

Emotion-Aware EEG Analysis for Alzheimer's Disease Detection Using Boosting and Deep Learning

Shynara Ayanbek¹, Abzal Issayev², Amandyk Kartbayev^{3*}

School of Information Technology and Engineering, Kazakh-British Technical University, Almaty, Kazakhstan^{1,2,3}

Faculty of Informatics, Masaryk University, Brno, Czech Republic¹

Yessenov Caspian University of Technology and Engineering, Aktau, Kazakhstan³

Abstract—Alzheimer's disease (AD) is a leading cause of dementia, yet its diagnosis remains challenging. EEG provides a noninvasive and cost-effective method for monitoring brain activity, which may reflect both cognitive decline and altered emotional states. In this study, an EEG-based pipeline was developed to classify AD using two approaches: an ensemble of boosting classifiers based on extracted features, and a deep convolutional neural network (CNN) applied to raw signals. A publicly available dataset was processed to extract time, frequency, and complexity features, with emotional brain dynamics implicitly reflected in the signals and considered during analysis. Five ensemble models (including CatBoost, LightGBM, and XGBoost) were optimized using Bayesian search. The CNN was trained separately and evaluated under cross-validation schemes. A balanced accuracy of 78.96% was achieved for AD detection using XGBoost, while the CNN reached 70.92% for Frontotemporal dementia. The study demonstrates that combining machine learning with EEG produces generalizable models for dementia detection and suggests that accounting for emotion-related variability may enhance diagnostic results.

Keywords—Alzheimer's disease; feature extraction; machine learning; CNN; boosting algorithms; deep learning

I. INTRODUCTION

Dementia has been recognized as a growing global health concern, with over 50 million individuals currently affected. This number is expected to increase to over 100 million by 2050. Among the various forms of dementia, Alzheimer's disease (AD) is the most common, representing around 60 to 80% of diagnosed cases. Frontotemporal dementia (FTD), although less prevalent, is one of the leading early-onset subtypes and is characterized by diverse clinical presentations [1]. The accurate and early differentiation between dementia subtypes remains a critical need, as effective clinical management, prognosis, and treatment decisions are heavily dependent on the correct diagnosis. However, the diagnostic process continues to pose difficulties, particularly due to the subjective nature of neuropsychological assessments and the reliance on advanced imaging techniques that may not be universally accessible.

Neuroimaging modalities such as magnetic resonance imaging (MRI) and positron emission tomography (PET) have been routinely used to support dementia diagnosis. PET scans, for instance, can reveal amyloid plaque accumulation and regional metabolic changes that are often associated with Alzheimer's pathology. Despite their utility, these imaging

approaches present limitations. They are often costly, involve limited access in certain clinical settings, and in the case of PET, expose patients to ionizing radiation. Furthermore, such methods provide only indirect and static assessments of brain function. The temporal resolution of MRI and PET is low, which restricts their ability to observe dynamic neural processes that may reflect cognitive and emotional states in real time. Because of these limitations, increasing attention has been given to alternative diagnostic tools that offer safe, affordable, and functionally informative assessments of brain activity [2].

Electroencephalography (EEG) has emerged as a promising technique in this context. EEG provides direct, high-temporal-resolution recordings of electrical brain activity and is widely available in clinical environments. Unlike imaging techniques, it captures fast-changing neural oscillations and is well suited for identifying abnormalities in functional connectivity and rhythm patterns associated with neurodegenerative disorders [3]. In dementia, characteristic changes have been consistently observed. These include a general slowing of brain rhythms, specifically, increased power in delta and theta bands and reduced power in alpha and beta bands. Additionally, lower signal complexity and decreased inter-regional coherence have been reported [4]. Such changes are often quantified using entropy, fractal dimension, and other non-linear measures. These alterations may reflect not only cognitive deterioration but also changes in emotional processing and brain state, which are often affected in dementia, particularly in FTD. As a result, EEG has been increasingly recognized for its potential to contribute to differential diagnosis and to detect subtle emotional or cognitive alterations that may not be visible through structural imaging [5].

The proposed methodology, as shown in Fig. 1, builds upon earlier research by incorporating more diverse feature sets, better validation practices, and state-of-the-art learning techniques. Previous EEG classification efforts have been reviewed extensively. Classical machine learning studies typically reported AD classification accuracies in the range of 75 to 85% when subject-aware validation was applied. For instance, Tzimourta *et al.* conducted a systematic review and found that traditional models based on handcrafted features often performed reasonably well under careful evaluation [6]. However, when improper validation was used, much higher but unreliable accuracies were observed. Goerttler *et al.* later demonstrated that incorporating a balanced set of spectral, spatial, and temporal features, along with grouped validation,

could achieve improved results. Their SVM model attained around 83.6% accuracy for AD classification [7].

Therefore, the present study seeks to answer the following research questions, aiming to advance dementia detection through signal dynamics modeling:

- How does enforcing subject-level separation through grouped cross-validation influence the perceived generalization ability of EEG-based dementia classifiers?
- Can modeling emotion-related variability in resting-state EEG signals improve the accuracy of machine learning approaches for early dementia detection?
- How do feature-based boosting methods and end-to-end CNNs differ in their ability to capture cognitive signatures associated with AD and FTD?

To address these limitations, a machine learning pipeline was developed, focusing on reproducibility and methodological rigor. A set of features was extracted, covering time-domain statistics, frequency-band characteristics, and signal complexity measures. These features were intended to capture various aspects of brain activity that are known to be affected in dementia and potentially linked to altered emotional states. By adopting a stratified grouped cross-validation framework, care was taken to ensure that all data from an individual subject was contained within a single fold. This was done to prevent data leakage and to simulate real-world diagnostic scenarios more closely.

In the machine learning stage, five boosting ensemble classifiers were trained: Extremely Randomized Trees, XGBoost, HistGradientBoosting, LightGBM, and CatBoost. These models were selected for their ability to handle structured data and for their success in many biomedical classification tasks. Each model was tuned using Bayesian optimization to identify the most effective combination of

parameters for the given data. Parallel to this feature-based approach, a CNN was also designed and trained directly on the raw signals. This allowed the model to learn discriminative patterns without the need for manual feature selection. This architecture was inspired by successful models used in other applications, such as psychiatric disorder classification.

The CNN was evaluated under both grouped and ungrouped validation schemes to assess the extent to which data leakage may affect deep learning performance. By comparing these evaluation strategies, it was possible to quantify the artificial performance gains introduced by improper splitting and to emphasize the importance of grouped evaluation. The inclusion of both ensemble and deep learning methods enabled a comprehensive comparison of approaches and demonstrated their complementary strengths.

As key contributions of this study, we present:

- A novel pipeline for dementia classification that combines both feature-based ensemble learning and deep learning directly from raw signals.
- One of the first systematic comparisons of a set of modern boosting algorithms and CNNs using grouped cross-validation to prevent data leakage.
- The feature engineering and incorporation of emotion-related variability as a potential factor influencing EEG signals and dementia detection performance.

The remainder of the study is organized as follows: Section II reviews related work. Section III presents the dataset, preprocessing steps, feature extraction pipeline, and the design of both the ensemble learning models and the convolutional neural network. Section IV details the experimental results and evaluation. Section V discusses the implications of the findings and the impact of validation strategies on reported performance. Finally, Section VI concludes the study, outlines its limitations, and suggests directions for future research.

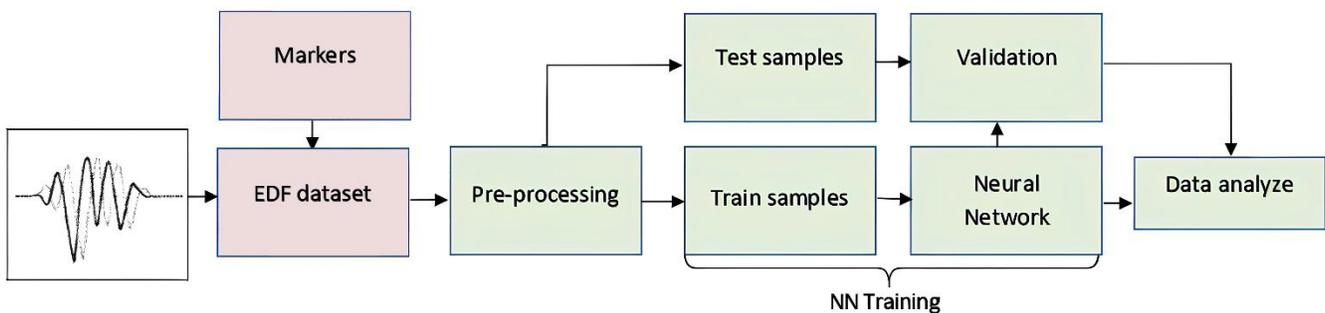


Fig. 1. Overview of the proposed methodology

II. RELATED WORKS

Computational methods have been developed to use EEG signals for automated dementia diagnosis, with early studies relying on classical machine learning techniques applied to manually extracted features. Among these, statistical and spectral descriptors such as band power and signal amplitude distributions were commonly used. Models such as Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), and Random Forests were frequently employed and achieved

moderate classification performance [8]. However, these works often exhibited certain limitations, with one of the most significant being improper model validation [9]–[11].

