

# EEG-Based Alzheimer’s Disease Detection Using Two-Branch Multi-Scale Deep Learning With Integrated Gradients-Guided Channel Selection

PATTERN PROCESSING LAB,  
SCHOOL OF COMPUTER SCIENCE AND ENGINEERING, UNIVERSITY OF AIZU



## INTRODUCTION

Early diagnosis of **Alzheimer’s disease (AD)** is critical for timely intervention. **Electroencephalography (EEG)** is a non-invasive, low-cost, portable alternative to neuroimaging, but existing methods face **high computational cost, complex representations, and limited deployability**. Standard clinical EEG uses 19 channels, which increases setup time, and selecting channels with test data can cause information leakage. We propose **Fused-MSCA-2B**, a lightweight two-branch multi-scale channel-attentive 1D CNN with **Integrated Gradients (IG)-guided channel reduction** from 19 to 12 channels.

## PURPOSE

This study aims to build an **accurate and compact EEG model** for **AD vs. healthy control (HC)** classification, to select informative channels with **validation-only IG**, and to evaluate it with **leakage-controlled recording-level cross-validation**.

## METHODOLOGY

The workflow (Fig. 1) has five stages: **EEG recording, preprocessing and splitting, the Fused-MSCA-2B model, XAI channel selection, and evaluation**. EEG is filtered, normalized per channel, and cut into **1-s windows (128 × 19)**. Two branches with different kernel sizes learn complementary temporal features that are fused for classification (Fig. 2). IG ranks channels on validation recordings only (Fig. 3), and a reduced-channel model is retrained from scratch and tested on unseen recordings.

## Datasets

Two public 19-channel EEG datasets sampled at **128 Hz** were used: Dataset-A (24 HC, 24 AD) and Dataset-B (12 HC, 80 AD, eyes-closed). After **MD5 duplicate removal** and a near-duplicate audit (maximum correlation 0.66), the revised cohort contains **113 unique recordings: 88 AD and 25 HC**. The **OpenNeuro ds004504** dataset (36 AD, 23 FTD, 29 HC) is used only as an external benchmark.

## Preprocessing and Evaluation

Signals are filtered with a **0.5–45 Hz Butterworth band-pass**, standardized per channel, and segmented into non-overlapping **1-s windows**. Evaluation uses **recording-level grouped 5-fold cross-validation** repeated over **5 random seeds**: all windows of one recording stay in the same partition, and window probabilities are averaged per recording. Zero train–test overlap was verified.

## Two-Branch Architecture

Each branch has two **multi-scale channel-attentive (MSCA) blocks** with parallel Conv1D paths, 1×1 fusion, residual projection, and **squeeze-and-excitation** attention. **Branch A** (24/48 filters, kernels 3-5-9 and 3-5-7) captures local patterns; **Branch B** (32/64 filters, kernels 5-9-15 and 3-5-9) captures broader ones. The embeddings are concatenated and classified using dense layers (96, 48) followed by a SoftMax output layer

## XAI-Guided Channel Selection

**Integrated Gradients** computed on validation recordings rank the 19 channels, and the **Top-k channels** (k = 5, 8, 10, 12, 15) are retained; test recordings are never used for ranking. **Temporal (T3, T4, T5, T6)**, parietal, and occipital channels showed the highest model-specific relevance (Fig. 6).

## RESULT AND CONCLUSION

**Fused-MSCA-2B Top-12 achieved 98.60% ± 1.18% accuracy** (sensitivity 99.07%, specificity 96.60%, ROC-AUC 0.9991) using **12 of 19 channels**, compared with 97.87% for the full 19-channel model (Tables 1 and 2). **A fixed Top-12 configuration selected from single-channel performance also achieved 98.40% ± 0.73% accuracy, further supporting the effectiveness of reduced-channel selection**. Top-12-IG outperformed Bottom-12-IG (95.38%) and Random-12 (97.32%) (Table 3), while Top-8 retained 96.30% accuracy. **Cross-dataset transfer was asymmetric** (86.4% and 89.6%). **Conclusion:** IG-guided Top-12 channel selection enables robust and compact EEG-based AD detection; however, larger multi-center validation is needed.

