# Utilizing Shuffle Attention for Subject-Wise Split for EEG-Based Alzheimer's Detection

**JASON LEE[1*], ABRAHAM N. REYES[1], ELIJAH GHOSSEIN[1]**

*[1]Decoded Brain, Pasadena City College, 1570 E. Colorado Blvd, Pasadena CA 91106, USA*
*[*]jlee265@go.pasadena.edu*

**Abstract:** Electroencephalography (EEG) has emerged as a noninvasive and cost-effective modality for the automated detection of Alzheimer's disease (AD), yet many deep learning approaches report inflated performance due to subject-level data leakage. In this study, we propose a convolutional neural network supplemented with a shuffle attention mechanism for binary classification of AD versus healthy controls, evaluated under strictly enforced subject-wise validation. Raw EEG recordings acquired in a routine clinical setting were denoised using artifact subspace reconstruction and independent component analysis, segmented into overlapping epochs, and transformed into time-frequency representations using the short-time Fourier transform. The proposed model integrates shuffle attention at the final convolutional stage to enhance discriminative spectro-temporal features while suppressing subject-specific noise. Model performance was assessed using both a subject-wise holdout split (70/15/15) and leave-one-subject-out (LOSO) cross-validation. Under subject-wise holdout validation, the model achieved an accuracy of 90.87%, recall of 89.47%, precision of 88.98%, F1 score of 89.23%, and specificity of 92.36%, while LOSO validation highlighted the impact of inter-subject variability. These results demonstrate that combining attention mechanisms with rigorous validation protocols yields clinically realistic performance estimates and supports the potential of EEG-based deep learning models for Alzheimer's disease classification.

**Keywords: machine learning, attention mechanism, deep learning, Alzheimer's disease, neuroscience, EEG**



## *Introduction*

Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterized by cognitive decline and substantial functional impairment, representing a major public health challenge worldwide [1]. Current diagnostic approaches rely primarily on clinical assessment supported by neuroimaging and cerebrospinal fluid biomarkers, which can be costly, invasive, or limited in accessibility [2]. These constraints have motivated growing interest in alternative modalities that are noninvasive, scalable, and suitable for repeated measurements.

Electroencephalography (EEG) offers a low-cost and noninvasive means of capturing neural activity with high temporal resolution and has been widely studied as a potential biomarker for neurodegenerative disease [3]. Prior work has reported alterations in EEG spectral power, functional connectivity, and signal complexity in patients with AD relative to healthy controls [3], [4]. Building on these observations, machine learning and deep learning techniques have been increasingly applied to EEG data to support automated dementia classification [5]. Many recent approaches employ time-frequency representations, such as short-time Fourier transform or wavelet-based spectrograms, in combination with convolutional neural networks to learn discriminative spectro-temporal features [6], [7].

Despite promising results, a growing body of methodological literature has highlighted limitations in the evaluation protocols used in EEG-based deep learning studies for Alzheimer's disease. In particular, epoch-wise or slice-wise data splitting can introduce subject-level data leakage because EEG segments from the same individual are strongly correlated. Models evaluated under such conditions may learn subject-specific signatures rather than disease-related patterns, leading to inflated performance estimates that do not generalize to unseen patients. Subject-wise validation strategies, including leave-one-subject-out and patient-level holdout splits, have therefore been recommended to obtain clinically realistic assessments of model performance [8].

In parallel, attention mechanisms have been incorporated into deep learning architectures to enhance feature discrimination by selectively emphasizing informative components of learned representations [9]. Channel-based and spatial attention modules have demonstrated performance gains in a variety of computer vision and biomedical signal processing tasks [10]. However, the application of attention mechanisms to EEG-based dementia classification remains limited, and many existing studies combine attention with validation strategies that do not adequately control for subject-level leakage.

In this study, we investigate a convolutional neural network augmented with a shuffle attention mechanism for binary classification of Alzheimer's disease versus healthy controls using EEG data. Time-frequency representations derived from short-time Fourier transform spectrograms are used as model inputs, and all evaluations are conducted under strictly enforced subject-wise validation protocols, including both holdout and leave-one-subject-out schemes. By combining attention-based feature refinement with rigorous validation, this work aims to provide a methodologically sound assessment of deep learning performance for EEG-based Alzheimer's disease classification. The development and execution of this workflow and pipeline were made possible by an open-source pedagogical framework designed to foster collaboration among engineers and researchers alike.

This research was conducted by researchers within Decoded Brain, an active student-led research organization of scientists and engineers focused on collaborative advancement of the brain. Inspired by the traditional peer-review system, engineers at Decoded Brain created a similar structure within the organization to best facilitate research projects and the peer-review process called Open Labs. This open-source software implements a continuous feedback pedagogy where student researchers submit structured weekly progress reports. The literature undergoes a weekly rigorous peer-evaluation process, producing review feedback from fellow research peers that includes technical guidance, directional adjustments, and overall research synthesis. Utilizing this ecosystem allowed the researchers to maintain and foster collaboration and objective peer validation throughout the lifecycle of this paper.

Beyond this specific study, this model is designed for scalability across diverse academic organizations and campuses, with the ultimate goal of fostering decentralized, active research communities. Because the foundational design of Open Labs is inherently domain-agnostic, it can be seamlessly adapted to any field of study. By implementing this framework within institutions—particularly community colleges—we aim to transform the research experience into an enjoyable, deeply engaging, and accessible pursuit, reinforcing that impactful and meaningful technical education is fundamentally a collaborative journey.

## *Methods*

Figure 1 outlines the four principal stages of the proposed pipeline: EEG preprocessing and artifact removal, feature extraction, and classification using the proposed model. Each step will be discussed in detail below to ensure reproducibility and validity.

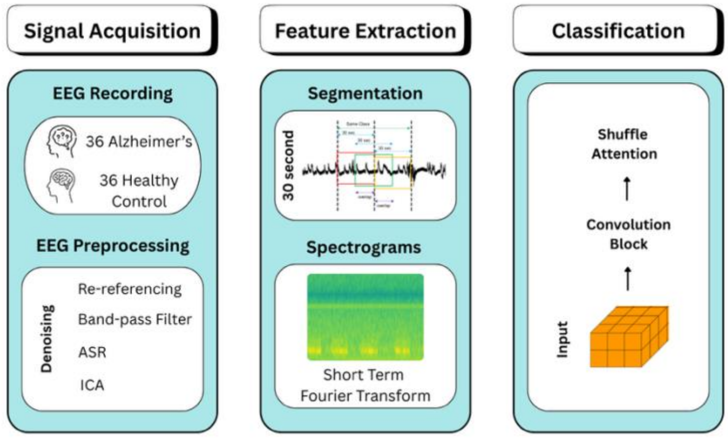



*Fig. 1. Diagram showing the main components of our pipeline. From left to right signal acquisition which includes patients and their description, signal denoising, window segmentation, and feature extraction, and finally classification.*

**About the EEG**

This study used recordings from the 2nd Department of Neurology at AHEPA General University Hospital in Thessaloniki, which are publicly available here. The dataset comprised 36 patients diagnosed with Alzheimer's disease (AD), 23 patients with frontotemporal dementia (FTD), and 29 healthy controls (HC). Cognitive status was assessed using the Mini-Mental State Examination (MMSE), with scores ranging from 0 to 30 and lower scores indicating greater cognitive impairment. Recordings were acquired and measured from routine EEG of patients in months, and the median value was 25, with the interquartile range (IQR) being 24-28.5 months. Concerning the AD group, no dementia-related comorbidities have been reported. Initial diagnosis of AD and FTD patients was performed according to the criteria provided by the Diagnostic and Statistical Manual of Mental Disorders, 3rd ed., revised (DSM-IIIR, DSM IV, ICD-10) and the National Institute of Neurological, Communicative Disorders and Stroke—Alzheimer's Disease and Related Disorders Association (NINCDS—ADRDA) [11], [12]. The average MMSE for the AD group was 17.75 (SD = 4.5), for the FTD group it was 22.17 (SD = 8.22), and for the CN group it was 30. The mean age of the AD group was 66.4 years (SD = 7.9), for the FTD group it was 63.6 (SD = 8.2), and for the CN group it was 67.9 (SD = 5.4).

