Journal of Alzheimer's Disease 95 (2023) 1667–1683
DOI 10.3233/JAD-230525
IOS Press

Inter-Hospital Advanced and Mild Alzheimer's Disease Classification Based on Electroencephalogram Measurements via Classical Machine Learning Algorithms

Alfonso Parreño Torres[a], Carlos Roncero-Parra[b], Alejandro L. Borja[a,*] and Jorge Mateo-Sotos[c]
[a]*School of Industrial Engineering, University of Castilla-La Mancha, Albacete, Spain*
[b]*School of Informatics, University of Castilla-La Mancha, Albacete, Spain*
[c]*Polytechnic School, University of Castilla-La Mancha, Cuenca, Spain*

Accepted 31 July 2023
Pre-press 14 September 2023

Abstract.
Background: In pursuit of diagnostic tools capable of targeting distinct stages of Alzheimer's disease (AD), this study explores the potential of electroencephalography (EEG) combined with machine learning (ML) algorithms to identify patients with mild or moderate AD (ADM) and advanced AD (ADA).
Objective: This study aims to assess the classification accuracy of six classical ML algorithms using a dataset of 668 patients from multiple hospitals.
Methods: The dataset comprised measurements obtained from 668 patients, distributed among control, ADM, and ADA groups, collected from five distinct hospitals between 2011 and 2022. For classification purposes, six classical ML algorithms were employed: support vector machine, Bayesian linear discriminant analysis, decision tree, Gaussian Naïve Bayes, K-nearest neighbor and random forest.
Results: The RF algorithm exhibited outstanding performance, achieving a remarkable balanced accuracy of 93.55% for ADA classification and 93.25% for ADM classification. The consistent reliability in distinguishing ADA and ADM patients underscores the potential of the EEG-based approach for AD diagnosis.
Conclusions: By leveraging a dataset sourced from multiple hospitals and encompassing a substantial patient cohort, coupled with the straightforwardness of the implemented models, it is feasible to attain notably robust results in AD classification.

Keywords: Alzheimer's disease, machine learning, EEG, feature extraction, ADM, ADA

INTRODUCTION

With the gradual aging of society, neurodegenerative disorders have attracted the attention of the scientific community. Dementia is considered to be the most frequent brain disorder, with more than 130 million people expected to be affected by 2050 [1, 2]. Alzheimer's disease (AD) accounts for approximately 70% of all diagnosed cases of dementia. Age is the main risk factor for AD, with memory loss and cognitive impairment being its most common symptoms [3]. Although a definitive diagnosis can only be made with a post mortem brain biopsy [4], the clinical diagnosis has traditionally been based on behavioral observations or neurocognitive tests such as the Mini-Mental State Examination (MMSE) [5]. However, neuropathological biomarkers for AD diagnosis are widely employed, and they are associated with a

*Correspondence to: Alejandro L. Borja, School of Industrial Engineering, University of Castilla-La Mancha, Albacete, Spain. Tel.: +34 926 052 884; E-mail: alejandro.lucas@uclm.es; ORCID: 0000-0003-2880-0678.

reduction of cerebrospinal fluid (CSF) amyloid-β1-42 (Aβ_{42}) and an increase in amyloid as revealed by positron emission tomography (PET) and hyperphosphorylated tau protein [6]. Another biomarker is the volumetric measurement of brain gray matter using magnetic resonance imaging (MRI). These techniques are limited, as general solutions for AD diagnosis, due to their cost or because they are either totally or partially invasive. In addition, these biomarkers do not directly reflect synaptic integrity or the state of neuronal signal transmitters, which are strongly related to cognitive processes [7].

Electroencephalography (EEG) is a technique that can address some of these issues, as it is a non-invasive, painless, cost-effective, can be found at any medical institution and shows the neuronal activity in real time [8]. Compared to MRI, EEG has a higher temporal resolution and an acceptable space resolution. EEG signals have demonstrated to be efficient in the diagnosis and evaluation of a wide range of neurological pathologies and brain disorders, including but not limited to epilepsy [9], depression [10], disorders of consciousness [11] and Parkinson's disease [12]. Usually for clinical diagnoses such as AD, EEG measurements are performed in the resting state (rsEEG) [13]. In the literature, a wealth of studies can be found that deal with the use of EEG signals as biomarkers [5, 7, 14, 15]. In general, patients with AD exhibit slow oscillations of brain activity, with a reduction in complexity in EEG signals as well as in synchrony [16]. Some of the studied biomarkers demonstrate a reduction in cortical connectivity and complexity across various frequency bands. Others are associated with a decrease in frontal delta/theta and posterior alpha rhythms, respectively, as well as a decrease in the amplitude of late positive potentials [17]. Feature extraction from EEG signals is the process of transforming raw data into numerical data that are smaller in size and retain important information. EEG feature extraction methods can be divided into two groups: univariate, in which measurements are taken from each channel separately, and multivariate, in which data are taken from two or more channels [18]. In addition, feature extraction can be implemented in the time domain, frequency domain or time-frequency domain [19].

The progression of AD can be categorized into three stages. The early stage, or mild cognitive impairment (MCI) precedes AD development and it implies only subtle symptoms in patients. However, individuals are often still able to live independently [20]. The next stage, which can be subdivided into mild and/or moderate AD (ADM), is characterized by a significant decline in cognitive function, as well as the loss of independence. Individuals may require more assistance with daily activities such as bathing and dressing. Finally, in the advanced AD (ADA) stage, patients have difficulty performing any type of task and they are typically completely dependent on others for their care.

In order to classify data obtained from EEG, machine learning (ML) techniques have emerged as a promising and powerful tool. These techniques offer considerable advantages for analyzing complex data, including pattern recognition, medical research or modeling, such as real-time implementation, finding non-linear relationships or fault tolerance [21]. ML algorithms can be classified into four categories: supervised, unsupervised, reinforcement, and deep learning [22]. Supervised learning algorithms are based on training the model using predefined data inputs and their corresponding output data. In unsupervised learning, the algorithm only uses data inputs, so the model is self-training. Reinforcement algorithms have an initial approach similar to unsupervised methods; however, if the model obtains correct outputs in the training process, it is rewarded. Finally, deep learning is based on simulating the human brain by emulating neural networks with the inclusion of a multilayer structure in which hidden layers are added between inputs and outputs [23, 24]. These algorithms use large amounts of data in the learning process.

Several studies have been published in recent years to identify AD with ML techniques. In [25], a review of ML methods for the identification of AD based on genetic data is presented. In some of the approaches, data obtained with MRI, PET, or CSF are included to enhance accuracy. This study presents different algorithms, such as Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), or Naïve Bayes (NB). The results achieved using ML with different and novel biomarkers are shown in [26], which concludes that the N-Methy-D-aspartate receptors (NMDAR) biomarker provides promising outcomes. The relationship between the amyloid-β biomarker and the measured EEG signals is studied in [16]. In [27], two features were extracted from EEG data, namely Bispectrum (BiS) and Continuous Wavelet Transform (CWT). In order to classify the signals, four different ML approaches were employed: Autoencoder (AE), Multilayer Perceptron (MP), LR, and SVM. The best performance was obtained with a multi-modal solution in which the two

feature types were combined and analyzed simultaneously, resulting in an accuracy rate of 96%. In order to identify signals of patients with and without AD from EEG recordings, a Takagi-Sugeno-Kang (TSK) fuzzy model approach based on latent factors with variational AE is proposed in [28]. While the results are very positive, the dataset is limited, consisting of only 40 patients. An alternative approach using a TSK classifier, wherein weighted and unweighted functional networks are considered, is shown in [29], obtaining an accuracy of 97.3% with a dataset of 60 patients. In [30], Artificial Neural Networks (ANNs) are used to classify patients with and without AD. For this purpose, they use signals obtained from EEG, MRI and both, obtaining an accuracy of 80%, 84% and 89%, respectively. An SVM classification algorithm using different feature selection techniques is proposed in [31], yielding an accuracy of 91%. In [32], the evaluation of the results achieved for AD classification using K-nearest neighbor (KNN) and two types of features, frequency-based and time-frequency-based, was presented. In addition, the influence of taking the measurements with eyes open and eyes closed is studied.

The majority of studies that have been conducted to classify patients with some stage of AD have been limited by the size of the databases used, which typically include around 50 patients. This is largely due to the difficulties associated with collecting and processing large amounts of data from patients with AD. However, in this paper, we present an analysis of the results achieved by various ML algorithms for the detection of two stages of AD, namely ADM and ADA, using a database of 668 patients. This database was collected from several hospitals located in five different cities, which provides a diverse and representative sample of patients with AD. By using a larger database, we were able to explore the performance of ML algorithms in a more realistic clinical setting. This allowed us to identify which algorithms are most effective at detecting early and late stages of AD. Overall, our results suggest that using a larger database can improve the accuracy and generalizability of ML models for AD detection, and could pave the way for more effective and personalized diagnosis and treatment of this devastating disease. Therefore, the key advantages of this paper can be summarized as follows:

- The large number of patients included and their inter-hospital nature, which contribute to more accurate and realistic results.
- The application of classical ML algorithms for ADM and ADA classification, that, when combined with a considerable dataset, leads to more robust conclusions than those obtained with limited data.

