity is provided during neurofeedback training occasionally referred to as neurotherapy which is a technique that enables users to control their brainwaves voluntarily via EEG [15]. Through the neurofeedback, patients can learn to control their own brain activity, which may enhance their cognitive performance and lessen symptoms. This technique has shown promising results in improving memory, concentration and cognitive abilities [15]. Furthermore, sleep difficulties are frequent in dementia and can accelerate cognitive deterioration. EEG may be used to track sleep patterns and direct therapies to enhance the quality of sleep in AD.

IV. Limitations

In this section, we discussed some of the different challenges faced while implementing machine learning algorithms on EEG data to detect dementia.

Some of the limitations encountered include the need for larger dataset. Typically, larger datasets are essential to be implemented on machine learning algorithms because they allow the model to recognize patterns and correlations in the data. Hence, there’s an important need to obtain a large dataset for dementia detection with EEG. Yet, this might be challenging since it frequently entails recruiting volunteers from varied populations [16], including those with various stages of dementia. This process takes much time and resources.

Also, dementia is a progressive condition, and accurately detecting and monitoring requires recording the temporal changes in EEG signals over time through collecting the EEG data from the same subject consistently. However, this process also requires a lengthy period of time [16].

Preprocessing and feature extraction and selection stages are required when analyzing EEG data for dementia diagnosis. Creating an automatic pipeline for EEG analysis simplifies this workflow, nevertheless, it is such a difficult process that requires great expertise in this field [16]. Additionally, it has evolved into a global public health issue, so it's critical to have varied and representative datasets that include EEG data from several sites and countries. However, gathering such data necessitates cross-country cooperation between researchers, healthcare institutions which is a challenging procedure.

Last but not least, EEG signals contain artifacts like eye blinks, muscle movements, and others that contaminate the EEG data. While signal preprocessing is performed to get rid of noise and acquire a clean signal, it might not completely eliminate them [16].

V. Future Research Directions

In this section, future research suggestions are going to be proposed that address the limitations mentioned before and therefore, have a potential to yield better results in detecting AD.

Certain studies have emphasized that deep learning (DL) algorithms have outperformed traditional machine learning approaches in a variety of fields [15]. Some of the reasons may include that deep learning has the ability to automatically learn features from raw EEG data, which enables DL models to extract features automatically from signals or images without the need for human intervention. Because EEG data contains complex temporal and spatial patterns that may be difficult to detect, using this method can produce more accurate results for brain disease detection [15]. Moreover, convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are examples of deep learning models that can be useful for pattern recognition tasks [12]. CNNs are effective at finding spatial patterns, and RNNs are effective at finding temporal patterns of EEG data [12].

Besides, when compared to typical supervised machine learning approaches, unsupervised learning techniques may offer some advantages in identifying Alzheimer's disease using EEG data. For instance, Generative Adversarial Networks (GANs) can be utilized to diagnose Alzheimer's disease using EEG data [14]. GANs can generate artificial EEG signals that can closely match real EEG signals in the time and frequency domain. In this case, data augmentation would be conducted to avoid over-fitting [14]. Similarly, unsupervised learning algorithms, such as auto-encoders, can learn non-linear data, hence effectively capture the complex behavior with non-linear dynamic properties that the EEG patterns exhibit [13]. They can be also capable of identifying these abnormal patterns without the use of labeled AD samples.

VI. Conclusion and Perspectives

In brief, Alzheimer's disease is the most prevalent form of dementia [2], which is characterized by a progressive decline in memory, thinking, behavior, and social skills. The focus of this research was to briefly describe the potential of combining EEG with machine learning. The pipeline for EEG data pre-processing, extraction, selection and classification was described, highlighting the need of collecting high-quality data with appropriate signal cleaning approaches, choosing only the relevant features in the signals to finally accurately classify the disease. This can pave the way for the development of non-invasive and effective technologies for Alzheimer's disease early diagnosis and intervention [6].

