

Article

Robust Covert Spatial Attention Decoding from Low-Channel Dry EEG by Hybrid AI Model

Doyeon Kim and Jaeho Lee * 

Department of Software, Duksung Women's University, Seoul 01369, Republic of Korea; ehds2022@duksung.ac.kr

* Correspondence: izeho@duksung.ac.kr

Abstract

Background: Decoding covert spatial attention (CSA) from dry, low-channel electroencephalography (EEG) is key for gaze-independent brain–computer interfaces (BCIs). **Methods:** We evaluate, on sixteen participants and three tasks (CSA, motor imagery (MI), Emotion), a four-electrode, subject-wise pipeline combining leak-safe preprocessing, multi-resolution wavelets, and a compact Hybrid encoder (CNN-LSTM-MHSA) with robustness-oriented training (noise/shift/channel-dropout and supervised consistency). **Results:** Online, the Hybrid All-on-Wav achieved 0.695 accuracy with end-to-end latency ~ 2.03 s per 2.0 s decision window; the pure model inference latency is ≈ 185 ms on CPU and ≈ 11 ms on GPU. The same backbone without defenses reached 0.673, a CNN-LSTM 0.612, and a compact CNN 0.578. Offline subject-wise analyses showed a CSA median Δ balanced accuracy (BAcc) of $+2.9\%$ (paired Wilcoxon $p = 0.037$; $N = 16$), with usability-aligned improvements (error $0.272 \rightarrow 0.268$; information transfer rate (ITR) $3.120 \rightarrow 3.240$). Effects were smaller for MI and present for Emotion. **Conclusions:** Even with simple hardware, compact attention-augmented models and training-time defenses support feasible, low-latency left–right CSA control above chance, suitable for embedded or laptop-class deployment.

Keywords: dry-EEG; covert spatial attention (CSA); wavelet features; data augmentation; hybrid encoder; LSTM; multi-head self-attention (MHSA); filter-bank common spatial pattern (FBCSP); Riemannian tangent space; subject-wise validation

1. Introduction

Wearable, low-channel dry electroencephalography (EEG) enables mobile brain–computer interfaces (BCIs) but faces inherent challenges in low signal-to-noise ratio, non-stationarity, and inter-subject variability. Compared with wet electrodes, dry sensors markedly shorten preparation time yet introduce electrode–skin impedance variability and increased motion susceptibility [1,2]. Portable headsets (e.g., Muse [3]) have demonstrated feasibility for event-related potentials (ERPs)—notably P300 [3] in oddball/reward paradigms—and N400 [4] in semantic processing tasks; effects are detectable though often smaller than with lab-grade systems [3,4]. Beyond ERPs, spectral markers (e.g., alpha/beta band activity) are measurable with low-density, consumer-grade headbands, supporting attention-related monitoring under practical constraints [5]. These findings delineate both the potential and the limits of dry, few-channel recordings for cognitive BCI.

Despite these constraints, dry, low-channel headsets offer clear advantages in accessibility and cost, enabling rapid setup and home-/field-based deployments for frequent use cases in assistive technology and daily-life applications [6]. Materials advances and



Academic Editor: Gianni D'Angelo

Received: 11 November 2025

Revised: 24 December 2025

Accepted: 27 December 2025

Published: 30 December 2025

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

flexible or (semi-) dry electrodes continue to improve comfort and contact stability, mitigating impedance and motion issues without forfeiting usability [4,5]. Accordingly, our study targets covert spatial attention decoding with a 4-channel dry EEG, emphasizing leakage-safe preprocessing, augmentation, and compact attention-augmented models to enhance robustness under real-world conditions.

