

EEG-based Alzheimer's disease detection using two-branch multi-scale deep learning with integrated gradients-guided channel selection

Jungpil Shin  · Najmul Hassan · Abu Saleh Musa Miah · Satoshi Nishimura

Received: 18 June 2026 / Accepted: 13 October 2026

© The Author(s) 2026

Abstract

Early diagnosis of Alzheimer's disease (AD) is critical for timely intervention and clinical monitoring. Electroencephalography (EEG) offers a non-invasive, low-cost, and portable alternative to neuroimaging. However, existing traditional and deep learning (DL)-based methods often face challenges related to computational cost, representational complexity, and practical deployment. In this work, we propose Fused-MSCA-2B, a two-branch multi-scale channel-attentive 1D convolutional neural network (CNN) for EEG-based AD detection. The model processes broadband multichannel EEG directly as temporal sequences and combines multi-scale temporal convolution, residual learning, squeeze-and-excitation channel attention, and feature-level branch fusion. Specifically, Branch A uses MSCA blocks with kernel configurations (3,5,9) and (3,5,7) to capture local and medium-range temporal dependencies, whereas Branch B uses (5,9,15) and (3,5,9) to capture broader and higher-level temporal patterns. The branch-level embeddings are fused and passed through dense classification layers to generate AD/HC posterior probabilities. A key component of the proposed model is an explainable AI-guided reduced-channel selection strategy. Integrated Gradients (IG) is used to rank EEG channel relevance from validation data, enabling the EEG montage to be reduced from 19 to 12 channels. We evaluated the proposed model using two benchmark datasets. The IG-guided Top-12 configuration achieved $98.60\% \pm 1.18\%$ recording-level accuracy while reducing the EEG montage from 19 to 12 channels. As a supporting ablation analysis, the Fixed Top-12 configuration achieved $98.40\% \pm 0.73\%$ accuracy, further confirming the effectiveness of the IG-guided reduced-channel selection strategy. The results demonstrate that the proposed XAI-guided Fused-MSCA-2B framework supports robust EEG-based AD detection with a compact and practical electrode configuration, showing strong performance with reduced channel sets.

Keywords Alzheimer's disease · Electroencephalography · EEG · Explainable artificial intelligence · Integrated Gradients · Reduced-channel EEG · 1D convolutional neural network · Channel attention · Recording-level classification

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.



Introduction

Alzheimer's disease (AD) is one of the most prevalent neurodegenerative disorders in the elderly population and is characterized by progressive deterioration of memory, cognition, and functional ability¹. Recent studies have also highlighted glial-cell nutrient sensing and neurobiological mechanisms involved in AD pathogenesis and precision intervention². Since AD-related pathological and functional changes may begin before severe clinical symptoms become apparent, early detection is important for timely intervention, treatment planning, disease monitoring, and patient-specific care^{3,4}. Conventional diagnostic procedures, including neuropsychological assessment, magnetic resonance imaging (MRI), positron emission tomography (PET), and cerebrospinal-fluid biomarker analysis, provide valuable clinical evidence, but they are often expensive, invasive, facility-dependent, or unsuitable for repeated large-scale screening^{5–7}. Recent electroencephalography (EEG) studies in related neurological and cognitive-state recognition tasks have shown the importance of preprocessing, connectivity-based feature extraction, and rigorous validation protocols for improving model reliability and generalization^{8,9}.

EEG has emerged as a promising alternative for AD assessment because it is non-invasive, portable, comparatively inexpensive, and capable of recording neural activity with high temporal resolution^{10,11}. Unlike structural or metabolic imaging, EEG directly captures functional brain dynamics and can provide complementary information for detecting disease-related neural abnormalities^{12,13}. These characteristics make EEG attractive for early-stage AD screening, continuous monitoring, and deployment in clinical or resource-constrained environments.

Despite these advantages, EEG-based AD detection remains a challenging machine-learning problem⁸. EEG signals are highly non-stationary and noisy, and EEG-based classification performance can be affected by preprocessing choices and feature representation^{9,14,15}. Recent EEG-AD research has also emphasized that limited dataset size⁸, inter-subject variability, subject-specific interference, and validation protocols are major factors affecting model generalization on unseen subjects^{16–18}. Therefore, strict subject-wise evaluation is essential; otherwise, models may learn subject-specific signatures rather than disease-related patterns, resulting in overestimated diagnostic performance^{17,18}.

Existing EEG-based AD methods can be broadly divided into handcrafted-feature approaches and deep learning approaches⁸. Conventional handcrafted-feature methods commonly use statistical descriptors, spectral power, relative band power, coherence, phase synchronization, entropy, and complexity measures^{19–21}. Recurrence-based complexity and graph-based connectivity features have also been explored for dementia-related EEG analysis^{22,23}. These features provide useful neurophysiological interpretability and have shown that AD-related EEG changes are associated with EEG rhythm slowing, reduced signal complexity, and altered synchronization among brain regions^{24,25}.

Several computational EEG signal-analysis methods have also been explored for AD detection, including information-theoretic, time-frequency, nonlinear-dynamics, recurrence-based, connectivity-based, and graph-based approaches^{8,26,27}. Earlier entropy-based and automated EEG studies further support the usefulness of signal-complexity and computational features for AD-related EEG analysis^{11,28,29}. Recent computational approaches^{30,31}, such as wavelet coherence, fractal dimension, quadratic entropy, recurrence-complexity analysis, wavelet energy, quantile graphs, and visibility graphs, further show that EEG dynamics contain discriminative information for separating dementia patients from healthy controls^{22,32,33}.

However, these handcrafted and computational methods still rely on predefined mathematical descriptors, selected frequency bands, electrode-specific analysis, feature-selection criteria, and feature-level classification. Their performance and computational efficiency may vary across feature type, signal length, preprocessing pipeline, and recording condition. This motivates compact deep models that can learn discriminative temporal representations directly from preprocessed EEG windows without requiring extensive handcrafted feature engineering.

Recent studies have applied deep learning (DL) approaches for EEG-based AD detection³¹. CNN and transformer-based methods have demonstrated promising performance³⁴. DL-based methods reduce the need for manual feature engineering by learning discriminative representations directly from EEG signals³⁵. DL-based

methods, including CNNs, graph neural networks, transformers, self-supervised models, and hybrid systems, have recently been applied to EEG-based AD detection^{16,36,37}. In addition, optimized DL studies for EEG-based fatigue detection further demonstrate the value of deep neural architectures for neurophysiological signal modeling^{14,38}. Among these approaches, ensemble and three-dimensional CNN methods have shown strong performance by constructing time-channel-frequency EEG representations and combining multiple classifiers. For example, Alghamdi et al.³⁵ integrate ensemble DL with multiple 3D-CNN classifiers and report high AD/healthy-control classification performance on two public EEG datasets. However, such approaches require artificial volumetric EEG representations and ensemble inference, which increase computational cost, memory requirements, and implementation complexity.

Another recent direction is large-scale EEG foundation modeling¹⁶. LEAD introduced a foundation-model pipeline for EEG-based AD detection using channel alignment, frequency alignment, self-supervised contrastive pretraining, unified fine-tuning, subject-independent evaluation, and subject-level voting¹⁶. Although this direction improves robustness against inter-subject variation, foundation-model pipelines typically require large-scale pretraining data, substantial computational resources, and complex training procedures. Therefore, compact architectures that can operate effectively without large pretraining pipelines or heavy ensemble structures remain important for practical EEG-based AD screening.

Standard clinical EEG recordings commonly use 19 scalp channels, but full-channel acquisition may increase preparation time, require trained personnel, and affect patient comfort³⁵. Reduced-channel EEG systems may therefore support faster setup, improved usability, and more portable screening. However, channel reduction must be performed carefully because using held-out test recordings for electrode selection, channel ranking, or reduced-montage optimization can introduce information leakage and lead to overestimated diagnostic performance^{17,18}. Therefore, we performed reduced-channel selection only during training and validation, excluding held-out test recordings from channel ranking and montage construction.

To address these limitations, this study proposes Fused-MSCA-2B, a lightweight two-branch multi-scale channel-attentive 1D CNN for EEG-based AD detection. Unlike 3D-CNN and ensemble-based approaches that transform EEG signals into artificial volumetric representations, the proposed model processes broadband multichannel EEG directly as temporal windows. In addition, an Integrated Gradients (IG) based channel-selection strategy is introduced to construct a compact EEG montage using validation recordings only. The framework is evaluated using recording-level grouped cross-validation, repeated-seed analysis, cross-dataset transfer, and an independent OpenNeuro benchmark. The key contributions of this work are summarized as follows:

- **Fused-MSCA-2B framework:** We propose the Fused-MSCA-2B framework, which integrates a two-branch network for multi-scale feature extraction using a channel-attentive mechanism. This architecture differs from conventional single-branch models and addresses limitations in feature learning for EEG signals, specifically in the context of AD detection.
- **Explainable AI (XAI)-guided reduced-channel selection:** We introduce an explainable-AI-guided reduced-channel selection strategy using IG. This approach identifies the most important EEG channels for the task, reducing the EEG montage from 19 to 12 channels. This selection enhances computational efficiency while maintaining model performance, making the framework more deployable in resource-constrained environments.
- **Performance evaluation using standard metrics:**
 - The Top-12 IG-selected channels in the Fused-MSCA-2B model achieved $98.60\% \pm 1.18\%$ accuracy, demonstrating strong performance in AD detection compared with other state-of-the-art methods.
 - After manually selecting the fixed top 12 channels based on their individual performance, the model achieved $98.40\% \pm 0.73\%$ accuracy, further supporting the importance of channel selection for optimizing model performance.

- In addition to accuracy, the model demonstrated high performance across multiple metrics, including sensitivity ($99.07\% \pm 1.53\%$), specificity ($96.60\% \pm 1.95\%$), F1-score ($99.09\% \pm 0.83\%$), and ROC-AUC (0.9991 ± 0.0020), establishing it as a robust and reliable method for AD detection.
- **Cross-dataset evaluation:** We conducted cross-dataset transfer experiments to examine the influence of source-dataset differences. The results revealed asymmetric transfer performance and threshold calibration, highlighting the need for calibration-aware external validation.
- **Robustness and practical deployability:** The proposed approach achieved high accuracy and showed robustness to inter-subject variability. The reduced-channel model configuration retained most of the performance benefits, making it particularly well-suited for further evaluation in clinical and portable EEG settings. The source code for the proposed framework is publicly available at: [GitHub](#).

The remainder of this paper is organized as follows. Section [Materials and methods](#) describes the EEG datasets, preprocessing steps, segmentation procedure, channel configuration, and external benchmark dataset. Section [Proposed method](#) presents the proposed Fused-MSCA-2B architecture, including the two-branch multi-scale channel-attentive design, feature fusion strategy, and validation-only Integrated Gradients-guided reduced-channel selection. Section [Experimental setup](#) explains the experimental setup, including the recording-level grouped cross-validation protocol, validation strategy, evaluation metrics, and selected channel patterns. Section [Results](#) reports the experimental results, including architecture ablation, reduced-channel analysis, control channel-selection experiments, cross-dataset evaluation, external OpenNeuro benchmarking, and comparison with prior studies. Section [Discussion](#) discusses the main findings, limitations, and generalization considerations. Finally, Section [Conclusion](#) concludes the study and outlines future research directions.

Materials and methods

This study used two publicly available EEG datasets for AD/HC classification. For clarity, the Quantile Graphs for EEG-Based Diagnosis of AD dataset is referred to as Dataset-A, and the Computational Methods of EEG Signals Analysis for AD Classification dataset is referred to as Dataset-B. These two datasets were previously merged in the EDL3DCNN study to form a 140-recording cohort, including 104 AD and 36 healthy control (HC) recordings³⁵. In the present study, we first followed the same merged-cohort setting for comparability. However, because Dataset-A and Dataset-B are related public EEG resources, we performed signal-level duplicate checking before preprocessing, segmentation, model training, channel selection, and evaluation. After removing exact duplicate EEG signals, the revised cohort contained 113 unique EEG recordings, including 88 AD and 25 HC recordings.