In many studies, EEG epochs (segments of continuous recordings) were randomly assigned to training and test sets using standard k-fold cross-validation, without accounting for subject identity [12]. This allowed data from the same individual to appear in both sets, inadvertently leaking subject-specific information and inflating model performance. In

contrast, when proper validation strategies, such as leave-one-subject-out or grouped k-fold cross-validation, were applied, performance often decreased considerably. This observation highlighted the need for more rigorous evaluation practices.

An example of this issue was demonstrated by Miltiadous *et al.*, where a Random Forest classifier trained on EEG data achieved nearly 99% accuracy for distinguishing AD from control subjects when evaluated using standard k-fold validation [13]. However, when a grouped validation method was used, accuracy dropped to approximately 78%, illustrating the significant impact of cross-validation strategy. Despite this, many existing studies continued to report only standard cross-validation results, often without an independent test set [14]. In addition to validation concerns, previous work frequently relied on a narrow range of features and default model parameters, leaving room for improvement through more thorough feature engineering and hyperparameter tuning. Modern ensemble learning methods such as gradient boosting algorithms and deep learning approaches have not been widely explored in this context, particularly for EEG-based dementia classification [15].

In one study, EEG signals were converted into spectrogram images and input into a CNN along with connectivity matrices, achieving high classification accuracy for three classes: AD, FTD, and healthy controls [16]. Recurrent neural networks, such as Long Short-Term Memory (LSTM) models, have also been applied, particularly for capturing the temporal dynamics of EEG signals [17]. While these models showed potential, their success was often limited by the size and quality of available datasets. Alessandrini *et al.*, for example, used an LSTM to classify multiple dementia types and achieved around 75.3% accuracy, suggesting that more data or improved architectures might be necessary [18].

Hybrid strategies have also been developed. Nour *et al.* proposed an ensemble of multiple CNNs, each trained on different input representations, and combined their outputs to improve robustness [19]. Jha *et al.* further enhanced classification performance by integrating clinical data with EEG features in a boosted ensemble model. In another approach, graph-based signal processing techniques were used to represent EEG data as networks, which were then analyzed using graph Fourier transforms and classified with support vector machines. Although this method achieved promising results in binary classification, its performance dropped significantly in multi-class settings, possibly due to increased complexity and noise [20]. Seo *et al.* [21] investigated emotion recognition in AD patients using EEG data, comparing multilayer perceptrons (MLP), SVM, and recurrent neural networks (RNN). Their findings suggested that classical machine learning methods, particularly MLP, could achieve promising accuracy, indicating the importance of affective state monitoring in dementia research. Extending beyond classical models, Gu *et al.* [22] provided a systematic review of deep learning applications in EEG-based brain-computer interfaces (BCI), highlighting the growing use of GANs and recurrent models for decoding complex emotional and cognitive patterns.

Recent studies have begun to bridge emotion processing and dementia. Dauwels *et al.* [23] demonstrated that EEG synchronization measures differ between AD and controls, while also noting that emotional tasks could amplify these distinctions. Meanwhile, Kumfor *et al.* [24] showed that emotional reactivity, as measured by EEG, diminishes progressively in dementia patients, suggesting an avenue for incorporating affective features into diagnostic models.

With respect to modeling approaches, Pillalamarri *et al.* [25] systematically evaluated CNN architectures for emotion recognition using EEG, demonstrating that even simple autoencoder models can outperform traditional classifiers when trained on raw signals. Boosting methods have also been explored: Chatterjee *et al.* [26] applied gradient boosting machines to EEG emotion datasets, reporting superior performance over Random Forest, especially when combining time-frequency features.

A promising hybrid approach was proposed by Iyer *et al.* [27], who integrated boosting models with CNNs in an ensemble framework for emotion-aware EEG classification. Their results indicated that blending handcrafted and learned representations could enhance generalization. Cope *et al.* [28] investigated the impact of emotional context on EEG dementia biomarkers, finding that incorporating emotional modulation improved the robustness of dementia detection models.

Across all these efforts, it has been consistently shown that EEG signals contain valuable information for detecting dementia. However, evaluation practices, feature diversity, and algorithm selection have a substantial impact on reported results [29]. Models validated using subject-independent methods tend to yield more modest but realistic accuracies in the range of 70 to 85%, while those using standard cross-validation often report inflated performance above 90%. These discrepancies highlight the need for careful methodological choices.

III. METHODOLOGY

A. Dataset

In this study, we utilize a publicly available EEG dementia dataset published by Miltiadous *et al.* [30], hosted on the OpenNeuro repository. The dataset comprises resting-state EEG recordings collected from 88 participants at a neurology clinic in Greece. These participants are categorized into three diagnostic groups: 36 individuals with probable Alzheimer's disease (AD), 23 with Frontotemporal dementia (FTD), and 29 cognitively normal elderly controls (CN).

All participants underwent comprehensive cognitive evaluation, including the Mini-Mental State Examination (MMSE), to assess the severity of cognitive impairment. The AD group had a mean MMSE score of 18 (standard deviation [SD] 4.5), indicating moderate impairment. The FTD group had a higher average MMSE score of 22.2 (SD 8.2), reflecting milder but variable impairment, while the control group had an average MMSE score of 30, indicating no cognitive decline. The age distribution across the three groups was comparable, with mean ages ranging between 66 and 67 years. However, there were differences in sex distribution: the AD group included a higher proportion of female participants, whereas

the FTD and control groups consisted mainly of males. This dataset provides a valuable resource for exploring EEG-based biomarkers in the differential diagnosis of dementia.

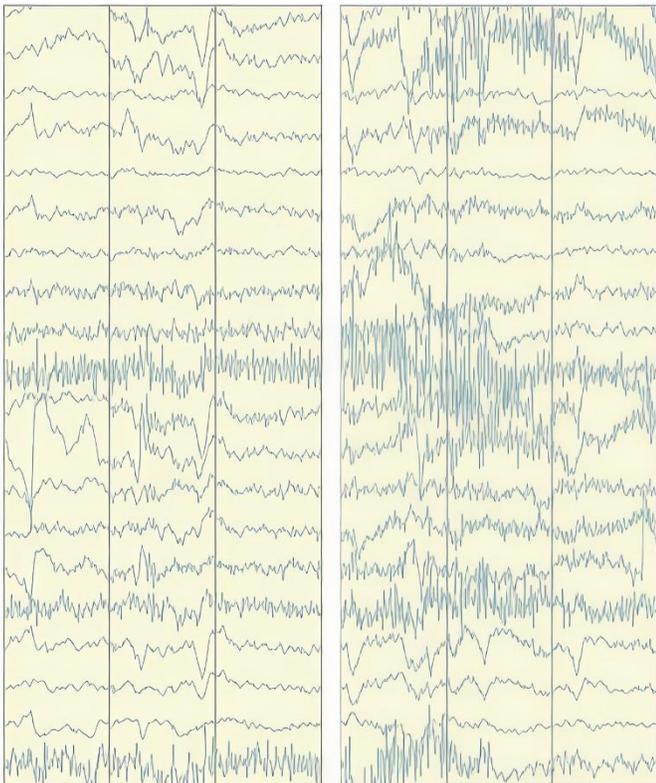


Fig. 2. Example of recorded signals in an awake resting state

EEG recordings were acquired from 19 scalp electrodes placed according to the international 10–20 system (channels: Fp1, Fp2, F7, F3, Fz, F4, F8, T3/T7, C3, Cz, C4, T4/T8, T5/P7, P3, Pz, P4, T6/P8, O1, O2). Two additional electrodes (A1, A2) served as reference leads during acquisition. Subjects were recorded in an awake resting state (eyes open, with minimal cognitive task) for several minutes, as shown in Fig. 2. The raw EEG signals were originally sampled at 500 Hz. As part of the dataset release, the authors provided data that had undergone some initial preprocessing: a band-pass filter from 0.5–45 Hz was applied (capturing the delta through low-gamma frequency range), and Artifact Subspace Reconstruction (ASR) was used to remove transient artifacts and high-amplitude noise burst. Furthermore, an Independent Component Analysis (ICA) with the ICLabel algorithm identified and removed components corresponding to eye-blink and muscle (jaw) artifacts. These steps attenuate common artifacts and yield cleaned multi-channel time series for each participant.