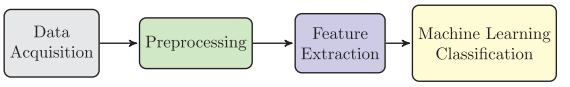

Fig. 1. Block diagram of the classification procedure.

The procedure followed to obtain the results is shown in Fig. 1 and it involves four steps: patient data acquisition, data processing, feature extraction, and finally, classification using ML algorithms.

The paper is organized as follows. The materials and equipment used to collect the data, as well as the source data from which the data were obtained, are shown in *Materials and Equipment* Section. *Methods* Section delivers a thorough overview of the ML techniques applied for classification, along with a description of the selected hyperparameters, the utilized feature extractions and the implemented validation process. In *Results* Section, the results achieved in the classification of patients are shown and analyzed. A comparison of the results with those obtained in other previous works is presented in *Discussion* Section. Finally, the main conclusions of the paper are detailed in *Conclusion* Section.

MATERIALS AND METHODS

EEG measurements

EEG signals were recorded using the brain vision system shown in Fig. 2 with sintered Ag/AgCl electrodes placed at a fixed distance of 20% and 10% between them, according to the configuration 10-20. The recording setup included 32 channels, and the signals were sampled at a rate of 500 Hz.

To minimize the effects of external interferences such as electrical noise, breathing or patient movements on the raw brain signal measurements, a two-step pre-processing approach was employed. The pre-processing consisted of a specific notch filter at 50 Hz to counteract powerline interference, complemented by a band pass filter with cutoff frequencies set at 0.5 Hz and 60 Hz, effectively isolating the frequency band of interest. The recorded EEG data were processed using Matlab on a computer

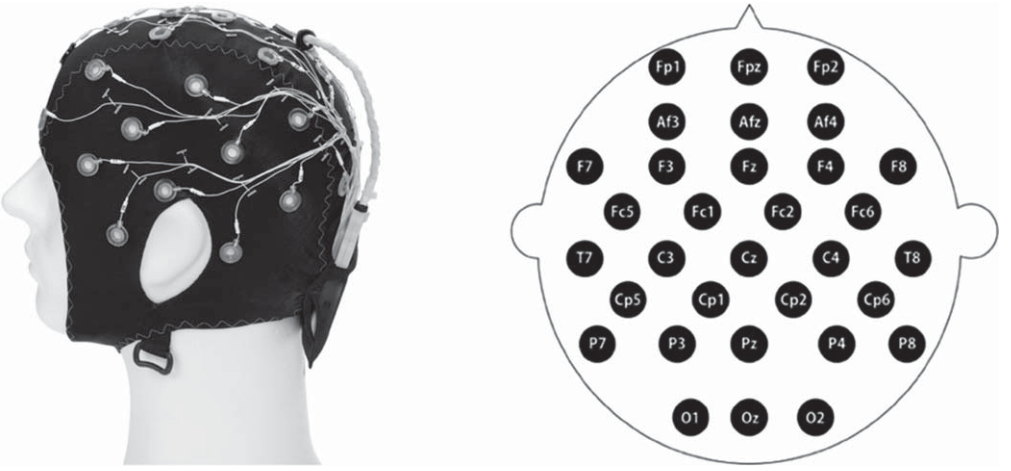

Fig. 2. Brain vision system and electrode positions [33].

with an Intel Xeon dual-core processor and 32 GB of RAM.

In order to obtain a proper extraction of EEG data, measurements were taken from patients who were instructed to keep their eyes closed and remain in a relaxed state, allowing their minds to freely roam with minimal cognitive demands. The acquisition of rsEEG recordings allows for the establishment of a baseline of cerebral activity, enabling a more precise characterization of spontaneous neuronal activity and intrinsic patterns. This approach enhances the detection of anomalies, as abnormal patterns tend to be more discernible in the absence of external stimuli. The measurements were consistently performed during the morning hours, specifically from 10:00 to 13:00. The recording time of the EEG was set up to 5 min. Also, the EEGs have been divided into rectangular windows of 10 s with an overlapping of 0.5 s. The optimal length of the window has been obtained by means of a parametric analysis between 2 and 20 s using Matlab software. Both manual and automated inspection of the measurements were conducted by the specialized physician in the respective field, ensuring the exclusion of EEG recordings with excessive noise.

Study population

This study presents EEG data collected between 2011 and 2022 from five hospitals located in different cities: Hospital Virgen de la Luz in Cuenca, Hospital General Universitario in Valencia, Hospital Rio Hortega in Valladolid, Hospital ClÃ-nico Universitario in Salamanca and Hospital Universitario Virgen de las Nieves in Granada. Consequently, the variety in the EGG data collection provides considerable relevance to the research and increases the reliability of the results analysis. The dataset includes EEG recordings from 668 subjects, which can be divided into three groups: 261 control subjects, 201 subjects diagnosed with ADM, and 206 subjects diagnosed with ADA. The patients were diagnosed with AD based on clinical criteria (MMSE and CDR scales) and the NINCDS/ADRDA criteria [34, 35].

The inclusion and exclusion criteria used in AD studies are as it follows. Inclusion Criteria: 1) Diagnosis of AD disease according to the established criteria of NIA-AA (National Institute on Aging-Alzheimer's Association) [34]. 2) Specific age range between 40 and 85, either to include only older adults or a broader age group. 3) Presence of cognitive symptoms and/or documented cognitive impairment through neuropsychological tests. 4) Capacity to provide informed consent and actively participate in the study. 5) Compliance with other specific study requirements, such as availability to attend scheduled follow-up visits or undergo additional tests. Exclusion Criteria: 1) Presence of other significant neurological or psychiatric diseases that may affect the evaluation or study outcomes. 2) Use of medications that may interfere with the results or are contraindicated for this particular study. 3) Presence of severe medical disorders or systemic diseases that may influence the evaluation or response to treatment. 4) History of severe traumatic brain injury or significant cere-

Table 1
Demographic characteristics of the dataset

Group	Subjects	Women (percentage)	Mean Age (y)	Mean Education (y)	Time from first symptoms (y)
Control	261	158 (60.5%)	70.4±12.4	8.5±5.3	–
ADM	201	123 (61.2%)	71.3±10.5	5.1±6.8	4.3±3.7
ADA	206	125 (60.6%)	72.6±11.3	4.9±6.2	5.6±3.5

brovascular diseases. 5) Excessive consumption of alcohol or drugs that may affect cognitive function.

After receiving an explanation of the study and its procedures, all the participants willingly gave their written informed consent. Additionally, the Clinical Research Ethics Committee of the Health Area of all hospitals granted approval for the study. The patients granted permission solely for the publication of the results, consequently restricting public availability. However, interested parties may request access to the datasets from the corresponding author, subject to reasonable inquiry. Table 1 displays the demographic characteristics of the dataset, including the number of subjects of each group, the % of females, the mean age (and its standard deviation), the level of education (and its standard deviation), and years from the first symptoms (and its standard deviation). It should be noted that the considerable size of the dataset offers several benefits compared to other studies in the literature that employed smaller samples. A large dataset reduces the variance in the training data, allowing the model to learn more efficiently and providing more accurate predictions. Furthermore, the algorithms can learn more diverse and representative patterns from the training data that are reflective of the real-world problem. Finally, overfitting is avoided during training, decreasing the likelihood of the model fitting solely to the training data without identifying the relevant features for classification.

Methods

This section provides a description of the six classical ML algorithms, including SVM, BLDA, DT, GNB, KNN, and RF. These algorithms are used for classifying patients into ADM, ADA, and control subjects. The main hyperparameters used for implementing the ML methods in MATLAB, as well as the feature extraction methods employed, are outlined. In addition, the section details the validation process and the different indicators used to evaluate the performance of the different algorithms.

ML algorithms

The first ML method used in the paper is SVM. It is a popular type of ML algorithm for classification, particularly useful when dealing with complex datasets that have non-linear boundaries. SVM separates data points by identifying the best hyperplane that maximizes the margin between the two classes, enabling effective classification of new data [36]. This makes it a versatile tool in many applications.

BLDA is a probabilistic ML algorithm based on Bayes' theorem, which entails computing the likelihood of a hypothesis given data [37]. The algorithm estimates the probability distribution of the features within each class and applies this information to classify new data. BLDA assumes that the data is normally distributed and that the covariance matrix of each class is shared. These assumptions render the algorithm valuable in analyzing high-dimensional datasets, where the number of features can exceed the number of samples.