Dataset description

Dataset-A³⁹ contains EEG recordings from 48 subjects, including 24 healthy elderly individuals and 24 probable AD patients. The signals were recorded from 19 scalp channels at a sampling rate of 128 Hz under eyes-open and eyes-closed conditions. In the original dataset organization, Groups A–B correspond to HC recordings, whereas Groups C–D correspond to AD recordings. Accordingly, SETA and SETB were treated as HC recordings, while SETC and SETD were treated as AD recordings. Dataset-B⁴⁰ contains 19-channel EEG recordings sampled at 128 Hz from 92 subjects, including 12 HC and 80 AD recordings. The diagnostic labels follow clinical criteria, including NINCDS, ADRDA, and DSM-III-R guidelines. To improve recording-condition consistency, only eyes-closed/blindfolded recordings were used in the present analysis. Table 1 summarizes Dataset-A, Dataset-B, the original merged cohort, and the revised cohort used in this study. The revised cohort refers to the final set of unique EEG recordings retained after signal-level duplicate checking. Although the original dataset descriptions

Table 1 Summary of the public EEG datasets, the original merged cohort, and the revised cohort used in this study.

Dataset/cohort	Sampling rate	Original duration	Channels	Groups and size	Recording condition	Total recordings
Dataset-A ³⁹	128 Hz	8 s	19 scalp channels	24 HC, 24 AD	Eyes-open and eyes-closed	48
Dataset-B ⁴⁰	128 Hz	Not specified	19 scalp channels	12 HC, 80 AD	Eyes-closed/blindfolded EEG	92
Original merged cohort ³⁵	128 Hz	–	19 scalp channels	36 HC, 104 AD	Mixed recording conditions	140
Revised cohort used in this study	128 Hz	1-s analysis windows	19 scalp channels	25 HC, 88 AD	Mixed after duplicate removal	113

report Dataset-A and Dataset-B in terms of subjects, universal participant identifiers are not consistently available across the two related public resources. Therefore, the revised cohort is reported as retained EEG recordings in the present study.

Signal-level duplicate checking

To address possible overlap between Dataset-A and Dataset-B, signal-level duplicate checking was performed before preprocessing, segmentation, model training, channel selection, and evaluation. Each loaded 19-channel EEG recording was converted into a consistent numerical representation and then encoded as a deterministic MD5 hash value. Recordings producing identical hash values were considered exact duplicate EEG signals. This process identified duplicate recordings across Dataset-A and Dataset-B, as well as internal duplicate recordings within Dataset-B. Therefore, only one copy of each unique EEG signal was retained. All revised main experiments were conducted using the resulting duplicate-free cohort.

Recording-level grouping and partition verification

Because universal participant identifiers are not consistently available across Dataset-A and Dataset-B, the revised evaluation is described as a recording-level grouped evaluation rather than a participant-level evaluation. Each retained EEG recording was assigned a unique group identifier derived from its dataset source and either its file name or patient folder name. All windows generated from the same retained recording were assigned to the same training, validation, or test partition. To verify the partitioning protocol, we generated a recording-level partition inventory for all seed-fold splits. For each split, the inventory recorded the number of training, validation, and test recordings, the AD/HC distribution, and the overlap counts between training, validation, and test partitions. The maximum overlap was zero for both recording identifiers and signal-level MD5 hashes across all seed-fold splits. This confirmed that no retained EEG recording appeared in more than one partition.

Exact and near-duplicate recording audit

Signal-level exact duplicate checking was first performed using deterministic MD5 hashes computed from the loaded 19-channel EEG recordings. This step identified exact duplicate EEG signals across Dataset-A and Dataset-B, as well as internal duplicate recordings. After removing exact duplicates, the final duplicate-free cohort contained 113 retained EEG recordings, including 88 AD and 25 HC recordings. Because MD5 hashing identifies only identical signals, we also performed a near-duplicate audit on the duplicate-free cohort. After applying the same preprocessing pipeline used for model training, pairwise similarity was computed across all retained EEG recordings using the mean absolute Pearson correlation across aligned channels. In total, 6,328 pairwise comparisons were performed among the 113 retained recordings. No pair exceeded the high-similarity thresholds of 0.99 or 0.995, and the maximum observed mean absolute channel correlation was 0.6634. This audit further reduced the likelihood of near-duplicate leakage due to cropped, scaled, or highly similar non-identical recordings.

Preprocessing

All preprocessing steps were performed after signal-level duplicate removal. Each EEG recording was first checked for invalid numerical values. Any NaN, positive infinite, or negative infinite values were replaced with finite values to avoid numerical instability during filtering and model training. The signals were then filtered using a 0.5–45 Hz Butterworth band-pass filter⁴¹. This filtering step reduces slow baseline drift and high-frequency noise while preserving the main EEG frequency components relevant to AD-related EEG analysis. After filtering, channel-wise standardization was applied independently to each EEG channel. Specifically, the mean of each channel was subtracted, and the result was divided by the corresponding channel standard deviation. This normalization reduces amplitude variability across channels and recordings and improves the stability of neural-network optimization.

Segmentation and channel configuration

Each preprocessed EEG recording was segmented into non-overlapping windows of 128 samples. Since both main datasets were sampled at 128 Hz, each window corresponds to one second of EEG activity. For the full-channel experiments, all 19 EEG channels were retained, and each window was represented as:

$$X \in \mathbb{R}^{128 \times 19}, \quad (1)$$

where 128 denotes the number of temporal samples, and 19 denotes the number of EEG channels. For reduced-channel experiments, only the selected EEG channel indices were retained, producing:

$$X_k \in \mathbb{R}^{128 \times k}, \quad (2)$$

where k denotes the number of selected channels and $k \in \{5, 8, 10, 12, 15\}$.

All data splitting was performed at the recording level. Therefore, all EEG windows generated from the same retained EEG recording were kept in the same partition, preventing recording-level overlap between training and testing. Final diagnostic performance was reported at the recording level by aggregating the posterior probabilities of all EEG windows belonging to the same retained recording.

External OpenNeuro ds004504 benchmark

The public OpenNeuro ds004504 EEG dataset was used as an external benchmark⁴². This dataset contains resting-state eyes-closed EEG recordings from 88 participants, including 36 AD patients, 23 FTD patients, and 29 healthy controls, recorded using 19 scalp electrodes. The dataset was used only for external pairwise evaluation and was not merged with the main AD/HC cohort.

Proposed method

Overview of the Fused-MSCA-2B framework

We introduce Fused-MSCA-2B, a fused multi-scale channel-attentive two-branch 1D CNN for EEG-based AD detection. The overall diagnostic workflow is shown in Fig. 1. The framework consists of EEG preprocessing, recording-level data partitioning, Fused-MSCA-2B feature extraction, validation-only IG-based channel selection, and recording-level evaluation. The model processes broadband multichannel EEG windows directly in temporal form. In the full-channel setting, each input window has a shape of 128×19 , where 128 represents the number of temporal samples, and 19 represents the number of EEG channels. In the reduced-channel setting,

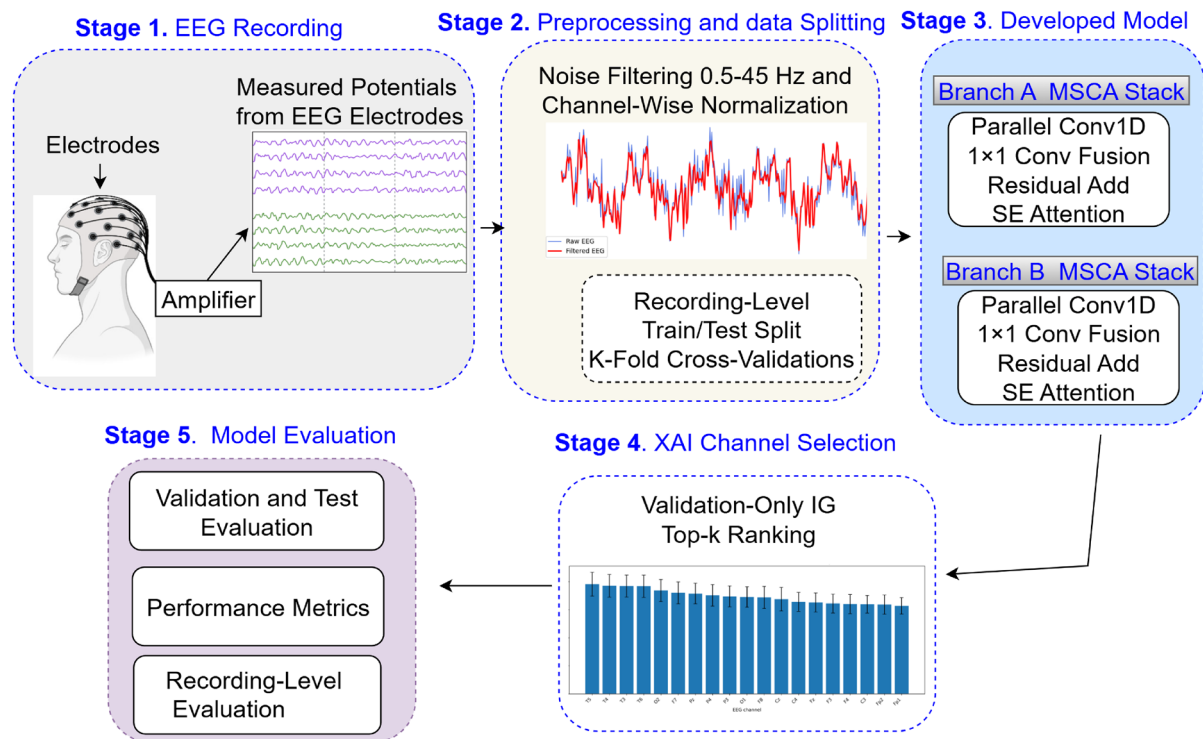


Figure 1 System schematic of the proposed XAI-guided diagnostic framework, including EEG preprocessing, recording-level partitioning, Fused-MSCA-2B feature extraction, validation-only IG-based channel selection, and recording-level evaluation.

the input shape becomes $128 \times k$, where k denotes the number of selected EEG channels. Temporal convolution is applied along the time axis, while inter-channel information is learned through convolutional filters and channel-attention mechanisms. As shown in Fig. 2(a), Fused-MSCA-2B contains two complementary model-level branches, denoted as Branch A and Branch B. Branch A is designed to capture local and medium-range temporal dependencies using compact kernel settings, whereas Branch B is designed to capture broader temporal dependencies using wider kernels and more filters. The asymmetric kernel and filter settings were selected to provide complementary temporal receptive fields. The two branches learn complementary EEG representations from the same input window. Their embeddings are then concatenated and passed through dense classification layers to generate AD/HC posterior probabilities.

Fig. 2(b) illustrates the internal structure of the multi-scale channel-attentive (MSCA) block. Each MSCA block contains parallel Conv1D paths with different kernel sizes, followed by kernel-size-1 Conv1D feature fusion, residual projection, ReLU activation, and squeeze-and-excitation (SE) channel attention. In this paper, Branch A and Branch B refer to the two model-level feature-extraction branches, whereas the parallel Conv1D paths refer only to the internal multi-scale operations inside each MSCA block.