For our analysis, we carried out additional preprocessing to standardize the data and segment it for learning. First, we re-referenced each recording to the average of all 19 channels (common average reference montage). This step subtracts the mean signal across electrodes at each time point, which can reduce global noise and emphasize localized activity. Next, we epoched each continuous recording into non-overlapping segments of 10 to 12 seconds duration. Each epoch at 500 Hz contains 6000 time points per channel. We chose this window

because prior research indicated that longer epochs (10 to 12s) improve dementia classification performance compared to shorter windows. In particular, Tzimourta *et al.* [31] found that length segments yielded higher accuracy for EEG-based AD detection than 4 or 5-second segments. After epoching, each participant's EEG is represented as a set of 10 to 12 s epochs (the number of epochs per subject depends on recording length; on average around 20 epochs per subject).

We then assigned labels to each epoch. Two labels were created: a group label indicating the subject of origin (so that all epochs from the same person share a unique ID), and a class label indicating the diagnosis (CN, AD, or FTD) of that subject. By tagging each epoch with a subject-group identifier, grouped splitting could be enforced in later steps to prevent leakage of person-specific patterns between training and testing. At this stage, the dataset was structured as a 3D array with dimensions (epochs \times channels \times timepoints) per class. A total of 4404 epochs were collected for the CN versus AD task, and 3366 epochs for the CN versus FTD task, where each epoch represented a multivariate time series.

Notably, although subjects were recorded in a nominal resting state, the brain's spontaneous activity during this period reflects ongoing internal cognitive processing. These intrinsic states may influence EEG patterns and add variability that reflects real-world conditions. As visualized in Fig. 3, differences in dynamics across epochs may partly arise from emotional fluctuations during the recording, suggesting that patients' active emotional states, while unprompted, still modulate the electrophysiological signals used for classification.

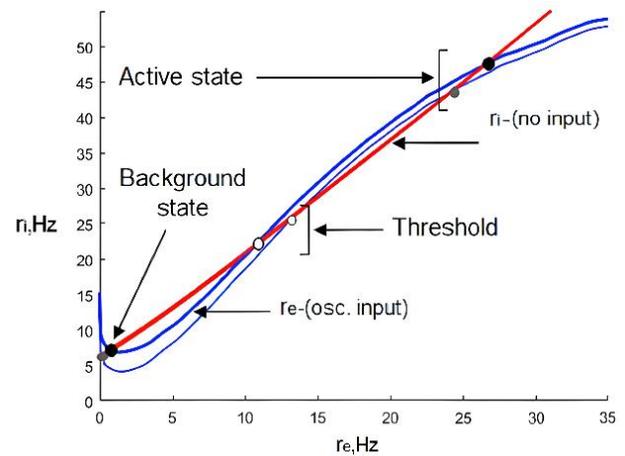


Fig. 3. Example of emotional fluctuations during the recording

A commonly used method for assessing emotional states in EEG data is event-related synchronization (ERS) or event-related desynchronization (ERD) within specific frequency bands in response to emotional stimuli. ERS and ERD are defined as increase or decrease, respectively, in the relative power of a particular rhythm, typically within the alpha, beta, or theta ranges [32]. These changes are considered indicative of underlying neural activation or inhibition in response to affective processing.

The strength of phase synchronization between frontal and right temporo-occipital electrodes has been shown to vary in relation to emotional arousal and tension. This modulation of synchronization reflects the brain’s dynamic adaptation to emotional states and can be captured using EEG-based measures. In the context of machine learning, the core objective becomes modeling the relationship between these patterns and emotional responses. This is achieved by training the model on labeled data, allowing it to learn statistical dependencies between input signals and affective outcomes.

Principal components corresponding to the largest eigenvalues are typically used to extract features that capture the strongest correlations between EEG activity and emotional state. These components are particularly valuable in identifying the neural substrates of emotion-related responses. In this study, such EEG-derived features, including power modulations and connectivity patterns, were considered during classification, as they may contribute to the distinction between cognitive decline and emotion-linked neural signatures. Relevant aspects of these dynamics are illustrated in Fig. 4.

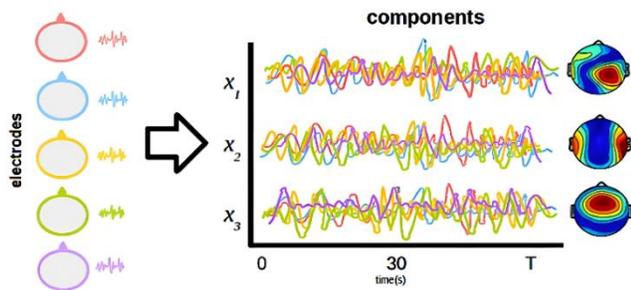


Fig. 4. Processing of emotional fluctuations based on emotion-linked neural signatures

Before feature extraction and modeling, we applied standard normalization. For the deep learning pipeline, each epoch’s time series were z-score normalized (channel-wise) to have zero mean and unit variance before feeding into the CNN, which helps to stabilize training. For the feature-based pipeline, we similarly standardized features (after extraction) by z-scoring each feature variable across the training set.

B. Feature Extraction

To enable the use of ensemble classifiers, each EEG epoch was transformed into a structured set of quantitative features. The feature engineering process was designed to capture relevant characteristics across three primary domains: time-domain statistics, frequency-domain power spectra, and non-linear measures of signal complexity. In order to reduce noise and dimensionality, the 19 original EEG channels were grouped into five anatomically informed regions of interest (ROIs), following common practices in prior studies.

These regions were categorized as frontal, temporal, central, parietal, and occipital. Within each ROI, signals from the respective electrodes were averaged to create a single representative time series. This regional averaging reduced the dimensionality from 19 to 5 signals per epoch, while also potentially improving signal-to-noise ratio by minimizing random fluctuations present in individual channels. Following this transformation, each EEG epoch was represented as a time series with dimensions 5×6000 , corresponding to 5 ROIs sampled over a 12-second window. From the ROI-aggregated signals, a total of 467 features were extracted for each epoch, as shown in Table I and Fig. 5. These features were identical in structure for classification tasks. The extracted features were grouped into three main categories:

1) Time-domain statistical features (15 per ROI) were computed to describe the amplitude and distribution characteristics of each signal. These included basic statistical measures such as mean, median, standard deviation, variance-to-mean ratio, minimum, maximum, and peak-to-peak amplitude. Additional descriptors included the interquartile range (IQR), root mean square (RMS) amplitude, and the sum of absolute differences between successive samples. Higher-order moments such as skewness and kurtosis were calculated to capture asymmetry and tail behavior in the signal distributions. Furthermore, three forms of mean absolute deviation (MAD1, MAD2, MAD3) were computed to quantify variability around central tendencies. These features have been commonly used in EEG classification tasks and are known to reflect relevant temporal dynamics in neural activity.

TABLE I. STRUCTURE OF THE FEATURE VECTOR PER EPOCH

Feature Category	Description	Features per ROI	ROIs	Total Features
Time-Domain Features	Mean, Median, Std Dev, Variance-to-Mean Ratio, Min, Max			
	Peak-to-Peak, IQR, RMS, Sum of Abs. Differences			
	Skewness, Kurtosis, MAD1, MAD2, MAD3	15	5	75
Frequency-Domain Features	Relative Band Power (Delta, Theta, Alpha, Beta, Gamma)	5	5	25
	Band Power Ratios (all unique pairs across RBPs)	—	—	300
Complexity Features	Approximate Entropy, Sample Entropy, Permutation Entropy			
	Spectral Entropy, SVD Entropy, DFA Exponent, Zero Crossings			
	Lempel-Ziv Complexity, Higuchi, Katz, Petrosian FD			
	Hjorth Mobility, Hjorth Complexity	15	5	75
Demographics	Age, Gender (One-hot: 0=Male, 1=Female)	—	—	2
Total				467

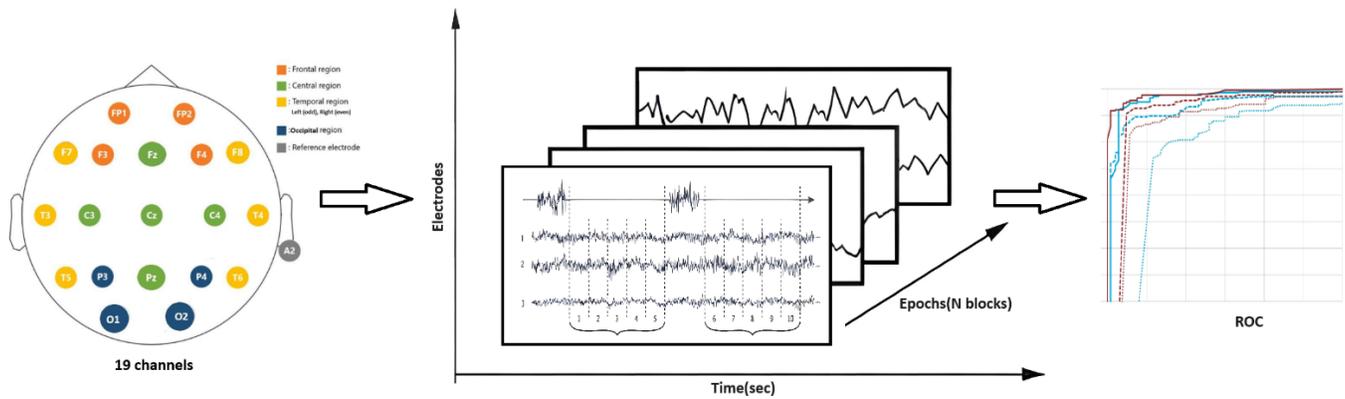


Fig. 5. The signal processing schema for feature extraction from each epoch

2) *Frequency-domain features* were derived using spectral decomposition tools, particularly the YASA toolbox. For each ROI, the relative band power (RBP) was computed for five canonical frequency bands: delta (0.5 to 4 Hz), theta (4 to 8 Hz), alpha (8 to 12 Hz), beta (12 to 30 Hz), and gamma (30 to 45 Hz). These power values were normalized as fractions of total power across the full frequency range. This procedure yielded 25 RBP features per epoch (5 bands across 5 ROIs). Additionally, power ratio features were calculated by forming all unique pairwise ratios between the 25 band-power features, resulting in 300 additional frequency features.

