# A Lightweight Subject-adaptive EEG Network for Driver Drowsiness Detection

Qien Hou*
Huaibei Vocational and Technical College, Huaibei 235000, China
*Corresponding email: houqien@163.com

**Abstract**

Electroencephalography (EEG)-based driver-drowsiness detection is hindered by cross-subject variability and the resource cost of multi-channel models. This article presents Lite dynamic vision with explicit-adaptive implicit matching knowledge distillation (Lite-DV-EAIM-KD), a lightweight, subject-adaptive, and interpretable framework for binary drowsiness recognition. Thirty-channel EEG is decomposed into six wavelet components and represented as raw and Euclidean-aligned views. Each view employs a shared convolutional trunk with component-specific normalization, temporal attention, and component attention. A larger dual-view teacher provides offline supervision during unlabeled target-subject adaptation, and only the compact two-view student is retained for inference. Leave-one-subject-out experiments on 2,022 balanced three-second EEG epochs from 11 subjects yield 83.400% mean accuracy, 83.330% macro-average F1 score (macro-F1), and 89.580% mean area under the curve (AUC). Compared with the non-distilled shared-component student, the proposed method gains 4.37 percentage points in accuracy while retaining 4,112 deployed parameters. It reduces parameters by 82.600% and median batch-one graphics processing unit (GPU) latency by 74.900% relative to the teacher. Attribution-guided occlusion further supports the faithfulness of the learned frequency and channel evidence. These results show that compact multi-component EEG modeling can preserve cross-subject discrimination when coupled with unlabeled target adaptation and explicit decision evidence.

**Keywords**

Driver drowsiness detection, Electroencephalography, Euclidean alignment, Knowledge distillation, Subject adaptation

## Introduction

Drowsy driving remains a material road-safety problem because vigilance declines gradually and its observable manifestations vary among drivers. Camera-based systems are convenient, but eye closure, facial appearance, illumination, and driving posture are indirect correlates of neural state. EEG provides a non-invasive, high-temporal-resolution measurement of vigilance-related neural dynamics and has therefore been widely used in sustained-attention driving paradigms [1]. Its practical use, however, is constrained by low signal-to-noise ratio, large subject-to-subject variability, and the need to decide on a resource-constrained sensing platform.

Recent work makes clear that the field has moved beyond selecting a classifier for a fixed dataset. On the algorithmic side, generative augmentation has been used to relieve class imbalance and data scarcity [2]. Cross-scenario/cross-subject adaptation combines adversarial representation learning with conditional distribution alignment, and cross-dataset recognition has adopted entropy-guided robust feature adaptation [3,4].

In 2025, Data-Model Parallelism (DP-MP) formulated cross-subject fatigue detection through prototype-aware adversarial learning and mix-up pairwise learning, explicitly addressing biological individual differences and noisy labels [5]. These studies confirm the importance of distribution shift, but the resulting training pipelines often involve adversarial branches, pair construction, or a substantial teacher model. They do not by themselves resolve the deployment question: What compact model? What evidence can be retained once adaptation is complete?

Model compactness is no longer a secondary concern. EEGNet showed that temporal, depthwise spatial, and separable convolutions can yield effective small EEG classifiers [6]. Most directly, Feng combined a separable

convolutional neural network (CNN) with a differentiable Gumbel-softmax channel-selection layer for real-time cross-subject drowsiness recognition, jointly optimizing electrode selection and classifier parameters [7]. In a separate embedded EEG drowsiness system, discrete wavelet transform (DWT)-based processing and a reduced electrode set were implemented on a Zynq system-on-chip, illustrating that accuracy, energy, and hardware resources must be considered together [8]. At the same time, channel selection and explanation are necessary to determine whether a model is exploiting physiologically plausible information rather than merely benefiting from a high-dimensional input. Thus, a useful cross-subject method should jointly account for subject shift, frequency-specific structure, online model size, and the faithfulness of its explanations. The public benchmark analyzed here contains 30-channel, 3-s EEG epochs from 11 subjects [9]. On this benchmark, the multi-channel InterpretableCNN reported 78.350% mean leave-one-subject-out (LOSO) accuracy, while a compact single-channel CNN reported 73.220% [10,11]. ID3RSNet subsequently reported 74.720% from one channel on an unbalanced version of the data, that number is not directly comparable with the balanced evaluation used here [12]. These studies establish that compact and interpretable EEG models are viable but leave room to investigate unsupervised target adaptation and efficiency in a unified protocol.

Therefore, it should combine three complementary mechanisms. Euclidean alignment (EA) to reduce covariance mismatch using unlabeled target data [13,14]. A within-view shared multi-band trunk avoids one neural network per frequency component, and knowledge distillation (KD) transfers the posterior structure of a larger component ensemble to the compact model [15].

The contributions are as follows: (1) Developing a dual-view, within-view shared-band EEG network with band-specific domain normalization and explicit temporal and band attention. (2) Using offline teacher-assisted adaptation on an unlabeled target-subject batch, retaining only two compact students for inference. (3) Reporting LOSO accuracy, macro-F1, AUC, latency, controlled ablations, and both attribution maps and deletion-based attribution faithfulness.

The protocol is deliberately described as transductive, not calibration-free: The held-out subject's complete unlabeled batch is used for EA and adaptation, while its labels are withheld until evaluation.