Two-branch multi-scale channel-attentive architecture

The input to the proposed model is a preprocessed EEG window represented as:

$$X \in \mathbb{R}^{T \times C}, \quad (3)$$

where X denotes one EEG window, T is the number of temporal samples, and C is the number of EEG channels. In this study, $T = 128$, corresponding to a 1-s EEG segment at a sampling frequency of 128 Hz. For full-channel

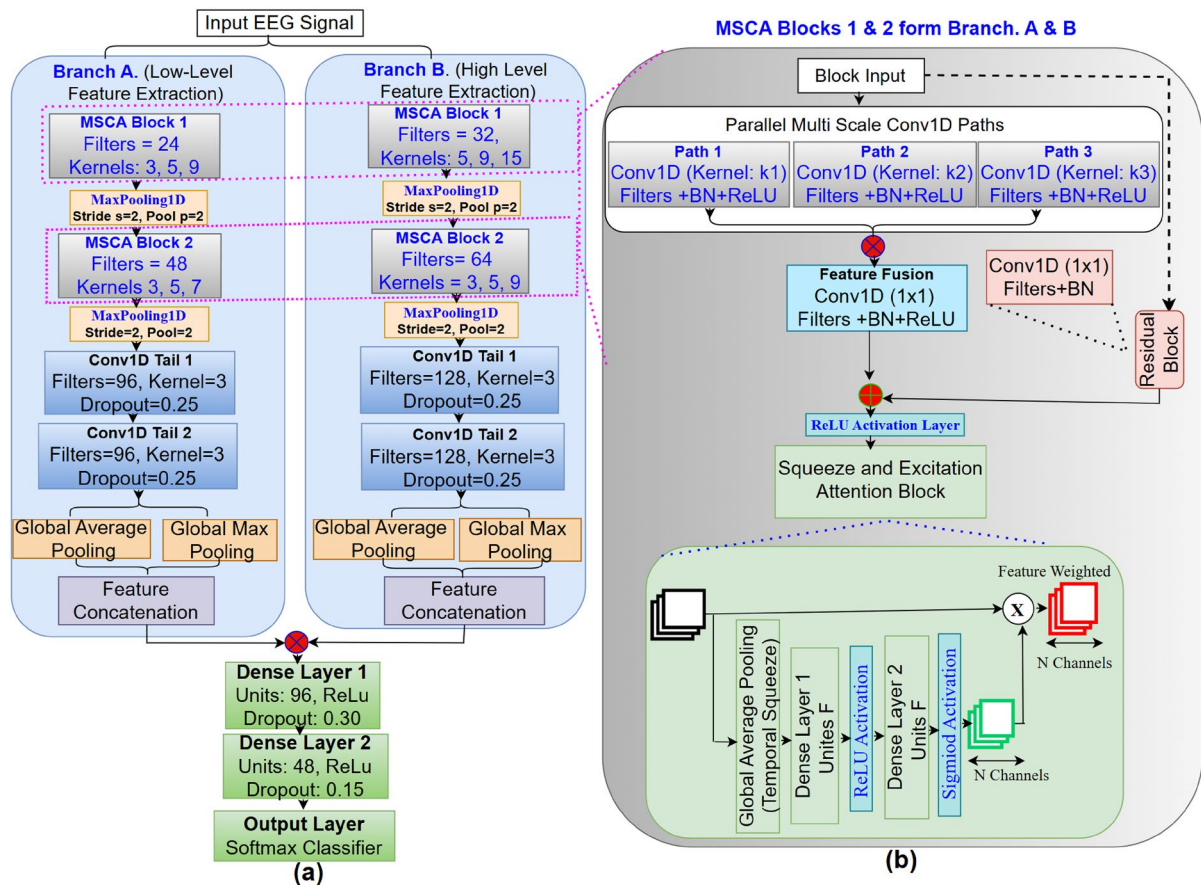


Figure 2 Detailed schematic of the proposed Fused-MSCA-2B architecture: (a) dual-branch macro-architecture and (b) internal structure of the Multi-Scale Channel-Attentive block with parallel multi-scale Conv1D paths.

experiments, $C = 19$. For reduced-channel experiments, $C = k$, where k is the number of selected EEG channels. The proposed architecture uses two asymmetric branches to learn complementary temporal EEG features. Both branches receive the same input window, but their filter numbers and kernel-size configurations are different. This design enables the model to capture EEG temporal patterns at multiple receptive-field scales. Branch-A focuses on compact local and medium-range temporal features, while Branch-B focuses on broader high-level temporal features. Each branch consists of two MSCA blocks. In each MSCA block, three parallel Conv1D paths are applied using different kernel sizes. For an input feature map x , the output of the m -th convolutional path is defined as:

$$Z_m = \delta \left(\text{BN} \left(\text{Conv1D}_{k_m}^F(x) \right) \right), \quad (4)$$

where Z_m denotes the output feature map of the m -th path, k_m is the kernel size, F is the number of filters, $\text{BN}(\cdot)$ denotes batch normalization, and $\delta(\cdot)$ denotes the ReLU activation function. The use of different kernel sizes allows the MSCA block to capture short-range, medium-range, and broader temporal EEG dependencies. Inside each MSCA block, the outputs of the parallel Conv1D paths are concatenated and fused using a kernel-size-1 Conv1D layer. This fusion operation integrates the multi-scale temporal features and controls the output dimensionality. A residual projection is then added to preserve useful low-level temporal information and improve gradient propagation⁴³. After residual fusion, squeeze-and-excitation (SE) channel attention is applied to recalibrate the fused feature maps⁴⁴.

Squeeze-and-excitation channel attention. As shown in Fig. 2(b), squeeze-and-excitation (SE) channel attention is applied after multi-scale feature fusion and residual addition. The purpose of this block is to adaptively recalibrate the fused feature maps by assigning higher weights to informative feature channels and lower weights to less relevant feature channels.

Given the residual-enhanced feature map Z_{res} , global average pooling is first applied along the temporal dimension to generate a compact feature-channel descriptor:

$$s = \text{GAP}(Z_{\text{res}}), \quad (5)$$

where s summarizes the global response of each learned feature channel. The descriptor is then passed through two fully connected layers. The first layer uses ReLU activation to learn a compact intermediate representation, and the second layer uses sigmoid activation to generate normalized feature-channel attention weights:

$$a = \sigma(W_2 \delta(W_1 s)), \quad (6)$$

where a denotes the learned attention-weight vector, W_1 and W_2 are trainable weight matrices, $\delta(\cdot)$ denotes ReLU activation, and $\sigma(\cdot)$ denotes sigmoid activation.

The learned attention weights are multiplied with the residual-enhanced feature map:

$$H = a \odot Z_{\text{res}}, \quad (7)$$

where H denotes the recalibrated output feature map and $\odot$ denotes channel-wise multiplication. This operation enables the MSCA block to emphasize discriminative temporal EEG feature channels before temporal downsampling is performed by max-pooling. Finally, max-pooling is used after the SE attention block to reduce the temporal resolution while retaining salient EEG features.

Branch-A: local and medium-range temporal feature extraction

Branch-A is designed to learn compact local and medium-range temporal EEG patterns. The first MSCA block in Branch-A uses 24 filters with kernel sizes (3, 5, 9). These kernel sizes allow the branch to capture short-term and medium-range EEG variations. The second MSCA block uses 48 filters with kernel sizes (3, 5, 7), which further refines the temporal representation using a slightly narrower multi-scale kernel range. After the two MSCA blocks, Branch-A applies two Conv1D tail layers, each with 96 filters and kernel size 3. Dropout rates of 0.25 and 0.25 are applied after the first and second tail layers, respectively, to reduce overfitting. These layers transform the extracted temporal patterns into a higher-level branch representation. Finally, global average pooling and global max pooling are applied in parallel to summarize the final Branch-A feature map:

$$b_A = [\text{GAP}(H_A); \text{GMP}(H_A)], \quad (8)$$

where b_A denotes the Branch-A embedding, H_A denotes the final feature map generated by Branch-A, and $\text{GAP}(\cdot)$ and $\text{GMP}(\cdot)$ denote global average pooling and global max pooling, respectively.

Branch-B: broader high-level temporal feature extraction

Branch-B receives the same preprocessed EEG window as Branch-A but uses a larger filter configuration and broader kernel settings. Therefore, Branch-B is designed to capture broader high-level temporal dependencies that may not be fully represented by Branch-A. The first MSCA block in Branch-B uses 32 filters with kernel sizes (5, 9, 15). Compared with Branch-A, these wider kernels provide a larger temporal receptive field and allow the branch to learn broader EEG temporal dependencies. The second MSCA block uses 64 filters with kernel sizes (3, 5, 9), which combines local and medium-range refinement with the broader first-stage representation.

After the two MSCA blocks, Branch-B applies two Conv1D tail layers, each with 128 filters and kernel size 3. A dropout rate of 0.25 is applied after each tail layer to reduce overfitting. Finally, global average pooling and global max pooling are applied in parallel to summarize the final Branch-B feature map:

$$b_B = [\text{GAP}(H_B); \text{GMP}(H_B)], \quad (9)$$

where b_B denotes the Branch-B embedding and H_B denotes the final feature map generated by Branch-B.

Feature fusion and classification

After Branch-A and Branch-B extract complementary temporal EEG representations, their embeddings are concatenated to form a fused feature vector:

$$h_{\text{fused}} = [b_A; b_B], \quad (10)$$

where h_{fused} denotes the fused feature representation. This fusion combines compact local and medium-range temporal features from Branch-A with broader high-level temporal features from Branch-B. The fused representation is passed through two dense layers. The first dense layer contains 96 units with ReLU activation and a dropout rate of 0.30. The second dense layer contains 48 units with ReLU activation and a dropout rate of 0.15. Finally, a softmax classifier generates the window-level posterior probabilities:

$$\hat{p} = \text{Softmax}(W_o h_2 + b_o), \quad (11)$$

where $\hat{p}$ denotes the predicted probability vector for the HC and AD classes, h_2 denotes the output of the second dense layer, and w_o and b_o are the trainable weight matrix and bias vector of the output layer. The final diagnostic decision is made at the recording level by aggregating the posterior probabilities of all EEG windows belonging to the same recording.

XAI-guided reduced-channel selection

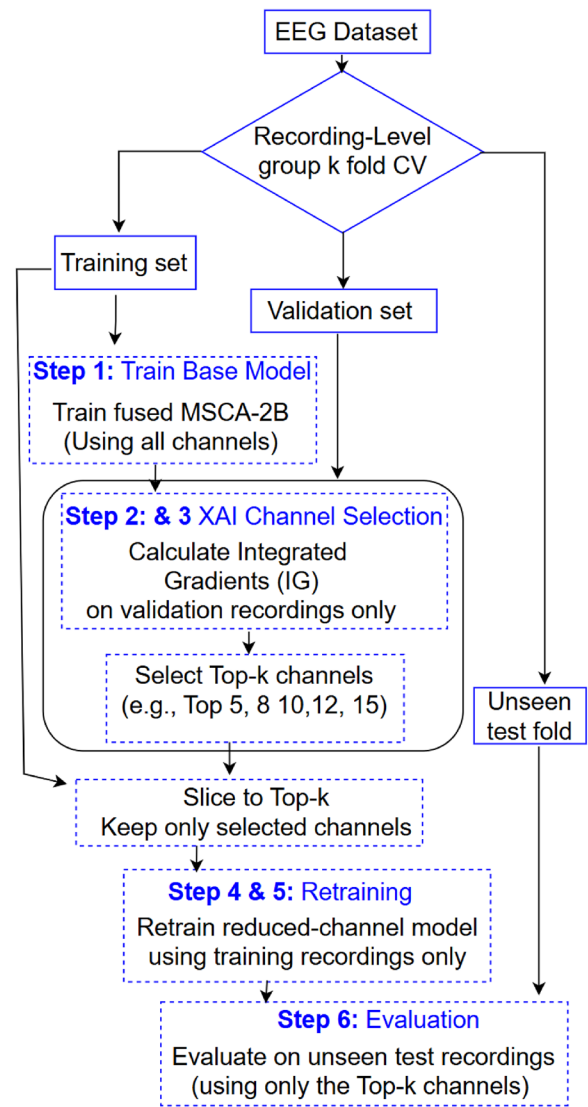
To evaluate whether strong classification performance can be maintained with fewer EEG electrodes, we introduce a validation-only XAI-guided reduced-channel selection strategy. Within each outer cross-validation fold, the training, validation, and test sets contain the same original 19 EEG channel names but different recordings. A full 19-channel Fused-MSCA-2B model is first trained using only the training recordings, while a recording-wise validation split is used for checkpoint selection and channel ranking. Figure 3 illustrates the Top- k channel-selection workflow. IG is computed only on validation recordings within the corresponding outer training fold to estimate EEG channel relevance⁴⁵. The held-out test recordings are not used for full-channel training, IG computation, channel ranking, model selection, or reduced-channel construction^{17,18}. After the Top- k channels are selected from the validation recordings, the reduced-channel inputs are constructed using the same selected channels. Each reduced-channel Fused-MSCA-2B model is then trained from scratch using only the selected channels from the training recordings and evaluated on the unseen test fold using the same selected channels.