References

[1] R. N. L. Lamptey, B. Chaulagain, R. Trivedi, A. Gothwal, B. Layek, and J. Singh, "A Review of the Common Neurodegenerative Disorders: Current Therapeutic Approaches and the Potential Role of Nanotherapeutics," Int J Mol Sci, vol. 23, no. 3, p. 1851, Feb. 2022.

[2] Mayo Clinic. "Alzheimer's disease - Symptoms and causes." Mayo Clinic.

[3] N. Chedid, J. Tabbal, A. Kabbara, S. Allouch, "The development of an automated machine learning pipeline for the detection of Alzheimer's Disease," Scientific Reports, vol. 12, pp. 18137, Oct. 28, 2022.

[4] N. Vivaldi, M. Caiola, K. Solarana, and M. Ye, "Evaluating Performance of EEG Data-Driven Machine Learning for Traumatic Brain Injury Classification," IEEE Transactions on Biomedical Engineering, vol. PP, pp. 1-1, Feb. 26, 2021.

[5] R. Akbari Movahed and M. Rezaeian, "Automatic Diagnosis of Mild Cognitive Impairment Based on Spectral, Functional Connectivity, and Nonlinear EEG-Based Features," Computational and Mathematical Methods in Medicine, vol. 2022, pp. 2014001, Aug. 11, 2022.

[6] D. Pirrone, E. Weitschek, P. Di Paolo, S. De Salvo, and M. C. De Cola, "EEG Signal Processing and Supervised Machine Learning to Early Diagnose Alzheimer's Disease," *Applied Sciences*, vol. 12, no. 11, p. 5413, May 2022,

[7] Z. M. Hira and D. F. Gillies, "A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data," Advances in Bioinformatics, vol. 2015, article ID 198363, 2015.

[8] X. Xie, J. W. Ho, C. Murphy, G. Kaiser, B. Xu, and T. Y. Chen, "Testing and Validating Machine Learning Classifiers by Metamorphic Testing," Journal of Systems and Software, vol. 84, no. 4, pp. 544-558, Apr. 1, 2011.

[9] I. Logunova, "Hyperparameter Tuning and Evaluation of ML Models," July 31, 2013.

[10] D. Bzdok, M. Krzywinski, and N. Altman, "Machine learning: Supervised methods, SVM and kNN," Nat. Methods, pp. 1–6, 2018.

[11] Y. Y. Song and Y. Lu, "Decision tree methods: applications for classification and prediction," Shanghai Arch Psychiatry, vol. 27, no. 2, pp. 130-135, Apr. 2015.

[12] I. Imran, "CNN-RNN Deep Learning Networks for Pattern Recognition Problems," in Proc. International Conference on Business Analytics for Technology and Security (ICBATS), Dubai, United Arab Emirates, pp. 1-5, 2023.

[13] R. Ferri, et al., "Stacked autoencoders as new models for an accurate Alzheimer's disease classification support using resting-state EEG and MRI measurements," Clin Neurophysiol, vol. 132, no. 1, pp. 232-245, Jan. 2021.

[14] T. Saleem, J. Tausifa, S. R. Zahra, F. Wu, A. Alwakeel, M. Alwakeel, F. Jeribi, and M. Hijji, "Deep Learning-Based Diagnosis of Alzheimer's Disease," Journal of Personalized Medicine, vol. 12, no. 5, p. 815, 2022.

[15] I. Vilou, A. Varka, D. Parisis, T. Afrantou, and P. Ioannidis, "EEG-Neurofeedback as a Potential Therapeutic Approach for Cognitive Deficits in Patients with Dementia," Life (Basel, Switzerland), vol. 13, no. 2, p. 365, 2023.

[16] R. Cassani, M. Estarellas, R. San-Martin, F. J. Fraga, and T. H. Falk, "Systematic Review on Resting-State EEG for Alzheimer's Disease Diagnosis and Progression Assessment," Disease Markers, vol. 2018