## Materials and methods

### *Dataset and evaluation protocol*

The analyzed Matrix LABORATORY (MATLAB) dataset comprises 2,022 balanced epochs (1,011 alert and 1,011 drowsy), acquired from 11 subjects. Each epoch has 30 scalp channels, 384 samples, a sampling rate of 128 Hz, and a duration of 3 s. The supplied labels are 0 for alerts and 1 for drowsy. It uses the supplied preprocessed epochs without additional resampling or temporal cropping. The channel order follows the accompanying authors' visualization code: Fp1, Fp2, F7, F3, Fz, F4, F8, FT7, FC3, FCz, FC4, FT8, T3, C3, Cz, C4, T4, TP7, CP3, CPz, CP4, TP8, T5, P3, Pz, P4, T6, O1, Oz, and O2.

LOSO is used throughout: All samples from one subject form the test set and all remaining subjects form the labeled source set. Every subject is held out once. Source training never uses target labels. The main score is the unweighted mean of the 11 subject-level scores, preventing subjects with more epochs from dominating the result. Accuracy, macro-F1, and Receiver Operating Characteristic (ROC) AUC are computed for each held-out subject. A pooled out-of-fold (OOF) score is additionally reported for the ROC curve and confusion matrix. All reported training runs use seed 7. The 95.000% bootstrap intervals resample the 11 subject scores and therefore quantify subject-level, rather than multi-seed, variation.

### *Dual-view multi-band student*

Figure 1 summarizes the signal path. Let $X\_iinR\^C\ x\ T$ denote epoch $i$, where $C = 30$ and $T = 384$. A five-level db4 discrete wavelet transform (DWT) reconstructs six components at the original temporal length: A5 (0-2 Hz), D5 (2-4 Hz), D4 (4-8 Hz), D3 (8-16 Hz), D2 (16-32 Hz), and D1 (32-50 Hz). Each view is consequently a tensor $VinR\^B\ x\ C\ x\ T$ with $B = 6$. The raw component tensor is processed in parallel with an EA component tensor. For subject $s$, EA uses the inverse square root of the time-normalized mean trial covariance,

$$\bar{R}_s = \frac{1}{N_s T} \Sigma_{i=1}^{Ns} X_i X_i^T \quad (1)$$

$$\tilde{X}_i = \bar{R}_s^{-1/2} X_i \quad (2)$$

where $N_s$is the number of epochs for subject$s$. Before inversion, $\bar{R}_s\^ - 1/2$ by eigendecomposition and floor

each eigenvalue at $10\hat{}6\,max\,(mean(lambda),1)$ is obtained. The transformation is applied independently to every subject. In each LOSO fold, the target covariance is estimated from all unlabeled target epochs only. Source and target trials are never jointly normalized. This makes the preprocessing transductive, consistent with the subsequent adaptation setting.

Each view has its own EEGNet-style trunk. Weights are shared across the six DWT components within that view, but not across the raw and EA views. The trunk applies to eight temporal filters of length 64, a depthwise spatial convolution covering all 30 electrodes with depth multiplier two, average pooling by four, and a depthwise-separable temporal convolution of length 16 followed by average pooling by eight.

Exponential linear unit (ELU) activation and dropout 0.35 follow the spatial and separable blocks to introduce nonlinear representation capability and mitigate the risk of network overfitting during training. Therefore, each band yields F = 16 feature maps at L = 12 reduced temporal positions. Band-domain batch normalization is deliberately not shared: Each component has its own running mean, running variance, scale, and bias, which permits frequency-dependent distribution adjustment without duplicating convolutional kernels and preserves independent statistical characteristics across different frequency bands.

For band $b$, let $M_binR\hat{}F\ x\ L$be the final feature map. Learned temporal query $q\_tinR\hat{}F$ forms $alpha\_b,l = softmax\_l(q\_t\hat{}TM\_b,:,l/sqrtF)$ and the band feature $h\_b = sum\_lalpha\_b,lM\_b,:,l$ . A second learned query $q\_b$ and a learned band prior $r\_b$ produce $beta\_b = softmax\_b(q\_b\hat{}Th\_b/sqrtF + r\_b)$.

The representation is $h = sum\_bbeta\_bh\_binR\hat{}16$ which is mapped to two logits by a linear head. At inference, raw and EA posterior probabilities are fused by an equal weight mean. This construction makes attention weights available directly from the decision path rather than deriving them only after training.

For source training, the student minimizes a supervised loss comprising the fused cross-entropy and an auxiliary per-band cross-entropy.

$$L_{sr}c = LCE(\hat{y},y) + 0.2\,LCE(\hat{y}band,y) \quad (3)$$

The auxiliary term encourages every band representation to remain class-discriminative before attention fusion. Both views are trained with Adam ( $beta\_1 = 0.9$ , $beta\_2 = 0.99$), a learning rate of$10\hat{} - 3$, batch size 50, dropout 0.35, and 50 epochs. A maximum-norm constraint is applied after each source-training update.