The DT algorithm constructs a tree-like model to represent decisions and their potential consequences [38, 39]. It starts by selecting the optimal feature that maximizes the separation of the data into different classes or groups. It subsequently divides the data based on the chosen feature and repeats this process recursively until a specified stopping criterion is reached. The algorithm is able to capture non-linear relationships between the features and the target variable.

GNB is based on Bayes' theorem and assumes that the features are independent of each other, given the class label. It models the probability distribution of each feature in each class using a Gaussian distribution, which means that the data is assumed to be normally distributed. When a new data is presented, the algorithm calculates the probability of the data point belonging to each class and assigns it to the class with the highest probability. GNB is a simple algorithm that is computationally efficient and requires relatively few training data [40].

KNN algorithm works by finding the K closest data points in the training set to a new data point and using the majority vote of their class labels to predict the class of the new data point. KNN is a non-parametric algorithm, which means that it does not make any assumptions about the underlying data distribution. It can handle both numerical and categorical data and is robust to noisy data. However, it can be computationally expensive, especially when dealing with large datasets. Nevertheless, the performance of the algorithm can be sensitive to the distance metric used to calculate the similarity between data points. Despite these limitations, KNN remains a popular and effective ML algorithm [41].

RF is a popular ensemble ML algorithm used for classification [42]. It is composed of multiple decision trees, where each tree is trained on a random subset of the data and a random subset of the features. The trees are then combined to make predictions, either by voting for the most common class in the case of classification, or by averaging the output of the individual trees in the case of regression. RF is a powerful algorithm that is able to capture non-linear relationships between the features and the target variable, and it is also robust to noisy data and overfitting. Additionally, it can handle high-dimensional data and can provide feature importance rankings. RF has become a widely used and well-studied algorithm in the field of ML, and it has found successful applications in various fields.

Hyperparameters and algorithm characteristics

In order to implement the ML algorithms in the Matlab toolbox, it is required to tune various parameters such as iteration limits, Kernel functions, tolerances, maximum number of splits and/or number of neighbors or instances. By means of a Bayesian approach, in which different combinations of values are applied to each algorithm, the optimum hyperparameters are assigned. This approach leverages Bayes' theorem to calculate the posterior probability distribution of the hyperparameters given the observed data, $p(y|x)$ [43]. It can be expressed as:

$$p(y|x) = \frac{p(x|y) * p(y)}{p(x)} \tag{1}$$

where $p(x|y)$ is the likelihood function, which quantifies the probability of observing the data given specific hyperparameter values, $p(y)$ refers to the prior probability distribution, reflecting the initial beliefs or knowledge about the hyperparameters and $p(x)$ is the marginal probability of the data, ensuring that the posterior distribution is properly normalized. The aforementioned optimization has been carried out employing the commercial software Matlab by a trial-and-error procedure, maintaining the values obtained through the application of the Bayes' theorem, which result in the highest level of accuracy in classification. In this regard the main hyperparameters fixed for each algorithm are detailed in Table 2.

Table 2
Parameters for the machine learning algorithms

ML Method	Parameters
SVM	*Kernel function:* Gaussian *Sigma* = Gaussian $C = 1$ *Numerical tolerance* = 0.001 *Iteration limit* = 100
BLDA	*Kernel function:* Bayesian
DT	*Minimum instances in leaves* = 4 *Minimum instances in internal nodes* = 6 *Maximum depth* = 100
GNB	*Usekernel:* False $fL = 0$ *Adjust* = 0
KNN	*Number of neighbours* = 20 *Distance metric:* Euclidean *Weight:* Uniform
RF	*Number of trees* = 20 *Maximum number of considered features:* unlimited *Maximum tree depth:* unlimited *Stop splitting nodes with maximum instances* = 5

For the SVM and BLDA algorithms, a kernel function is employed, and the iteration limit for the SVM is set at 100. The DT algorithm uses a minimum of 4 instances in leaves and 6 instances in internal nodes. The GNB algorithm assumes a Gaussian distribution with a correction factor Laplace (fL) and an adjust of 0. For KNN, the number of neighbors is set at 20, using a Euclidean distance metric. Finally, the Random Forest algorithm is adjusted with a number of trees of 20, unlimited for the maximum number of considered features and tree depth and considering a limit of 5 maximum instances before the nodes stop splitting.

Feature extraction

Feature extraction is a fundamental process that involves identifying and selecting the most relevant attributes from EEG data to represent them more efficiently and effectively. This process contributes to data dimensionality reduction, enhanced model performance and increased interpretability of the data.

The extracted features are utilized to train classification models. In ML classification methods, the following features have been employed [44]: Detrended Fluctuation Analysis (DFA), Approximate Entropy (ApEn), Hurst exponent (HE), Higuchi fractal dimension (HFD), Lyapunov exponent (LE), and EEG band power. The techniques that exhibit lower computational load are ApEn and EEG band power. For the remaining methods, specific implementations were employed to alleviate their computational burden. Consequently, these features are calculated fast and allow for an interpretation, almost, in real-time, of the studied EEG. In our specific scenario, the analysis of each patient's data was efficiently computed within the order of seconds using the computer described in *EEG Measurements* Section.

Detrended fluctuation analysis

DFA is a feature extraction technique in signal processing that focuses on characterizing the presence of long-term correlations within data [45]. The DFA follows an iterative approach that comprises the following steps:

1. Calculation of the cumulative profile of the original time series, $x(k)$, where N represents its length:
$$y(k) = \sum_{i=1}^{k} \left[x(i) - \left(\frac{1}{N} \sum_{i=1}^{k} x(i) \right) \right] \quad (2)$$
2. Detrending the cumulative profile by fitting a linear function, $y_n(k)$, to remove the trend, where $\tilde{y}(k)$ is the detrended profile.
3. Division of $\tilde{y}(k)$ into non-overlapping segments of equal length, N, and calculation of the local fluctuations, $F(n, N)$ within each segment:
$$F(n, N) = \sqrt{\frac{1}{N} \sum_{k=1}^{N} [y(k) - \tilde{y}(k)]^2} \quad (3)$$
4. Calculation of $F(n, N)$ by averaging the local fluctuations across all segments for a given segment length. Then, plotting the logarithm of the segment against the logarithm of $F(n, N)$ to observe the scaling behavior. The slope of the linear regression corresponds to the fluctuation exponent α, which characterizes the presence of long-term correlations in the data.

DFA values around 0.5 suggests a random or uncorrelated series, while values greater than 0.5 indicate the presence of long-term correlations, making the series more predictable.

Approximate entropy

ApEn provides a measure of the irregularity or complexity of a time series, where higher values indicate greater complexity and lower ApEn values suggest more regularity [46]. The algorithm involves the following steps [47]:

1. Choose a pattern length, m, and a tolerance level, r. Then, embed the time series into a higher-dimensional space using a sliding window approach.
2. Calculation of the Euclidean distances, d_{ij}, between pairs of vectors, $X(i)$ and $X(j)$, within the tolerance level:
$$d_{ij} = \sqrt{\sum_{k=0}^{m-1} (X(i+k) - X(j+k))^2} \quad (4)$$
3. The probability that vectors remain close in the next embedding step is calculated by dividing the number of vector pairs within the specified tolerance by the total number of vector pairs.
4. Repeat the process for different pattern lengths and calculate the average probabilities. The $ApEn_m$ for a specific pattern length, m, is computed as:
$$ApEn_m = \ln \left(\frac{Average\ Probability(m)}{Average\ Probability(m+1)} \right) \quad (5)$$
5. Compare the results with a reference value obtained from a surrogate time series.

Hurst exponent

Hurst exponent quantifies the long-term memory and predictability of time series data, represented as $x(i)$, with, N, as the length of the series [48]. Initially, the deviation from the mean, $\bar{x}(n)$, is defined as the difference between each data point and the mean value, considering the first k data elements:

$$W_k = (x_1 + x_2 + \cdots + x_k) - k\bar{x}(n) \quad (6)$$

Next, the difference between the maximum and minimum values of the deviations, denoted as $R(n)$, and the standard deviation, $S(n)$, are calculated for each n. The ratio $\frac{R(n)}{S(n)}$ exhibits an increasing trend that can be described by a power law relationship, expressed

as:

$$\frac{R(n)}{S(n)} = C \times n^H \tag{7}$$

where C is a constant and the Hurst exponent can be calculated as:

$$H = \frac{\log\left(\frac{R(n)}{S(n)}\right)}{\log n} \tag{8}$$

The Hurst exponent value ranges between 0 and 1, wher 0.5 represents a random or uncorrelated time series. Values greater than 0.5 indicate a higher level of predictability and correlation [49].