EEG recordings were acquired using a Nihon Kohden EEG-2100 system with 19 scalp electrodes (Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, PZ, P4, T6, 01, and 02) and two mastoid reference electrodes (A1, A2), placed according to the international 10-20 system. Recordings were conducted in a routine clinical setting with participants seated and eyes closed. Signals were sampled at 500 Hz with a resolution of 10 uV/mm. Recording durations averaged 13.5 min for the AD group (range, 5.1-21.3 min), 12.0 min for the FTD group (range, 7.9-16.9 min), and 13.8 min for the HC group (range, 12.5-16.5 min), yielding total recording times of 485.5 min (AD), 276.5 min (FTD), and 402.0 min (HC).

The study was approved by the Scientific and Ethics Committee of AHEPA University Hospital, Aristotle University of Thessaloniki, under protocol number 142/12-04-2023. Informed consent was obtained from all 5 subjects involved in the study, and all participants were anonymized, and personal information has not been disclosed, following GDPR restrictions.

The dataset contains the following: (1) The dataset description.json file that contains information regarding the authors of the dataset, the acknowledgment of the research project that made this work possible, the DOI, the BIDS version of the dataset, the license under which it is published, and the ethics approval statement. (2) The participants.json file contains attributes of the participants such as ID, gender, age, diagnosis group, and MMSE score. (3) The participants.tsv contains a tab-separated file with participant information. (4) A folder system named sub-0XX for each respective participant that contains their EEG recording information such as placement scheme, reference electrode, channel count, sampling frequency, and recording duration. The folder also contains EEG recordings in a .set file format, which is one of four BIDS-allowed EEG formats. Finally, it is worth mentioning that files (2) and (3) contain the same information; thus, researchers do not need to examine both files.

**Cleaning**

EEG recordings are contaminated by non-neural artifacts that obscure underlying brain activity and must be attenuated prior to analysis to ensure reliable model performance [13]. Many of these artifacts are not readily identifiable through visual inspection and instead require signal processing-based detection and removal [14]. The most common of these artifacts are line noise, muscle movement, brain and eye artifacts, and others caused by recording equipment [15]. Although preprocessing is essential for cleaning EEG signals before feature extraction and classification, there is no universally optimal algorithm or standardized method for all machine learning tasks. The impact of filtering, rereferencing, and artifact removal varies between datasets and models; researchers must experiment with different combinations to achieve the desired performance depending on the experimental context. Moreover, reviews of artifact removal methods conclude no single algorithm completely removes all artifact types, and no quality metric exists to assess preprocessing pipelines across studies, necessitating researchers to select and tune preprocessing steps based on dataset and model [16], [17].

The EEG preprocessing steps are as follows. First, signals were put through a Butterworth band-pass filter with a frequency range of 0.5 to 45 Hz [18]. Then, the Artifact Subspace Reconstruction (ASR) routine, an automatic artifact rejection method that can remove transient or high-amplitude signal components, indicates and separates dirty data periods that exceed the maximum acceptable 0.5-second window and standard deviation of 17 (considered a conservative window). The signal is then reconstructed based on the remaining components [19], [20]. Finally, the ICA method (RunICA algorithm) was used to convert the 19 EEG signals to 19 independent components. ICA components that were classified as "eye artifacts" or "jaw artifacts" by the automatic classification routine "ICLabel" in the EEGLAB platform were automatically rejected, and the EEG signal was reconstituted from the remaining components [21]. Although patients were recording in a resting state, it is worth noting that eye artifacts were still found in some EEG recordings.

**Segmentation**

The EEG recordings used in this study are relatively long, with durations of up to 21.3 min, which complicates direct feature extraction. Shorter epochs are typically employed for classification on EEG-based data, in the context of AD and FTD studies, because of their effectiveness for observing cognitive defects [22], [23]. However, there is no universally adopted epoch size; it is essential to test varying window lengths to identify the optimal size for each dataset and task [24]. To address this, a sliding window approach was applied to segment the continuous signals into shorter, fixed-length epochs suitable for downstream processing.

The literature expresses that signals can be divided into epochs ranging from a few to 30s and can include overlap [25]. Based on this, preprocessed signals are segmented into 30s with 50% overlap [26]. A relatively high degree of overlap has been associated with improved classification performance by increasing the effective number of training samples while preserving temporal continuity [27]. Furthermore, segmenting the data with sufficient overlap is necessary to minimize data loss and provide the model a large enough data pool to train on, without introducing redundant or noisy information that may negatively impact performance. Without overlap, segments are seen as independent and may lose transitional information.

**Feature Extraction**

The primary goal of feature extraction is to identify meaningful and distinctive features from preprocessed signals and to create a feature vector or matrix. It is intended to be performed after segmenting rather than on the full EEG signal, which can improve classification accuracy. Creating these feature vectors or matrices reduces the data size, thereby increasing training time, and enhances the model's accuracy.

A wide range of EEG-derived biomarkers has been explored in machine learning studies for the automated detection of Alzheimer's disease and for discriminating between dementia subtypes, including AD and frontotemporal dementia. These biomarkers can be time-domain (statistical) features [22], spectral features using Welch's spectral analysis [28], complexity features using Hjorth parameters and spectral entropy [29], [30], or coherence analysis features such as spectral coherence [31]. In the present study, time-frequency representations obtained via the short-time Fourier transform (STFT) were extracted and analyzed, as several recent studies have demonstrated the utility of STFT spectrograms for Alzheimer's disease classification [6], [32], [33].

STFT spectrograms are generated by applying a sliding time window, typically with overlap, to the signal and computing the Fourier transform of each resulting time segment. This windowing process determines the trade-off between temporal and frequency resolution in the resulting time-frequency representation. Narrower windows provide higher temporal resolution, whereas wider windows yield improved frequency resolution [34], [35]. Therefore, the parameters for an STFT spectrogram are the window function, such as Hann and Gaussian, window size, frequency samples, and degree of overlap. The magnitude is calculated by taking the absolute value to produce the final spectrogram output [36], [37]. Taking the magnitude also converts the complex numbers from the Fourier coefficients into real non-negative numbers, to ensure compatibility with modern machine learning frameworks without compromising on pertinent time-frequency feature patterns. The output would be a time-frequency matrix with the number of frequency bins in the y-axis or height, and time steps in the x-axis or width of the matrix [38]. In summary, the chosen parameters are 30s epochs with 50% overlap to segment the preprocessed EEG signal, after which STFT spectrograms were computed with a 256 sample size and 128 time step, based on a sampling rate of 500 Hz.