Consider a hands-free joystick controlled by EEG: covert spatial attention (CSA) is a stringent but practical target under the constraints of dry, low-channel EEG. Posterior alpha-band activity exhibits contralateral (retinotopic) lateralization—attending to the left (right) hemifield yields decreased alpha power over right (left) occipito-parietal sites—providing an index of the locus and timing of attention [7–9]. Because our four-channel dry montage does not include occipito-parietal electrodes, we interpret such canonical posterior patterns as providing theoretical motivation, while expecting only attenuated and indirect sensitivity in our recordings. However, these effects are subtle at the single-trial level and heterogeneous across participants, with variability in alpha topography and effect size, and decoding relies on small amplitude differences and temporally narrow windows [8,10,11].

Accordingly, our objective is not to replicate the canonical occipito-parietal topography with a sparse consumer montage, but to test a deployment-motivated feasibility question: whether a commodity, low-setup, four-sensor dry headset can provide binary left–right control in a gaze-independent paradigm. Under this constraint, any sensitivity to CSA-related cortical dynamics is expected to be attenuated and indirect, potentially reflecting a mixture of distributed neural effects and volume-conducted activity rather than direct posterior retinotopic alpha sources.

These challenges are amplified in a 4-channel dry-EEG setting: higher and more variable skin–electrode impedance, motion/electrooculogram (EOG) susceptibility, and limited spatial sampling suppress signal-to-noise ratio (SNR), and they can complicate cross-subject generalization [6,12,13]. Even in BCI paradigms where dry electrodes succeed, performance often shows large inter-individual spread, underscoring the need for robust preprocessing and models tailored to few-channel data [14].

Nevertheless, successful CSA decoding would enable concrete applications of gaze-independent, hands-free control: communication spellers and yes/no interfaces for users with impaired eye control, bedside assessments in disorders of consciousness, and attention-steered human–computer interaction in environments where manual or gaze input is impractical [15–18]. Our work therefore targets CSA under a 4-channel dry-EEG constraint and emphasizes leakage-safe preprocessing, augmentation, and compact attention-augmented architectures to address these realities.

Methodological pitfalls can inflate reported accuracies in EEG decoding. In particular, treating overlapping windows as independent samples creates strong temporal autocorrelation between train and test folds when trials are segmented (e.g., 2 s windows with 1 s shift), leading to leakage and optimistic performance estimates. Empirically, *k*-fold cross-validation (CV) on windowed EEG can overestimate true accuracy by large margins compared to trial-/block-wise splits, and “segment-based” holdout can yield near-perfect test scores that collapse to chance when evaluated on previously unseen subjects [19,20]. These issues generalize across passive-BCI settings and have been highlighted as a systemic form of data leakage in machine learning (ML)-based science [21].

A second, complementary problem is reliance on theoretical chance levels (e.g., 50% for two classes) for significance testing. With small sample sizes, the null distribution of accuracy is wide; accuracies far above 50% can occur by chance, inflating *p*-values if one compares against a fixed theoretical level instead of an empirical (permutation-based or exact-binomial) null [22]. Accordingly, we adopt trial/subject-wise CV and label-

permutation testing (with folds preserved) to obtain calibrated p -values and confidence intervals [19,20,22].

For reporting under balanced two-class designs, balanced accuracy (macro recall)—the mean of per-class recalls—is a robust summary that down-weights class-imbalance and variance from window counts. It is well-defined for binary and multiclass settings and widely recommended for small, imbalanced biomedical datasets [23,24].