Higuchi fractal dimension

HFD is a feature extraction technique used to quantify the complexity and self-similarity of time series data. It involves the successive steps [50, 51]:

1. Divide the time series into multiple subseries of equal length, x_k^m, where k and m are the length and the initial time value, respectively.
2. Connect the data elements within each subseries to form curves.
3. Approximate the length of each curve using the Higuchi method as:

$$L(m,k) = \frac{\left[\left(\sum_{i=1}^{H} |x[m+ik] - x[m+(i-1)k]|\right)\frac{N-1}{H}\right]}{k} \tag{9}$$

where H is a normalization factor that can be calculated as:

$$H = int\left(\frac{N-m}{k}\right) \tag{10}$$

4. Compute the average length of the curves obtained from all subseries, where K is the total number of subseries, as:

$$L(K) = \frac{1}{K}\sum_{m=1}^{K} L(m,k) \tag{11}$$

5. The HFD value is calculated by dividing the logarithm of $L(k)$ by the logarithm of k.

HFD value ranges between 1 and 2. A higher HFD value indicates a more complex and irregular time series, suggesting greater self-similarity and fractal properties.

Lyapunov exponent

The Lyapunov exponent is used to quantify the sensitivity to initial conditions and chaotic behavior in time series analysis. A positive LE indicates sensitive dependence on initial conditions and chaotic behavior, suggesting unpredictability and complexity. A negative LE indicates stability or convergence, while a zero suggests regular or periodic behavior [52]. The calculation procedure follows these steps:

1. Define an initial state or seed point and generate a nearby trajectory by perturbing the initial state.
2. Calculate the divergence between the trajectories of the perturbed and original states over time.
3. Estimate the LE, λ, as the average rate of divergence of nearby trajectories with the following expression:

$$\lambda = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\ln\frac{|\delta(n)|}{|\delta(0)|} \tag{12}$$

where $\delta(n)$ represents the difference between the perturbed trajectory, $x(n)$, and the original trajectory, $x(0)$.

EEG band power

EEG band power feature extraction quantifies the power distribution across different frequency bands in brain signals [53]. After applying a Fourier transform to convert the EEG signals from the time domain to the frequency domain, the frequency spectrum is divided into four bands such as delta (0.54 Hz), theta (48 Hz), alpha (813 Hz), and beta (1330 Hz). In order to obtain the power spectrum of each band, the Welch method was implemented [44]. The relative power for each band is calculated as:

$$\delta = \frac{1}{P}\sum_{f=0.5\ Hz}^{4\ Hz} p_f; \qquad \theta = \frac{1}{P}\sum_{f=4\ Hz}^{8\ Hz} p_f \tag{13}$$

$$\alpha = \frac{1}{P}\sum_{f=8\ Hz}^{13\ Hz} p_f; \qquad \beta = \frac{1}{P}\sum_{f=13\ Hz}^{30\ Hz} p_f \tag{14}$$

where P and p_f are the total power and the power spectrum, respectively.

Validation procedure

The validation of the proposed ML methods was carried out through a K-fold cross-validation process

to assess its predictive capability [54]. This validation approach was chosen due to its ability to improve the reliability, efficiency and generalization of model evaluation. K-fold cross-validation offers a more dependable estimation of model performance compared to a single train-test partition, as it mitigates the impact of a specific data split. Furthermore, K-fold validation is computationally less demanding compared to leave-one-out cross-validation (LOOCV), which requires creating a fold for each instance in the dataset. In this paper, a value of K equal to 10 was used, and the input dataset was divided into 70% for training and 30% for testing purposes. In order to prevent overfitting, the cross-validation analysis was conducted without sharing data between the training and validation groups.

Performance indicators

To evaluate the individual performance of each of the chosen ML methods, various metrics have been used to measure classification results. Standard indicators, considering TP as true positive, TN as true negative, FP as false positive, and FN as false negative, include sensitivity (also known as recall), specificity, precision (also known as positive predictive value), and negative predictive value. These parameters are defined as $TP/(TP+TN)$, $TN/(TN+FP)$, $TP/(TP+FP)$, and $TN/(TN+FN)$, respectively. The AUC, which is a typical value used to visualize how the ML classifiers performs, is also provided. In addition, the BA and F1 score [55], are calculated as:

$$BA = \frac{Sensitivity + Specificity}{2} \tag{15}$$

$$F1 = 2\frac{Precision \cdot Sensitivity}{Precision + Sensitivity} \tag{16}$$

In addition to the standard indicators and AUC, Matthew's correlation coefficient (MCC), degenerated Yonden's index (DYI) and Cohen's Kappa (CK) have been used to evaluate the performance of the chosen ML algorithms [56]. MCC measures the correlation between predicted and actual classifications, DYI is a modified version of the Jaccard index, where 1 indicates perfect agreement and 0 indicates complete disagreement and CK measures inter-rate agreement. These indicators can be calculated as follows:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{17}$$

$$DYI = \sqrt{Sensitivity \cdot Specificity} \tag{18}$$

$$CK = \frac{2(TP \cdot TN - FN \cdot FP)}{(TP+FP)(FP+TN)+(TP+FN)(FN+TN)} \tag{19}$$

RESULTS

The performance of the six ML classifiers is analyzed in this section. The classification task involves a 2-class problem, in which patients must be categorized into either ADM or ADA compared to the control subject group. Table 3 reports the ADM versus control group classification performance. In this scenario, the GNB algorithm gives the least favorable outcomes across the diverse metrics applied. Nonetheless, the remaining ML methods achieve high values considering the number of patients involved. The most remarkable performance is observed with the RF method, which outperforms its counterparts by a notable margin. The balanced accuracy stands at 93.25%.

On the other hand, Table 4 shows the performance indicator values obtained for the classification of ADA and control individuals. The outcomes are quite comparable to those in the previous case. The worst classification are obtained with GNB method and the most favorable results, with an accuracy of 93.55%, are obtained with the RF algorithm. As it can be observed from these results, very small differences between ADA and ADM patients compared to controls were found in the accuracy during the classification.

In order to provide a graphical representation of the performance indicators for each of the evaluated ML methods, two radar plots are generated. Figure 3a shows the radar plot for the classification of patients with ADM, with a zoomed-in view in the range of 60 to 100% to emphasize the differences between each method. As expected from the results shown in Table 3, the RF model achieved values closest to the maximum performance indicators. Similarly, Fig. 3b shows the radar plot for the classification of patients with ADA versus control subjects, where the RF method continues to yield the highest values. It should be noted that, for the RF algorithm, only the MCC and kappa indices achieved values below 90%.

Finally, the ROC curves obtained for each classification are presented. The ROC curve is a widely used

Table 3
Performance indicators for ADM versus control classification

ML Method	Balanced Acuraccy (%)	Recall (%)	Specificity (%)	Precision (%)	Negative Predictive Value (%)	AUC (%)	F1 (%)	MCC (%)	DYI (%)	Kappa (%)
SVM	89.61±0.84	89.72±0.79	89.5±0.74	88.97±0.9	88.76±0.81	0.89±0.02	89.34±0.7	79.51±0.82	89.61±0.73	79.78±0.75
BLDA	85.25±0.82	85.36±0.75	85.15±0.71	84.65±0.78	84.45±0.88	0.81±0.02	85.03±0.83	75.65±0.75	85.25±0.82	75.93±0.80
DT	88.16±0.76	88.27±0.72	88.06±0.69	87.53±0.65	87.33±0.61	0.88±0.02	87.91±0.66	78.23±0.69	88.16±0.72	78.49±0.71
GNB	77.51±0.93	77.61±0.92	77.42±0.88	76.96±0.85	76.78±0.89	0.77±0.02	77.28±0.94	68.78±0.88	77.51±0.89	69.01±0.91
KNN	90.36±0.67	90.47±0.66	90.25±0.71	89.71±0.65	89.54±0.64	0.90±0.01	90.09±0.70	80.18±0.64	90.36±0.65	80.44±0.68
RF	93.25±0.44	93.36±0.49	93.14±0.47	92.58±0.43	92.36±0.42	0.93±0.01	92.97±0.44	82.74±0.41	93.25±0.42	83.01±0.48