The STFT is computed by the integral:

$$X(\tau, \omega) = \int_{-\infty}^{\infty} x(t) w(t-\tau) e^{-i\omega t} dt$$

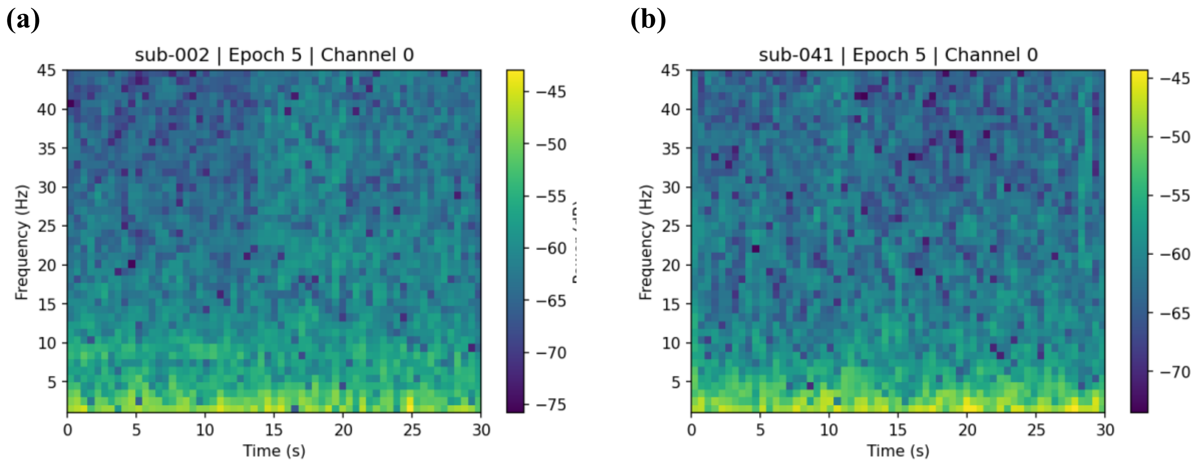
where $X(\tau, \omega)$ represents the STFT result, outputting the signal's frequency ($\omega$) for a specific time window ($\tau$).

Following the STFT, the resulting time-frequency matrix $X(m, k)$ must be processed to eliminate complex-valued coefficients by taking the magnitude to analyze the signal energy. The magnitude is computed as:

$$|X(m, k)| = \sqrt{\mathrm{Re}\big(X(m, k)\big)^2 + \mathrm{Im}\big(X(m, k)\big)^2}$$

where $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ are the real and imaginary components of the complex STFT output.

After computing the STFT spectrogram for each epoch, we restricted the frequency range from 0.5 to 45 hz to focus on physiologically meaningful EEG activity. This band encompasses the canonical EEG frequency ranges (delta, theta, alpha, beta, and gamma) associated with cognitive and pathological processes, while attenuating slow baseline drifts below 0.5 Hz and high-frequency components dominated




by line noise and electromyographic artifacts. Similar band-limiting strategies have been widely adopted in prior EEG studies to improve signal quality and analytical robustness. [39], [28]

*Fig. 2. Displayed are images of Alzheimer's patient's spectrograms (a) and healthy control (b) after applying a frequency mask between 0.5 and 45 Hz.*

**Data Normalization**

After extracting STFT spectrograms, two data transformation methods were evaluated: min-max normalization and z-score standardization. Min-max normalization scales the data output range from $A_{min}$ to $A_{max}$ while standardization transforms the data by centering it around the mean and scaling based on the standard deviation. Data transformation plays an important role in ensuring consistency and comparability of features. This step ensures that the model learns patterns relevant to the task, rather than being influenced by scale differences in the input data. In our approach, we applied scaling methods across epochs to ensure each epoch contributed equally to model training, preventing differences from different segments from dominating the learned representations and model outcome. [23], [40], [41], [42]. After testing both data transformation methods, z-score standardization had better accuracy against min-max normalization (90.75% vs. 87.91% accuracy, respectively).

Z-Score Standardization is calculated as:

$$z = \frac{x - \mu}{\sigma}$$

The Min-Max normalization is calculated as:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

**Classification**

This section discusses the proposed CSA-net and the algorithms employed to optimally tune the model.

**Model**

The CSA-net model first receives the input $X_{\mathrm{i}} \in \mathbb{R}^{(B \times C \times H \times W)}$, where B denotes the batch size for the neural network, C denotes the number of channels, and H and W denote the height and width of the STFT spectrogram, respectively. To enable two-dimensional convolutional processing, the input tensor is reshaped along the channel dimension, thereby collapsing the height and channel dimensions, resulting in a tensor shape of $X_{\mathrm{i}} \in \mathbb{R}^{(B \times (C \cdot H) \times W)}$. The reshaped tensor is used as the input for the convolutional block, which includes 2 two-dimension convolutional layers, instance normalization, Relu activation and dropout followed by a shuffle attention block. Finally, the output is flattened along the channel dimension, resulting in a tensor shape of $X_{\mathrm{i}} \in \mathbb{R}^{B \times (C \cdot H \cdot W)}$ to be used in the feed-forward network (FFN) layer, which determines the class of the input. The detailed model architecture is summarized in Figure 3.

| Layer | Type | Input | Parameter | Output |
|---|---|---|---|---|
| A1 | Input | | | [B, 19, 45, 60] |
| A2 | Reshape | A1 | | [B, 19*45, 60] |
| C1 | Conv2d | A2 | Kernel = [5,5] | [B, 32, 853, 53] |
| | Conv2d | C1 | Kernel = [5,5] | |
| | InstanceNorm2d | | Features = 128 | |
| | ReLU | | | |
| | MaxPool2d | | | [B, 32, 426, 26] |
| | Dropout | | 0.5 | |
| C2 | Conv2d | C1 | Kernel = [5,5] | [B, 128, 419, 39] |
| | Conv2d | C2 | Kernel = [5,5] | |
| | InstanceNorm2d | | Features = 128 | |
| | ReLU | | | |
| | MaxPool2d | | | [B, 128, 209, 19] |
| | Dropout | | 0.5 | |
| CH1 | torch.chunk | C2 | chunk = 128 / g | [B, 128 / g, 209, 19] |
| K1, K2 | torch.chunk | CH1 | chunk = 128 / 2g | [B, 128 / 2g, 209, 19] |
| GAP | nn.AdaptiveAvgPool2d | K1 | size = [1,1] | [B, 128 / 2g, 1, 1] |
| C3 | GroupNorm | K2 | Layers = 128 / 2g | [B, 128 / 2g, 209, 19] |
| | Conv2d | | Kernel = [3,3], Padding = Kernel // 2, Stride = [1,1] | |
| CC1 | torch.concat | K1, K2 | | [B, 128 / g, 209, 19] |
| CC2 | torch.concat | CC1 | | [B, 128, 209, 19] |
| SA | torch.channel_shuffle | CC2 | | [B, 128, 209, 19] |

*Fig. 3. The Architecture of CSA-Net.*

Early stoppage was performed to determine the best number of epochs (number of epochs represents the number of times data moves within the model in a forward and backward direction) [43]. The dataset was partitioned into training, validation, and test sets in a 70%, 15%, and 15% split, respectively, with model performance on the validation set evaluated after each epoch. Training was terminated when the validation loss failed to improve for three consecutive epochs, at which point the model parameters corresponding to the lowest validation loss were retained.