These ratios allowed the model to capture relative changes between frequency bands and brain regions, which are particularly relevant in conditions such as Alzheimer’s disease, where increased theta and delta power relative to alpha and beta have been frequently observed. In total, 325 frequency-domain features were generated per epoch.

3) *Complexity features* (15 per ROI) were calculated to quantify the regularity and unpredictability of the EEG signals. These were derived using the AnTroPy and EEGLib libraries and included entropy-based metrics (approximate, sample, permutation, spectral, and SVD entropy), fractal dimensions (Higuchi’s, Katz’s, Petrosian’s), Hjorth parameters (mobility and complexity), Lempel-Ziv complexity, zero-crossing counts, and the detrended fluctuation analysis (DFA) exponent [33]. These measures provided insight into the non-linear, dynamical structure of EEG signals and were especially relevant in dementia, where reductions in signal complexity and entropy are typically observed. Across 5 ROIs, a total of 75 complexity features were extracted per epoch.

To account for demographic influences, two additional features, age and sex, were appended to each feature vector. These variables are recognized as risk factors for dementia, with age being a primary determinant of AD prevalence and possible sex-related differences observed in EEG patterns [34]. Gender was encoded as a binary feature (0 for male, 1 for female). The MMSE scores, despite their strong correlation with dementia severity, were deliberately excluded to prevent target leakage. Inclusion of MMSE would risk artificially

boosting classification accuracy due to its near-linear relationship with the diagnosis.

C. Ensemble Learning Models

Five ensemble machine learning classifiers were trained on the extracted EEG features: Extra Trees (Extremely Randomized Trees), XGBoost (Extreme Gradient Boosting), HistGradientBoosting (Histogram-based Gradient Boosting), LightGBM (Light Gradient Boosting Machine), and CatBoost (Categorical Boosting). These models were selected due to their proven performance in handling structured data and their capacity to process high-dimensional feature sets effectively. As each model included several hyperparameters—such as the number of estimators, tree depth, and learning rate—model tuning was required to ensure optimal performance.

To minimize overfitting and enable fair model selection, stratified grouped 5-fold cross-validation (CV) was applied. Initially, the full dataset was divided into a training set and an independent hold-out test set. This partition was made by assigning 20% of subjects from each class to the test set, with the remaining 80% allocated to training. Within the 80% training set, model selection and hyperparameter tuning were conducted using stratified grouped 5-fold CV.

Subjects in the training set were split into five folds, each containing entire subject recordings while preserving class balance. For each fold, a model was trained on four folds and validated on the remaining one. This procedure was repeated across all five folds, with performance metrics averaged to assess model quality, preventing leakage of subject-specific patterns between training and validation subsets. Given that false positives and false negatives both have clinical consequences in dementia diagnosis, balanced accuracy was considered more suitable than metrics such as F1-score.

Hyperparameter optimization was performed for each classifier using Bayesian search with the HyperOpt library [35]. Unlike brute-force grid search, Bayesian optimization evaluates a sequence of hyperparameter combinations, using past results to choose the next combination in an informed manner. Reasonable search ranges for key parameters were defined based on prior research and exploratory trials. For example, for the tree-based models we allowed up to 1000+ trees and depths up to 30, and for learning rates we searched on a log-scale from 0.001 to 0.1.

HyperOpt's Tree-structured Parzen Estimator algorithm then suggested new hyperparameter sets likely to improve performance [36]. The search process was limited to 50 iterations per model, and the optimal configurations were selected based on mean balanced accuracy. Final models were retrained on the full training set and evaluated on the held-out test set for the tasks.

D. CNN Model

In parallel with the feature-based ensemble approach, a deep CNN was developed to classify EEG signals directly from raw time-series input, as shown in Fig. 6. This model operated without the need for handcrafted feature extraction, instead learning temporal representations from the multi-channel EEG data. A compact 1D convolutional architecture was adopted, designed to capture meaningful patterns while limiting the number of parameters to reduce the risk of overfitting. The structure was inspired by prior works in EEG-based deep learning, with modifications introduced to accommodate the scale and characteristics of the dataset used in this study.

Each input to the CNN was a single EEG epoch with dimensions 19×6000 , corresponding to 19 channels and 6000 time points (12 seconds at 500 Hz). The architecture was organized into three functional stages: feature extraction, dimensionality reduction, and classification. In the first stage, a sequence of one-dimensional convolutional layers was applied along the time axis of each channel. The initial convolutional layer consisted of five filters with a kernel size of 3 and a stride of 1, allowing the model to learn local temporal features on short (~6 ms) segments of the signal. Each convolutional layer was followed by a LeakyReLU activation to introduce non-linearity, batch normalization to stabilize training, and a pooling layer to downsample the temporal dimension.

across four convolutional blocks. Both max-pooling and average-pooling were used in alternating layers, enabling the network to extract both peak-oriented and trend-based features. As the signal passed through successive blocks, the temporal dimension was reduced by a factor of 16, while the depth of feature maps increased. By the final convolutional layer, abstract features representing temporal dynamics in the EEG were extracted.

Dimensionality reduction was then performed using a global average pooling layer. Rather than flattening the entire output volume into a long vector, the global average pooling layer computed the mean value across each feature map's time dimension, summarizing the temporal activity into a compact feature vector. This served as a bottleneck layer and reduced the number of parameters, improving generalization and training efficiency.

In the classification stage, a dense output layer with a single neuron was employed, followed by a sigmoid activation function to produce a probability score between 0 and 1. This score represented the model's estimated probability that the input epoch belonged to the positive class. Model training was conducted using the Adam optimizer, with a learning rate of 0.001. The loss function used was binary cross-entropy, which was appropriate for the binary classification setting. To fairly evaluate performance and examine the effects of data leakage, two cross-validation strategies were applied.

1) *In the first strategy*, stratified ungrouped 15-fold cross-validation was used. Each epoch was treated as an independent sample, and folds were created by randomly assigning epochs while maintaining class balance. This method did not ensure separation by subject, allowing data from the same individual to appear in both training and validation sets. As a result, it served to illustrate the inflated performance that can result from improper validation practices.

2) *In the second strategy*, stratified grouped 15-fold cross-validation was employed. In this approach, folds were constructed at the subject level, with all epochs from each subject assigned to a single fold. The CNN was trained on data from 14 groups of subjects and validated on the remaining group. This ensured subject-wise separation between training and validation and matched the evaluation protocol used for the ensemble models. The choice of 15 folds (instead of five) was made to increase the amount of training data per fold and to stabilize performance estimates, especially given the data demands of deep learning models.

Early stopping was implemented during training to reduce overfitting. If the validation loss failed to improve for three consecutive epochs, training was halted. A maximum of 10 epochs per fold was permitted, although early stopping typically occurred between epochs 5 and 8. Training was performed using mini-batches of size 28 per GPU (effectively 56 per step when two GPUs were used). These parameters (learning rate, batch size, and early stopping patience) were selected based on empirical testing to ensure convergence without overfitting.