Table 4
Performance indicators for ADA versus control classification

ML Method	Balanced Acuraccy (%)	Recall (%)	Specificity (%)	Precision (%)	Negative Predictive Value (%)	AUC (%)	F1 (%)	MCC (%)	DYI (%)	Kappa (%)
SVM	89.26±0.76	89.37±0.82	89.16±0.79	88.62±0.74	88.42±0.86	0.89±0.02	88.99±0.88	79.2±0.84	89.26±0.86	79.46±0.84
BLDA	84.59±0.79	84.69±0.86	84.49±0.87	83.98±0.78	83.79±0.81	0.84±0.02	84.33±0.82	75.05±0.77	84.59±0.81	75.32±0.74
DT	88.52±0.66	88.63±0.71	88.42±0.76	87.89±0.67	87.68±0.67	0.88±0.02	88.26±0.64	78.55±0.65	88.52±0.69	78.81±0.71
GNB	78.38±0.89	78.48±0.83	78.29±0.94	77.82±0.88	77.64±0.95	0.78±0.02	78.15±0.84	69.55±0.91	78.38±0.87	69.78±0.86
KNN	90.78±0.65	90.89±0.68	90.67±0.61	90.13±0.66	89.92±0.69	0.9±0.01	90.51±0.66	80.55±0.67	90.78±0.66	80.82±0.68
RF	93.55±0.38	93.67±0.43	93.44±0.39	92.89±0.45	92.67±0.42	0.93±0.01	93.27±0.48	83.01±0.47	93.55±0.43	83.29±0.45

statistical tool that enables visualization of the performance of a binary classification algorithm. It displays sensitivity on the y-axis and the false positive rate (1-specificity) on the x-axis, allowing evaluation of the ability of the model to distinguish between groups and selection of the optimal decision threshold that maximizes the accuracy of the model. Values close to 1 in ROC curve represent a perfect predictive capability of the algorithm. The ROC curves for ADM and ADA classification are shown in Figs. 4 and 5, respectively. Once again, the RF algorithm achieves considerably higher results than the rest of the implemented ML methods.

DISCUSSION

The objective of this section is to compare the results obtained in this analysis with other studies. The realization of a comparison with other studies can be challenging and may lead to erroneous conclusions, primarily due to the lack of a standardized method for defining preprocessing, feature extraction and dataset size. For instance, the use of a specific feature extraction method from EEG signals can significantly improve classification accuracy, and/or using reduced dataset may result in overfitting during the training of classification algorithms, leading to high accuracy values. An option for artificially increasing the dataset size is through techniques of data augmentation. However, these methods cannot emulate the variability and realism of data collected from patients. Therefore, it should be noted that the results obtained in this work were achieved with a considerably larger dataset than the works referenced in this section.

The results obtained in this paper have been compared with various methods proposed in recent years. Table 5 shows this comparison, highlighting the outcomes achieved in the classification between control patients and those with ADM, as well as between control patients and those with ADA. Additionally, the sample size, the feature extraction technique and the algorithms used for classification have been included. It should be noted that the number of patients included in these studies is considerably smaller than that used in our work. In addition to the dataset size, the nature of the data, collecting from five different hospitals, enhances the representativeness of the data.

We will initially address the studies that aim to distinguish between control patients and those with ADM. The classification accuracy of the approaches proposed in [57, 58] is considerably lower. In [57], a validation procedure based on LOOCV is performed (similar to [58]) and data augmentation techniques are used to increase the training dataset. Furthermore, this study introduces a multiclass approach that involves classifying the two stages of patients with ADM, resulting in a decrease in the classification accuracy to 69.3%. Slightly higher accuracy or precision than those obtained in our work are achieved in [59, 60]; however, it should be noted that the number of patients in these works is reduced, consisting of 41 and 38 patients, respectively. A significantly different method based on graph theory is employed in [59]. In [60], excellent results are achieved by utilizing LOOCV validation, although implementing it for large datasets poses a considerable computational burden. Regarding the classification methods for ADA, the approaches proposed in [61–63] yield accuracies substantially lower than those achieved in our work. However, it should be emphasized that the dataset employed in [62] is appreciably larger compared to the other studies presented in Table 5, even though the number of patients is less than half of that utilized in this paper. A significantly lower accuracy of 72.2% is achieved in [61] when utilizing the features obtained through DFT instead of wavelets. In [62], a 5-fold cross-validation is employed, yielding relatively modest accuracy for the binary classification problems used. In [27], a 96.95% accuracy is achieved with a sample of 189 patients by utilizing a multi-layer perceptron with one hidden layer and extracting features using a multi-modal vector based on continuous wavelet transform and bispectrum. This accuracy is achieved in the two-class approach (ADM versus Control). However, when considering the multiclass approach (ADM versus MCI versus Control), this accuracy decreases to 89.22%. Finally, an accuracy of 90.3% with a dataset composed of 74 patients is presented in [64]. This study does not find any significant differences in classification between males and females. Its focus is on distinguishing between patients with ADA and MCI.

In addition to the methods included in Table 5, other previous studies have been found that employ ML techniques for AD diagnosis. In [65], seven feature extraction techniques and nine ML algorithms for the classification of AD are studied using EEG data. The accuracy results for classification are very high (around 99%), but the sample size is reduced: 51 patients for mild AD versus ADM classification and 86 patients for mild AD and ADM versus control classification. In order to distinguish between three

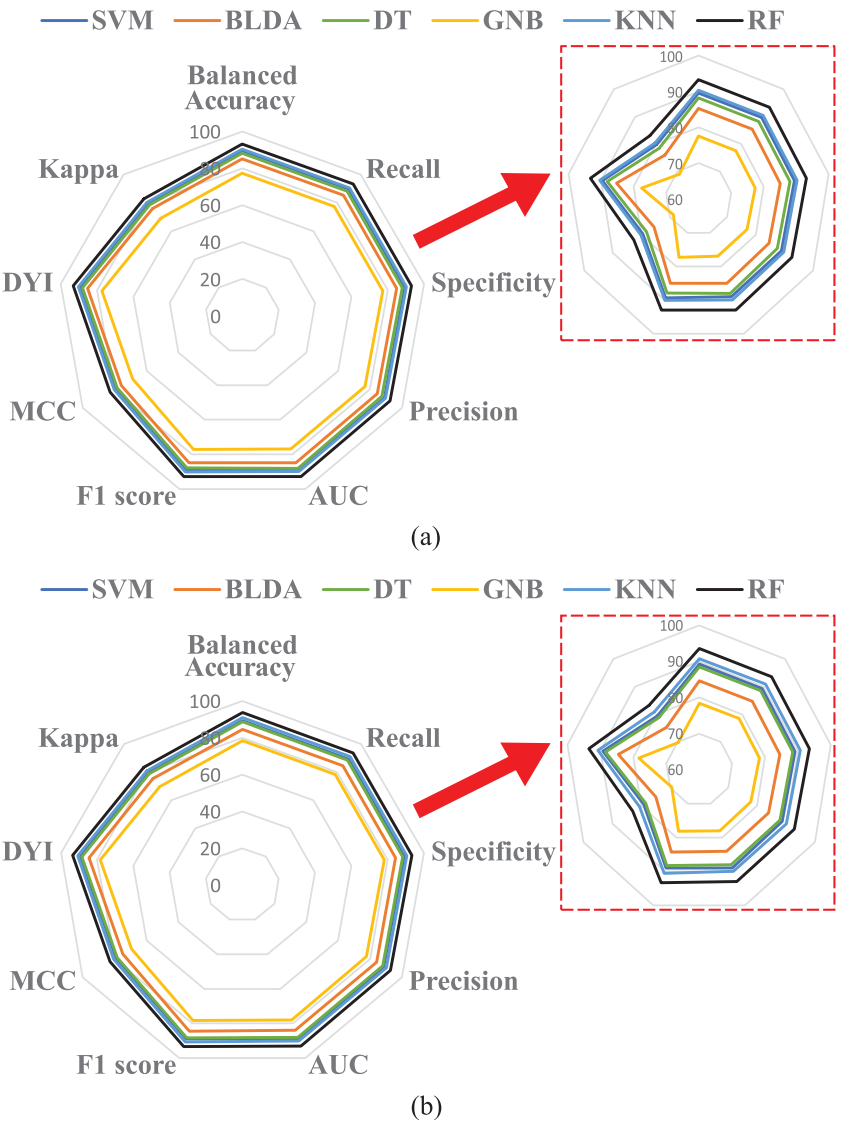

Fig. 3. Radarplot of (a) ADM, and (b) ADA patients classification for the six implemented ML methods.

categories of problems (MCI, control, and ADA), a Power Spectrum Density (PSD) feature extraction with a SVM classifier is proposed in [66]. The dataset consists of 161 patients, divided into the three groups to be classified, obtaining the best results for the MCI+ADA versus control subjects classification, achieving an 86.6% for the F1 score. This value is slightly lower than that obtained in our work. A dataset with 40 patients is classified with TSK fuzzy model obtaining a best accuracy of 98.1% when the Delta, Theta, Alpha and Beta frequency band energy features are employed as input. In [31], various feature extraction methods are analyzed with an SVM classifier for a group of 22 patients with an accuracy of 91.18%. Also, it is important to note that the proposed methods could be easily implemented in any system /computer /microcontroller that meets the minimum operational requirements. In our specific scenario, the analysis of each patient's data was efficiently computed within the order of seconds using the computer described in the paper. Additionally, an interface could be created to facilitate and enable the analysis of data and results.