For an input component x_i , IG is defined as:

$$\text{IG}_i(X) = (X_i - X_i^{(0)}) \int_{\alpha=0}^1 \frac{\partial F(X^{(0)} + \alpha(X - X^{(0)}))}{\partial X_i} d\alpha, \quad (12)$$

where $X^{(0)}$ is the baseline input and $F(\cdot)$ denotes the AD-class output score. The integral is approximated using a finite number of interpolation steps. For an EEG input $X \in \mathbb{R}^{T \times C}$, the channel relevance score is computed by averaging the absolute attribution values across validation windows and temporal samples:

Figure 3 Validation-only XAI-guided Top- k channel-selection workflow.



$$I_c = \frac{1}{NT} \sum_{n=1}^N \sum_{t=1}^T |IG_{n,t,c}|, \quad (13)$$

where N is the number of validation windows and $T = 128$ is the number of temporal samples.

The EEG channels are ranked according to I_c . For each Top- k setting, the selected channel subset s_k is used to form:

$$X_k = X[:, s_k] \in \mathbb{R}^{T \times k}. \quad (14)$$

This protocol provides a leakage-controlled estimate of reduced-channel generalization performance. However, because the selected Top- k channels may vary across folds and seeds, the validation-guided Top- k setting should be interpreted as a reduced-channel feasibility analysis rather than a finalized fixed electrode montage for direct deployment.

Experimental setup

Implementation details

To improve reproducibility, the main architectural and training hyperparameters of the proposed Fused-MSCA-2B model are summarized in Table 2. All experiments were implemented using Python and TensorFlow/Keras on a GPU-enabled workstation equipped with an NVIDIA GeForce RTX 5090 GPU with 32 GB of memory. The model was trained using the Adam optimizer with categorical cross-entropy loss and an initial learning rate of 8×10^{-4} . The learning rate was adaptively updated during training using a ReduceLROnPlateau strategy, as detailed in Table 2. Early stopping and checkpoint selection were based on recording-level balanced accuracy on the validation set.

Table 2 Implementation details and key hyperparameters of the proposed Fused-MSCA-2B framework.

Component	Setting
Input window length	128 samples, corresponding to 1 s at 128 Hz
Input shape	$128 \times C$, where $C = 19$ for full-channel input and $C = k$ for reduced-channel input
Preprocessing filter	0.5–45 Hz Butterworth band-pass filter, order 6
Segmentation	Non-overlapping 1-s windows; window size = 128, stride = 128
Architecture	Two-branch multi-scale channel-attentive 1D CNN
Branch-A MSCA block 1	24 filters; kernel sizes (3, 5, 9)
Branch-A MSCA block 2	48 filters; kernel sizes (3, 5, 7)
Branch-A tail layers	Conv1D layers with 96 filters and kernel size 3
Branch-A dropout	0.25
Branch-B MSCA block 1	32 filters; kernel sizes (5, 9, 15)
Branch-B MSCA block 2	64 filters; kernel sizes (3, 5, 9)
Branch-B tail layers	Conv1D layers with 128 filters and kernel size 3
Branch-B dropout	0.25
MSCA block operations	Parallel Conv1D paths, batch normalization, ReLU, 1×1 Conv1D fusion, residual projection, SE attention, max pooling
Pooling before fusion	Global average pooling and global max pooling for each branch
Feature fusion	Concatenation of Branch-A and Branch-B embeddings
Dense classifier	Dense 96 with dropout 0.30; Dense 48 with dropout 0.15; softmax output
Optimizer	Adam
Learning rate	8×10^{-4}
Loss function	Categorical cross-entropy
Batch size	32
Maximum epochs	20 for Fused-MSCA-2B
Checkpoint criterion	Recording-level balanced accuracy on the validation set
Cross-validation	Recording-level grouped 5-fold cross-validation
Repeated evaluation	Five random seeds: 42, 52, 62, 72, and 82
Validation split	20% of the outer-training recordings
Reduced-channel settings	$k \in \{5, 8, 10, 12, 15\}$
Integrated Gradients steps	32 interpolation steps
IG target class	AD class
IG ranking data	Validation recordings only; held-out test recordings excluded
Recording-level prediction	Mean posterior probability aggregation across all windows from the same recording
Learning-rate strategy	Adaptive learning-rate reduction using ReduceLROnPlateau
Scheduler monitor	Validation balanced accuracy
Learning-rate reduction factor	0.5
Scheduler patience	3 epochs
Minimum learning rate	1×10^{-6}

Recording-level grouped cross-validation and multi-seed protocol

A recording-level five-fold cross-validation protocol was used to prevent recording-level train-test leakage. The revised cohort contained 113 unique EEG recordings, including 88 AD and 25 HC recordings. Recording identifiers were checked in every outer fold. Across all 25 outer train-test splits from five folds repeated across five random seeds, the training-test recordings overlap was zero. Within each seed-level five-fold evaluation, all 113 recordings appeared exactly once in the held-out test partition. The complete evaluation was repeated using five independent random seeds: 42, 52, 62, 72, and 82. For each seed, metrics were first averaged across the five folds, and final results were reported as mean ± standard deviation across the five seed-level means.

Validation strategy

Within each outer training fold, a recording-level stratified validation split was used for checkpoint selection, early stopping, and XAI-guided channel ranking. The outer training recordings were further divided into training and validation subsets using a 20% validation split at the recording level. Model checkpoints were selected using recording-level balanced accuracy on the validation set. For reduced-channel experiments, channel ranking followed the validation-only IG protocol described in Section [XAI-guided reduced-channel selection](#). Held-out test recordings were not used for model selection, channel ranking, or channel-subset construction.

Evaluation metrics

All primary metrics were computed from recording-level predictions. Window-level posterior probabilities were averaged across all windows belonging to the same recording, and the following metrics were reported: accuracy, balanced accuracy, sensitivity, specificity, F1-score, Matthews correlation coefficient, and ROC-AUC.

Selected channel patterns

The selected Top-12 channel sets varied across folds and seeds because each fold used different training and validation recordings, and each seed introduced different initialization, partitioning, and training dynamics. This variation is expected in repeated recording-level grouped cross-validation and indicates that multiple related EEG channel subsets can provide discriminative information. Across representative folds, temporal, parietal, and occipital electrodes were frequently selected. Temporal electrodes, including T3, T4, T5, and T6, appeared repeatedly, while parietal electrodes such as P3, P4, and Pz, and occipital electrodes such as O1 and O2, were also selected. These patterns suggest distributed model relevance but should not be interpreted as definitive neurophysiological biomarkers. Table 3 shows representative fold-specific Top-12 channel selections from one grouped cross-validation run.

Table 3 Representative fold-specific Top-12 EEG channels selected by validation-set IG in one grouped cross-validation run on the revised cohort.

Fold	Selected Top-12 Channels by IG Rank
Fold 1	O2, T3, T5, T6, Fz, Fp1, Pz, C3, T4, F7, Cz, P3
Fold 2	Cz, T5, T4, F7, F8, F4, P4, C4, T3, Fp2, T6, O2
Fold 3	T4, F7, T3, O1, T6, T5, P4, Fz, F8, O2, P3, Cz
Fold 4	T3, Fz, T5, F7, C4, T4, F8, O2, F4, Fp1, P3, T6
Fold 5	T5, O1, Pz, P4, T3, F3, T4, T6, O2, C3, F8, Fz

Results

We conducted a comprehensive experimental evaluation to assess the effectiveness of the proposed Fused-MSCA-2B framework. The evaluation included architecture ablation, reduced-channel ablation, control channel-selection analysis, cross-dataset testing, external OpenNeuro benchmarking, and comparison with state-of-the-art methods.

Recording-level architecture, parameter-matched, and compact-channel ablation

To evaluate the effectiveness of the proposed architecture, we conducted an ablation study comparing the proposed Fused-MSCA-2B model with single-branch, parameter-matched, and module-ablated variants. The original MSCA-Single model uses one multi-scale channel-attentive branch, whereas Fused-MSCA-2B uses two complementary MSCA branches with feature-level fusion. The asymmetric kernel and filter settings in the two branches were designed to capture complementary temporal receptive fields. Branch A uses relatively compact kernels and fewer filters to capture local and medium-range EEG temporal patterns, whereas Branch B uses wider kernels and more filters to capture broader temporal dependencies. To address the possible effect of model capacity, we further included a parameter-matched single-branch baseline, denoted as MSCA-Single-PM, where PM indicates parameter-matched. This model contains a comparable or larger number of trainable parameters than the fused variants and was used to examine whether the observed performance was mainly due to increased parameter count. In addition, we evaluated a no-SE variant of the proposed model, denoted as Fused-MSCA-2B w/o SE, to assess the contribution of the squeeze-and-excitation channel-attention module. For reduced-channel analysis, validation-set IG rankings obtained from the full-channel Fused-MSCA-2B model were used to automatically rank the 19 EEG channels according to their contribution to AD classification. Based on these IG-derived rankings, multiple Top- k reduced-channel subsets were evaluated:

$$k \in \{5, 8, 10, 12, 15\}. \quad (15)$$

Among the evaluated IG-guided Top- k subsets, Top-12 achieved the strongest performance-oriented reduced-channel result. Therefore, Top-12 was used as the compact-channel configuration for the architecture, parameter-matched, and module-ablation comparison. Table 4 compares the baseline, parameter-matched, no-SE, and proposed fused models under both full-channel and Top-12 reduced-channel settings. The parameter-matched MSCA-Single-PM baseline had more trainable parameters than the fused variants, but it achieved lower performance than Fused-MSCA-2B under both full-channel and Top-12 settings. This suggests that the performance of the proposed framework is not explained solely by parameter count. The no-SE variant also achieved competitive performance, indicating that the dual-branch multi-scale fusion structure itself contributes meaningfully to classification performance. At the same time, the original Fused-MSCA-2B with SE provided a more stable and balanced overall performance, particularly in the Top-12 reduced-channel setting. Therefore, SE attention was retained as a supportive channel-recalibration component. Overall, the proposed Fused-MSCA-2B is interpreted as an effective and competitive two-branch architecture in which model capacity, dual-branch fusion, and SE-based channel attention jointly contribute to the final performance.

Recording-level XAI-guided channel selection analysis

Table 5 summarizes the recording-level performance of the XAI-guided reduced-channel Fused-MSCA-2B models. The Full-19 model is included as the full-channel reference. Among the reduced-channel configurations, Top-12 achieved the strongest performance-oriented result, with the highest accuracy, F1-score, and MCC. Top-15 achieved slightly higher balanced accuracy and ROC-AUC, but required more electrodes. Therefore, Top-12 was selected as the main reduced-channel configuration because it provided the best overall balance between

Table 4 Recording-level architecture, parameter-matched, and module-ablation analysis under full-channel and Top-12 reduced-channel settings. Results are reported as mean $\pm$ standard deviation across five independent random seeds.