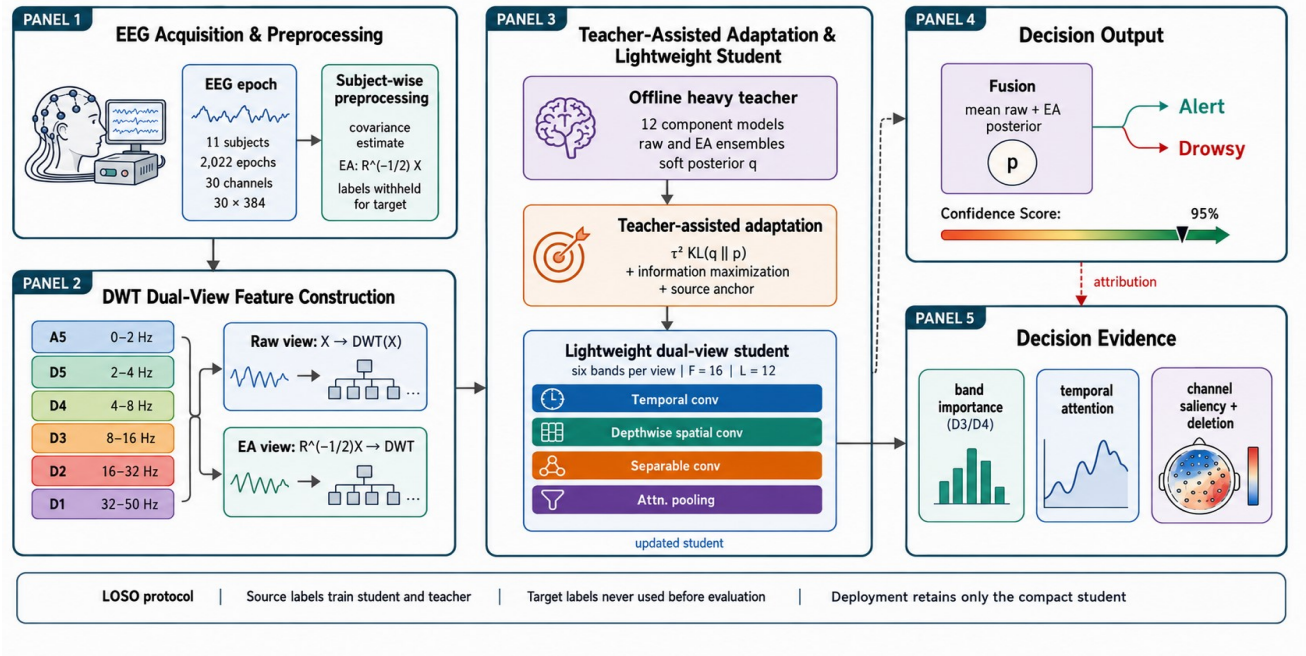


Figure 1. End-to-end workflow of Lite-DV-EAIM-KD.

Each 30-channel, 3-s EEG epoch is processed through subject-wise EA and a six-component DWT filter bank. Raw and EA component tensors are processed by the lightweight dual-view student, which shares convolutional weights across frequency components within each view and combines temporal and band attention. A 12-model teacher is used only during offline, transductive adaptation on the held-out subject's

unlabeled batch. At deployment, only the compact student is retained to fuse alert/drowsy posteriors and produce band, temporal, and channel evidence.

***Teacher-assisted target adaptation***

The offline teacher comprises two six-component EEGNet ensembles, one raw and one EA, totaling 12 independently trained component models. It is not retained for deployed inference. For target epoch $i$ in view $v$, its corresponding ensemble supplies a two-class posterior $q\_i,v$ . The compact student produces logits $z\_i,v$ and posterior $p\_i,v = softmax(z\_i,v)$ . During offline target-subject adaptation, the student objective is:

$$L = T^2 DKL(softmax(log\ q/T) \parallel softmax(z/T)) + 0.01\ LIM + \lambda Lanchor \quad (4)$$

Where $T = 2$ and the teacher posterior is temperature-softened in log-probability space because the stored teacher output is a probability rather than a logit. For a target batch of $n$ epochs, $L\_IM = H(p\_i) - H(n\hat{}-1 sum\_ip\_i)$, averaged over epochs in its first term. It reduces uncertain individual predictions while preventing the trivial one-class solution through marginal prediction diversity.

The anchor is the mean parameter-wise, squared deviation from the source model, $L\_anchor = mean\_j\ ||theta\_j - theta\_src,j\backslash rVert\_2\hat{}2/ |theta\_j|$, with $lambda = 10\hat{}-3$. No target label is used in Eq. eq: loss. We optimize all student parameters for 100 Adam steps at $10\hat{}-3$, then average the two adapted view posteriors. This protocol is explicitly transductive: The complete unlabeled target batch is available for EA and adaptation, whereas no target labels are available.

***Baselines, metrics, and efficiency measurement***

The external literature baselines in Table 1 are reported only when their dataset and protocol are sufficiently close to the balanced benchmark. Their original input configurations are preserved in the table: The compact CNN is single channel, whereas InterpretableCNN and the present method use 30 channels. Consequently, those numbers provide context rather than a controlled head-to-head comparison. Controlled comparisons use the same LOSO folds, source-training data, DWT components, and lightweight architecture. The internal variants are Raw + CNN-self-attention-BiLSTM (CSBN), EA + IM, dual view without KD, KD raw, KD EA, and the full KD dual-view model. The heavy Dual View-EAIM teacher is also evaluated on the same folds. Accuracy is the fraction of correctly classified epochs, macro-F1 is the unweighted mean of class-wise F1 scores, and AUC is computed from the fatigue posterior. For paired comparisons we use each held-out subject as one observation. Reported p values therefore test consistency of an effect across subjects, rather than independence of individual epochs. Batch-one latency is measured after 30 warm-up calls and over 200 repeated calls on a desktop GPU, and median latency is reported. The DWT and EA preprocessing times are excluded so that the measured value isolates neural-network inference. Parameter counts include the two deployed student views but exclude the offline teacher.