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

Table 1: Architecture, parameter-matched, and module ablation (Full-19 vs. Top-12).

| 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 ± 1.34        | 97.75 ± 1.14        | 97.90 ± 1.91        | <b>97.60 ± 2.19</b> | 98.54 ± 0.97        | 0.9491 ± 0.0271        | <b>0.9996 ± 0.0010</b> |
| Top-5                 | 5        | 73.68%        | 256,704        | 96.27 ± 1.28        | 95.95 ± 2.06        | 96.30 ± 0.99        | 95.60 ± 3.21        | 97.50 ± 0.85        | 0.9058 ± 0.0365        | 0.9948 ± 0.0050        |
| Top-8                 | 8        | 57.89%        | 260,880        | 96.30 ± 0.96        | 94.90 ± 1.76        | 97.46 ± 2.08        | 92.33 ± 4.96        | 97.56 ± 0.68        | 0.9048 ± 0.0216        | 0.9983 ± 0.0017        |
| Top-10                | 10       | 47.37%        | 263,664        | 97.21 ± 1.44        | 97.05 ± 0.86        | 97.23 ± 2.21        | 96.87 ± 1.88        | 98.04 ± 1.10        | 0.9362 ± 0.0250        | 0.9975 ± 0.0036        |
| Top-12                | 12       | <b>36.84%</b> | <b>266,448</b> | <b>98.60 ± 1.18</b> | 97.84 ± 1.15        | <b>99.07 ± 1.53</b> | 96.60 ± 1.95        | <b>99.09 ± 0.83</b> | <b>0.9617 ± 0.0269</b> | 0.9991 ± 0.0020        |
| Top-15                | 15       | 21.05%        | 270,624        | 98.41 ± 0.76        | <b>98.06 ± 0.51</b> | 98.59 ± 1.53        | 97.53 ± 2.33        | 98.93 ± 0.56        | 0.9597 ± 0.0140        | 0.9995 ± 0.0011        |

Table 2: Top-k reduced-channel ablation of XAI-guided Fused-MSCA-2B.

## REFERENCES

[1]. Hassan, N., Miah, A. S. M. & Shin, J. Residual-based multi-stage deep learning framework for computer-aided Alzheimer’s disease detection. J. Imaging 10, 141 (2024).  
[2]. Alghamdi, A. M. et al. A novel approach hybrid of ensemble learning and 3-D CNN: early-stage diagnosis of Alzheimer’s disease using EEG signals. Sci. Rep. 15, 35893 (2025).  
[3]. 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 (2023).