**Convolutional block**

Convolutional layers form the core computational mechanism of convolutional neural networks, enabling systematic feature extraction across images and matrices. Each convolution operation applies a learnable kernel to local regions of the input STFT spectrogram, computes element-wise products, and aggregates the results to produce an output value at the corresponding spatial location. Repeating this operation across the entire input yields a feature map that encodes localized spectro-temporal patterns relevant for downstream processing [44].

To capture local spectro-temporal patterns while preserving signal structure, two convolutional layers were employed with a kernel size of $5 \times 5$, stride = 1, and padding = 1. This configuration enables the extraction of localized frequency-time relationships from the input matrix while ensuring that boundary regions are processed consistently with central regions. The use of kernels larger than a single time or frequency point allows the network to model interactions across neighboring frequency bands and adjacent time steps. Collectively, these convolutional operations reduce the effective dimensionality of the representation for subsequent layers while exposing salient spectro-temporal features.

**Instance Normalization**

EEG recordings exhibit significant variability across subjects due to differences in recording sessions and acquisition hardware. This variability introduces domain-specific shifts in signal distributions that can negatively impact the generalization performance of deep learning models [45]. Conventional normalization methods such as batch normalization or layer normalization can remove subject-specific or clinically relevant characteristics or introduce data leakage across patients, thereby reducing the discriminative power of the model [46]. To mitigate these effects, instance normalization was applied to preserve epoch-level feature structure while maintaining discriminative capacity across classes and subjects. [47], [48].

**Shuffle Attention**

Attention mechanisms, motivated by principles of the human visual system, enable neural networks to selectively emphasize informative components of the input while suppressing less relevant features [49]. This narrow focus and biasing of pertinent information facilitates the understanding of the model to create more accurate predictions. In natural language processing, attention allows models to weight words or phrases according to their contextual importance, whereas in image classification it highlights spatial regions or feature channels that contribute most strongly to the prediction [50].

Shuffle attention integrates two widely used attention paradigms in computer vision - channel attention and spatial attention - to enhance feature representations through complementary aggregation and transformation operations [51]. In this architecture, the input feature map is partitioned into subgroups, within which channel and spatial attention are jointly applied. The resulting features are then aggregated and concatenated, followed by a channel shuffle operation that promotes information exchange across feature groups.

**Grouping**

Given an input feature map $X \in \mathbb{R}^{C \times H \times W}$ where $C, H, W$ denote the channel, height, and width, respectively, the shuffle attention module partitions $X$ into $G$ groups along the channel dimension, yielding sub-tensors $X = [X_1, \ldots, X_g]$, $X_k \in \mathbb{R}^{\frac{C}{G} \times H \times W}$. Each group $X_k$ is then evenly divided along the channel dimension into two subcomponents, generating $X_{k1}, X_{k2} \in \mathbb{R}^{\frac{C}{2G} \times H \times W}$. This partitioning enables the parallel application of channel attention and spatial attention within each group, allowing the model to capture complementary information related to feature importance ("what") and spatial relevance ("where").

$$X_k = [X_{k1}, X_{k2}], \; where \; X_{k1}, X_{k2} \in \mathbb{R}^{\frac{C}{2G} \times H \times W}$$

**Channel Attention**

To capture channel-specific information, global average pooling is applied to compute channel-wise summary statistics, which are then passed through a fully connected layer to generate adaptive channel weights. A sigmoid activation maps these weights to the [0, 1] range, and the resulting coefficients are applied multiplicatively to the original feature map $X_{k1}$.

$$V_{k1} = Sigmoid\left(FC\left(AvgPool(X_{k1})\right)\right)$$
$$Z_{k1} = X_{k1} \odot V_{k1}$$

**Spatial Attention**

Spatial attention complements channel attention by emphasizing where informative regions are located within the feature map. Group normalization is first applied to $X_{k2}$ to generate spatial-wise statistics, followed by a convolutional layer with a 3 × 3 kernel to model local spatial dependencies. A sigmoid activation is then applied to generate spatial attention weights, which are multiplied element-wise with the original feature map $X_{k2}$.

$$M_{k2} = Sigmoid\left(Conv_{3\times3}\left(GN(X_{k2})\right)\right)$$
$$Z_{k2} = X_{k2} \odot M_{k2}$$

**Concatenation and Channel Shuffling**

The outputs of the channel and spatial attention modules are concatenated to form a processed feature group $Z_k$. All $Z_k$ groups are then concatenated along the channel dimension to reconstruct a feature map with the same dimensionality as the original input $X$ . A final channel shuffle operation is applied to facilitate information exchange across groups.

$$Z_k = [Z_{k1}, Z_{k2}]$$
$$Z_{out} = CS([Z_1, Z_{2,} \dots Z_G])$$

Unlike approaches that insert attention modules after each convolutional layer, the shuffle attention mechanism was applied only after the final convolutional block. Applying attention at every stage can increase computational and memory overhead while providing limited performance gains, particularly when early-layer features are noisy and weakly semantic. Restricting attention to higher-level feature maps allows the model to emphasize salient spectro-temporal patterns extracted by the convolutional layers, thereby enhancing discriminative representations with reduced complexity [52].

**Feed Forward Layer**

For classification, the flattened feature representation was passed to a fully connected layer followed by a sigmoid cross-entropy activation to produce binary class predictions. Batch size was set to 64, learning was 0.0004 and L2 regularization weight decay was set to 0.0001. Model optimization was performed using AdamW, in which weight decay is decoupled from the gradient-based parameter updates. This decoupling prevents interference with adaptive learning rates and provides more consistent regularization compared with standard Adam, supporting improved generalization. [53].

## *Results and Discussion*

**Experimental Setup**

The time-frequency and signal processing steps, including, EEG cleaning, segmentation and feature extraction were implemented in Python 3.12 MNE library. The deep learning model used was trained and validated with PyTorch 2.4, and the evaluation functions were implemented using the Scikit-Learn library.

**Results**

Raw EEG recordings acquired in a routine clinical setting were denoised, segmented into overlapping epochs, and finally transformed into time-frequency representations using the short-time Fourier Transform for binary classification of Alzheimer's Disease against healthy patients. The dataset was then partitioned at the subject level into training, validation, and test sets using a 70%, 15%, and 15% split, respectively, to prevent data leakage. Data leakage occurs when information from the same subject appears across multiple data splits, leading to inflated performance estimates. In EEG-based Alzheimer's classification, random epoch-wise splitting introduces patient-level (subject-wise) leakage because epochs from the same individual are highly correlated. To mitigate this issue, all splits were performed at the patient level, ensuring that no epochs from a given subject were shared across training, validation, or testing sets [5]. Young et al. [54] found that data leakage is the most powerful and recurrent source of performance inflation in deep-learning (DL) studies of Alzheimer's disease (AD). Deep learning models often claim high accuracy, exceeding 95%, yet fail in clinical translation because neural networks excel at detecting subtle patterns within features. When splitting the dataset by epoch from the same patient, the model is more likely to memorize patient-specific signatures rather than disease-related features. In essence, the model remembers patients rather than pathology, a form of overfitting that produces inflated results but has zero clinical utility. Yagis et al. [55] evaluated a model with the same dataset under two different validation methods: slice-wise splitting and patient-wise splitting, yielding 94% and 66% accuracy, respectively. This 28-percentage-point difference is attributed solely to the data split. These inflated performance metrics can cause serious clinical implications by creating unrealistic expectations among clinicians, patients, funding bodies and an erosion of trust in AI solutions following real-world failures. Healthcare systems may prematurely invest in AI tools to assist in diagnosis that ultimately fail when deployed on real-world data, while patients and families desperate for early diagnosis may be given false hope about the realistic capabilities of AI-assisted diagnosis. This ultimately undermines the trust and legitimacy of AI advancements and predictions and delays the implementation of these tools in a clinical setting.