***Interpretability and statistical analysis***

Band importance is obtained from learned band attention. Channel evidence uses gradient-times-input saliency, $|partial\ p\_y/partial X\ o\ X|$, averaged across time and normalized within each trial. Time evidence combines 12-bin temporal attention with time-resolved saliency. To obtain joint frequency-spatial evidence, saliency is averaged over time before normalization over the full band-channel matrix. All class-conditioned attribution summaries use correctly classified trials only. Faithfulness is assessed with cumulative deletion. For each trial, bands or channels are ranked by attribution, the top $k$ components are zeroed in both views, and the decrease in the original predicted-class confidence is compared with equal-size random masks. By using $kin\backslash 1,2,3,4,5,6\backslash$ for bands and $kin\backslash 1,3,5,10\backslash$ for channels, it is found that four random masks are averaged per trial. We report one-sided Wilcoxon signed-rank tests across subjects. Explanation stability is quantified by pairwise Spearman correlations of subject-level attribution ranks and Top-$k$ overlap, separately for alert and fatigue. Raw-EA complementarity is measured by sample-level Spearman correlation between the two view-specific channel-saliency vectors.

**Experimental results**

***Main performance and efficiency***

Table 1 reports the principal results. The full Lite-DV-EAIM-KD model reaches 83.400% mean LOSO accuracy, 83.330% macro-F1, and 89.580% mean subject AUC. Pooled OOF accuracy is 84.370% because the larger held-out subjects receive greater weight in that

statistic.

The full student exceeds the heavy teacher by 0.66 percentage points in mean accuracy. However, the paired bootstrap interval of the difference is [-0.01, 1.37] percentage points.

Thus, the evidence supports preserved performance with a slightly higher meaning, rather than a definitive claim of stable superiority over the teacher. Relative to the 30-channel InterpretableCNN, the proposed model improves the reported mean accuracy by 5.050 percentage points under the balanced benchmark. It also exceeds the 80.840% accuracy and 79.650% F1 reported by the IEEE Sensors Journal Gumbel-softmax CNN by 2.56 and 3.68 percentage points, respectively.

This is contextual rather than a controlled comparison:

The comparison with the teacher therefore remains the appropriate controlled test of model compression, as it isolates performance differences caused by parameter reduction rather than external confounding experimental factors. The student has 4,112 deployed parameters (2,056 per view), versus 23,640 for the teacher.

Although both paths evaluate all six DWT components and therefore have nearly identical multiply-accumulate counts, the shared vectorized implementation lowers median batch-one GPU latency from 10.68 ms to 2.69 ms, demonstrating that parameter-sharing optimizations can deliver substantial inference speed gains without altering the core computational workload.

This reduction reflects parameter sharing and parallel component processing, not a reduction in signal-preprocessing cost, since the core signal transformation procedures remain unchanged and still demand comparable computational overhead for input preparation.

DWT and EA are deliberately excluded from the latency benchmark and must be included in a future end-to-end embedded evaluation, which will offer more realistic runtime measurements under practical hardware deployment constraints for real-world application scenarios.

Table 1. LOSO performance and deployment cost on the balanced 2,022-epoch dataset.

| Method | Protocol | Accuracy (%) | F1 | AUC | Params. | Lat. (ms) |
|---|---|---|---|---|---|---|
| LogPower + GNB | CF LOSO | 72.680 | / | / | / | / |
| Compact int. CNN | CF LOSO, 1 ch | 73.220 | / | / | / | / |
| InterpretableCNN | CF LOSO, 30 ch | 78.350 | / | / | / | / |
| Gumbel-softmax CNN | CF LOSO, learned ch | 80.840 | 79.650 | / | / | / |
| Heavy Dual View-EAIM | Unlab. targets adapt | 82.740 | 82.670 | 89.16 | 23,640 | 10.68 |
| Lite shared band, no KD | Unlab. targets adapt | 79.030 | 78.910 | 85.53 | 4,112 | / |

***Ablation analysis***

The ablation variants isolate the effect of input view and target distillation, although they are component variants rather than a fully orthogonal factorial design. Such partial-orthogonal setup still enables us to disentangle key contributing factors without excessive experimental overhead. Figure 2 presents mean scores and per-subject accuracy from Raw + CSBN to a non-distilled dual view raises mean accuracy from 76.760% to 79.030%, indicating that the two representations provide useful complementary information even before teacher supervision. This observable performance gain confirms that multiview input itself brings discriminative benefits, independent of the knowledge transferred from pre-trained teacher models.

Distillation improves the dual-view student by a further 4.370 percentage points, with corresponding macro-F1 and AUC increases of 4.42 and 4.05 points. The full model has 10 wins, one tie, and zero losses relative to the non-distilled dual view over the 11 held-out subjects

(p=0.00253, one-sided Wilcoxon). The EA-only distilled variant is stronger than the raw-only distilled variant, but the full fusion remains best on the mean scores, supporting the use of both views.