In the context of Alzheimer classification employing electroencephalography, it can be found that come specific electrodes or frequency bands can contribute more significantly [67, 68]. Several electrode locations have shown promise in AD classification. For

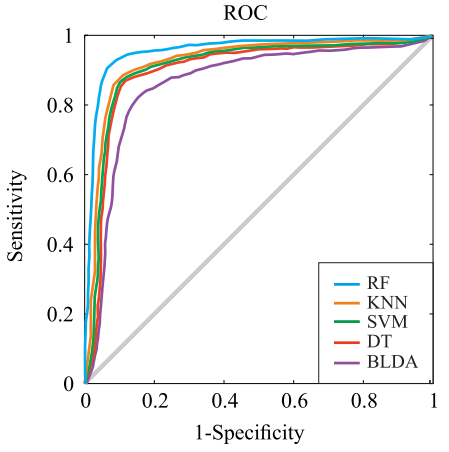

Fig. 4. ROC for ADM classification for ML methods.

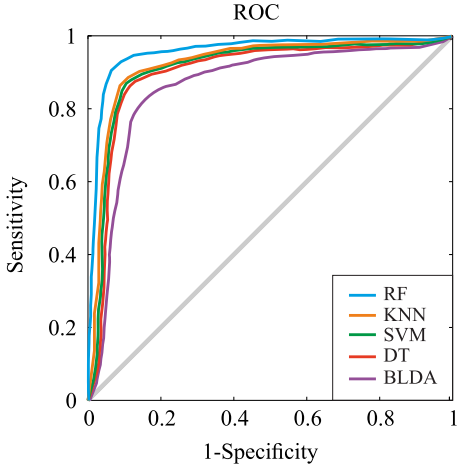

Fig. 5. ROC for ADA classification for ML methods.

example, the temporal regions (T3, T4, T5, T6), frontal regions (F3, F4, F7, F8), and parietal regions (P3, P4, Pz) are often investigated due to their relevance to cognitive function and pathology in AD. In addition, certain frequency bands in the EEG signal have been associated with cognitive processes affected by AD. The most commonly investigated frequency bands include increased delta power (particularly in posterior regions of the brain), alterations in theta activity (especially in the temporal and frontal regions), reduced alpha power (particularly in posterior regions), and changes in beta activity (particularly in frontal and parietal regions).

In addition age [69, 70], sex [71], and education [72, 73] can have some effect on the recorded electroencephalogram and classification results. Age is a significant risk factor for AD, and it can affect the EEG patterns associated with the disease. EEG signals of older individuals, especially those with AD, may exhibit increased theta and delta activity, reduced alpha power, and altered functional connectivity. Also, sex differences in AD have been observed, with higher prevalence in women. EEG studies have also reported sex-related differences in brain activity and connectivity. Including sex as a factor in classification models can help account for these differences and potentially improve the accuracy of classification. Finally, education level has been associated with cognitive reserve and can impact the manifestation and progression of AD. Individuals with higher education levels may exhibit different EEG patterns compared to those with lower education levels. Including education as a feature in classification models can help capture these differences and potentially improve the accuracy of classification. Thus, it is imperative for future studies to investigate the influence of these factors on the outcomes derived from the patient data. Furthermore, it remains crucial to maintain a keen focus on examining the impact and influence that distinct regions of brain activity could have in forecasting AD.

While ML approaches have shown promise in EEG classification for AD, there are several limitations to consider. These limitations include Labeling and Data Variability as obtaining accurate and reliable labels for EEG data in AD can be challenging. Also, Inter- and Intra-Subject Variability is an important issue as EEG signals can exhibit significant inter- and intra-subject variability due to individual differences, age, medications, and comorbidities. This variability can make it challenging to distinguish between AD and normal aging or other cognitive disorders solely based on EEG features. It requires careful consideration and additional validation to ensure robust and reliable classification. Finally, selecting relevant features from EEG signals is a critical step in ML models. However, identifying the most informative features that capture disease-related changes can be challenging. Feature extraction techniques may overlook subtle but important changes in the EEG signal, limiting the model's ability to capture the full spectrum of disease-related patterns.

Conclusion

In this paper, six well-known ML algorithms have been implemented for the detection of patients with AD through EEG data. The classification has been carried out for two stages in the AD devel-

Table 5
Comparison with previous works

Reference	C vs. ADM Accuracy	C vs. ADA Accuracy	Dataset Size	Validation	Feature Extraction	Problem Type	Classification
[27]	–	96.95 %	189	k-fold	Continuous Wavelet Transform and Bispectrum	2-class and 3-class	AE, MP, LR and SVM
[57]	79 %	–	54	LOOCV+ Bootstraping	Spectral power, coherence and amplitude-modulation features	2-class and 3-class	SVM
[58]	88 (Precision)	–	21	LOOCV	Relative power, Hjorth complexity and spectral entropy	2-class	SVM and LR
[59]	94 %	–	41	–	Graph Theory	2-class	Directed Transfer Function (DTF)
[60]	96 %	–	38	LOOCV	Multiband cepstral and lacstral analysis via DWT	2-class	ANN
[61]	–	83.3 %	109	LOOCV	Discrete Fourier Transform (DFT) and Wavelet	2-class	DT
[62]	–	72.43 %	301	–	Continuous Wavelet Transform and Band power ratio	2-class	LR, NB, SVM, RF and Random Undersampling (RUS) boosting
[63]	–	78.5 %	28	LOOCV and k-fold	Statistical and spectral	2-class	DT, RF, ANN, SVM, NB and KNN
[64]	–	90.3 %	74	–	MUSICEWT	2-class	EPNN
Our work	93.25 %	93.55 %	668	k-fold	EEG band power, HE, HFD, LE DFA and ApEn	2-class	SVM, BLDA, DT, GNB, KNN, and RF

opment: ADM and ADA. The database consists of 668 patients: 261 control subjects, 201 subjects diagnosed with ADM, and 206 subjects diagnosed with ADA. The data have been collected from five different hospitals located in five different cities that offers a more confident sample. The six implemented ML algorithms are SVM, BLDA, DT, GNB, KNN, and RF. The best results have been achieved for the RF algorithm for both ADM and ADA classification with accuracy values better than 93%.

In addition, the results obtained in this study show that it is possible to differentiate between patients with ADA or ADM and controls with high accuracy independently of the stage of the disease, i.e., intermediate or advanced. The classification outcomes for both disease states exhibit remarkable similarity. Finally, it should be noted that the algorithms studied in this paper could be integrated into medical equipment to aid healthcare professionals in providing better diagnoses for elderly patients with symptoms that are typical of ADA and ADM.

ACKNOWLEDGMENTS

The authors have no acknowledgments to report.

FUNDING

The authors have no funding to report.

CONFLICT OF INTEREST

The authors have no conflict of interest to report.

DATA AVAILABILITY

The data supporting the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

REFERENCES

[1] Grueso S, Viejo-Sobera R (2021) Machine learning methods for predicting progression from mild cognitive impairment to Alzheimer's disease dementia: A systematic review. *Alzheimers Res Ther* **13**, 162.

[2] Abbott A (2011) Dementia: A problem for our age. *Nature* **475**, S2-S4.

[3] Breijyeh Z, Karaman R (2020) Comprehensive review on Alzheimer's disease: causes and treatment. *Molecules* **25**, 5789.

[4] Paraskevaidi M, Martin-Hirsch PL, Martin FL (2018) Progress and challenges in the diagnosis of dementia: a critical review. *ACS Chem Neurosci* **9**, 446-461.

[5] Horvath A (2018) EEG and ERP biomarkers of Alzheimer's disease a critical review. *Front Biosci* **23**, 183-220.

[6] Jafari Z, Kolb BE, Mohajerani MH (2020) Neural oscillations and brain stimulation in Alzheimer's disease. *Prog Neurobiol* **194**, 101878.

[7] Babiloni C, Arakaki X, Azami H, Bennys K, Blinowska K, Bonanni L, Bujan A, Carrillo MC, Cichocki A, Frutos-Lucas J, Percio CD, Dubois B, Edelmayer R, Egan G, Epelbaum S, Escudero J, Evans A, Farina F, Fargo K, Fernández A, Ferri R, Frisoni G, Hampel H, Harrington MG, Jelic V, Jeong J, Jiang Y, Kaminski M, Kavcic V, Kilborn K, Kumar S, Lam A, Lim L, Lizio R, Lopez D, Lopez S, Lucey B, Maestú F, McGeown WJ, McKeith I, Moretti DV, Nobili F, Noce G, Olichney J, Onofrj M, Osorio R, Parra-Rodriguez M, Rajji T, Ritter P, Soricelli A, Stocchi F, Tarnanas I, Taylor JP, Teipel S, Tucci F, Valdes-Sosa M, Valdes-Sosa P, Weiergräber M, Yener G, Guntekin B (2021) Measures of resting state EEG rhythms for clinical trials in Alzheimer's disease: Recommendations of an expert panel. *Alzheimers Dement* **17**, 1528-1553.