**Table 1: Model Evaluation Metrics**

| Metric | Equation |
|---|---|
| Accuracy | $\frac{TP + TN}{TP + TN + FP + FN}$ |
| Recall | $\frac{TP}{TP + FN}$ |
| Precision | $\frac{TP}{TP + FP}$ |
| F1 | $\frac{TN}{TN + FP}$ |
| Sensitivity | $2 \times \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$ |

Using subject-wise validation, the proposed model achieved an accuracy, recall, precision, F1 score, and specificity of 90.87%, 89.47%, 88.98%, 89.23%, and 92.36%, respectively. A confusion matrix is a metric table used to display and evaluate the findings of a classification algorithm by calculating statistical measures [56]. Figure 4 displays a confusion matrix of how well our binary classification (AD vs HC) performed against actual labels, revealing True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) [57]. Although higher accuracies exceeding 95% have been reported in studies employing epoch-level splitting, such estimates are susceptible to subject-level data leakage. In contrast, the performance obtained here reflects a more conservative but clinically realistic evaluation of model generalization.

**Table 2: Performance Comparison of Machine Learning Models**

| Model | Accuracy | Recall | Precision | F1 | Sensitivity |
|---|---|---|---|---|---|
| PSA-Net | 82.55% | 65.29% | 91.15% | 76.08% | 92.30% |
| CBAM | 85.71% | 88.71% | 79.90% | 84.07% | 83.50% |
| Coordinate | 87.47% | 85.12% | 85.36% | 85.24% | 89.20% |
| CNN + LSTM | 84.06% | 70.78% | 89.58% | 79.08% | 93.9% |
| 2d CNN | 84.24% | 80.25% | 82.28% | 81.25% | 87.20% |
| **CSA-NET** | **90.87%** | **89.47%** | **88.98%** | **89.23%** | **92.36%** |

Early stopping was applied to mitigate overfitting and retain the model corresponding to the lowest validation loss. Validation loss was monitored after each epoch with a patience of three epochs. As shown in Figure 5, training loss decreased monotonically, indicating effective optimization of the model parameters. Validation loss initially decreased in parallel with training loss, suggesting adequate generalization; however, after 11 epochs, validation loss plateaued while training loss continued to decrease, indicating the onset of overfitting [58]. The limited number of training epochs is consistent with the modest dataset size and substantial inter-subject variability typical of biomedical EEG data.

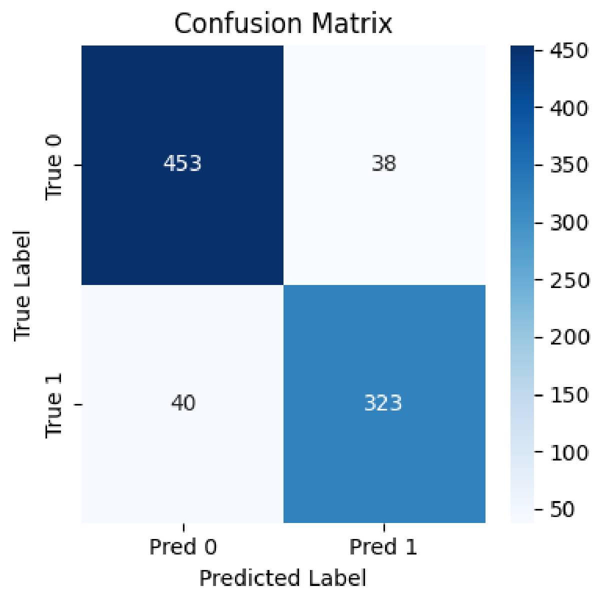



*Fig. 4. Confusion matrix with the x-axis representing predicted labels and the y-axis representing the true or actual label.*

To assess the effectiveness of the proposed approach, the model was compared with several established architectures and attention mechanisms, as summarized in Table 3. Shuffle attention was evaluated against CBAM (convolutional block attention mechanism) [10], CA (coordinate attention) [59], PSA-net (point-wise spatial attention network) [60], along with a convolutional neural network with a long-short-term-memory (CNN-LSTM) module [61] to account for sequential dependencies in STFT spectrograms. Hyperparameters, including learning rate, early stopping criteria, L2 weight decay, and dropout rate, were held constant across all experiments. In addition, an ablation experiment using a two-dimensional convolutional neural network without attention resulted in reduced performance, supporting the contribution of the shuffle attention module.

The leave-one-subject-out (LOSO) validation strategy was used to further evaluate the model's ability to generalize to unseen individuals. Under this scheme, data from one subject were held out for testing while the remaining subjects' data were used for training, and this process was repeated iteratively for each subject. Performance metrics were then aggregated across all folds, providing a direct assessment of subject-level generalization [62]. Using LOSO validation, the model achieved an accuracy of 81.43%, recall of 83.85%, precision of 81.21%, and F1 score of 84.71%. The performance gap observed between holdout validation and LOSO cross-validation reflects the substantial inter-subject variability inherent in EEG data and the associated challenge of robust subject-level discrimination.

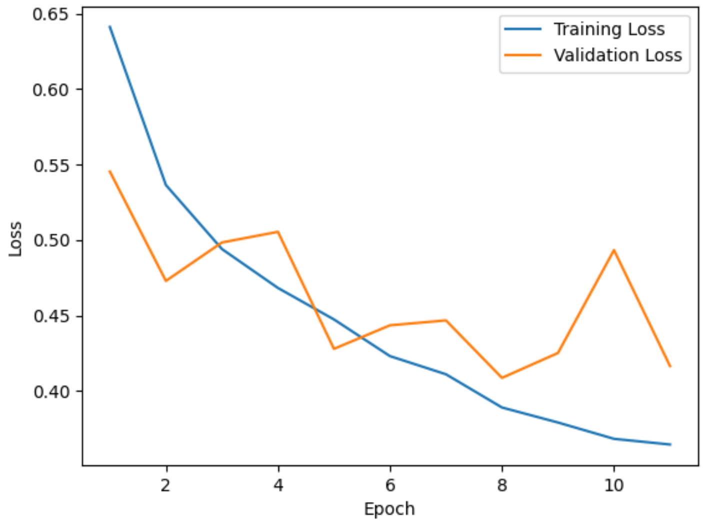



*Fig. 5. Training and validation loss over training epochs.*

**Discussion**

The proposed methodology integrates a shuffle attention mechanism at the output of the convolutional feature extractor to discriminate between Alzheimer's disease (AD) and healthy control (HC) subjects. The workflow begins with raw EEG recordings acquired under routine clinical conditions, with no reported dementia-related comorbidities. Signals were denoised using Artifact Subspace Reconstruction and independent component analysis to remove non-neural artifacts. The cleaned recordings were segmented into 30s epochs with 50% overlap, after which short-time Fourier transform (STFT) spectrograms were computed and band-limited to 0.5-45 Hz to obtain time-frequency representations. A deep neural network comprising two convolutional layers, a shuffle attention module, and a fully connected classifier was then trained. Model performance was evaluated using both leave-one-subject-out (LOSO) cross-validation and a 70/15/15 subject-wise holdout split.