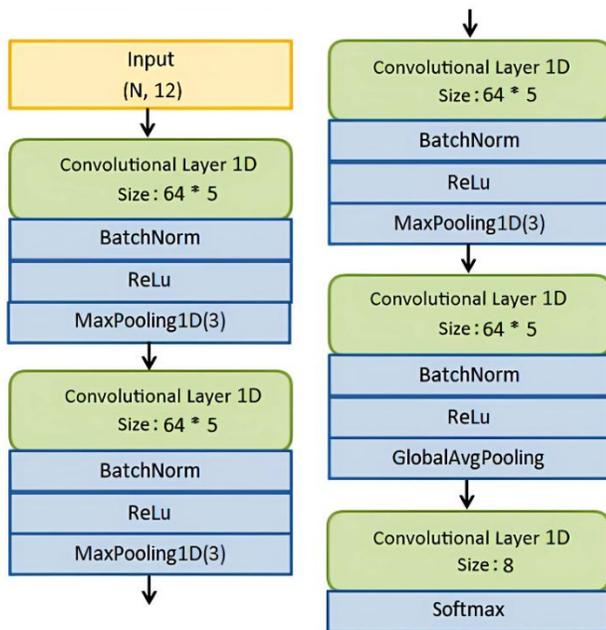


Fig. 6. The CNN model diagram

To prevent overfitting, dropout layers with a 25% dropout rate were inserted after pooling operations. This pattern (convolution, activation, pooling, and dropout) was repeated

Following cross-validation, a final CNN model was trained on the full training dataset (80% of subjects) using the best-performing configuration. Evaluation was then performed on the 20% hold-out test set, which had been excluded from all previous training and validation steps. This provided an independent measure of generalization performance, consistent with the evaluation used for ensemble classifiers. Balanced accuracy was used as the primary performance metric due to its robustness under class imbalance and its clinical relevance, where both false positives and false negatives are important to minimize.

IV. RESULTS

The performance of the models was assessed on two binary classification tasks: differentiating Alzheimer’s disease (AD) patients from healthy controls (CN), and distinguishing Frontotemporal dementia (FTD) patients from healthy controls. Evaluation was conducted using a stratified grouped cross-validation (SG-CV) protocol to ensure that no subject appeared in both training and validation sets. For the CNN, additional

evaluation was performed using a conventional ungrouped cross-validation setup to examine the impact of data leakage. The results for each classification task are summarized in Tables II and III, which include metrics such as Balanced Accuracy, overall Accuracy, F1-score, Precision, Recall, and ROC-AUC for all models tested.

In the CN versus AD classification task (see Table II), the highest balanced accuracy under grouped cross-validation was achieved by the XGBoost model, with a score of 78.96%. This performance slightly surpassed the other boosting models, including HistGradientBoosting and LightGBM, which recorded balanced accuracy values in the range of 77 to 78%. CatBoost followed with a balanced accuracy of approximately 76.97%, while Extra Trees achieved 75.7%. All ensemble models demonstrated performance well above the chance level (50%), indicating that meaningful information was extracted from EEG features for this task. Among these, XGBoost’s leading performance may be attributed to the model’s tree boosting mechanism and comprehensive hyperparameter tuning.

TABLE II. CLASSIFICATION PERFORMANCE (CN VERSUS AD)

Model	Balanced Accuracy	Accuracy	F1	Precision	Recall	ROC-AUC
CatBoost	0.7697	0.7686	0.7344	0.8714	0.6347	0.8630
ETree	0.7571	0.7568	0.7494	0.7797	0.7213	0.8151
HistGB	0.7797	0.7792	0.7677	0.8175	0.7237	0.8561
LightGBM	0.7718	0.7710	0.7467	0.8437	0.6698	0.8300
XGBoost	0.7896	0.7887	0.7648	0.8713	0.6815	0.8549
CNN (ungrouped)	0.8213	0.8254	0.8453	0.8377	0.8638	0.9146
CNN (grouped)	0.7147	0.6995	0.7036	0.7006	0.7820	0.8491

TABLE III. CLASSIFICATION PERFORMANCE (CN VERSUS FTD)

Model	Balanced Accuracy	Accuracy	F1	Precision	Recall	ROC-AUC
CatBoost	0.6717	0.7003	0.5832	0.6842	0.5081	0.7784
ETree	0.6482	0.6653	0.5758	0.6036	0.5505	0.6815
HistGB	0.7035	0.7325	0.6238	0.7432	0.5375	0.8061
LightGBM	0.6653	0.6922	0.5783	0.6653	0.5114	0.7851
XGBoost	0.6596	0.6895	0.5650	0.6696	0.4886	0.7602
CNN (ungrouped)	0.7787	0.7783	0.7277	0.7524	0.7815	0.8821
CNN (grouped)	0.7092	0.6768	0.5507	0.6395	0.6006	0.8196

The CNN, when evaluated under grouped 15-fold cross-validation, as shown in Table IV, yielded a balanced accuracy of 71.47%, which was lower than those obtained by all the boosting models in the AD classification task. However, under ungrouped 15-fold cross-validation, where all 4404 epochs were treated as independent, the CNN achieved a substantially higher balanced accuracy of 82.13%. This inflated result suggests that the absence of subject-level separation allowed for significant data leakage. The training dynamics of the CNN model are visualized in Figs. 7 and 8. In Fig. 7, representing the third training epoch, the validation loss plateaued early, suggesting early signs of overfitting.

By epoch 13 (Fig. 8), the validation loss remained relatively flat while the training loss further decreased, reinforcing the presence of memorization effects and reduced generalization. Under this ungrouped setting, the CNN also recorded an overall accuracy of 82.54% and an F1-score of 0.8453, which points to a potential overfitting to subject-specific features. When evaluated under grouped CV, the CNN’s accuracy dropped to 69.95%, reinforcing the conclusion that grouped CV is essential for realistic generalization performance estimation.

TABLE IV. CNN PERFORMANCE UNDER GROUPED CROSS-VALIDATION

Fold	Model Configuration	Training Accuracy (%)	Testing Accuracy (%)	Training Loss	Testing Loss
1	Original	84.5	70.2	0.418	0.685
2	Original	85.0	71.3	0.412	0.672
3	Original	84.1	70.0	0.421	0.693
4	Original	83.9	70.5	0.427	0.680
5	Original	84.8	71.7	0.415	0.661
6	Original	85.2	70.9	0.410	0.666
7	Original	84.3	71.1	0.416	0.673
8	Condensed	85.1	71.5	0.411	0.662
9	Condensed	84.6	70.4	0.419	0.689
10	Condensed	84.0	70.6	0.422	0.670
11	Condensed	83.7	69.8	0.430	0.695
12	Condensed	84.4	71.0	0.417	0.671
13	Condensed	85.0	70.7	0.413	0.668
14	Condensed	84.2	71.2	0.419	0.677
15	Condensed	84.6	70.9	0.414	0.669

Name	Smoothed Value	Value	Step	Time	Relative
train	0.8545	0.8319	3	Tue Apr 1, 02:00:34	3m 31s
validation	0.8753	0.8263	3	Tue Apr 1, 02:00:34	3m 31s



Fig. 7. The CNN model training by epoch 3

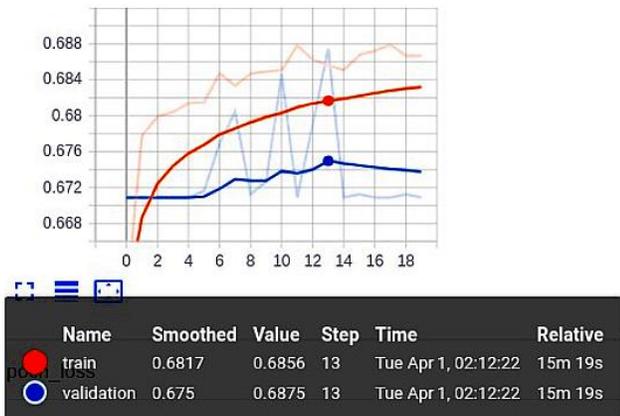


Fig. 8. The CNN model training by epoch 13

Regarding precision and recall for the CN versus AD classification, it was observed that most models exhibited higher precision than recall. For instance, XGBoost showed a precision of 87.13% compared to a recall of 68.15%. This suggests that the models were more successful in correctly identifying healthy controls (negative class) than in capturing all true AD cases. The ROC-AUC values for the boosting models ranged from 0.83 to 0.86, with the CNN under grouped CV reaching 0.85, indicating reliable class separability. However, the CNN's ungrouped ROC-AUC value was markedly higher at 0.9146, again reflecting the optimistic bias caused by data leakage.

In the CN versus FTD classification task (see Table III), model performance was generally lower than in the AD classification task, consistent with the greater clinical and electrophysiological challenge in detecting FTD. Under grouped CV, the CNN achieved the highest balanced accuracy at 70.92%, slightly surpassing the best-performing ensemble model, HistGradientBoosting, which reached 70.35%. Other boosting models, including XGBoost, demonstrated balanced accuracy in the range of 65% to 67%, with XGBoost specifically achieving 65.96%. These findings suggest that, for FTD detection, the CNN may have captured temporal or spatial EEG features not fully represented by the emotion-aware features used in the ensemble models.