[8] Klepl D, He F, Wu M, Marco MD, Blackburn DJ, Sarrigiannis PG (2022) Characterising Alzheimer's Disease With EEG-Based Energy Landscape Analysis. *IEEE J Biomed Health Inform* **26**, 992-1000.

[9] Subasi A, Kevric J, Canbaz MA (2017) Epileptic seizure detection using hybrid machine learning methods. *Neural Comput Appl* **31**, 317-325.

[10] de Aguiar Neto FS, Rosa JLG (2019) Depression biomarkers using non-invasive EEG: A review. *Neurosci Biobehav Rev* **105**, 83-93.

[11] Engemann DA, Raimondo F, King J-R, Rohaut B, Louppe G, Faugeras F, Annen J, Cassol H, Gosseries O, Fernandez-Slezak D, Laureys S, Naccache L, Dehaene S, Sitt JD (2018) Robust EEG-based cross-site and cross-protocol classification of states of consciousness. *Brain* **141**, 3179-3192.

[12] Waninger S, Berka C, Karic MS, Korszen S, Mozley PD, Henchcliffe C, Kang Y, Hesterman J, Mangoubi T, Verma A (2020) Neurophysiological biomarkers of Parkinson's disease. *J Parkinsons Dis* **10**, 471-480.

[13] Hampel H, Toschi N, Babiloni C, Baldacci F, Black KL, Bokde ALW, Bun RS, Cacciola F, Cavedo E, Chiesa PA, Colliot O, Coman C, Dubois B, Duggento A, Durrleman S, Ferretti M, George N, Genthon R, Habert MO, Herholz K, Koronyo Y, Koronyo-Hamaoui M, Lamari F, Langevin T, Lehéricy S, Lorenceau J, Neri C, Nisticò R, Nyasse-Messene F, Ritchie C, Rossi S, Santarnecchi E, Sporns O, Verdooner SR, Vergallo A, Villain N, Younesi E, Garaci F, Lista S (2018) Revolution of Alzheimer precision neurology. Passageway of systems biology and neurophysiology. *J Alzheimers Dis* **64**, S47-S105.

[14] Dauwels J, Vialatte F, Cichocki A (2010) Diagnosis of Alzheimer's disease from EEG signals: Where are we standing? *Curr Alzheimer Res* **7**, 487-505.

[15] Cassani R, Estarellas M, San-Martin R, Fraga FJ, Falk TH (2018) Systematic review on resting-state EEG for Alzheimer's disease diagnosis and progression assessment. *Dis Markers* **2018**, 5174815.

[16] Gaubert S, Raimondo F, Houot M, Corsi MC, Naccache L, Sitt JD, Hermann B, Oudiette D, Gagliardi G, Habert MO, Dubois B, Fallani FV, Bakardjian H, Epelbau S (2019) EEG evidence of compensatory mechanisms in preclinical Alzheimer's disease. *Brain* **142**, 2096-2112.

[17] Babiloni C, Blinowska K, Bonanni L, Cichocki A, Haan WD, Percio CD, Dubois B, Escudero J, Fernández A, Frisoni G, Guntekin B, Hajos M, Hampel H, Ifeachor E, Kilborn K, Kumar S, Johnsen K, Johannsson M, Jeong J, LeBeau F, Lizio R, da Silva FL, Maestú F, McGeown WJ, McKeith I, Moretti DV, Nobili F, Olichney J, Onofrj M, Palop JJ, Rowan M, Stocchi F, Struzik ZM, Tanila H, Teipel S, Taylor JP, Weiergräber M, Yener G, Young-Pearse T, Drinkenburg WH, Randall F (2020) What electrophysiology tells us about Alzheimer's disease: A window into the synchronization and connectivity of brain neurons. *Neurobiol Aging* **85**, 58-73.

[18] Ouchani M, Gharibzadeh S, Jamshidi M, Amini M (2021) A review of methods of diagnosis and complexity analysis of Alzheimer's disease using EEG signals. *Biomed Res Int* **2021**, 5425569.

[19] Shi Y, Ma Q, Feng C, Wang M, Wang H, Li B, Fang J, Ma S, Guo X, Li T (2022) Microstate feature fusion for distinguishing AD from MCI. *Health Inf Sci Syst* **10** 16.

[20] Siuly S, Alcin OF, Kabir E, Sengur A, Wang H, Zhang Y, Whittaker F (2020) A new framework for automatic detection of patients with mild cognitive impairment using resting-state EEG signals. *IEEE Trans Neural Syst Rehabil Eng* **28**, 1966-1976.

[21] Nichols JA, Chan HWH, Baker MAB (2018) Machine learning: Applications of artificial intelligence to imaging and diagnosis. *Biophys Rev* **11**, 111-118.

[22] Luján MÁ, Jimeno M, Mateo-Sotos J, Ricarte J, Borja AL (2021) A survey on EEG signal processing techniques and machine learning: Applications to the neurofeedback of autobiographical memory deficits in schizophrenia. *Electronics* **10**, 3037.

[23] Luján MÁ, Mateo-Sotos J, Aranda AT, Borja AL (2022) EEG based schizophrenia and bipolar disorder classification by means of deep learning methods. *J Biomed Eng Biosci* **9**, 1-5.

[24] Luján MA, Torres AM, Borja AL, Santos JL, Sotos JM (2022) High-precise bipolar disorder detection by using radial basis functions based neural network. *Electronics* **11**, 343.

[25] Rowe TW, Katzourou IK, Stevenson-Hoare JO, Bracher-Smith MR, Ivanov DK, Escott-Price V (2021) Machine learning for the life-time risk prediction of Alzheimer's disease: A systematic review. *Brain Commun* **3**, fcab246.

[26] Chang C-H, Lin C-H, Lane H-Y (2021) Machine learning and novel biomarkers for the diagnosis of Alzheimer's disease. *Int J Mol Sci* **22**, 2761.

[27] Ieracitano C, Mammone N, Hussain A, Morabito FC (2020) A novel multi-modal machine learning based approach for automatic classification of EEG recordings in dementia. *Neural Netw* **123**, 176-190.

[28] Li K, Wang J, Li S, Yu H, Zhu L, Liu J, Wu L (2021) Feature extraction and identification of Alzheimer's disease based on latent factor of multi-channel EEG. *IEEE Trans Neural Syst Rehabil Eng* **29**, 1557-1567

[29] Yu H, Lei X, Song Z, Liu C, Wang J (2020) Supervised network-based fuzzy learning of EEG signals for Alzheimer's disease identification. *IEEE Trans Fuzzy Syst* **28**, 60-71.

[30] Ferri R, Babiloni C, Karami V, Triggiani AI, Carducci F, Noce G, Lizio R, Pascarelli MT, Soricelli A, Amenta F, Bozzao A, Romano A, Giubilei F, Percio C D, Stocchi F, Frisoni GB, Nobili F, Patanè L, Arena P (2021) Stacked autoencoders as new models for an accurate Alzheimer's disease classification support using resting-state EEG and MRI measurements. *Clin Neurophysiol* **132**, 232-245.

[31] Trambaiolli LR, Spolaôr N, Lorena AC, Anghinah R, Sato JR (2017) Feature selection before EEG classification supports the diagnosis of Alzheimer's disease. *Clin Neurophysiol* **128**, 2058-2067.

[32] Durongbhan P, Zhao Y, Chen L, Zis P, Marco MD, Unwin ZC, Venneri A, He X, Li S, Zhao Y, Blackburn DJ, Sarrigiannis PG (2019) A dementia classification framework using frequency and time-frequency features based on EEG signals. *IEEE Trans Neural Syst Rehabil Eng* **27**, 826-835.

[33] Luján MÁ, Mateo-Sotos J, Santos JL, Borja AL (2023) Accurate neural network classification model for schizophrenia disease based on electroencephalogram data. *Int J Mach Learn Cybern* **14**, 861-872.

[34] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack Jr. CR, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, Mohs RC, Morris JC, Rossor MN, Scheltens P, Carrillo MC, Thies B, Weintraub S, Phelps CH (2011) The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 263-269.

[35] Dubois B, Feldman HH, Jacova C, DeKosky ST, Barberger-Gateau P, Cummings J, Delacourte A, Galasko D, Gauthier S, Jicha G, Meguro K, O'Brien J, Pasquier F, Robert P, Rossor M, Salloway S, Stern Y, Visser PJ, Scheltens P (2007) Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS-ADRDA criteria. *Lancet Neurol* **6**, 734-746.