Model	Purpose	Channels	Params	Acc. (%)	Bal. Acc. (%)	Sensitivity (%)	Specificity (%)	F1 (%)	MCC	ROC-AUC
MSCA-Single Full	Single-branch baseline	19	153,710	96.15 $\pm$ 3.30	95.14 $\pm$ 3.42	96.48 $\pm$ 3.82	93.80 $\pm$ 3.03	97.23 $\pm$ 2.51	0.9113 $\pm$ 0.0680	0.9967 $\pm$ 0.0041
MSCA-Single-PM Full	Parameter-matched single branch	19	285,796	94.32 $\pm$ 3.84	94.35 $\pm$ 4.76	94.30 $\pm$ 3.29	94.40 $\pm$ 6.69	95.79 $\pm$ 3.10	0.8812 $\pm$ 0.0720	0.9968 $\pm$ 0.0039
Fused-MSCA-2B w/o SE Full	Module ablation	19	273,954	98.21 $\pm$ 0.89	96.58 $\pm$ 1.67	99.56 $\pm$ 0.99	93.60 $\pm$ 3.58	98.86 $\pm$ 0.57	0.9498 $\pm$ 0.0240	0.9995 $\pm$ 0.0011
Fused-MSCA-2B Full	Proposed two-branch model	19	276,192	97.87 $\pm$ 1.34	97.75 $\pm$ 1.14	97.90 $\pm$ 1.91	97.60 $\pm$ 2.19	98.54 $\pm$ 0.97	0.9491 $\pm$ 0.0271	0.9996 $\pm$ 0.0010
MSCA-Single Top-12	Single-branch baseline	12	149,678	97.20 $\pm$ 2.40	95.93 $\pm$ 3.66	98.12 $\pm$ 1.78	93.73 $\pm$ 6.52	98.15 $\pm$ 1.58	0.9274 $\pm$ 0.0607	0.9992 $\pm$ 0.0011
MSCA-Single-PM Top-12	Parameter-matched single branch	12	279,748	96.83 $\pm$ 2.30	96.26 $\pm$ 2.45	97.32 $\pm$ 2.81	95.20 $\pm$ 4.38	97.84 $\pm$ 1.66	0.9252 $\pm$ 0.0479	0.9955 $\pm$ 0.0035
Fused-MSCA-2B w/o SE Top-12	Module ablation	12	264,210	98.40 $\pm$ 0.75	97.55 $\pm$ 0.48	99.10 $\pm$ 0.95	96.00 $\pm$ 0.00	98.98 $\pm$ 0.50	0.9558 $\pm$ 0.0190	0.9991 $\pm$ 0.0021
Fused-MSCA-2B Top-12	Proposed two-branch model	12	266,448	98.60 $\pm$ 1.18	97.84 $\pm$ 1.15	99.07 $\pm$ 1.53	96.60 $\pm$ 1.95	99.09 $\pm$ 0.83	0.9617 $\pm$ 0.0269	0.9991 $\pm$ 0.0020

Table 5 Recording-level Top-k reduced-channel ablation of the XAI-guided Fused-MSCA-2B model. Results are reported as mean $\pm$ standard deviation across five independent random-seed runs.

Model	Channels	Reduction	Params	Acc. (%)	Bal. Acc. (%)	Sensitivity (%)	Specificity (%)	F1 (%)	MCC	ROC-AUC
Fused-MSCA-2B Full-19	19	0.00%	276,192	97.87 $\pm$ 1.34	97.75 $\pm$ 1.14	97.90 $\pm$ 1.91	97.60 $\pm$ 2.19	98.54 $\pm$ 0.97	0.9491 $\pm$ 0.0271	0.9996 $\pm$ 0.0010
Top-5	5	73.68%	256,704	96.27 $\pm$ 1.28	95.95 $\pm$ 2.06	96.30 $\pm$ 0.99	95.60 $\pm$ 3.21	97.50 $\pm$ 0.85	0.9058 $\pm$ 0.0365	0.9948 $\pm$ 0.0050
Top-8	8	57.89%	260,880	96.30 $\pm$ 0.96	94.90 $\pm$ 1.76	97.46 $\pm$ 2.08	92.33 $\pm$ 4.96	97.56 $\pm$ 0.68	0.9048 $\pm$ 0.0216	0.9983 $\pm$ 0.0017
Top-10	10	47.37%	263,664	97.21 $\pm$ 1.44	97.05 $\pm$ 0.86	97.23 $\pm$ 2.21	96.87 $\pm$ 1.88	98.04 $\pm$ 1.10	0.9362 $\pm$ 0.0250	0.9975 $\pm$ 0.0036
Top-12	12	36.84%	266,448	98.60 $\pm$ 1.18	97.84 $\pm$ 1.15	99.07 $\pm$ 1.53	96.60 $\pm$ 1.95	99.09 $\pm$ 0.83	0.9617 $\pm$ 0.0269	0.9991 $\pm$ 0.0020
Top-15	15	21.05%	270,624	98.41 $\pm$ 0.76	98.06 $\pm$ 0.51	98.59 $\pm$ 1.53	97.53 $\pm$ 2.33	98.93 $\pm$ 0.56	0.9597 $\pm$ 0.0140	0.9995 $\pm$ 0.0011

classification performance and electrode reduction. The Top-8 configuration also provided a compact trade-off by using only 8 of 19 EEG channels while maintaining strong classification performance. Therefore, Top-8 is best interpreted as a compact reduced-channel option, whereas Top-12 represents the main performance-oriented reduced-channel configuration.

Figure 4 presents the seed-specific recording-level confusion matrices for the Top-12 model. The results show stable AD/HC classification across repeated runs, with only a small number of misclassified recordings in non-perfect seeds. Figure 5 shows the seed-specific recording-level ROC curves for the Top-12 model. The curves are concentrated near the upper-left corner, indicating strong discriminative ranking across repeated runs.

Control ablation of IG-guided Top-12 channel selection

To verify whether the IG-guided Top-12 subset is informative rather than arbitrary, five channel configurations were evaluated: Full-19, Top-12-IG, Random-12, Fixed-Top12-SC, and Bottom-12-IG, as summarized in Table 6. Random-12 used 30 randomly selected 12-channel subsets per seed and fold, whereas Fixed-Top12-SC used

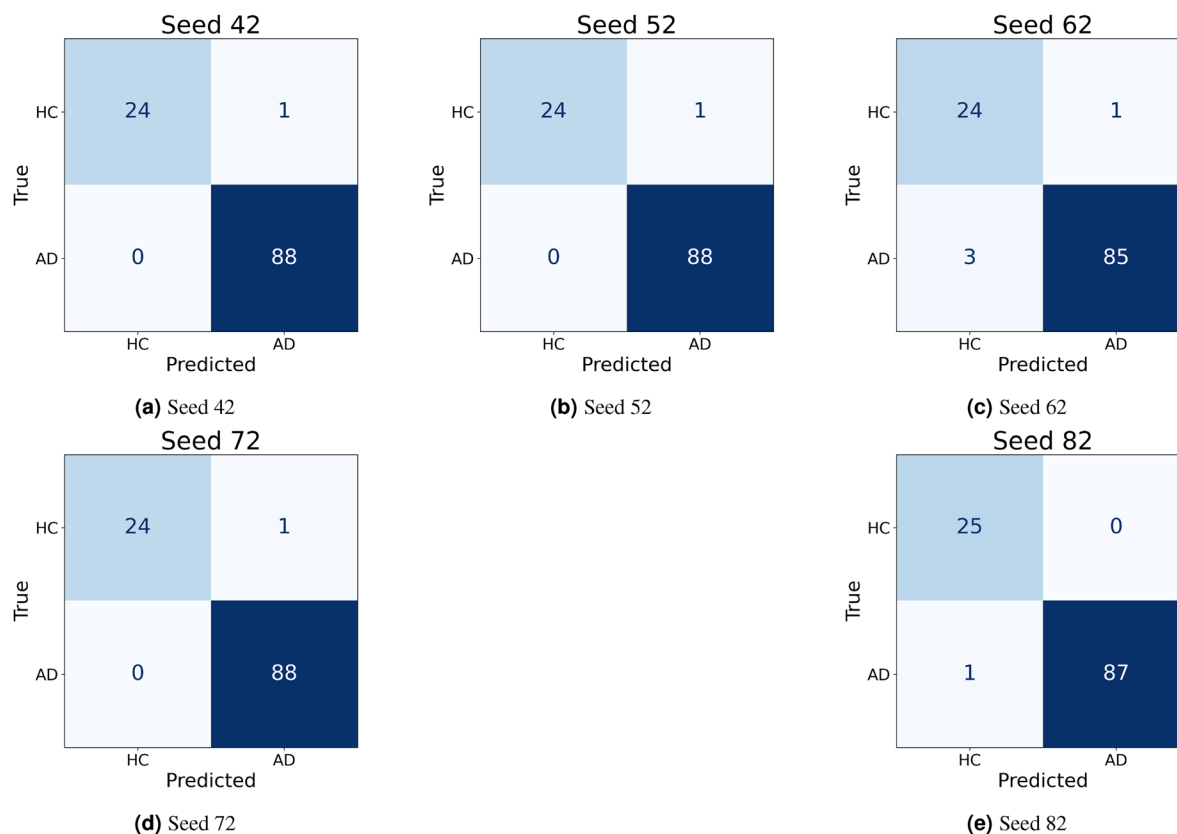


Figure 4 Seed-specific recording-level confusion matrices for the proposed Fused-MSCA-2B Top-12 model. Each panel corresponds to one independent random-seed run evaluated using recording-level grouped five-fold cross-validation over the same 113 unique EEG recordings.

Figure 5 Seed-specific recording-level ROC curves for the proposed Fused-MSCA-2B Top-12 model across five independent random-seed runs.

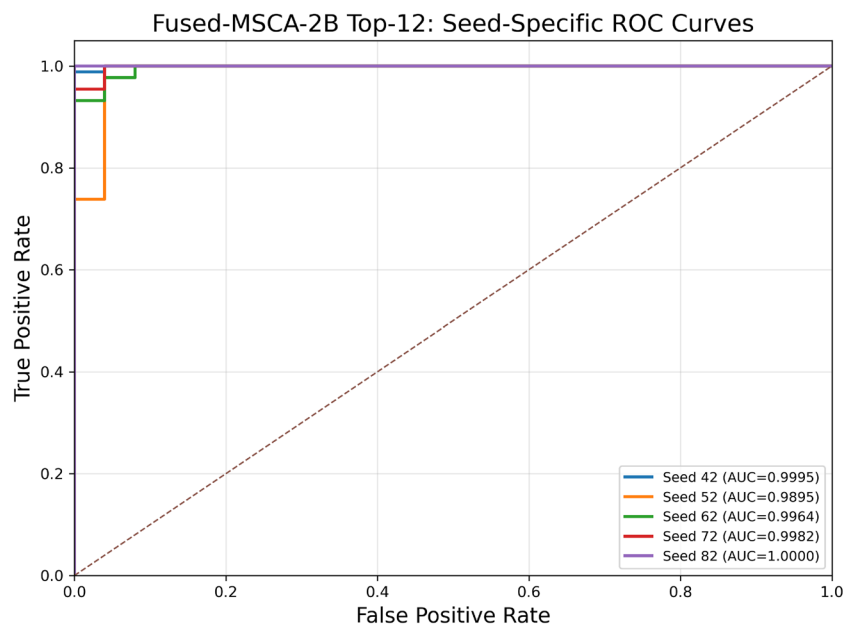


Table 6 Control ablation comparing Full-19 and Top-12-IG with Random-12, Fixed-Top12-SC, and Bottom-12-IG channel controls.

Channel set	Acc. (%)	Bal. Acc. (%)	Sens. (%)	Spec. (%)	F1 (%)	MCC	ROC-AUC
Bottom-12-IG	95.38 ± 2.61	93.88 ± 4.15	96.56 ± 1.84	91.20 ± 7.16	96.99 ± 1.71	0.8731 ± 0.0770	0.9863 ± 0.0091
Random-12	97.32 ± 0.63	96.72 ± 0.68	97.81 ± 0.67	95.63 ± 1.02	98.23 ± 0.43	0.9309 ± 0.0154	0.9968 ± 0.0017
Fixed-Top12-SC	98.40 ± 0.73	96.25 ± 1.73	100.00	92.50	99.00	0.9507	0.9979
Full-19	97.87 ± 1.34	97.75 ± 1.14	97.90 ± 1.91	97.60 ± 2.19	98.54 ± 0.97	0.9491 ± 0.0271	0.9996 ± 0.0010
Top-12-IG	98.60 ± 1.18	97.84 ± 1.15	99.07 ± 1.53	96.60 ± 1.95	99.09 ± 0.83	0.9617 ± 0.0269	0.9991 ± 0.0020

Table 7 Recording-level OOF statistical comparison of Top-12-IG with Full-19, Bottom-12-IG, and Random-12 configurations. Note: Mean differences are reported as Top-12-IG minus the comparison configuration.