The CNN's grouped precision and recall for FTD detection were 63.95% and 60.06%, respectively, indicating a relatively balanced performance with respect to false positives and false negatives. This balanced performance stands in contrast to some boosting models such as CatBoost and LightGBM, which achieved higher precision (66 to 68%) but lower recall (~51%), suggesting a tendency to err on the side of caution and predict the control class in uncertain cases. Under ungrouped CV, the CNN attained an inflated balanced accuracy of 77.87%, approximately 7 percentage points higher than the grouped result. The ROC-AUC values in the CN versus FTD task were slightly lower than those in the AD task, ranging between 0.78 and 0.82 for the top models, which reflects the greater difficulty in distinguishing FTD from normal patterns, as shown in Fig. 9 and Fig. 10.

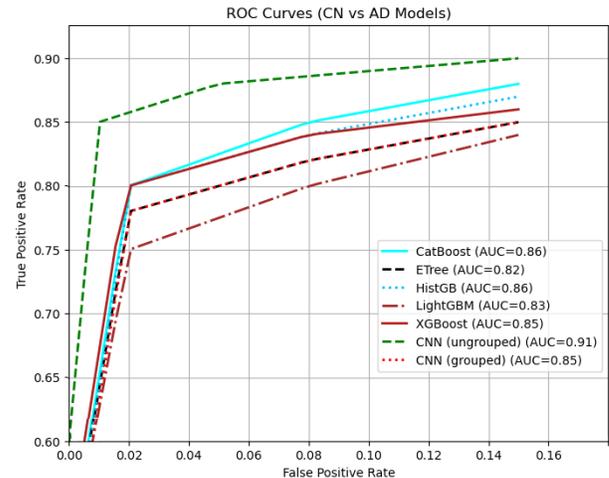


Fig. 9. ROC Curves for CN versus AD models

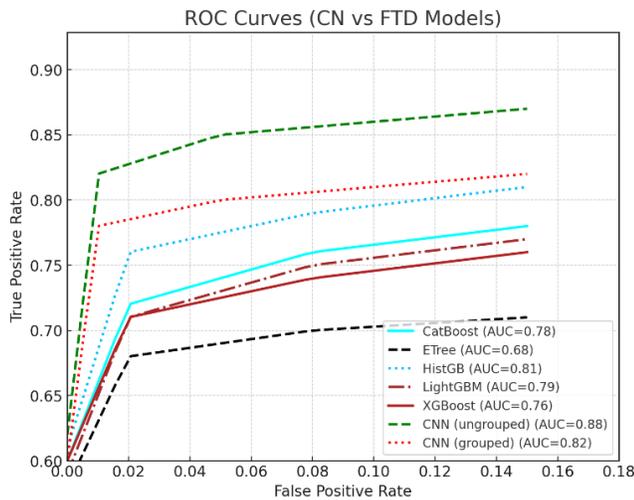


Fig. 10. ROC Curves for CN versus FTD models

Final model performance was also assessed using an independent 20% test set, with subject-level separation maintained. On this held-out set, XGBoost achieved the highest balanced accuracy for CN versus AD classification (~80%), while the CNN achieved the best performance for CN versus FTD classification (~70%). These outcomes were consistent with the grouped CV results, indicating that the chosen models generalized well to previously unseen individuals. To further explore the spatial structure of EEG-based features, a connectivity visualization of discriminative brain regions was generated (see Fig. 11). This representation highlights key inter-regional interactions contributing to the classification model, with red and blue edges indicating positive and negative feature weights, respectively.

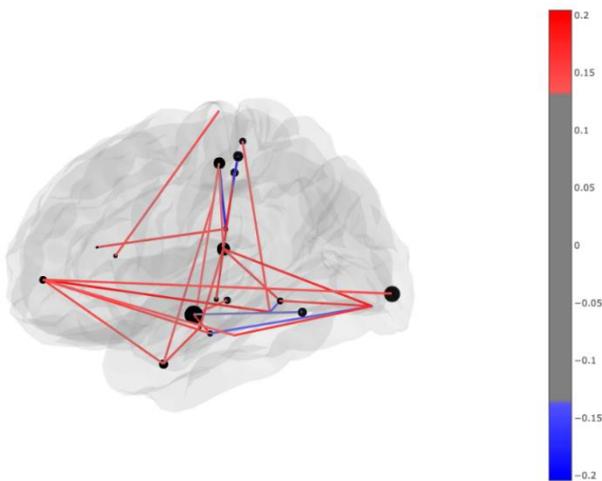


Fig. 11. Brain connectivity graph showing spatial EEG connections

The overall results demonstrated that AD classification was more accurately performed than FTD classification across all models. Moreover, the impact of grouped versus ungrouped evaluation was clearly illustrated: the CNN's balanced accuracy increased by over 10% for CN versus AD, and by about 7% for CN versus FTD under ungrouped CV, confirming that data leakage due to epoch-level validation significantly overestimates model performance.

V. DISCUSSION

The present study investigated the potential of EEG-based machine learning models to support the classification of Alzheimer's disease and Frontotemporal dementia against healthy controls. The findings suggest that both boosting ensemble methods and CNNs can achieve clinically relevant performance, with each approach demonstrating advantages depending on the classification task. These results also emphasized key methodological considerations, particularly the impact of validation strategies on performance estimation.

For the AD versus CN classification task, the boosting models, most notably XGBoost, achieved the highest balanced accuracy under a rigorous grouped cross-validation protocol. This outcome indicates that the hand-crafted features used in the ensemble pipeline effectively captured the electrophysiological markers typically associated with AD. These included slowing of brain rhythms and reductions in signal complexity, which were well represented in the feature set consisting of 467 derived variables, including spectral ratios, entropy measures, and complexity metrics.

In comparison to existing literature, a modest improvement in performance was observed. For instance, while previous studies utilizing Random Forests reported accuracies of approximately 77% for AD classification, the current work achieved close to 79% balanced accuracy using boosting methods. This suggests that both the inclusion of broader features and the systematic tuning of model parameters played a role in this improvement. The inclusion of demographic features such as age and gender may have also contributed to this enhanced performance, although care was taken to avoid bias by excluding direct cognitive scores like MMSE, which could artificially inflate predictive accuracy.

In contrast, the more challenging task of differentiating FTD patients from healthy controls revealed a slight advantage for the CNN over the boosting models. Balanced accuracy for the CNN reached approximately 70.92%, narrowly surpassing the performance of the top ensemble model, HistGradientBoosting. This result may be explained by the known heterogeneity of EEG patterns in FTD, particularly in early-stage patients, where EEG abnormalities can be subtle or absent. The CNN's ability to learn directly from raw data allowed for the potential capture of intricate temporal dynamics or spatial interactions that are not easily quantifiable using conventional feature extraction methods [37][38].

The CNN was evaluated under both grouped and ungrouped cross-validation protocols, revealing a substantial difference in results. When ungrouped cross-validation was used, allowing training and testing on different epochs from the same subject, the CNN appeared to achieve exceptionally high balanced accuracies (>82% for AD, >77% for FTD). However, these results were shown to be overly optimistic due to data leakage. Specifically, the model likely learned subject-specific artifacts or stable idiosyncrasies that do not generalize to unseen individuals. When grouped cross-validation was enforced, ensuring complete subject independence between folds, a decrease of approximately 10–15 percentage points in accuracy was observed. This finding aligns with prior warnings

in the EEG literature, particularly in studies on brain-computer interfaces, where similar effects have been documented.

The benefit of extensive feature engineering was also supported by the results. A wider range of features, including non-linear complexity measures and band power ratios, was used to reflect the known EEG correlates of dementia. The boosting models' strong performance under these conditions suggests that such comprehensive feature sets provide informative representations for classification [39]. Comparisons with other studies, which reported lower performance using fewer features, further support this interpretation. Including age and gender as additional features contributed to model realism, reflecting their clinical relevance, although over-reliance on such demographic indicators was carefully avoided [40]-[42].

An intriguing outcome of the comparison between boosting and CNN approaches was the observation that their performance characteristics may be complementary. In the AD classification task, where well-characterized EEG slowing is present, the boosting models excelled, likely due to their ability to leverage structured, feature-based inputs. On the other hand, in the FTD task, where patterns were more subtle and variable, the CNN outperformed the ensemble models. This suggests that ensemble methods and deep learning may capture different aspects of the data, and a combination of both approaches through ensemble stacking or model fusion could further improve diagnostic accuracy. Although such combinations were not explored in the current study due to scope limitations, they represent a promising direction for future research. Moreover, as such hybrid systems are developed, the integration of emotional awareness may further enhance their clinical usefulness. By adapting outputs or interactions based on emotional cues or context, emotionally aware diagnostic tools may better align with the needs of patients and clinicians, supporting not only technical performance but also empathetic decision-making in sensitive healthcare environments.