[36] Ghosh S, Dasgupta A, Swetapadma A (2019) A study on support vector machine based linear and non-linear pattern classification. *2019 International Conference on Intelligent Sustainable Systems (ICISS)*, Palladam, India, pp. 24-28.

[37] Zhou W, Liu Y, Yuan Q, Li X (2013) Epileptic seizure detection using lacunarity and bayesian linear discriminant analysis in intracranial EEG. *IEEE Trans Biom Eng* **60**, 3375-3381.

[38] Swain PH, Hauska H (1977) The decision tree classifier: Design and potential. *IEEE Trans Geosci Electronics* **15**, 142-147.

[39] Kushwah JS, Kumar A, Patel S, Soni R, Gawande A, Gupta S (2022) Comparative study of regressor and classifier with decision tree using modern tools. *Mater Today Proc* **56**, 3571-3576.

[40] Jahromi AH, Taheri M (2017) A non-parametric mixture of Gaussian naive Bayes classifiers based on local independent features. *2017 Artificial Intelligence and Signal Processing Conference (AISP)*, Shiraz, Iran, pp. 209-212.

[41] Peterson LE (2009) K-nearest neighbor. *Scholarpedia* **4**, 1883.

[42] Breiman L (2001) Random forests. *Mach Lear* **45**, 5-32.

[43] Martinez-Cantin R (2019) Funneled bayesian optimization for design, tuning and control of autonomous systems. *IEEE Trans Cybern* **49**, 1489-1500.

[44] Mateo-Sotos J, Torres AM, Santos JL, Quevedo O, Basar C (2021) A machine learning-based method to identify bipolar disorder patients. *Circuits Syst Signal Process* **41**, 2244-2265.

[45] Jospin M, Caminal P, Jensen EW, Litvan H, Vallverdu M, Struys MMRF, Vereecke HEM, Kaplan DT (2007) Detrended fluctuation analysis of EEG as a measure of depth of anesthesia. *IEEE Trans Biom Eng* **54**, 840-846.

[46] Pincus SM, Huang W-M (1992) Approximate entropy: Statistical properties and applications. *Commun Stat Theory Methods* **21**, 3061-3077.
[47] Yentes JM, Hunt N, Schmid KK, Kaipust JP, McGrath D, Stergiou N (2012) The appropriate use of approximate entropy and sample entropy with short data sets. *Ann Biom Eng* **41**, 349-365.
[48] Nguyen-Ky T, Wen P, Li Y (2014) Monitoring the depth of anaesthesia using Hurst exponent and Bayesian methods. *IET Signal Process* **8**, 907-917.
[49] Lal GJ, Gopalakrishnan EA, Govind D (2019) Glottal activity detection from the speech signal using multifractal analysis. *Circuits Syst Signal Process* **39**, 2118-2150.
[50] Esteller R, Vachtsevanos G, Echauz J, Litt B (2001) A comparison of waveform fractal dimension algorithms. *IEEE Trans Circuits Syst I Fundam Theory Appl* **48**, 177-183.
[51] Hosseinifard B, Moradi MH, Rostami R (2013) Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal. *Comput Methods Programs Biomed* **109**, 339-345.
[52] Valenza G, Allegrini P, Lanatà A, Scilingo EP (2012) Dominant Lyapunov exponent and approximate entropy in heart rate variability during emotional visual elicitation. *Front Neuroeng* **5**, 3.
[53] Alkan A, Kiymik MK (2006) Comparison of AR and Welch methods in epileptic seizure detection. *J Med Syst* **30**, 413-419.
[54] Kelleher JD, Namee BM, D'Arcy A (2015) *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*, MIT Press.
[55] Hicks SA, Strümke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, Parasa S (2022) On evaluation metrics for medical applications of artificial intelligence. *Sci Rep* **12**, 5979.
[56] Rácz, Bajusz, Héberger (2019) Multi-level comparison of machine learning classifiers and their performance metrics. *Molecules* **24**, 2811.
[57] Cassani R, Falk T (2019) Alzheimer's disease diagnosis and severity level detection based on electroencephalography modulation spectral "patch" features. *IEEE J Biomed Health Inform* **24**, 1982-1993.
[58] Perez-Valero E, Morillas C, Lopez-Gordo MA, Carrera-Muñoz I, López-Alcalde S, Vílchez-Carrillo RM (2022) An automated approach for the detection of Alzheimer's disease from resting state electroencephalography. *Front Neuroinform* **16**, 924547.
[59] Afshari S, Jalili M (2017) Directed functional networks in Alzheimer's disease: Disruption of global and local connectivity measures. *IEEE J Biomed Health Inform* **21**, 949-955.
[60] Rodrigues PM, Bispo BC, Garrett C, Alves D, Teixeira JP, Freitas D (2021) Lacsogram: A new EEG tool to diagnose Alzheimer's disease. *IEEE J Biomed Health Inform* **25**, 3384-3395.
[61] Fiscon G, Weitschek E, Cialini A, Felici G, Bertolazzi P, Salvo SD, Bramanti A, Bramanti P, Cola MCD (2018) Combining EEG signal processing with supervised methods for Alzheimer's patients classification. *BMC Med Inform Decis Mak* **18**, 35.
[62] Ding Y, Chu Y, Liu M, Ling Z, Wang S, Li X, Li Y (2022) Fully automated discrimination of Alzheimer's disease using resting-state electroencephalography signals. *Quant Imaging Med Surg* **12**, 1063-1078.
[63] Miltiadous A, Tzimourta KD, Giannakeas N, Tsipouras MG, Afrantou T, Ioannidis P, Tzallas AT (2021) Alzheimer's disease and frontotemporal dementia: A robust classification method of EEG signals and a comparison of validation methods. *Diagnostics* **11**, 1437.
[64] Amezquita-Sanchez JP, Mammone N, Morabito FC, Marino S, Adeli H (2019) A novel methodology for automated differential diagnosis of mild cognitive impairment and the Alzheimer's disease using EEG signals. *J Neurosci Methods* **322**, 88-95.
[65] AlSharabi K, Salamah YB, Abdurraqeeb AM, Aljalal M, Alturki FA (2022) EEG signal processing for Alzheimer's disorders using discrete wavelet transform and machine learning approaches. *IEEE Access* **10**, 89781-89797.
[66] Sadegh-Zadeh S-A, Fakhri E, Bahrami M, Bagheri E, Khamsehashari R, Noroozian M, Hajiyavand AM (2023) An approach toward artificial intelligence Alzheimer's disease diagnosis using brain signals. *Diagnostics* **13**, 477.
[67] Vialatte FB, Dauwels J, Maurice M, Musha T, Cichocki A (2011) Improving the specificity of EEG for diagnosing Alzheimer's disease. *Int J Alzheimers Dis* **2011**, 259069.
[68] Jeong J (2004) EEG dynamics in patients with Alzheimer's disease. *Clin Neurophysiol* **115**, 1490-1505.
[69] Nebel RA, Aggarwal N, Barnes L, Gallagher A, Goldstein J, Kantarci K, Mallampalli M, Mormino E, Scott L, Yu W, Maki P, Mielke M (2018) Understanding the impact of sex and gender in Alzheimer's disease: A call to action. *Alzheimers Dement* **14**, 1171-1183.
[70] Rosende-Roca M, Abdelnour C, Esteban E, Tartari JP, Alarcon E, Martínez-Atienza J, González-Pérez A, Sáez ME, Lafuente A, Buendia M, Pancho A, Aguilera N, Ibarria M, Diego S, Jofresa S, Hernández I, López R, Gurruchaga MJ, Tárraga L, Valero S, Ruiz A, Marquié M, Boada M (2021) The role of sex and gender in the selection of Alzheimer patients for clinical trial pre-screening. *Alzheimers Res Ther* **13**, 1-13.
[71] Falahati F, Ferreira D, Soininen H, Mecocci P, Vellas B, Tsolaki M, Kłoszewska I, Lovestone S, Eriksdotter M, Wahlund LO, Simmons A, Westman E (2015) The effect of age correction on multivariate classification in Alzheimer's disease, with a focus on the characteristics of incorrectly and correctly classified subjects. *Brain Topogr* **29**, 296-307.
[72] Rosselli M, Uribe IV, Ahne E, Shihadeh L (2022) Culture, ethnicity, and level of education in Alzheimer's disease. *Neurotherapeutics* **19**, 26-54.
[73] Seyedsalehi A, Warrier V, Bethlehem RAI, Perry BI, Burgess S, Murray GK (2022) Educational attainment, structural brain reserve and Alzheimer's disease: A Mendelian randomization analysis. *Brain* **146**, 2059-2074.