Comparison	Metric	Mean difference	95% CI	FDR-corrected p
Top-12-IG vs Full-19	Accuracy	0.0000	[0.0000, 0.0000]	1.0000
Top-12-IG vs Full-19	Balanced Accuracy	0.0000	[0.0000, 0.0000]	1.0000
Top-12-IG vs Bottom-12-IG	Accuracy	0.0088	[0.0000, 0.0265]	1.0000
Top-12-IG vs Bottom-12-IG	Balanced Accuracy	0.0057	[0.0000, 0.0181]	1.0000
Top-12-IG vs Random-12	Accuracy	0.0000	[0.0000, 0.0000]	1.0000
Top-12-IG vs Random-12	Balanced Accuracy	0.0000	[0.0000, 0.0000]	1.0000

the 12 strongest channels derived from the single-channel-only ablation. For IG-based configurations, channel ranking was computed only from validation recordings within each fold to avoid test-data leakage. The Full-19 and Top-12-IG values were reused from the main performance tables, while Random-12, Fixed-Top12-SC, and Bottom-12-IG were evaluated through additional control experiments.

Table 6 shows that Top-12-IG achieved the strongest overall performance-efficiency trade-off among the evaluated reduced-channel configurations. Bottom-12-IG showed lower performance than Top-12-IG, indicating that the lowest-ranked IG channels were less informative for the classification task. Random-12 also performed strongly, suggesting that AD-related information is distributed across multiple EEG channels; however, it remained below Top-12-IG on the main recording-level metrics. Fixed-Top12-SC achieved competitive performance, showing that a stable 12-channel configuration derived from single-channel ranking can also preserve strong classification performance.

Overall, this ablation supports the usefulness of validation-set IG-guided channel selection while also showing that Top-12-IG should be interpreted as a strong reduced-channel configuration rather than the only informative 12-channel subset. Because the original montage contains 19 channels, the Top-12 and Bottom-12 subsets are not fully disjoint. Therefore, Bottom-12-IG should be interpreted as a low-importance control rather than a completely independent electrode set.

Although Table 6 reports repeated cross-validation mean ± standard deviation, these repeated seed-fold values were not treated as independent samples for statistical testing. Therefore, an additional retained-recording-level out-of-fold (OOF) analysis was performed for statistical comparison.

Recording-level OOF statistical analysis. Because universal participant identifiers are not consistently available across the public EEG datasets, the statistical analysis was performed at the retained-recording level rather than the participant level. Repeated seed-fold evaluations reuse the same retained recordings across different random partitions; therefore, seed-fold results were not treated as independent statistical samples. We performed a recording-level out-of-fold (OOF) prediction analysis. For each configuration, OOF predictions from five seeds and five folds were aggregated at the retained-recording level by averaging the AD posterior probability. Confidence intervals were estimated using bootstrap resampling over the 113 retained recordings. Paired effect sizes were reported as metric differences between configurations, and multiple comparisons were corrected using the false discovery rate (FDR) procedure⁴⁶. Table 7 reports the retained-recording-level OOF statistical comparison of Top-12-IG with Full-19, Bottom-12-IG, and Random-12 configurations.

Table 8 Cross-dataset generalization results of the full-channel Fused-MSCA-2B model. Thresholds were selected without using target test labels. Results are reported as mean $\pm$ standard deviation across five independent random seeds.

Train $\rightarrow$ Test	Thr.	Acc. (%)	Bal. Acc. (%)	Sensitivity (%)	Specificity (%)	F1 (%)	MCC	ROC-AUC
Dataset-A $\rightarrow$ Dataset-B	0.500	86.36 $\pm$ 10.38	75.96 $\pm$ 12.77	90.26 $\pm$ 13.97	61.67 $\pm$ 30.39	91.50 $\pm$ 7.56	0.532 $\pm$ 0.191	0.8954 $\pm$ 0.0946
Dataset-B $\rightarrow$ Dataset-A	0.500	89.60 $\pm$ 4.56	89.94 $\pm$ 4.27	98.33 $\pm$ 3.73	81.54 $\pm$ 11.67	90.29 $\pm$ 3.80	0.813 $\pm$ 0.077	0.9885 $\pm$ 0.0131

Table 9 Single-dataset recording-level results of the full-channel Fused-MSCA-2B model. Results are reported as mean $\pm$ standard deviation across five random seeds.

Dataset	Acc. (%)	Bal. Acc. (%)	Sens. (%)	Spec. (%)	F1 (%)	MCC	ROC-AUC (%)
Dataset-A only	96.31 $\pm$ 7.09	96.30 $\pm$ 7.11	96.60 $\pm$ 9.87	96.00 $\pm$ 8.17	96.12 $\pm$ 7.73	0.9306 $\pm$ 0.1367	99.36 $\pm$ 1.89
Dataset-B only	97.28 $\pm$ 3.31	92.48 $\pm$ 12.64	98.95 $\pm$ 2.46	86.00 $\pm$ 25.77	98.45 $\pm$ 1.86	0.8671 $\pm$ 0.2170	99.87 $\pm$ 0.67

The OOF analysis showed high retained-recording-level performance across the evaluated channel configurations. Top-12-IG achieved the same OOF accuracy and balanced accuracy as Full-19 and Random-12, and slightly higher OOF accuracy and balanced accuracy than Bottom-12-IG. However, no paired OOF comparison remained statistically significant after FDR correction. Therefore, the reduced-channel results are interpreted as strong and competitive ablation findings rather than evidence of statistically superior or universally optimal channel selection.

Cross-dataset generalization and calibration analysis

To examine source-dataset effects, the full-channel Fused-MSCA-2B model was evaluated under cross-dataset transfer. In each experiment, one dataset was used for source training and validation, while the other was held out entirely for target testing. The experiment was repeated across five independent random seeds, and the default threshold of 0.5 was used without using target test labels. Table 8 summarizes the recording-level results. Dataset-A $\rightarrow$ Dataset-B achieved 86.36% $\pm$ 10.38% accuracy and 75.96% $\pm$ 12.77% balanced accuracy. Dataset-B $\rightarrow$ Dataset-A achieved 89.60% $\pm$ 4.56% accuracy and 89.94% $\pm$ 4.27% balanced accuracy. The Dataset-B $\rightarrow$ Dataset-A direction showed stronger threshold transfer, whereas Dataset-A $\rightarrow$ Dataset-B showed lower specificity and higher variability. These results suggest that the learned representation can transfer across the two source datasets to some extent, but threshold calibration and source-dataset effects remain important.

These results indicate that cross-dataset transfer is feasible but not fully symmetric. The lower specificity and higher variability in the Dataset-A $\rightarrow$ Dataset-B direction suggest that calibration learned from one source dataset may not transfer directly to another dataset. Therefore, calibration-aware validation and larger independent cohorts remain necessary before deployment.

To examine possible source-dataset effects, we first evaluated the full-channel Fused-MSCA-2B model within Dataset-A and Dataset-B separately using the same recording-level grouped evaluation protocol. As shown in Table 9, the model achieved high recording-level performance within both datasets. Dataset-A-only evaluation achieved an accuracy of 96.31% and balanced accuracy of 96.30%, while Dataset-B-only evaluation achieved an accuracy of 97.28% and balanced accuracy of 92.48%. The lower specificity and higher variability for Dataset-B-only evaluation indicate that dataset imbalance and source-specific characteristics may still affect model calibration.

External OpenNeuro ds004504 benchmark and prior comparison

The proposed framework was further evaluated on the independent OpenNeuro ds004504 EEG dataset⁴². This dataset contains resting-state EEG recordings from healthy controls, AD patients, and frontotemporal dementia (FTD) patients. It was used only as an external benchmark and was not merged with the main AD/HC cohort. Three pairwise tasks were evaluated: AD/HC, FTD/HC, and AD/FTD. For the external OpenNeuro benchmark,

exploratory task-specific settings were reported because the external dataset differs from the main AD/HC cohort in population, diagnostic labels, and recording characteristics. For AD/HC, the Top-12 Fused-MSCA-2B model with broadband 1-s epochs and no functional-connectivity branch was used to maintain consistency with the main reduced-channel analysis. For FTD/HC, a broadband 1-s Top-8 Fused-MSCA-2B setting without a functional-connectivity branch was evaluated. For AD/FTD, a delta-band 5-s Top-15 setting with an additional functional-connectivity branch was evaluated. These OpenNeuro settings should be interpreted as exploratory external feasibility evaluations rather than final optimized configurations. Table 10 compares the proposed OpenNeuro results with representative prior studies^{42,47,48} evaluated on the same dataset, including relative-band-power, functional-connectivity, and Riemannian-geometry-based approaches. Recent Transformer-based studies have also evaluated cross-subject EEG-based AD classification on this dataset⁴⁷.

For AD/HC, the proposed Top-12 Fused-MSCA-2B setting achieved higher accuracy than the relative-band-power baseline and the stacked ensemble, while achieving a higher ROC-AUC than the compared prior methods that reported ROC-AUC. Its accuracy was lower than the FgMDM result, but its ROC-AUC was higher. For FTD/HC, the proposed Top-8 setting improved over the relative-band-power baseline and stacked ensemble⁴⁸ and approached the FgMDM⁴⁸ accuracy, although its ROC-AUC remained lower. For AD/FTD, the proposed Top-15 FC setting showed moderate hard-classification performance, indicating that disease-versus-disease discrimination remains more challenging than disease-versus-control classification. Overall, these results should be interpreted cautiously because preprocessing, feature representation, task definition, validation protocol, and prediction-aggregation strategy differ across studies.

Comparison with state-of-the-art methods

To ensure a fair comparison, the main state-of-the-art comparison was restricted to EEG-based studies using the same or closely related datasets. Therefore, we compare the proposed model with EDL3DCNN on the merged AD/HC cohort and with prior OpenNeuro ds004504 studies on the external pairwise benchmark. Table 11 compares the proposed Fused-MSCA-2B framework with the reference EDL3DCNN method³⁵. EDL3DCNN constructs a time-channel-frequency tensor and combines multiple 3D-CNN classifiers, whereas the proposed framework directly processes EEG as multichannel temporal windows using a compact two-branch 1D CNN. For transparency, Table 11 reports both the original merged-cohort results and the revised-cohort results after signal-level duplicate removal.

Table 10 External OpenNeuro ds004504 benchmark and comparison with representative prior studies.

Study	Feature / Setting	Task	Model	Acc. (%)	Sensitivity (%)	Specificity (%)	F1 (%)	ROC-AUC (%)
Miltiadous et al. ⁴²	Relative band power	AD/HC	Random Forest	77.01	78.32	80.94	75.31	–
Miltiadous et al. ⁴²	Relative band power	FTD/HC	MLP	73.12	63.00	78.63	72.82	–
Mlinaric et al. ⁴⁸	Multiple FC measures	AD/HC	Stacked ensemble	73.85	69.44	79.31	74.63	81.80
Mlinaric et al. ⁴⁸	Multiple FC measures	FTD/HC	Stacked ensemble	71.15	69.57	72.41	68.09	71.36
Mlinaric et al. ⁴⁸	Multiple FC measures	AD/FTD	Stacked ensemble	69.49	73.91	66.67	65.38	65.10
Mlinaric et al. ⁴⁸	α -CSD	AD/HC	FgMDM	84.62	77.78	93.10	84.85	88.70
Mlinaric et al. ⁴⁸	α -Cross-covariance	FTD/HC	FgMDM	82.69	73.91	89.66	79.07	86.36
Mlinaric et al. ⁴⁸	δ -Cross-correlation	AD/FTD	FgMDM	64.41	52.21	72.22	53.33	70.65
Wang et al. ⁴⁷	Spatiotemporal Transformer fusion	Dataset-I binary task	STFormer	75.17	75.23	74.67	76.37	–
Wang et al. ⁴⁷	Spatiotemporal Transformer fusion	AD/FTD/HC	STFormer	54.79	51.68	76.61	50.29	–
Proposed	Broadband 1-s, Top-12, no FC	AD/HC	Fused-MSCA-2B	78.15	73.50	84.27	77.42	89.22
Proposed	Broadband 1-s, Top-8, no FC	FTD/HC	Fused-MSCA-2B	80.55	76.00	86.48	79.60	81.46
Proposed	Delta 5-s, Top-15, FC branch	AD/FTD	Fused-MSCA-2B + FC	66.06	60.00	70.36	63.04	64.07

Table 11 Recording-level comparison with the state-of-the-art method on the original merged cohort and the revised cohort. The revised-cohort results are used as the main results in this study.