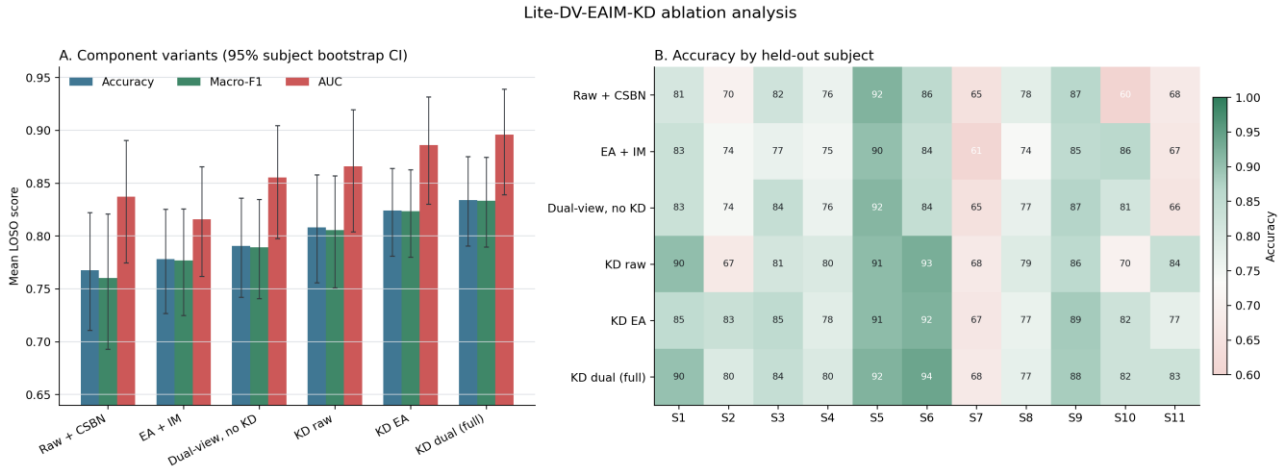


Figure 2. Ablation results (A: Mean accuracy, macro-F1, and AUC with 95.000% subject bootstrap intervals. B: Accuracy for each held-out subject. KD denotes teacher-assisted knowledge distillation).

Table 2. Ablation summary (mean subject scores).

| Variant | Accuracy (%) | Macro-F1 (%) | AUC (%) |
|---|---|---|---|
| Raw + CSBN | 76.760 | 76.020 | 83.710 |
| EA + IM | 77.790 | 77.690 | 81.560 |
| Dual view, no KD | 79.030 | 78.910 | 85.530 |
| KD raw | 80.800 | 80.550 | 86.570 |
| KD EA | 82.410 | 82.330 | 88.570 |

***ROC curve and confusion matrix***

Figure 3 visualizes the subject-level and pooled OOF ROC curves. Mean subject AUC and pooled OOF AUC use different weighting schemes and are reported separately. The left panel shows each subject in a thin line, their macro mean in a thick line, and the inter-subject standard deviation.

The full model has a mean subject AUC of 0.896. The broad dispersion is informative: Subject 7 has an AUC of 0.655, whereas subjects 5 and 6 achieve 0.985 and 0.997, respectively. Thus, subject variation remains the dominant limitation.

At the default argmax operating point, the pooled confusion matrix is 885 and 126; 190 and 821, with rows corresponding to true alert and true fatigue. This yields 87.540% alert specificity and 81.210% fatigue sensitivity. In safety-oriented deployment, an independently selected operating threshold may prioritize fatigue sensitivity, and the test-set threshold must not be tuned and re-reported as an unbiased result.

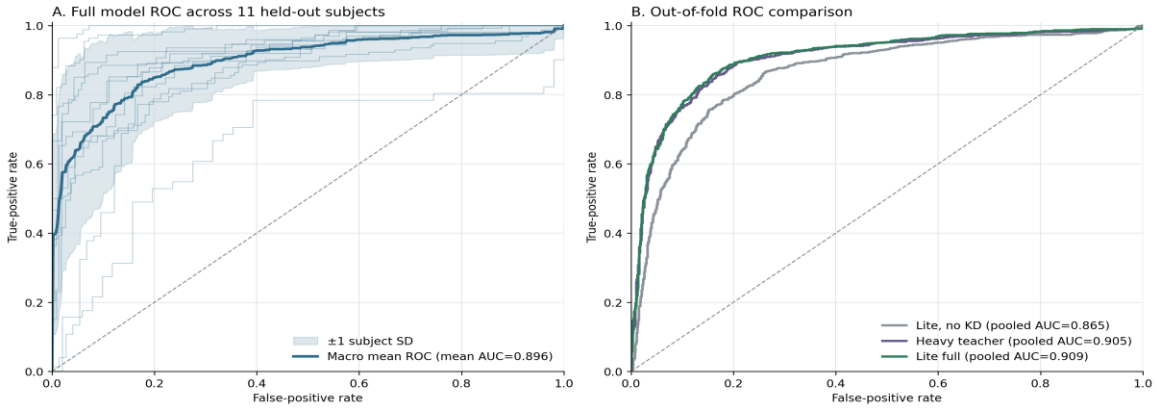


Figure 3. ROC analysis (A: The full model across the 11 held-out subjects. B: Pooled OOF ROC curves comparing the non-distilled student, the heavy teacher, and the full lightweight student).

***Attribution faithfulness, interactions, and stability***

The mean band attention is highest in D4 (4-8 Hz; 30.220% for fatigue and 28.280% for alert) and D3 (8-16 Hz; 27.420% and 26.850%, respectively). Fatigue-minus-alert temporal evidence is distributed across multiple windows, particularly in D4 and D3, rather than concentrated in a single transient. At channel level, the largest state-averaged saliency values are CPz (5.071%), Cz (4.869%), Oz (4.589%), FT7 (4.424%), Pz (4.116%), and T5 (4.075%).