Prior studies have investigated EEG-based dementia detection using a range of machine learning approaches [63]. The most advanced and recent studies usually propose a deep neural network architecture that first processes the time-domain signal through a time-frequency transform such as a Wavelet Transform [64] or Continuous Wavelet Transform [65] and then utilizes the capabilities of a Convolutional layer to extract salient features or neural networks such as autoencoders to efficiently learn data representations while down-sampling [66]. Although attention mechanisms have been incorporated in some studies to enhance feature discrimination [65], [67], many rely on epoch-wise or slice-wise data partitioning, which introduces subject-level data leakage and limits clinical interpretability. To the best of our knowledge, this study is the first to combine a shuffle attention mechanism with a convolutional neural network under a strictly enforced subject-wise validation protocol. By emphasizing salient time-frequency features while suppressing subject-specific noise patterns, the proposed approach jointly addresses feature selection and validation rigor, contributing to improved robustness and potential clinical relevance. Although some studies report classification accuracies exceeding 95%, such performance is typically obtained using epoch-wise or slice-wise data splitting and is therefore susceptible to subject-level data leakage. In contrast, the more conservative performance observed here better reflects real-world clinical conditions, in which models are evaluated on previously unseen patients. The performance reduction observed under LOSO cross-validation highlights the substantial inter-subject variability inherent in EEG data and reinforces the necessity of subject-wise validation in neuropsychological deep learning studies.

**Limitations**

Several limitations should be acknowledged. First, although the dataset is sufficient for methodological evaluation, its size limits the model's ability to generalize broadly. Validation on larger and more diverse cohorts is required to establish robustness across populations. Second, the proposed convolutional architecture depends on a fixed electrode configuration. EEG recordings are acquired using predefined montages, and the learned convolutional kernels implicitly assume a consistent channel arrangement and input dimensionality. Consequently, models trained on a specific electrode setup may not generalize to datasets acquired with different electrode configurations, restricting applicability across heterogeneous EEG systems.

An additional limitation concerns the choice of window segmentation (30s 21 with 50% overlap), which differs from the shorter epoch durations (typically < 5 s) commonly used in EEG-based machine learning studies [68], [69]. Shorter windows are often adopted due to the classifier's inability to capture long-range dependencies and the size of the dataset, which necessitates the need to generate a lot of training samples from a small duration of EEG recordings. In the present study, longer temporal context was accommodated through the use of 5 × 5 convolutional kernels and multiple convolutional layers, which expand the receptive field and enable the extraction of longer-range spectro-temporal features.

Finally, although this work emphasizes algorithmic development, clinical applicability requires validation on larger datasets collected across multiple medical centers. Future studies should include more diverse patient populations and additional diagnostic categories, such as mild cognitive impairment and other neurological conditions, to better reflect real-world clinical scenarios. Addressing these limitations is essential for improving cross-dataset robustness and advancing the deployment of EEG-based deep learning models in clinical practice.

## *Conclusion*

This study evaluated the effectiveness of integrating a shuffle attention mechanism into a convolutional deep learning architecture for binary classification of Alzheimer's disease versus healthy controls. Model performance was assessed using a clinical EEG dataset acquired at AHEPA General Hospital of Thessaloniki, Greece, under a strictly enforced subject-wise validation protocol. The proposed approach achieved an accuracy of 90.87%, recall of 89.47%, precision of 88.98%, F1 score of 89.23%, and specificity of 92.36%, indicating effective extraction of discriminative time-frequency features from EEG signals. These results contribute to the growing literature on EEG-based Alzheimer's disease classification using machine learning methods and highlight the importance of combining attention mechanisms with rigorous validation strategies.

**Acknowledgements.** This work was supported by Open Labs, an open-source pedagogical framework and continuous-feedback software designed to foster a community of creative thinkers and researchers. We would like to express our eternal gratitude to Decoded Brain, an active student-led research organization of scientists and engineers whose ecosystem and creation of Open Labs made the development and execution of this research possible.

**Disclosures.** The authors declare no conflicts of interest.

## References


[1] GBD 2019 Dementia Forecasting Collaborators et al., “Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: An analysis for the Global Burden of Disease Study 2019,” Lancet Public Health, vol. 7, no. 2, pp. 105–125, 2022 Jan 6. doi: 10.1016/S2468-2667(21)00249-8.

[2] C. R. J. Jr et al., “Revised criteria for diagnosis and staging of Alzheimer’s disease: Alzheimer’s Association Workgroup,” Alzheimer’s & Dementia, vol. 20, no. 8, pp. 5143–5169, 2024, 2024 Jun 27. doi: 10.1002/alz.13859.

[3] J. Jeong, “EEG dynamics in patients with Alzheimer’s disease,” Clinical Neurophysiology, vol. 115, no. 7, pp. 1490–1505, 2004. doi: 10.1016/j.clinph.2004.01.001.

[4] M. H. F. Fischer, I. C. Zibrandtsen, P. Høgh, and C. S. Musaeus, “Systematic review of EEG coherence in Alzheimer’s Disease,” Journal of Alzheimer’s Disease, vol. 91, no. 4, pp. 1261–1272, 2023. doi: 10.3233/JAD-220508

[5] G. Brookshire et al., “Data leakage in deep learning studies of translational EEG,” Frontiers in Neuroscience, vol. 18, p. 1 373 515, 2024. doi: 10.3389/fnins.2024.1373515.

[6] M. A. Bravo-Ortiz et al., “SpectroCVT-Net: A convolutional vision transformer architecture and channel attention for classifying Alzheimer’s disease using spectrograms,” Computers in Biology and Medicine, vol. 181, p. 109 022, 2024. doi: 10.1016/j.compbiomed.2024.109022.

[7] M. Khosravi, H. Parsaei, K. Rezaee, and M. S. Helfroush, “Fusing convolutional learning and attention-based Bi-LSTM networks for early Alzheimer's diagnosis from EEG signals towards IoMT,” Scientific Reports, vol. 14, p. 26 002, 2024. doi: 10.1038/s41598-024-77876-8.

[8] F. D. Pup, A. Zanola, L. F. Tshimanga, A. Bertoldo, L. Finos, and M. Atzori, “The role of data partitioning on the performance of EEG-based deep learning models in supervised cross-subject analysis: A preliminary study,” Computers in Biology and Medicine, vol. 196, p. 110 608, 2025, 2025 Jul 1. doi: 10.1016/j.compbiomed.2025.110608.

[9] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 7132–7141. doi: 10.1109/CVPR.2018.00745.

[10] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “CBAM: Convolutional block attention module,” Proceedings of the European conference on computer vision (ECCV), 2018, pp. 3–19. [Online]. https://arxiv.org/abs/1807.06521.

[11] C. C. Bell, “DSM-IV: Diagnostic and statistical manual of mental disorders,” pp. 272–828, Sep. 14, 1994. doi:10.1001/jama.1994.03520100096046.

[12] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M. Stadlan, “Clinical diagnosis of Alzheimer’s disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer’s Disease,” pp. 939–944, 1984. doi: 10.1212/wnl.34.7.939.

[13] E. Jovi˘ci'c, A. Jovi'c, and M. Cifrek, “Impact of EEG signal preprocessing methods on machine learning models for affective disorders,” in Proceedings of the 47th ICT and Electronics Convention (MIPRO 2024), K. Skala and V. Mornar, Eds., Croatian Society for Information, Communication and Electronic Technology – MIPRO, Rijeka, Croatia, 2024, pp. 1374–1379. doi:10.1109/MIPRO60963.2024.10569172.