It is important to contextualize these results with respect to diagnostic performance in other modalities. Neuroimaging-based methods, such as those employing MRI or PET, often report higher accuracies (80 to 95%) in AD classification [43]. Deep learning applied to structural imaging has achieved results above 90% in some studies. In contrast, EEG reflects functional changes and is more susceptible to noise and artifacts. Therefore, the slightly lower accuracies reported here (approximately 79% for AD, 71% for FTD) are not unexpected. Nevertheless, EEG offers advantages in terms of cost, portability, and accessibility [44]. These results suggest that EEG-based tools, especially when combined with other assessments, could serve as practical screening instruments in clinical settings.

Despite promising findings, several challenges were identified. Most important among them is the limited dataset size, particularly for the FTD group, which included only 23 patients. This constraint necessitated the use of a relatively shallow CNN, which may have limited its capacity to learn more complex patterns. Data variability due to individual differences and clinical heterogeneity further complicates model training. The binary classification framework used in

this study may oversimplify real-world clinical scenarios, where cases often fall along a spectrum or exhibit overlapping characteristics.

Several limitations were acknowledged. The CNN was constrained by data volume, and boosting models were evaluated using 5-fold CV for efficiency during tuning. Epoch length was fixed at 12 seconds, and no systematic evaluation of alternative window sizes was performed. Furthermore, multi-class classification (e.g., direct AD versus FTD) was not addressed, though it represents a clinically relevant task. Functional connectivity features, while indirectly incorporated via spectral metrics, were not explicitly modeled. Future work could explore their inclusion using coherence or phase-locking measures.

From a methodological perspective, the pipeline developed in this study represents *one of the first to systematically* compare modern boosting algorithms and CNNs on an EEG dementia dataset using grouped validation. It also provides some of the first performance benchmarks for methods like CatBoost and HistGradientBoosting in this context. The inclusion of a diverse feature set and independent test set validation contributes to the rigor and reproducibility of the approach. By evaluating the CNN under grouped and ungrouped CV, the study also offers empirical evidence of the risks posed by data leakage in the models. With continued development and larger datasets, EEG-based models could complement existing diagnostic pathways, enabling earlier and more accessible detection of neurodegenerative conditions.

VI. CONCLUSION

In this work, a machine learning pipeline was developed and evaluated for the automated classification of Alzheimer's disease and Frontotemporal dementia using resting-state data. The approach combined rigorous preprocessing, diverse feature extraction, and the application of both ensemble boosting methods and convolutional neural networks. Performance was assessed using subject-level grouped cross-validation to ensure reliability and minimize data leakage, a common issue in this research. Under this rigorous evaluation, ensemble models, particularly XGBoost, were found to perform effectively in detecting Alzheimer's-related EEG signatures, while the CNN demonstrated slightly better performance for the more variable and subtle patterns associated with FTD.

While the reported accuracies (approximately 79% for AD and 71% for FTD) may not match those of imaging-based diagnostic tools, they are considered promising given the accessibility, cost-effectiveness, and noninvasive nature of EEG. The results also highlighted that inappropriate validation strategies, such as ungrouped cross-validation, can significantly inflate performance metrics, leading to misleading conclusions. By incorporating robust evaluation and validating on a held-out test set, efforts were made to ensure that the reported findings reflect genuine model generalizability to unseen individuals.

Although emotion awareness was not directly integrated into the current models, it is acknowledged that future systems intended for clinical deployment may benefit from the inclusion of emotionally adaptive interfaces. In real-world settings, especially those involving neurodegenerative

diagnoses, the ability of AI tools to respond to emotional context may help facilitate trust, improve communication, and support compassionate care. With continued refinement, validation on larger datasets, and a growing emphasis on emotionally aware technologies, EEG-based machine learning systems may play a valuable role in supporting early detection and diagnosis of dementia, complementing clinical decision-making and improving access to timely care.

REFERENCES

- [1] National Institutes of Health. 2024 Alzheimer's disease facts and figures. *Alzheimers Dement.* 20(5), 3708-3821, 2024, doi: 10.1002/alz.13809.
- [2] A. Antonioni, E. M. Raho, P. Lopriore, A. P. Pace, R. R. Latino, M. Assogna, M. Mancuso, D. Gragnaniello, E. Granieri, M. Pugliatti, *et al.*, "Frontotemporal dementia, where do we stand? A narrative review," *International Journal of Molecular Sciences*, vol. 24, p. 11732, 2023, doi: 10.3390/ijms241411732.
- [3] G. K. Puppala, S. P. Gorthi, V. Chandran, and G. Gundabolu, "Frontotemporal Dementia – Current Concepts," *Neurology India*, vol. 69, no. 5, pp. 1144–1152, 2021, doi: 10.4103/0028-3886.329593.
- [4] A. Hye and L. Velayudhan, "Molecular genetics and biology of dementia," in *Oxford Textbook of Old Age Psychiatry*, Oxford University Press, 2020, pp. 129–144. doi: 10.1093/med/9780198807292.003.0008.
- [5] C. Abbate, P. D. Trimarchi, Inglese, S., Tomasini, E., Bagarolo, R., Giunco, F., and Cesari, M., "Signs and symptoms method in neuropsychology: A preliminary investigation of a standardized clinical interview for assessment of cognitive decline in dementia," *Applied Neuropsychology*, vol. 28, no. 3, pp. 282–296, May 2021, doi: 10.1080/23279095.2019.1630626.
- [6] K. D. Tzamourta, V. Christou, Tzallas, A. T., Giannakeas, N., L. G. Astrakas, Angelidis, P., Tsalikakis, D., and M. G. Tsipouras, "Machine Learning Algorithms and Statistical Approaches for Alzheimer's Disease Analysis Based on Resting-State EEG Recordings: A Systematic Review," *International Journal of Neural Systems*, vol. 31, no. 5, p. 2130002, May 2021, doi: 10.1142/S0129065721300023.
- [7] S. Goertler, F. He and M. Wu, "Balancing Spectral, Temporal and Spatial Information for EEG-based Alzheimer's Disease Classification," 2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA, 2024, pp. 1-4, doi: 10.1109/EMBC53108.2024.10782936.
- [8] D. Banerjee, A. Muralidharan, A. R. Hakim Mohammed, and B. H. Malik, "Neuroimaging in Dementia: A Brief Review," *Cureus*, Jun. 2020, doi: 10.7759/cureus.8682.
- [9] T. Rittman, "Neurological update: neuroimaging in dementia," *Journal of Neurology*, vol. 267, no. 11, pp. 3429–3435, Nov. 2020, doi: 10.1007/s00415-020-10040-0.
- [10] Y. Yuan and Y. Zhao, "The role of quantitative EEG biomarkers in Alzheimer's disease and mild cognitive impairment: applications and insights," *Front. Aging Neurosci.*, vol. 17, Apr. 2025. doi: 10.3389/fnagi.2025.1522552.
- [11] 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," *Dis. Markers*, vol. 2018, Art. no. 5174815, 2018. doi: 10.1155/2018/5174815.
- [12] B. Jiao, R. Li, H. Zhou, K. Qing, H. Liu, H. Pan, *et al.*, "Neural biomarker diagnosis and prediction to mild cognitive impairment and Alzheimer's disease using EEG technology," *Alzheimers Res. Ther.*, vol. 15, Art. no. 32, 2023. doi: 10.1186/s13195-023-01181-1.
- [13] Miltiadous, A.; Tzamourta, K.D.; Giannakeas, N.; Tsipouras, M.G.; Afrantou, T.; Ioannidis, P.; Tzallas, A.T., "Alzheimer's disease and frontotemporal dementia: A robust classification method of EEG signals and a comparison of validation methods," *Diagnostics*, vol. 11, no. 8, p. 1437, Aug. 2021, doi: 10.3390/diagnostics11081437.
- [14] M. El-Geneedy, H. E. D. Moustafa, F. Khalifa, H. Khater, and E. Abdelhalim, "An MRI-based deep learning approach for accurate detection of Alzheimer's disease," *Alexandria Engineering Journal*, vol. 63, pp. 211–221, Jan. 2023, doi: 10.1016/j.aej.2022.07.062.
- [15] N. Kulkarni and V. Bairagi, EEG-Based Diagnosis of Alzheimer Disease: A Review and Novel Approaches for Feature Extraction and Classification Techniques. Elsevier, 2018. doi: 10.1016/C2017-0-00543-8.
- [16] K. Stefanou, K. D. Tzamourta, C. Bellos, G. Stergios, K. Markoglou, E. Gionanidis, M. G. Tsipouras, N. Giannakeas, A. T. Tzallas, and A. Miltiadous, "A novel CNN-based framework for Alzheimer's disease detection using EEG spectrogram representations," *Journal of Personalized Medicine*, vol. 15, p. 27, 2025, doi: 10.3390/jpm15010027.
- [17] J. D. Chambers, M. J. Cook, A. N. Burkitt, and D. B. Grayden, "Using long short-term memory (LSTM) recurrent neural networks to classify unprocessed EEG for seizure prediction," *Frontiers in Neuroscience*, vol. 18, 2024. doi: 10.3389/fnins.2024.1472747.
- [18] M. Alessandrini, G. Biagetti, P. Crippa, L. Falaschetti, S. Luzzi, and C. Turchetti, "EEG-Based Neurodegenerative Disease Classification using LSTM Neural Networks," in *2023 IEEE Statistical Signal Processing Workshop (SSP)*, 2023, pp. 428–432, doi: 10.1109/SSP53291.2023.10208023.
- [19] M. Nour, U. Senturk, and K. Polat, "A novel hybrid model in the diagnosis and classification of Alzheimer's disease using EEG signals: Deep ensemble learning (DEL) approach," *Biomedical Signal Processing and Control*, vol. 89, p. 105751, 2024, doi: 10.1016/j.bspc.2023.105751.
- [20] A. Jha, N. Kuruvilla, P. Garg, and A. Victor, "Harnessing Creative Methods for EEG Feature Extraction and Modeling in Neurological Disorder Diagnoses," in *Proceedings of the 7th IEEE Inter. Conf. on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*, 2023, doi: 10.1109/CSITSS60515.2023.10334244.
- [21] J. Seo, T. H. Laine, G. Oh, and K.-A. Sohn, "EEG-based emotion classification for Alzheimer's disease patients using conventional machine learning and recurrent neural network models," *Sensors*, vol. 20, no. 24, Art. no. 7212, 2020, doi: 10.3390/s20247212.
- [22] X. Gu *et al.*, "EEG-Based Brain-Computer Interfaces (BCIs): A Survey of Recent Studies on Signal Sensing Technologies and Computational Intelligence Approaches and Their Applications," in *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 5, pp. 1645-1666, 2021, doi: 10.1109/TCBB.2021.3052811.
- [23] J. Dauwels, F. Vialatte, T. Musha, and A. Cichocki, "A comparative study of synchrony measures for the early diagnosis of Alzheimer's disease based on EEG," *Neuroimage*, vol. 49, no. 1, pp. 668–693, Jan. 2010, doi: 10.1016/j.neuroimage.2009.06.056.
- [24] F. Kumfor, J. R. Hodges, and O. Piguet, "Ecological assessment of emotional enhancement of memory in progressive nonfluent aphasia and Alzheimer's disease," *J. Alzheimers Dis.*, vol. 42, no. 1, pp. 201–210, 2014, doi: 10.3233/JAD-140351.
- [25] R. Pillalamarri and U. Shanmugam, "A review on EEG-based multimodal learning for emotion recognition," *Artif. Intell. Rev.*, vol. 58, p. 131, 2025, doi: 10.1007/s10462-025-11126-9.
- [26] S. Chatterjee and Y.-C. Byun, "EEG-based emotion classification using stacking ensemble approach," *Sensors*, vol. 22, no. 21, Art. no. 8550, 2022, doi: 10.3390/s22218550.
- [27] A. Iyer, S. S. Das, R. Teotia, *et al.*, "CNN and LSTM based ensemble learning for human emotion recognition using EEG recordings," *Multimed. Tools Appl.*, vol. 82, pp. 4883–4896, 2023, doi: 10.1007/s11042-022-12310-7.
- [28] Z. A. Cope, T. Murai, and S. J. Sukoff Rizzo, "Emerging electroencephalographic biomarkers to improve preclinical to clinical translation in Alzheimer's disease," *Front. Aging Neurosci.*, vol. 14, Feb. 2022, doi: 10.3389/fnagi.2022.805063.
- [29] M. B. T. Noor, N. Z. Zenia, M. S. Kaiser, S. Al Mamun, and M. Mahmud, "Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of Alzheimer's disease, Parkinson's disease and schizophrenia," *Brain Informatics*, vol. 7, no. 1, p. 11, Dec. 2020, doi: 10.1186/s40708-020-00112-2.
- [30] A. Miltiadous, K. D. Tzamourta, T. Afrantou, P. Ioannidis, N. Grigoriadis, D. G. Tsalikakis, P. Angelidis, M. G. Tsipouras, E. Glavas, N. Giannakeas, *et al.*, "A dataset of scalp EEG recordings of