The fatigue-minus-alert difference is strongest at Oz (+0.562 percentage points), O1 (+0.476), O2 (+0.397), CPz (+0.254), and Pz (+0.224), whereas FT7, F8, and F7 contribute relatively more to alert decisions. Figure 4 visualizes these values using the standard 10-20 montage, allowing readers to intuitively compare spatial attribution patterns across brain regions immediately. Interpolation is only a display aid, and all statistics use the 30 measured electrodes to avoid misleading inference from smoothed interpolated signals. The first two maps show state-specific attribution, and the third shows fatigue minus alert. These spatial plots help distinguish region-specific neural contributions under distinct mental states. The channel order and electrode locations follow the author's supplied visualization code and the standard 10-20 montage to guarantee reproducibility for subsequent replication or secondary analysis.

The band-channel interaction analysis in Figure 5 refines this interpretation. Each cell represents joint importance after temporal aggregation, calculated on correctly predicted trials and averaged across held-out subjects.

Alert evidence is strongest for D3-CPz (1.377%), D4-Cz (1.335%), and D3-FT7 (1.317%). Fatigue evidence is concentrated in D3-Oz (1.751%), D3-CPz (1.642%), D3-O1 (1.641%), D3-T5 (1.579%), and D3-Pz (1.540%). Thus, the model does not rely on a generic low-frequency effect: It combines D3/D4 activity with central-parietal and occipital electrodes.

We tested attribution faithfulness by progressively masking model-ranked components and comparing the corresponding confidence decrease with equal-size random masks. Table 3 and Figure 6 show that removing the two most salient bands decreases predicted-class confidence by 11.13 percentage points, versus 4.43 points for random bands (11/11 subject-level wins, p =0.00049). Model-ranked components produce a larger predicted-class confidence decrease than equal-size random masks. Shading denotes the standard error across 11 held-out subjects.

Removing the five most salient channels decreases confidence by 6.32 points, versus 4.04 for random channels (10/11 wins, p=0.00342). The larger targeted deletion curves support comprehensiveness of the explanations. They do not establish neural causality because masking creates out-of-distribution inputs.

Attribution stability was assessed using correctly predicted trials only. Across held-out subjects, band rankings have mean Spearman correlations of 0.942 for alert and 0.921 for fatigue, and the top two bands overlap completely.

Channel rankings are more subject-dependent (0.366 and 0.460, respectively), although their top five overlap remains 0.338. This pattern indicates stable spectral evidence together with subject-specific spatial evidence. Finally, the mean sample-level channel-saliency correlation between raw and EA views is 0.214 (0.199 for alert and 0.229 for fatigue), supporting their complementary rather than redundant contributions.

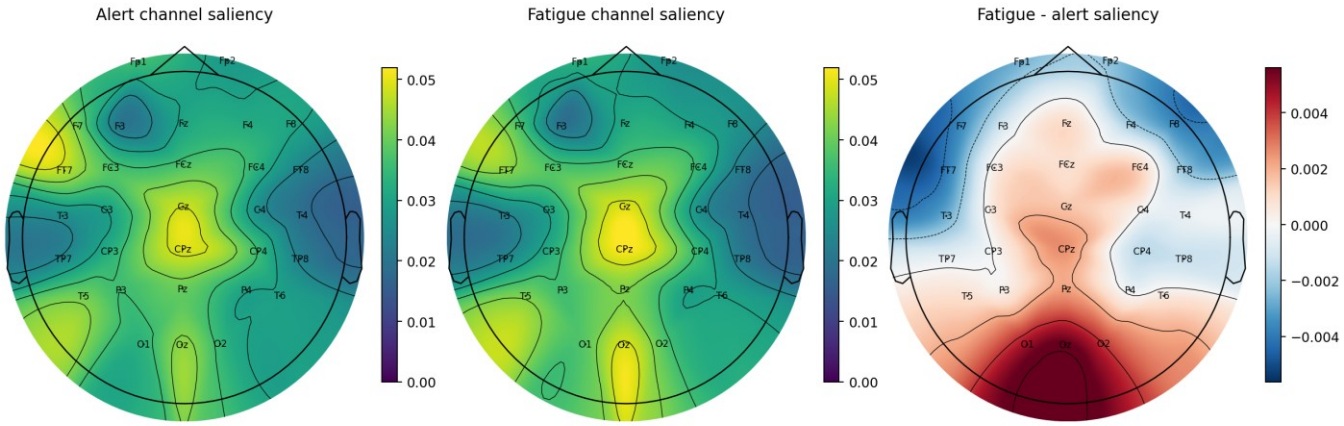


Figure 4. Normalized gradient-times-input channel saliency.