To address these challenges, we adopt a deployment-oriented pipeline that explicitly targets dry-EEG constraints—variable electrode–skin impedance, motion susceptibility, and occasional channel loss—using a multiresolution front end, compact encoders, and robust training. In parallel, we instantiate filter bank common spatial pattern (FBCSP) and Riemannian tangent-space with linear discriminant analysis (LDA) pipelines as subject-wise baselines under the same four-channel dry-EEG constraints to quantify how far well-established shallow methods can go. First, we extract discrete-wavelet features and apply leakage-safe normalization (fit on training folds only), which stabilizes non-stationary spectra and reduces contamination across folds; multiresolution representations are well-established for EEG while fold-wise preprocessing mitigates known leakage modes in decoding pipelines [1,12,25]. Second, we employ strong training-time augmentation matched to dry-EEG nuisances—(i) amplitude noise/scaling to mimic impedance drift, (ii) partial channel dropout to simulate transient contact loss, and (iii) small temporal shifts to emulate movement and timing jitter—following evidence that noise/warp/jitter policies and channel-corruption training improve generalization in low-channel EEG [26–28]. Third, we use a lightweight Hybrid encoder that stacks multi-head self-attention (MHSA) atop an LSTM backbone, so the model can (i) retain sequence memory on short windows and (ii) re-weight informative time–channel patches when some channels degrade; recent EEG studies support compact CNN/LSTM hybrids with attention for improved robustness under limited channels and noisy settings [28–30]. Finally, our subject-wise evaluation with four electrodes and short windows shows a statistically significant median improvement over a clean baseline while maintaining a balanced operating point; Section 4 details ablations for wavelet features, augmentation policies, and encoder variants [20,31].

Figure 1 summarizes the online pipeline considered throughout this paper: raw EEG is streamed from the headset, preprocessed with leakage-safe normalization and wavelet feature extraction, then passed to a compact Hybrid encoder to decide the attentional direction (Left vs. Right). Beyond statistical defensibility, the design is deployment-oriented: by targeting on-device/near-device inference on embedded Systems-on-Chip (SoCs), it reduces round-trip latency and dependence on network connectivity, improving feedback immediacy, privacy, and reliability—key factors for user acceptance in wearable health and consumer BCI products [32–35]. Such edge execution aligns with non-medical/consumer BCI trajectories (gaming, AR/VR, hands-free interfaces) and health-adjacent use cases (home monitoring, assistive communication), thereby strengthening pathways from research prototypes to commercial devices [6,36]. Recent demonstrations of embedded EEG classification using lightweight CNN/LSTM families (e.g., EEGNet-class models on Jetson-class hardware) further support feasibility within the compute/energy envelope of wearables and mobile gateways [35,37].

The remainder of the paper is organized as follows. Section 2 reviews related work on wearable/dry EEG, CSA decoding, compact neural encoders, and robustness strategies (augmentation and leakage control). Section 3 details the dataset, preprocessing, feature families, and model configurations. Section 4 reports subject-wise results, ablations, cross-task comparisons, and benchmarking against classical FBCSP + LDA and Riemannian tangent-space LDA baselines. Section 5 discusses limitations and potential confounds (e.g., limited posterior coverage and residual ocular contamination) and outlines future valida-

tion directions. Section 6 concludes with implications for online systems and directions for personalization and adaptive normalization.

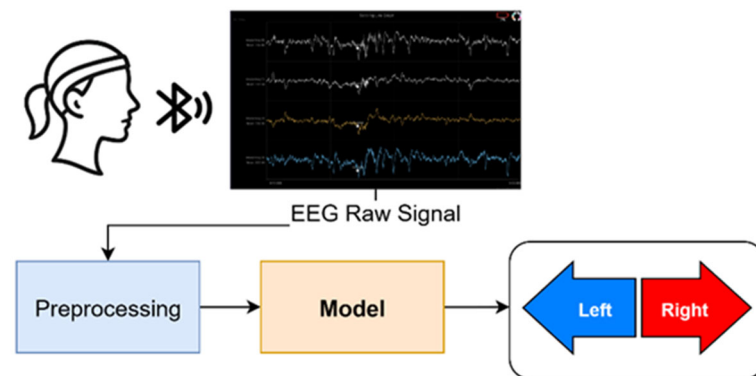


Figure 1. Online CSA decoding pipeline under four dry electrodes. Raw EEG is streamed (Bluetooth), preprocessed with leakage-safe normalization and wavelet features, fed to a compact Hybrid model (CNN, LSTM with MHSA), and yields a Left/Right decision.