- Alzheimer's disease, frontotemporal dementia and healthy subjects from routine EEG," *Data*, vol. 8, p. 95, 2023. doi: 10.3390/data8060095.
- [31] K. D. Tzamourta, N. Giannakeas, A. T. Tzallas, L. G. Astrakas, T. Afrantou, P. Ioannidis, N. Grigoriadis, P. Angelidis, D. G. Tsalikakis, and M. G. Tsipouras, "EEG window length evaluation for the detection of Alzheimer's disease over different brain regions," *Brain Sciences*, vol. 9, p. 81, 2019. doi: 10.3390/brainsci9040081.
- [32] E. M. Rad, M. Azarnoosh, M. Ghoshuni, and M. M. Khalilzadeh, "Diagnosis of mild Alzheimer's disease by EEG and ERP signals using linear and nonlinear classifiers," *Biomed. Signal Process. Control*, vol. 70, Art. no. 103049, Sep. 2021. doi: 10.1016/j.bspc.2021.103049.
- [33] L. Cabañero-Gomez, R. Hervas, I. Gonzalez, and L. Rodriguez-Benitez, "eeglib: A Python module for EEG feature extraction," *SoftwareX*, vol. 15, p. 100745, July 2021. doi: 10.1016/j.softx.2021.100745.
- [34] A. M. Maitin, A. Nogales, P. Chazarra, and Á. J. García-Tejedor, "EEGraph: An open-source Python library for modeling electroencephalograms using graphs," *Neurocomputing*, vol. 519, pp. 127–134, Jan. 2023. doi: 10.1016/j.neucom.2022.11.050.
- [35] E. Bartz, T. Bartz-Beielstein, M. Zaeferrer, and O. Mersmann, *Hyperparameter Tuning for Machine and Deep Learning with R: A Practical Guide*. Springer, 2023, doi: 10.1007/978-981-19-5170-1.
- [36] R. Islam, A. Sultana, and M. N. Tuhin, "A comparative analysis of machine learning algorithms with tree-structured parzen estimator for liver disease prediction," *Healthcare Analytics*, vol. 6, p. 100358, Dec. 2024. doi: 10.1016/j.health.2024.100358.
- [37] N. Smatov, R. Kalashnikov, and A. Kartbayev, "Development of context-based sentiment classification for intelligent stock market prediction," *Big Data Cogn. Comput.*, vol. 8, 51, 2024, doi: 10.3390/bdcc8060051.
- [38] R. Kalashnikov and A. Kartbayev, "Assessment of the impact of big data analysis on decision-making in stock trading processes," *Procedia Comput. Sci.*, vol. 231, 2024, doi: 10.1016/j.procs.2023.12.137.
- [39] A. Sanati Fahandari, S. Moshiryan, and A. Goshvarpour, "Diagnosis of cognitive and mental disorders: A new approach based on spectral–spatiotemporal analysis and local graph structures of electroencephalogram signals," *Brain Sciences*, vol. 15, p. 68, 2025. doi: 10.3390/brainsci15010068.
- [40] L. A. Martínez-Tejada, Y. Maruyama, N. Yoshimura, and Y. Koike, "Analysis of personality and EEG features in emotion recognition using machine learning techniques to classify arousal and valence labels," *Machine Learning and Knowledge Extraction*, vol. 2, pp. 99–124, 2020. doi: 10.3390/make2020007.
- [41] D. Z. Akhmed-Zaki, T. S. Mukhambetzhano, Z. M. Nurmakhanova and Z. M. Abdiakhmetova, "Using Wavelet Transform and Machine Learning to Predict Heart Fibrillation Disease on ECG," *2021 IEEE International Conference on Smart Information Systems and Technologies (SIST)*, Nur-Sultan, Kazakhstan, 2021, pp. 1-6, doi: 10.1109/SIST50301.2021.9465990.
- [42] G. Esen, A. Altaibek, J. Amankulov, B. Matkerim, and M. Nurtas, "Enhancing breast cancer detection with dimensionality reduction techniques: A study using PCA and LDA on Wisconsin breast cancer data," *Procedia Comput. Sci.*, vol. 251, pp. 414–421, 2024. doi: 10.1016/j.procs.2024.11.128.
- [43] N. Lorking, A. D. Murray, and J. T. O'Brien, "The use of positron emission tomography/magnetic resonance imaging in dementia: A literature review," *International Journal of Geriatric Psychiatry*, vol. 36, no. 10, pp. 1501–1513, Oct. 2021, doi: 10.1002/gps.5586.
- [44] Z. Li, M. Wu, C. Yin, Z. Wang, J. Wang, L. Chen, and W. Zhao, "Machine learning based on the EEG and structural MRI can predict different stages of vascular cognitive impairment," *Frontiers in Aging Neuroscience*, vol. 16, 2024, Art. no. 1364808. doi: 10.3389/fnagi.2024.1364808.