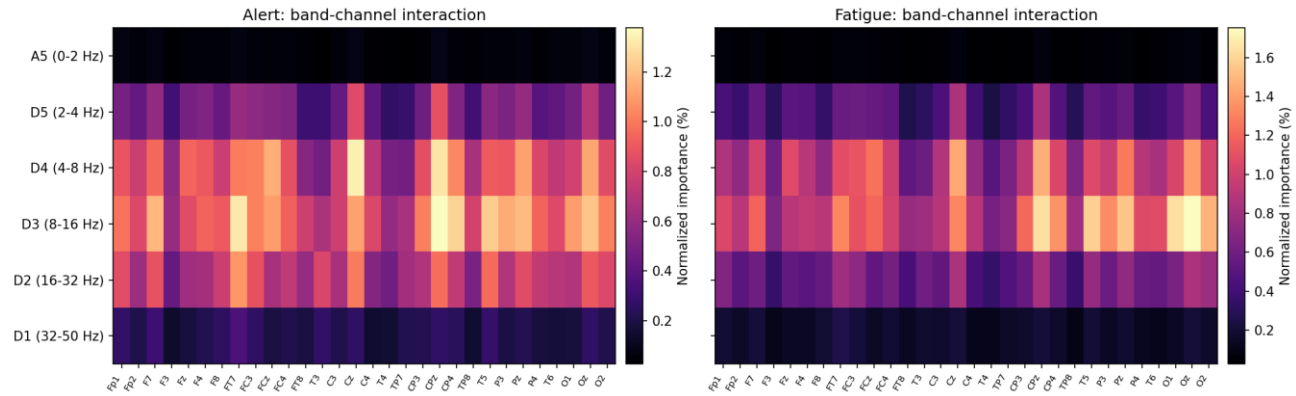


Figure 5. Band-channel interaction evidence from normalized gradient-times-input attribution.

Table 3. Cumulative deletion faithfulness.

| Mask | k | Ranked | Random | Wins | p |
|---|---|---|---|---|---|
| Bands | 2 | 11.13 | 4.43 | 11/11 | 0.00049 |
| Bands | 3 | 17.10 | 6.28 | 11/11 | 0.00049 |
| Channels | 5 | 6.32 | 4.04 | 10/11 | 0.00342 |
| Channels | 10 | 11.39 | 7.36 | 10/11 | 0.00098 |

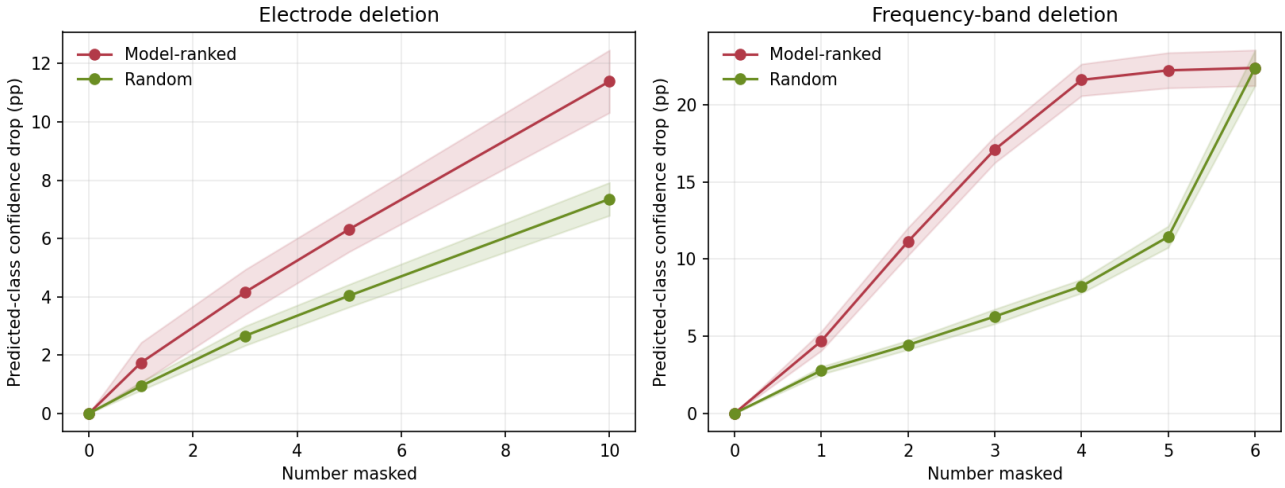


Figure 6. Cumulative deletion test of explanation faithfulness.

## Conclusion

This study shows that a compact two-view sensor model can retain and slightly improve the mean performance of a larger multi-model teacher on cross-subject EEG drowsiness recognition.

The most important empirical outcome is not only the 83.400% mean accuracy, but also the ablation pattern: A shared multi-band student without distillation loses accuracy, while target-aware KD restores and exceeds the teacher-level mean with 82.600% fewer deployed parameters.

The latency reduction is meaningful for online monitoring once DWT and EA are available. The result suggests that a large ensemble-like teacher can be useful as an offline adaptation instrument even when it is unsuitable for the final device.

Several limitations should guide interpretation. First, the adaptation protocol assumes access to an unlabeled target-subject batch and is therefore not calibration-free, and a true zero-shot domain-generalization setting may have lower performance. This makes comparison with the calibration-free Gumbel-softmax channel-selection CNN informative but non-controlled.

Second, results use one random seed, and the poor subject-7 performance indicates that robustness to extreme subject shift requires further work.

Third, external comparisons have different input channels, class balancing, and adaptation assumptions, so the table should not be treated as a standardized

leaderboard.

Fourth, the present latency excludes preprocessing and was measured on a desktop GPU, rather than an embedded platform. Finally, the spatial and frequency attributions are decision evidence for this model, not causal biomarkers, masking tests establish relative model dependence but do not establish a neurophysiological mechanism.