2. Related Work and Motivations

Brain–computer interfaces (BCIs) translate neural activity into control signals for communication or device control. Among noninvasive options, EEG remains attractive for everyday use owing to low cost, portability, and millisecond temporal resolution [38]. Within EEG BCIs, exogenous paradigms deliver robust online performance by eliciting stereotyped responses with structured stimulation. In the P300 speller, a matrix of symbols is presented and rows/columns flash in randomized sequences; the attended item evokes a target ERP whose intersection (row $\times$ column) yields the selection [39,40]. In steady-state visual evoked potentials (SSVEPs), repetitive visual flicker at fixed frequencies drives narrowband responses; canonical correlation analysis (CCA) and its filter-bank variant (FBCCA), together with task-related component analysis (TRCA), constitute well-established baselines with extensive validation in online spellers [41–43]. These designs achieve high accuracy/information transfer rate (ITR) but can induce visual fatigue and require sustained gaze to flickering targets, motivating gaze-independent approaches [15,16,44,45].

By contrast, CSA is an endogenous paradigm that minimizes overt gaze shifts yet yields subtler and more variable neural signatures. A large body of work shows that posterior alpha-band activity lateralizes contralaterally to the attended hemifield, but the single-trial effect is modest and heterogeneous across participants and tasks [7,10,11,46]. Recent decoding studies further indicate that informative CSA signals are distributed across raw activity and alpha power with non-stationary dynamics, complicating generalization—especially in few-channel settings [10,11]. This motivates methods that (i) represent transient, burst-like spectral events rather than assuming sustained rhythms, and (ii) leverage compact sequence/attention encoders that can re-weight informative time–channel patches under channel sparsity and noise [47–49]. Our study builds on these insights to target CSA decoding with low-channel dry EEG, complementing exogenous P300/SSVEP systems where visual load or gaze demands limit long-term usability [15,16,44,45].

The neurophysiological basis of CSA features two principal EEG markers. First, N2pc is a posterior, contralateral ERP negativity peaking ≈ 200 – 300 ms after array onset, maximal over lateral occipito-parietal sites (e.g., PO7/PO8, P7/P8). It indexes spatially selective target selection among competing items: targets in the right (left) field elicit a larger negativity at left (right) posterior electrodes [50,51]. Second, posterior alpha-band (≈ 8 – 12 Hz) activity shows retinotopic lateralization during covert orienting: relative desynchronization (power decrease) contralateral to the attended hemifield and/or synchronization (power

increase) ipsilateral to unattended locations, providing a time-resolved index of attentional allocation [7,52,53].

Contemporary work nuances interpretation: lateralized alpha can reflect target-signal enhancement (priority at the attended location) and/or distractor suppression (inhibitory filtering at irrelevant locations); evidence suggests functionally separable contributions of these mechanisms across tasks and individuals [53,54]. In practice, single-trial alpha/N2pc effects are modest and heterogeneous, and informative patterns are distributed across raw activity and alpha power with non-stationary dynamics, helping to explain CSA's lower SNR and greater between-subject dispersion relative to exogenous ERPs/VEPs—especially under low-channel, wearable recordings [10,11,21,22].

Notwithstanding these challenges, CSA enables hands-free, gaze-independent interaction because it requires neither intense visual flicker nor peripheral motor activity. However, the weaker and more idiosyncratic neural effects motivate models and training strategies that emphasize robustness and subject-invariant representation learning under limited labeled data—a common constraint in out-of-lab EEG where per-subject recording time and trials are modest [55,56]. In this context, self-supervised pretraining and data-efficient fine-tuning are increasingly advocated to exploit abundant unlabeled EEG while reducing label demands [55,56].