Model	Evaluation Setting	Architecture	Input Representation	Channels	Channel Selection	Acc. (%)	ROC-AUC
EDL3DCNN ³⁵	Original merged cohort	Ensemble 3D-CNN	Time–channel–frequency EEG tensor	19	No	99.02	Not reported
Fused-MSCA-2B Full	Original merged cohort	Two-branch 1D CNN	Temporal EEG window	19	No	98.86 ± 0.39	1.0000 ± 0.0000
Fused-MSCA-2B Top-8	Original merged cohort	XAI-guided two-branch 1D CNN	Reduced temporal EEG window	8	Validation-set IG	99.29 ± 0.71	0.9995 ± 0.0011
Fused-MSCA-2B Full	Revised cohort	Two-branch 1D CNN	Temporal EEG window	19	No	97.87 ± 1.34	0.9996 ± 0.0010
Fused-MSCA-2B Top-12	Revised cohort	XAI-guided two-branch 1D CNN	Reduced temporal EEG window	12	Validation-set IG	98.60 ± 1.18	0.9991 ± 0.0020
Fused-MSCA-2B Top-8	Revised cohort	XAI-guided two-branch 1D CNN	Reduced temporal EEG window	8	Validation-set IG	96.30 ± 0.96	0.9983 ± 0.0017

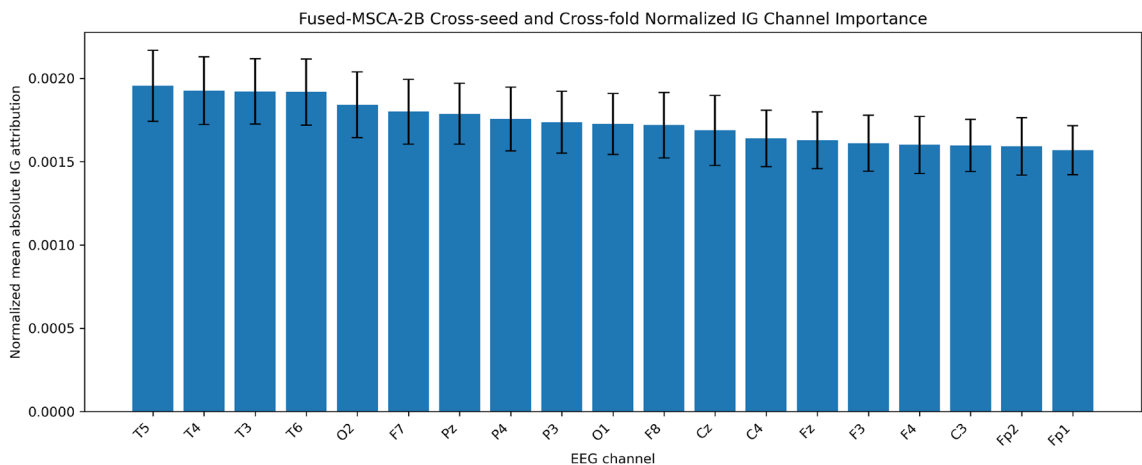


Figure 6 Cross-seed and cross-fold normalized Integrated Gradients channel-level importance for the full-channel Fused-MSCA-2B model with respect to the AD-class output neuron. Bars show the mean attribution across five seeds and five folds, and error bars indicate the standard error of the mean. Higher values indicate stronger model-specific relevance to the AD classification output.

Direct numerical comparisons remain limited because preprocessing, validation strategy, fold construction, duplicate handling, and prediction-aggregation protocols may differ across studies. The original merged-cohort results are included only for transparency and comparison with prior work that used the 140-recording merged setting. After signal-level duplicate removal, the revised cohort provides a stricter recording-level evaluation setting. Under this setting, Fused-MSCA-2B Top-12 achieved 98.60% ± 1.18% recording-level accuracy, while Top-8 retained strong compact-channel performance with 96.30% ± 0.96% accuracy. The main advantage of the proposed framework is its combination of direct temporal EEG modeling, validation-only IG-guided channel selection, reduced-channel evaluation, and repeated-seed recording-level assessment.

Explainability analysis

We applied IG to the trained fold-specific Fused-MSCA-2B models to examine model decision behavior. Attributions were computed with respect to the AD output neuron, and channel-level relevance was obtained by averaging the absolute attribution values over temporal samples and validation windows. Figure 6 shows the cross-seed and cross-fold normalized IG channel-level importance. The attribution values were averaged across five seeds

Table 12 Recording-level single-channel-only ablation results on the revised cohort using recording-balanced sample weights. Results are reported as mean values across five independent random seeds and ranked by balanced accuracy.

Channel Set	Acc. (%)	Bal. Acc. (%)	Sens. (%)	Spec. (%)	F1 (%)	MCC	ROC-AUC
Only-T5	84.98	82.27	86.51	78.03	86.56	0.6601	0.9123
Only-T6	77.52	76.26	77.55	74.96	76.74	0.5490	0.8677
Only-T3	82.50	75.14	89.87	60.41	87.22	0.5482	0.9320
Only-T4	84.62	75.07	91.71	58.42	90.09	0.5166	0.9098
Only-Pz	77.47	71.94	82.80	61.09	79.41	0.4852	0.9276
Only-O1	71.95	70.12	73.66	66.58	72.13	0.4034	0.8184
Only-C3	83.89	69.89	94.49	45.30	89.24	0.4636	0.8902
Only-P3	80.23	68.44	89.11	47.78	85.36	0.4270	0.9335
Only-F3	75.49	68.10	80.98	55.22	78.03	0.3912	0.8649
Only-P4	80.13	67.89	88.04	47.73	82.30	0.4383	0.9112
Only-O2	77.87	67.66	85.26	50.07	81.27	0.3588	0.8165
Only-F4	82.30	66.57	93.98	39.17	89.01	0.3495	0.9020
Only-C4	83.72	65.99	97.74	34.24	90.31	0.4095	0.9163
Only-Cz	71.14	63.76	77.93	49.60	75.81	0.3072	0.8382
Only-F8	78.58	61.92	91.78	32.05	86.86	0.2602	0.8237
Only-F7	75.09	60.30	87.58	33.01	80.77	0.2625	0.8430
Only-Fp1	73.53	59.96	85.34	34.59	79.66	0.1979	0.8379
Only-Fz	81.07	58.78	98.17	19.39	88.98	0.2391	0.9206
Only-Fp2	77.49	56.78	93.45	20.11	84.81	0.1716	0.8198

and five folds. Higher attribution values were mainly observed for temporal electrodes, including T5, T4, T3, and T6, followed by occipital and parietal electrodes such as O2, Pz, P4, P3, and O1. This pattern is consistent with the Top-12 channel-selection results and the single-channel-only ablation.

Figure 6 summarizes model-attribution patterns across repeated training and evaluation runs, providing a more stable visualization than a single-seed summary. However, IG values should still be interpreted as model-specific relevance scores rather than direct neurophysiological biomarkers. Overall, the explainability results suggest distributed EEG relevance across temporal, parietal, and occipital regions, but they should not be interpreted as evidence of clinical causality or definitive biomarkers.

Model efficiency

The full-channel Fused-MSCA-2B model contains 276,192 trainable parameters as summarized in Table 5. The Top-12 configuration contains 266,448 trainable parameters, corresponding to a 3.53% reduction, whereas the Top-8 configuration contains 260,880 trainable parameters, corresponding to a 5.54% reduction. This modest parameter reduction occurs because most parameters are located in convolutional and fusion layers rather than input-dependent layers. The main efficiency benefit is reduced acquisition complexity. By using 12 of 19 channels, the Top-12 model reduces the electrode montage by 36.84% while achieving the strongest reduced-channel performance. In comparison, the Top-8 model uses only 8 of 19 channels and reduces the electrode montage by 57.89%, providing a more compact trade-off. These reduced-channel settings can simplify electrode placement, shorten setup time, and improve patient comfort.

Single-channel-only ablation

To examine individual electrode informativeness, we trained Fused-MSCA-2B model and evaluated using one EEG channel at a time. Each configuration was evaluated using the same recording-level five-fold cross-validation protocol across five independent random seeds. Recording-balanced sample weights were used during training to reduce the effect of class imbalance. Table 12 summarizes the recording-level results ranked by balanced

accuracy. T5 achieved the strongest single-channel performance, followed by T6, T3, T4, and Pz. The strongest individual electrodes were mainly located in temporal and parietal regions, which is broadly consistent with the channel patterns observed in the IG-guided reduced-channel analysis. However, all single-channel configurations remained below the Top-12 multi-channel configuration, indicating that robust classification depends on distributed multi-electrode information rather than one dominant electrode.

Discussion

The experimental results show that Fused-MSCA-2B can support EEG-based AD classification with a reduced-channel setting. The framework models broadband EEG directly as a multichannel temporal signal, avoiding the construction of an artificial time-channel-frequency tensor and reducing representational complexity. The reduced-channel experiments indicate that validation-set IG can identify compact channel subsets that retain strong recording-level performance. The control and single-channel ablations further suggest that the Top-12 configuration benefits from distributed multi-electrode information rather than a single dominant electrode. A major methodological strength is the leakage-controlled evaluation protocol. Channel ranking was performed only on validation recordings; held-out test recordings were excluded from model selection and reduced-montage construction, and each reduced-channel model was retrained from scratch. Repeating the recording-level grouped evaluation across five random seeds further reduced split-specific bias. However, the validation-guided Top- k configurations should be interpreted as reduced-channel feasibility analyses rather than finalized fixed clinical montages. The explainability analysis showed higher model relevance in temporal, parietal, and occipital electrodes. These findings support the XAI-guided reduced-channel strategy but should be interpreted as model-specific attribution patterns rather than clinical biomarkers or evidence of biological causality. The cross-dataset analysis revealed asymmetric transfer performance between the two source–target directions. This indicates that calibration and source-dataset effects remain important considerations for external deployment. The OpenNeuro benchmark also showed that disease-versus-disease classification, especially AD/FTD, remains more challenging than disease-versus-control classification. Compared with the 3D-CNN-based approach³⁵, the proposed framework emphasizes direct temporal EEG modeling, validation-only reduced-channel selection, repeated-seed recording-level evaluation, and reduced-channel design. However, direct comparison with prior studies remains limited by differences in preprocessing and evaluation protocols. Several limitations remain, including the relatively small and class-imbalanced merged cohort, source-dataset differences, and reduced AD/FTD performance on the external benchmark. Larger independent and multicenter evaluations are therefore needed before clinical deployment.

Conclusion

This work presented Fused-MSCA-2B, an XAI-guided reduced-channel two-branch 1D CNN framework for EEG-based AD detection. The model combines multi-scale temporal convolution, residual learning, squeeze-and-excitation attention, and feature-level branch fusion to process multichannel EEG directly in temporal form. A validation-only Integrated Gradients strategy was used to construct reduced-channel EEG configurations, and reduced-channel models were evaluated only on held-out test recordings. The full-channel Fused-MSCA-2B model achieved $97.87\% \pm 1.34\%$ accuracy. Among the reduced-channel configurations, Top-12-IG achieved the strongest recording-level performance, with an accuracy of $98.60\% \pm 1.18\%$, while reducing the EEG montage from 19 to 12 channels. The Top-8 configuration also provided a compact trade-off by using only 8 of 19 channels, achieving $96.30\% \pm 0.96\%$ accuracy. The experimental analysis showed that Top-12-IG was comparable to Full-19 across all evaluated recording-level metrics and outperformed Bottom-12-IG and Random-12 in several key metrics. These findings support the usefulness of validation-only IG-guided channel selection while also

indicating that AD-related discriminative information is distributed across multiple EEG channels. Overall, the proposed framework provides an effective EEG-based AD detection approach by combining multi-scale temporal feature learning, residual learning, channel attention, feature-level fusion, and explainable reduced-channel selection. However, the present findings are based on public retrospective EEG datasets. Future work will include larger independent cohorts, multi-center prospective validation, calibration-aware threshold selection, domain-adaptation analysis, fixed-montage validation, and clinical investigation of selected channel patterns across different recording systems, populations, and acquisition protocols.