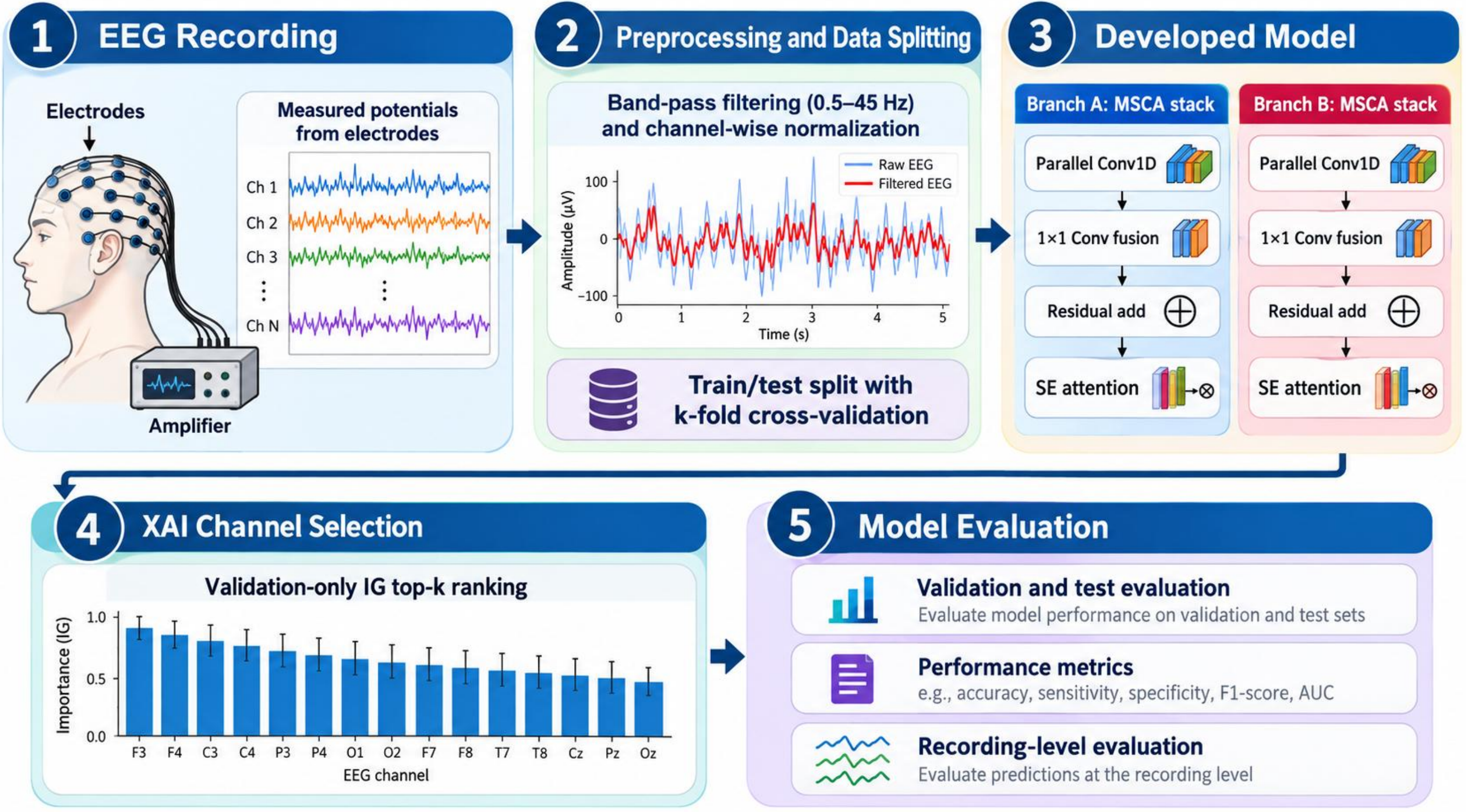


Fig. 1: System schematic of the XAI-guided diagnostic framework.