From an application standpoint, a deployment-oriented CSA pipeline (compact models, on/near-device inference) can translate to lower latency, higher privacy, and better reliability in real-world products, enabling smart-home control, robotic/Unmanned Aerial Vehicle (UAV) teleoperation, and AR-based interfaces without burdensome gaze/flicker requirements [15,57–60]. Such scenarios directly support user autonomy—especially for people with restricted eye or hand movement—by offering gaze-independent communication and control pathways [15,57].

Wearable, four-channel dry electrode EEG—our target use case—reduces setup time and improves comfort but introduces higher and more variable skin–electrode impedance, increased motion sensitivity, and reduced spatial coverage. A comprehensive review details these materials/mechanics–signal-quality trade-offs and discusses how active front-ends can mitigate high-impedance degradation [1].

Validation with consumer four-channel headsets (e.g., Muse) shows feasibility for ERPs and spectral markers but also clarifies when gaps vs. research-grade systems emerge. With Muse-class devices, P300 in oddball/reward tasks and N400 in semantic tasks are detectable though typically smaller than lab systems [3,4]. Resting-state comparisons further report elevated low-frequency power and across-device spectral misalignment relative to research-grade amplifiers, reflecting fit/motion sensitivity and montage constraints [61]. In controlled targets, dry vs. wet systems yield similar P3b amplitude/topography and low-band spectra, yet wet shows a marginal single-trial classification edge (all electrophysiology metrics correlate $r \approx 0.54$ – 0.89) [13]. As a quantitative BCI example, a c-VEP system with dry electrodes achieved ~76% accuracy (≈ 46 bit/min) on average, whereas gel-based recordings under matched analysis reached ~96% (≈ 144 bit/min); even with matched electrode subsets offline, gel remained higher ($\sim 84\%$, ≈ 112 bit/min) [14].

These observations motivate our robustness-first, data-efficient design for CSA with four channels: leakage-safe normalization to stabilize variable impedances; augmentations that explicitly mimic dry-EEG nuisances (amplitude drift, partial channel loss, micro-shifts); and a compact Hybrid encoder resilient to limited spatial coverage. Importantly, improving dry electrode reliability brings deployment advantages—faster setup, better comfort, and higher user acceptance—reported in clinical-style comparisons where participants preferred the dry headset while rsEEG/ERP quality remained comparable overall [36,62].

On the algorithmic side, classical EEG pipelines established strong MI/ERP/SSVEP baselines via filter-bank spatial filtering and covariance-space modeling. FBCSP optimizes subject-specific sub-bands and remains a durable reference for MI decoding [63]. Riemannian approaches treat trial covariance as symmetric positive definite (SPD) manifold points and classify either by minimum distance to the Riemannian mean (MDM) or by tangent-space (TS) projection followed by a linear classifier such as LDA (TS-LDA), often yielding strong out-of-the-box generalization [64]. In this work, we adopt FBCSP + LDA and Riemannian TS-LDA as representative shallow baselines under the same subject-wise, four-channel dry-EEG constraints used for our Hybrid model.

The deep-learning era introduced ShallowConvNet/DeepConvNet [65] and EEGNet [48]. ShallowConvNet emphasizes band-limited temporal convolution followed by depthwise spatial filtering and log-nonlinearities well-suited to oscillatory rhythms. DeepConvNet stacks deeper temporal-spatial convolutions to learn hierarchical features. EEGNet adopts depthwise-separable convolutions to factor temporal and spatial filtering, achieving competitive performance with compact parameter counts across multiple paradigms.

More recently, convolution-Transformer hybrids such as CTNet [66] combine local convolutions with MHSA to capture long-range temporal dependencies and cross-channel relations; an attractive property for CSA with few channels [66]. Broader EEG-Transformer studies corroborate that attention mechanisms efficiently model temporal context and inter-channel interactions in low-channel regimes [67–69].