Author contributions J.S., N.H., and A.S.M.M contributed to the conceptualization and project development. J.S., N.H., and A.S.M.M. prepared the original draft of the manuscript. J.S., N.H., A.S.M.M., and N.S. contributed to review and editing. J.S. and N.S. supervised the study. J.S. acquired funding. All authors reviewed and approved the final manuscript.

Funding This work was supported by the Competitive Research Fund of The University of Aizu, Japan.

Data availability The primary EEG datasets analyzed in this study are publicly available through the Open Science Framework at <https://osf.io/s74qf/> and <https://osf.io/2v5md/>. The external EEG dataset used for validation is publicly available through OpenNeuro under accession number ds004504 at <https://openneuro.org/datasets/ds004504>.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Hassan, N., Musa Miah, A. S. & Shin, J. Residual-based multi-stage deep learning framework for computer-aided alzheimer's disease detection. *Journal of Imaging***10**, 141 (2024).
2. Yan, Q. et al. Glial cell nutrient sensing: mechanisms of nutrients regulating alzheimer's pathogenesis and precision intervention. *Critical Reviews in Food Science and Nutrition***66**, 2335–2351 (2026).
3. Hassan, N., Miah, A. S. M., Suzuki, K. & Shin, J. 3d mri-based alzheimer's disease detection using modified 3d resnet framework. In *Proceedings of the Seventh Asia Digital Image Processing Conference (ADIP 2025)*, vol. 14169, 63–75 (SPIE, Tokyo, Japan, 2025). 20–22 December 2025.
4. Sifat, M. et al. 3d mri-based alzheimer's disease classification using multi-modal 3d cnn with leakage-aware subject-level evaluation. In *2026 International Conference on Power, Electronics, Communications, Computing, and Intelligent Infrastructure (PECCII)*, 1–6, <https://doi.org/10.1109/PECCII70991.2026.11661973> (2026).
5. Breijyeh, Z. & Karaman, R. Comprehensive review on alzheimer's disease: Causes and treatment. *Molecules***25**, 5789. <https://doi.org/10.3390/molecules25245789> (2020).
6. Hassan, N., Miah, A. S. M., Okuyama, Y. & Shin, J. Neurological disorder recognition via comprehensive feature fusion by integrating deep learning and texture analysis. *IEEE Open Journal of the Computer Society* (2025).
7. Siddiqua, A., Oni, A. M., Miah, A. S. M., Hassan, N. & Shin, J. Automatic alzheimer's disease diagnosis using xgboost with shap for explainability. In *Proceedings of the 2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)*, 1–6 (IEEE, 2025).
8. Alidadi, M. et al. Eeg-based alzheimer's disease detection: Past, present, and future. *Measurement* 120096, (2025).
9. Chen, J. et al. Entropy-informed deep residual network for nonlinear eeg dynamics in fine-grained driver state recognition. *Applied Soft Computing* 115532 (2026).
10. Abd Rahman, F., Othman, M. F. & Shaharuddin, N. A. Analysis methods of eeg signals: a review in eeg application for brain disorder. *Jurnal Teknologi (Sciences & Engineering)***72** (2015).

11. Kulkarni, N. & Bairagi, V. *EEG-based diagnosis of Alzheimer disease: a review and novel approaches for feature extraction and classification techniques* (Academic Press, 2018).
12. Ieracitano, C., Mammone, N., Bramanti, A., Hussain, A. & Morabito, F. C. A convolutional neural network approach for classification of dementia stages based on 2D-spectral representation of EEG recordings. *Neurocomputing***323**, 96–107. <https://doi.org/10.1016/j.neucom.2018.09.071> (2019).
13. Rana, M. J. *et al.* An explainable eeg framework for alzheimer's and frontotemporal dementia diagnosis using pairwise csp and random forest. In *2025 13th International Conference on Awareness Science and Technology (iCAST)* 1–8 (IEEE, 2025).
14. Chen, J. *et al.* A review of eeg-based driver fatigue detection: Nonlinear dynamics, brain networks, and deep learning advances. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery***16**, e70099 (2026).
15. Chen, J., Yuan, G. & Wei, C. Characterizing fatigue-related neural signatures via multi-band eeg features and a three-stage feature optimization strategy for driver fatigue detection. *Measurement Science and Technology***37**, 306108. <https://doi.org/10.1088/1361-6501/ae8d4b> (2026).
16. Wang, Y., Huang, N., Mammone, N., Cecchi, M. & Zhang, X. LEAD: Large foundation model for EEG-based alzheimer's disease detection (2025). [arXiv:2502.01678](https://arxiv.org/abs/2502.01678).
17. Brookshire, G. *et al.* Data leakage in deep learning studies of translational eeg. *Frontiers in neuroscience***18**, 1373515 (2024).
18. Del Pup, F. *et al.* The role of data partitioning on the performance of eeg deep learning models. *Computers in Biology and Medicine* (2025).
19. Fahimi, G., Tabatabaei, S. M., Fahimi, E. & Rajebi, H. Index of theta/alpha ratio of the quantitative electroencephalogram in alzheimer's disease: A case-control study. *Acta Medica Iranica***55**, 502–506 (2017).
20. Garn, H. *et al.* Quantitative EEG markers relate to alzheimer's disease severity in the prospective dementia registry austria (PRODEM). *Clinical Neurophysiology***126**, 505–513 (2015).
21. Azami, H. *et al.* Multiscale fluctuation-based dispersion entropy and its applications to neurological diseases. *IEEE Access***7**, 68718–68733. <https://doi.org/10.1109/ACCESS.2019.2918560> (2019).
22. Zheng, H., Xiong, X. & Zhang, X. Multi-threshold recurrence rate plot: A novel methodology for eeg analysis in alzheimer's disease and frontotemporal dementia. *Brain Sciences***14**, 565. <https://doi.org/10.3390/brainsci14060565> (2024).
23. Vicchietti, M. L., Ramos, F. M., Betting, L. E. & Campanharo, A. S. L. O. Computational methods of eeg signals analysis for alzheimer's disease classification. *Scientific Reports***13**, <https://doi.org/10.1038/s41598-023-32664-8> (2023).
24. Horvath, A. *et al.* Eeg and erp biomarkers of alzheimer's disease: a critical review. *Frontiers in Bioscience***23**, 183–220 (2018).
25. Wang, R. *et al.* Power spectral density and coherence analysis of alzheimer's eeg. *Cognitive Neurodynamics***9**, 291–304 (2015).
26. Abásolo, D., Hornero, R., Gómez, C., García, M. & López, M. Analysis of eeg background activity in alzheimer's disease patients with lempel-ziv complexity and central tendency measure. *Medical engineering & physics***28**, 315–322 (2006).
27. Adeli, H., Ghosh-Dastidar, S. & Dadmehr, N. Alzheimer's disease: models of computation and analysis of eegs. *Clinical EEG and Neuroscience***36**, 131–140 (2005).
28. Abásolo, D., Hornero, R., Espino, P., Alvarez, D. & Poza, J. Entropy analysis of the eeg background activity in alzheimer's disease patients. *Physiological Measurement***27**, 241 (2006).
29. Sharma, M., Patel, S. & Acharya, U. R. Automated detection of abnormal eeg signals using localized wavelet filter banks. *Pattern Recognition Letters***133**, 188–194 (2020).
30. Acharya, M., Deo, R. C., Barua, P. D., Devi, A. & Tao, X. Graph empirical mode decomposition and multiscale feature extraction for eeg-based classification of alzheimer's disease and frontotemporal dementia. *Computer Methods and Programs in Biomedicine* 109302 (2026).
31. Akbar, F. *et al.* Unlocking the potential of eeg in alzheimer's disease research: Current status and pathways to precision detection. *Brain Research Bulletin***223**, 111281 (2025).
32. Campanharo, A. S. L. O., Doescher, E. & Ramos, F. M. Application of quantile graphs to the automated analysis of eeg signals. *Neural Processing Letters***52**, 5–20 (2020).
33. Pineda, A. M., Ramos, F. M., Betting, L. E. & Campanharo, A. S. L. O. Quantile graphs for eeg-based diagnosis of alzheimer's disease. *PLOS ONE***15**, e0231169 (2020).
34. Dharia, S. Y., Liu, Q., Smith, S. D. & Valderrama, C. E. Dual-transformer cross-attention framework for alzheimer's disease detection via dpte-guided eeg channel selection and multi-modal integration. *Biomedical Signal Processing and Control***112**, 108390 (2026).
35. Alghamdi, A. M. *et al.* A novel approach hybrid of ensemble learning and 3-d cnn mechanism: early-stage diagnosis of alzheimer's disease using eeg signals. *Scientific Reports***15**, 35893. <https://doi.org/10.1038/s41598-025-19727-8> (2025).
36. Shan, X. *et al.* Spatial-temporal graph convolutional network for alzheimer classification based on brain functional connectivity imaging of electroencephalogram. *Human Brain Mapping***43**, 5194–5209 (2022).

37. Klepl, D., He, F., Wu, M., Blackburn, D. J. & Sarrigiannis, P. G. Adaptive gated graph convolutional network for explainable diagnosis of alzheimer's disease using EEG data. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* (2023).
38. Chen, L. et al. Brain power mapping with optimized deep learning for eeg-based pilot fatigue detection. *Biomedical Signal Processing and Control* **114**, 109284 (2026).
39. Eeg-based diagnosis of alzheimer's disease. Open Science Framework, <https://osf.io/s74qf/>. Accessed: 2025-11-30.
40. Computational methods of eeg signals analysis for alzheimer's disease classification. Open Science Framework, <https://osf.io/2v5md/>. Accessed: 2025-11-30.
41. Hussin, S. F., Birasamy, G. & Hamid, Z. Design of butterworth band-pass filter. *Politeknik & Kolej Komuniti Journal of Engineering and Technology* **1**, 32–46 (2016).
42. Miltiadous, A. et al. A dataset of scalp EEG recordings of alzheimer's disease, frontotemporal dementia and healthy subjects from routine EEG. *Data* **8**, 95. <https://doi.org/10.3390/data8060095> (2023).
43. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778 (2016).
44. Hu, J., Shen, L. & Sun, G. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7132–7141 (2018).
45. Sundararajan, M., Taly, A. & Yan, Q. Axiomatic attribution for deep networks. In *Proceedings of the 34th International Conference on Machine Learning*, 3319–3328 (2017).
46. Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N. & Golani, I. Controlling the false discovery rate in behavior genetics research. *Behavioural brain research* **125**, 279–284 (2001).
47. Wang, S. et al. Stformer: A transformer for spatiotemporal feature fusion in cross-subject eeg-based alzheimer's disease detection. *IRBM* 100965, <https://doi.org/10.1016/j.irbm.2026.100965> (2026).
48. Mlinarič, T., Van Den Kerchove, A., Barinaga, Z. I. & Van Hulle, M. M. Eeg-based classification of alzheimer's disease and frontotemporal dementia using functional connectivity. *Scientific Reports* (2026).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Jungpil Shin¹  · Najmul Hassan¹ · Abu Saleh Musa Miah¹ · Satoshi Nishimura¹

✉ Jungpil Shin
jpshin@u-aizu.ac.jp

¹ School of Computer Science and Engineering, The University of Aizu, Fukushima, Aizuwakamatsu 965-0006, Japan