Future work will evaluate sequential test-time adaptation, robust threshold selection for fatigue sensitivity, embedded inference including preprocessing, and external validation across acquisition systems. It should also assess whether the stable D3/D4 spectral evidence persists when electrode number, sensor type, and road scenario vary. In conclusion, Lite-DV-EAIM-KD provides a compact and interpretable route to multi-channel EEG driver-drowsiness sensing, while making the target-adaptation assumption explicit.

**Funding**

This work was supported by the 2025 Anhui Provincial Quality Engineering Project for Higher Education (Grant No. 2025gspaikc077), the Key Natural Science Research Project of the Anhui Provincial Department of Education (No. 2025AHGXZK31227), and the Huaibei Vocational and Technical College Quality Engineering Project (Grant No. 2024jyxm-02).

**Data availability**

The balanced EEG driver-drowsiness dataset is publicly available from Fig share. The implementation used for the reported lightweight model, the LOSO fold results, trained deployment weights, and the analysis artifacts are retained with this study. The principal experiment is implemented in run_lightweight_interpretable_ experiment.py. Fold-level results are stored in results_lightweight_interpretable_v2/fold_results.

Json. Access to the local experiment workspace can be provided by the corresponding author on reasonable request.

**Reproducibility statement**

All 11 LOSO folds use the supplied epoch labels only for source training and final scoring. The reported primary run uses random seed 7, 50 source-training epochs, batch size 50, learning rate 10 ^ -3, and 100 target-distillation updates. The complete unlabeled target batch is used only for subject-wise EA and transductive adaptation. Hardware latency is reported separately from DWT and EA preprocessing, as detailed in Section III-D.

**Acknowledgements**

The author would like to show sincere thanks to those technicians who have contributed to this research.

**Conflicts of Interest**

The author declares no conflict of interest.

**References**

[1] Al-Shargie, F., Tariq, U., Hassanin, O., Mir, H., Babiloni, F., Al-Nashash, H. (2019) Brain connectivity analysis under semantic vigilance and enhanced mental states. *Brain Sciences*, 9(12), 363.

[2] He, L., Zhang, L., Sun, Q., Lin, X. (2024) A generative adaptive convolutional neural network with attention mechanism for driver fatigue detection with class-imbalanced and insufficient data. *Behavioural Brain Research*, 464, 114898.

[3] Luo, Y., Liu, W., Li, H., Lu, Y., Lu, B. L. (2024) A cross-scenario and cross-subject domain adaptation method for driving fatigue detection. *Journal of Neural Engineering*, 21(4), 046004.

[4] Yuan, L., Cui, J., Li, R., Zheng, Z., Siyal, M. Y., Yi, Z. (2024) Entropy-guided robust feature domain adaptation for electroencephalogram-based cross-dataset drowsiness recognition. *Engineering Applications of Artificial Intelligence*, 137, 109153.

[5] He, X., Li, H., Yu, P., Wu, H., Chen, B. (2025) DP-MP: a novel cross-subject fatigue detection framework with DANN-based prototypical representation and mix-up pairwise learning. *Journal of Neural Engineering*, 22(2), 026049.

[6] Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., Lance, B. J. (2018) EEGNet: a compact convolutional neural network for EEG-based brain - computer interfaces. *Journal of Neural Engineering*, 15(5), 056013.

[7] Feng, W., Wang, X., Xie, J., Liu, W., Qiao, Y., Liu, G. (2024) Real-time EEG-based driver drowsiness detection based on convolutional neural network with gumbel-softmax trick. *IEEE Sensors Journal*, 25(1), 1860-1871.

[8] Zayed, A., Trabes, E., Tarrillo, J., Ben Khalifa, K., Valderrama, C. (2025) Efficient embedded system for drowsiness detection based on EEG signals: features extraction and hardware acceleration. *Electronics*, 14(3), 404.

[9] Pillalamarri, R., Shanmugam, U. (2026) Enhancing cybersecurity via an emotion aware framework: leveraging EEG and speech signal fusion through correlation-based networks. *Journal on Information Security*, 2026(1), 11.

[10] Cui, J., Lan, Z., Sourina, O., Müller-Wittig, W. (2022) EEG-based cross-subject driver drowsiness recognition with an interpretable convolutional neural network. *IEEE Transactions on Neural Networks and Learning Systems*, 34(10), 7921-7933.

[11] Cui, J., Lan, Z., Liu, Y., Li, R., Li, F., Sourina, O., Müller-Wittig, W. (2022) A compact and interpretable convolutional neural network for cross-subject driver drowsiness detection from single-channel EEG. *Methods*, 202, 173-184.

[12] Feng, X., Guo, Z., Kwong, S. (2025) ID3RSNet: cross-subject driver drowsiness detection from raw single-channel EEG with an interpretable residual shrinkage network. *Frontiers in Neuroscience*, 18, 1508747.

[13] He, H., Wu, D. (2019) Transfer learning for brain - computer interfaces: a Euclidean space data alignment approach. *IEEE Transactions on Biomedical Engineering*, 67(2), 399-410.

[14] Junqueira, B., Aristimunha, B., Chevallier, S., de Camargo, R. Y. (2024) A systematic evaluation of Euclidean alignment with deep learning for EEG decoding. *Journal of Neural Engineering*, 21(3), 036038.

[15] Cowley, B. R., Stan, P. L., Pillow, J. W., Smith, M. A. (2026) Compact deep neural network models of the visual cortex. *Nature*, 652(8111), 947-954.