Our Hybrid encoder differs from prior CNN-Transformer hybrids (e.g., CTNet [66]) in three deployment-driven aspects tailored to four-channel dry EEG: (i) a lightweight LSTM backbone preserves short-window sequence memory before a single MHSA block, reducing parameter/compute overhead compared to stacking multiple Transformer encoders; (ii) wavelet, leakage-safe front-end stabilizes non-stationary spectra before learning; and (iii) training-time channel/noise/shift augmentations align the representation with dry-EEG nuisances. In our subject-wise CSA evaluation, this configuration yields statistically significant median gains over clean baselines (Section 4), indicating advantages under few-channel constraints.

Because artifact contamination is a dominant failure mode for few-channel dry EEG, prior work explored standardized, lightweight cleaning complemented by model-level robustness. PREP [70] addresses referencing and line noise reproducibly via line-frequency removal, bad-channel detection/interpolation, and (robust) average reference; Autoreject [71] learns per-channel peak-to-peak rejection thresholds via CV and repairs trials by local interpolation when neighbors are reliable; ICLabel [72] supports scalable independent Component level annotation (brain, eye, muscle, heart, line-noise, channel-noise, other) after independent component analysis (ICA). In low-density montages (≤ 32 ch), ICA-based steps are less reliable or infeasible, and average-reference/bad-channel mapping are less stable due to sparse spatial sampling; pipelines purpose-built for low-electrode EEG (e.g., HAPPILEE [73]) therefore avoid ICA, emphasize notch/high-pass, robust re-referencing, channel/segment rejection, and transparent quality control (QC) reports. In our four-channel setting, we thus favor notch + fold-wise normalization, trial/segment rejection over interpolation when neighbors are insufficient, and no ICA (hence ICLabel is not used), while relying on model-side robustness (Section 4) to absorb residual artifacts. These design choices follow recent evaluations noting method selection should depend on density and task, with automated pipelines (e.g., RELAX-Jr) emerging for low-channel data [74].

Complementary to cleaning, physiology-aware data augmentation can improve generalization in low-data regimes. Systematic reviews emphasize that augmentation gains are task- and dataset-dependent, with the largest benefits when perturbations reflect plausible nuisances rather than ad-hoc recipes [26,31]. Guided by dry-EEG failure modes, we employ

(i) amplitude scaling/noise to mimic impedance drift, (ii) small temporal shifts to emulate motion/timing jitter, and (iii) partial channel dropout to simulate transient contact loss—policies consistent with recent findings on corrupted-channel training and attention-based dynamic spatial filtering improving robustness in few-channel settings [27,28]. While GAN-based synthesis can help in some contexts, it often increases compute/complexity and risks distribution shift; our lightweight, parametric augmentations remain edge-friendly and physiologically grounded for four-channel dry EEG [75–77].

A complementary thread targets representation-level robustness. Self-supervised learning (SSL) pretrains an encoder on large unlabeled EEG to learn subject-/device-robust features that transfer with limited labels: a pretext objective drives invariance to benign variability (e.g., montage, amplitude, timing) while preserving task-relevant structure. Contrastive pretraining maximizes agreement between two augmented views of the same trial and pushes apart different trials, yielding embeddings that are stable across subjects and hardware [56,78]. BENDR [56] operationalized this with a Transformer backbone and contrastive objectives, improving cross-dataset and cross-device generalization after light fine-tuning.

Evaluation practice underpins trustworthy claims. Treating highly overlapping windows as independent inflates the effective sample size and can leak temporally correlated data across folds, overestimating accuracy; reliance on theoretical chance levels (e.g., 50%) further risks spurious significance on small samples [19,22]. For symmetric two-class designs, balanced accuracy (macro recall) is a robust summary and admits principled uncertainty estimates [23]. Finally, small-sample neuroimaging exhibits large error bars and unstable rankings under cross-validation, motivating subject-wise splits, leakage-safe normalization, and uncertainty-aware reporting (e.g., permutation p -values, confidence intervals) [79].