[14] I. Stancin, M. Cifrek, and A. Jovic, “A review of EEG signal features and their application in driver drowsiness detection systems,” p. 3786, May 30, 2021. doi: 10.3390/s21113786.

[15] R. Ranjan, B. C. Sahana, and A. K. Bhandari, “Ocular artifact elimination from electroencephalography signals: A systematic review,” Biocybernetics and Biomedical Engineering, vol. 41, no. 3, pp. 960–996, 2021. doi: 10.1016/j.bbe.2021.06.007.

[16] R. Kessler, A. Enge, M. A. Skeide, et al., “How EEG preprocessing shapes decoding performance,” Communications Biology, vol. 8, no. 1, p. 1039, 2025, PMCID: PMC12246244.

[17] M. K. Islam, A. Rastegarnia, and Z. Yang, “Methods for artifact detection and removal from scalp EEG: A review,” Neurophysiologie Clinique / Clinical Neurophysiology, vol. 46, no. 4-5, pp. 287–305, 206. doi: 10.1016/j.neucli.2016.07.002

[18] A. M. Alghamdi, M. U. Ashraf, A. A. Bahaddad, K. A. Almarhabi, W. A. Al Shehri, and A. Daraz, “A novel approach hybrid of ensemble learning and 3-d CNN mechanism: Early-stage diagnosis of Alzheimer’s disease using EEG signals,” Scientific Reports, vol. 15, p. 35 893, 2025. doi: 10.1038/s41598-025-19727-8.

[19] P. Anders, H. M¨uller, N. Skjæret-Maroni, B. Vereijken, and J. Baumeister, “The influence of motor tasks and cut-off parameter selection on artifact subspace reconstruction in EEG recordings,” Medical & Biological Engineering & Computing, vol. 58, pp. 2673–2683, 2020. doi: 10.1007/s11517-020-02252-3.

[20] M. Plechawska-W'ojcik, P. Augustynowicz, M. Kaczorowska, E. Zabielska-Mendyk, and D. Zapała, “The influence assessment of artifact subspace reconstruction on the EEG signal characteristics,” Applied Sciences, vol. 13, no. 3, p. 1605, 2023. doi: 10.3390/app13031605.

[21] A. Delorme and S. Makeig, “EEGLab: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis,” Journal of Neuroscience Methods, vol. 134, no. 1, pp. 9–21, 2004. doi: 10.1016/j.jneumeth.2003.10.009.

[22] A. Miltiadous et al., “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, 2021. doi: 10.3390/diagnostics11081437.

[23] B. Mouazen et al., “Transparent EEG analysis: Leveraging autoencoders, Bi-LSTMs, and SHAP for improved neurodegenerative diseases detection,” Sensors, vol. 25, no. 18, p. 5690, 2025. doi: 10.3390/s25185690.

[24] K. D. Tzimourta et al., “EEG window length evaluation for the detection of Alzheimer’s disease over different brain regions,” Brain Sciences, vol. 9, no. 4, p. 81, 2019. doi: 10.3390/brainsci9040081.

[25] M. Bedoin, B. Dorizzi, J. Boudy, K. Kinugawa, and N. Houmani, “Multi-scale probabilistic score fusion for enhancing Alzheimer’s disease detection using EEG,” in Proceedings of the 18th International Joint Conference on Biomedical Engineering Systems and Technologies – BIOSIGNALS, International Joint Conference on Biomedical Engineering Systems and Technologies, SciTePress, 2025, pp. 741–751. doi: 10.5220/0013169700003911.

[26] L. Pu, K. M. Lion, M. Todorovic, and W. Moyle, “Portable EEG monitoring for older adults with dementia and chronic pain: A feasibility study,” Geriatric Nursing, vol. 42, no. 1, pp. 124–128, 2021. doi: 10.1016/j.gerinurse.2020.12.008

[27] B. S. Falih, M. K. Sabir, and A. Aydın, “Impact of sliding window overlap ratio on EEG-based ASD diagnosis using brain hemisphere energy and machine learning,” Applied Sciences, vol. 14, no. 24, p. 11 702, 2024. doi: 10.3390/app142411702.

[28] Y. Senkaya, C. Kurnaz, and F. Ozbilgin, “Enhancing Alzheimer’s diagnosis with machine learning on EEG: A spectral feature-based comparative analysis,” Diagnostics, vol. 15, no. 17, p. 2190, 2025. doi: 10.3390/diagnostics15172190.

[29] M. S¸ eker, Y. ¨Ozbek, G. Yener, and M. S. ¨Ozerdem, “Complexity of EEG dynamics for early diagnosis of Alzheimer’s disease using permutation entropy neuromarker,” Computers in Biology and Medicine, vol. 206, p. 106 116, 2021. doi: 10.1016/j.cmpb.2021.106116.

[30] M. S. Safi and S. M. M. Safi, “Early detection of Alzheimer’s disease from EEG signals using Hjorth parameters,” Journal of Neuroscience Methods, vol. 65, p. 109 000, 2020. doi: 10.1016/j.jneumeth.2020.109000.

[31] B. Oltu, M. F. Aks¸ahin, and S. Kibaro˘glu, “A novel electroencephalography based approach for Alzheimer’s disease and mild cognitive impairment detection,” Biomedical Signal Processing and Control, vol. 63, p. 102 223, 2020. doi: 10.1016/j.bspc.2020.102223.

[32] S. Siuly, M. N. A. Tawhid, Y. Li, R. Acharya, M. Tariq Sadiq, and H. Wang, “Investigating brain lobe biomarkers to enhance dementia detection using EEG data,” Cognitive Computation, vol. 17, no. 2, p. 90, 2025. doi: 10.1007/s12559-025-10447-9.

[33] K. Rezaee and M. Zhu, “Diagnose Alzheimer’s disease and mild cognitive impairment using deep Cascade-net and handcrafted features from EEG signals,” Biomedical Signal Processing and Control, vol. 99, p. 106 895, 2025. doi: 10.1016/j.bspc.2024.106895

[34] M. C. Ng, J. Jing, and M. B. Westover, “A primer on EEG spectrograms,” Journal of Clinical Neurophysiology, vol. 39, no. 3, pp. 177–183, 2022. doi: 10.1097/WNP.0000000000000736.

[35] Wikipedia contributors, Short-time fourier transform, https : //en.wikipedia.org/wiki/Short-time_Fourier_transform, Accessed: 2025-12-28, 2025.

[36] O. Biari, M. Teshnehlab, and A. Mansouri, “A deep learning based framework for Alzheimer’s disease diagnosis using EEG signals,” Pattern Recognition Letters, vol. 133, pp. 172–178, 2020. doi: 10.1016/j.patrec.2020.03.003.

[37] S. A. Moti and S. Sanei, “Automated detection of Alzheimer’s disease using brain network nodes and deep learning,” Biomedical Signal Processing and Control, vol. 70, p. 102 960, 2021. doi: 10.1016/j.bspc.2021.102960.

[38] O. Karabiber Cura, H. S. Ture, and A. Akan, “Detection of Alzheimer’s Dementia by Using Deep Time-Frequency Feature Extraction,” Electrica, vol. 24, no. 1, pp. 109–118, 2024. doi: 10.5152/electrica.2023.23029.