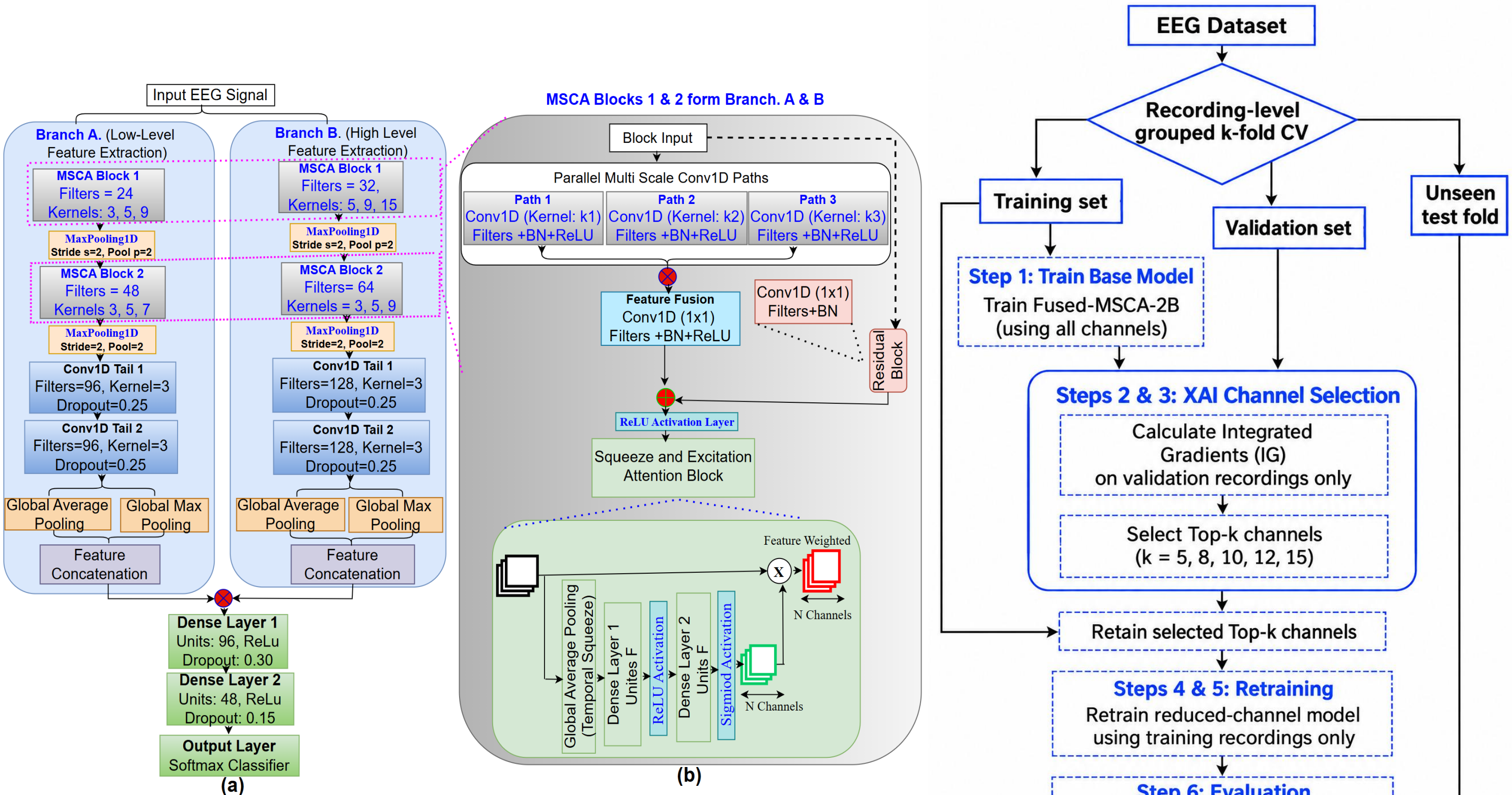


Fig. 2: Fused-MSCA-2B architecture: (a) dual-branch network, (b) MSCA block.

Fig. 3: Validation-only IG-guided Top-k channel selection.

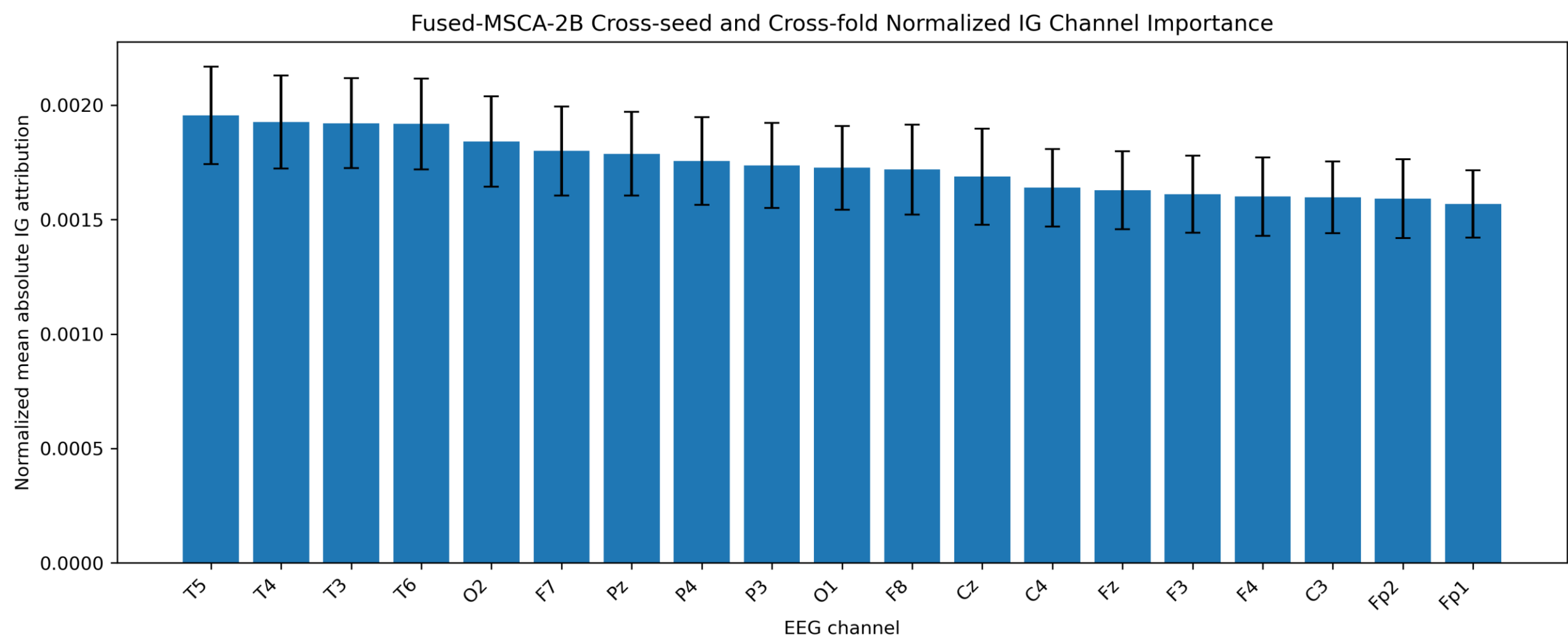


Fig. 6: Cross-seed and cross-fold normalized IG channel importance.