In summary, prior art indicates that (i) CSA is compelling for real-world, gaze-independent BCI yet harder to decode than MI/ERP/SSVEP because its endogenous markers (e.g., contralateral posterior alpha; modest single-trial effects) are weaker and more variable [7,10,11]. (ii) With four-channel dry EEG, higher/variable contact impedance, motion sensitivity, and reduced spatial sampling raise the premium on denoising and robustness; head-to-head studies and device comparisons further quantify when gaps to wet, research-grade systems widen [1,12,13,61]. (iii) Attention-enhanced, compact architectures—e.g., convolution–Transformer hybrids such as CTNet [66] and EEG-Transformer families [67–69]—better capture long-range temporal context and cross-channel relations under few-channel constraints, improving data efficiency and edge deployability. (iv) SSL offers a principled route to mitigate subject/device drift via contrastive pretraining on unlabeled EEG (e.g., BENDR [56]) with broad evidence for gains under label scarcity and noise; paired with rigorous, leakage-safe evaluation (subject-wise splits, empirical nulls), it guards against overestimated accuracies from overlapping windows and theoretical chance baselines [19,22,79,80]. Taken together, these insights motivate our focus on multiresolution features, attention-augmented compact encoders, and physiology-aware augmentation under subject-wise protocols tailored to four-channel dry EEG for CSA, with the downstream impact of lower-latency, privacy-preserving, and more user-acceptable edge BCI. In addition, we explicitly benchmark our Hybrid model against classical FBCSP + LDA and Riemannian TS-LDA baselines to quantify the remaining gap between shallow and compact deep approaches under the same dry, low-channel constraints [63,64].

3. Methods

3.1. Setup and Task

Recordings were acquired with a Muse-2 headband (InteraXon) sampling at 256 Hz using four dry sensors at TP9, AF7, AF8, and TP10 with a common reference/ground at

FPz: see Figure 2, consistent with the manufacturer’s specifications and prior technical descriptions of the Muse family [81–83]. Dry electrodes substantially reduce preparation time and improve user comfort; however, they exhibit higher and more variable skin–electrode impedance and greater motion susceptibility than gel systems, which increases non-stationary noise and brief contact-loss events [4,84,85]. Independent validations of the Muse platform have demonstrated feasibility for event-related potentials and spectral analyses (e.g., robust N200/P300 and frequency-band measures) using portable hardware [1,82], while comparative studies also document conditions where consumer-grade dry systems—including Muse models—lag research-grade amplifiers, motivating conservative preprocessing and robustness-oriented modeling in the present work [61,86]. Sixteen right-handed healthy adults (normal or corrected-to-normal vision) gave written informed consent in accordance with the Declaration of Helsinki; sessions were conducted in a dim, quiet room with a viewing distance of ~60–70 cm [87]. Live monitoring during acquisition used MuseLab v1.9.5 InteraXon [88], which offered oscilloscope-style inspection and logging; we cite it as software in the references list for reproducibility.