[39] C. S. Nayak and A. C. Anilkumar, Normal EEG Waveforms, 1st ed. StatPearls Publishing, 2025, StatPearls; Last Update: August 3, 2025.

[40] P. Arpaia et al., “Assessing the role of EEG biosignal preprocessing to enhance multiscale fuzzy entropy in Alzheimer’s disease detection,” Biosensors, vol. 15, no. 6, p. 374, 2025. doi: 10.3390/bios15060374.

[41] K. V. Saboo et al., “Unsupervised machine-learning classification of electrophysiologically active electrodes during human cognitive task performance,” Scientific Reports, vol. 9, p. 17 390, 2019. doi: 10.1038/s41598-019-53925-5.

[42] D. Eesha, M. Nagaraju, D. Divya, and S. Kashyap Reddy, “Cognitive load classification using feature masked autoencoding and electroencephalogram signals,” in Proceedings of the 3rd International Conference on Futuristic Technology (INCOFT) - Volume 2, INSTICC, SciTePress, 2025, pp. 624–634, ISBN: 978-989-758-763-4. doi: 10.5220/0013599000004664.

[43] B. M. Hussein and S. M. Shareef, “An empirical study on the correlation between early stopping patience and epochs in deep learning,” ITM Web of Conferences, vol. 64, p. 01 003, 2024. doi: 10.1051/itmconf/20246401003.

[44] I. Pacal, D. Karaboga, A. Basturk, B. Akay, and U. Nalbantoglu, “A comprehensive review of deep learning in colon cancer,” Computers in Biology and Medicine, vol. 126, p. 104 313, 2020. doi: 10.1016/j.compbiomed.2020.104003.

[45] A. Apicella, F. Isgr'o, A. Pollastro, and R. Prevete, “On the effects of data normalization for domain adaptation on EEG data,” Computers in Biology and Medicine, vol. 123, p. 106 205, 2023. doi: 10.1016/j.engappai.2023.106205.

[46] Y. Wu and J. Johnson, “Rethinking “batch” in batchnorm,” 2021. doi: 10.48550/arXiv.2105.07576.

[47] N. Francis and G. Vadivu, “ReHA-Net: A ReVIN–hybrid attention network with multiscale convolution for robust EEG artifact removal in brain–computer interfaces,” Scientific Reports, vol. 15, p. 28 855, 2025. doi: 10.1038/s41598-025-28855-0.

[48] D. Ulyanov, A. Vedaldi, and V. Lempitsky, “Instance normalization: The missing ingredient for fast stylization”, arXiv preprint arXiv:1607.08022, 2016. doi: 10.48550/arXiv.1607.08022.

[49] D. Soydaner, “Attention mechanism in neural networks: Where it comes and where it goes,” Neural Computing and Applications, vol. 34, no. 16, pp. 13 371–13 385, 2022. doi: 10.1007 / s00521 -022-07366-3.
[50] B. Kotipalli, “The role of attention mechanisms in enhancing transparency and interpretability of neural network models in explainable AI,” Accessed 2025-12-28, Master’s thesis, Harrisburg University of Science and Technology, 2024.
[51] Q.-L. Zhang and Y.-B. Yang, “SA-Net: Shuffle attention for deep convolutional neural networks,” in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2021, pp. 2235–2239. doi: 10.1109/ICASSP39728.2021.9414568.
[52] N. Kapila, J. Glattki, and T. Rathi, Cnntention, “Can CNNs do better with attention?”, arXiv preprint arXiv:2412.11657, 2024. doi: 10.48550/arXiv.2412.11657.
[53] I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” in Proceedings of the 7th International Conference on Learning Representations (ICLR), 2019. doi: 10.1109/ICASSP39728.2021.9414568
[54] S. Gates, L. Y. Garcia, and A. Salardini, “Data leakage in deep learning for Alzheimer’s disease diagnosis: A scoping review of methodological rigor and performance inflation,” Diagnostics, vol. 15, no. 18, p. 2348, 2025. doi: 10.3390/diagnostics15182348.
[55] E. Yagis et al., “Effect of data leakage in brain MRI classification using 2D convolutional neural networks,” Scientific Reports, vol. 11, p. 22 544, 2021. doi: 10.1038/s41598-021-01681-w.
[56] M. N. J. R and P. R, “Performance analysis of text classification algorithms using confusion matrix,” International Journal of Engineering and Technical Research, vol. 6, 2016.
[57] M. Heydarian, T. E. Doyle, and R. Samavi, “MLCM: Multi-label confusion matrix,” IEEE Access, vol. 10, pp. 19 083–19 095, 2022. doi: 10.1109/ACCESS.2022.3151048.
[58] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016, http : / / www . Deeplearningbook.org.
[59] Q. Hou, D. Zhou, and J. Feng, “Coordinate Attention for Efficient Mobile Network Design,” 2021, pp. 13 713–13 722.
[60] A. Balagopal et al., “PSA-Net: Deep learning based physician style-aware segmentation network for post-operative prostate cancer clinical target volume”, arXiv preprint arXiv:2102.07880, 2021. doi: 10.48550/arXiv.2102.07880.
[61] S. Baldo-J'unior et al., “A bioinspired multimodal CNN-LSTM network for EEG analysis of patients in coma,” Sensors, vol. 25, no. 22, p. 6981, 2025. doi: 10.3390/s25226981.
[62] R. Saini et al., Evaluating the generalizability of EEG-based AI models in Alzheimer’s and dementia diagnosis, medRxiv preprint, 2025. doi: 10.1101/2025.09.10.25334048.
[63] K. D. Tzimourta et al., “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. 05, p. 2 130 002, 2021. doi: 10.1142/S0129065721300023.
[64] K. D. Tzimourta, T. Afrantou, N. Giannakeas, P. Ioannidis, and M. G. Tsipouras, “Analysis of electroencephalographic signals complexity regarding Alzheimer’s disease,” Computers & Electrical Engineering, vol. 76, pp. 198–212, 2019. doi: 10.1016/j.compeleceng.2019.03.018.
[65] F. B. Arikan, D. Cetintas, and M. Yildirim, “A deep learning approach to Alzheimer’s diagnosis using EEG data: Dual-attention and Optuna-Optimized SVM,” Biomedicines, vol. 13, no. 8, p. 2017, 2025. doi: 10.3390/biomedicines13082017.
[66] S. Fouladi, A. A. Safaei, N. Mammone, F. Ghaderi, and M. J. Ebadi, “Efficient deep neural network for classification of Alzheimer’s disease and mild cognitive impairment from scalp EEG recordings,” Cognitive Computation, vol. 14, no. 4, pp. 1247–1268, 2022. doi: 10.1007/s12559-022-10033-3.

[67] D. Truong, M. A. Khalid, and A. Delorme, Deep learning applied to EEG data with different montages using spatial attention, arXiv preprint arXiv:2310.10550, 2023. doi: 10.48550/arXiv.2310.10550.
[68] O. ¨Ozdenizci, S. Eldeeb, A. Demir, D. Erdo˘gmus¸, and M. Akc¸akaya, “EEG-based texture roughness classification in active tactile exploration using adversarial invariant representation learning,” Neural Networks, vol. 142, pp. 312–327, 2021. doi: 10.1016/j.neunet.2021.09.014.
[69] 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, p. 7212, 2020. doi: 10.3390/s20247212.