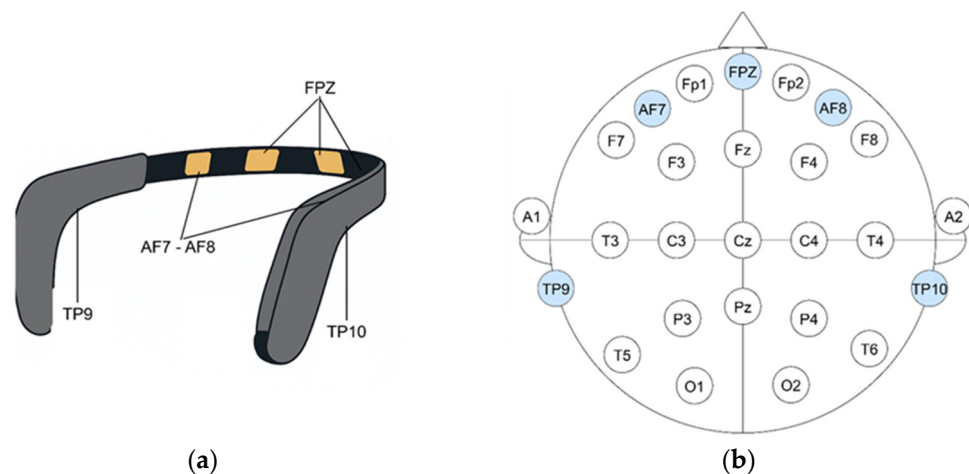


Figure 2. Dry electrode montage and corresponding 10–20 scalp locations. (a) Schematic of the four-channel dry-electrode montage (TP9/AF7/AF8/TP10) with the reference at FPz; (b) Corresponding electrode positions on the international 10–20 system, highlighting TP9/AF7/AF8/TP10 and the FPz reference.

We recorded three binary paradigms to probe complementary neural mechanisms and to stress-test generalization across effect sizes. CSA required attending left versus right while avoiding eye movements; in our protocol, a spoken cue (“left”/“right”) initiated a 7-s attention hold with eyes closed (15 trials per side; inter-trial rest ≥ 5 s). To reduce ocular and facial EMG contamination, participants kept their eyes closed throughout each CSA trial and were instructed to minimize blinks, jaw tension, and any intentional gaze shift. After the spoken cue, we allowed a short settling period before analyzing the sustained attention interval, aiming to avoid cue-evoked startle/blink transients and initial orienting microsaccades. Furthermore, unlike visual search tasks, the eyes-closed protocol minimizes sustained directional gaze shifts since there is no visual target to fixate, thereby reducing the likelihood of systematic EOG contamination in the sustained attention interval. During acquisition, the operator continuously monitored the AF7/AF8 channels—where ocular artifacts are most prominent—and repeated trials when clear blinks/saccades or motion bursts were observed. A canonical marker of CSA is posterior alpha lateralization contralateral to the attended hemifield; however, because our four-channel dry montage lacks occipito-parietal coverage, sensitivity to this canonical posterior pattern is expected to be attenuated and indirect. Classical visual-cue CSA also elicits the

N2pc ERP, but our eyes-closed auditory cues do not. We do not expect a phase-locked N2pc and therefore rely primarily on induced alpha dynamics [50,63,89,90]. Motor imagery (MI) trials presented a central arrow on a white background (left/right) cueing 5-s kinesthetic imagery of movement toward the cued side (21 trials per side; rest ≥ 5 s); MI is known to produce contralateral μ/β ERD with post-trial beta rebound (ERS), which typically yields stronger, more distributed spectral changes than CSA in low-channel settings [91,92]. The Emotion paradigm displayed 5-s images from two categories—pleasant “cute” animals versus unpleasant “disgust” insects—balanced across 10 items per class; this design follows established affective EEG paradigms using standardized pictures or film excerpts and commonly targets band-power modulations linked to valence/arousal [93,94]. For timing, symbolic cues appeared immediately (stimulus onset asynchrony (SOA) ≈ 0 ms) and epochs were time-stamped to cue onset to enable consistent windowing; in ERP-based BCIs, manipulating SOA can alter amplitude and accuracy, but here we fixed cue timing to reduce variability across subjects [95,96]. No online task-performance feedback was given during any paradigm (only instructions and rest prompts) to avoid feedback-induced learning or bias effects during data collection; feedback is known to modulate BCI performance and learning dynamics and was therefore excluded at acquisition time [97].

Live QC was performed with MuseLab v1.9.5 (InteraXon) [88], an oscilloscope-style viewer used to continuously inspect the four channels for line-noise drift, brief contact-loss events, ocular artifacts (blinks/saccades), jaw-clench EMG, and gross motion (Figure 3). Contaminated trials were immediately repeated, and QC notes were time-stamped to build an artifact log (event type/time/affected channel), a practice recommended in ERP/EEG reporting guidelines to minimize label noise and to support transparent post-hoc auditing [98,99].

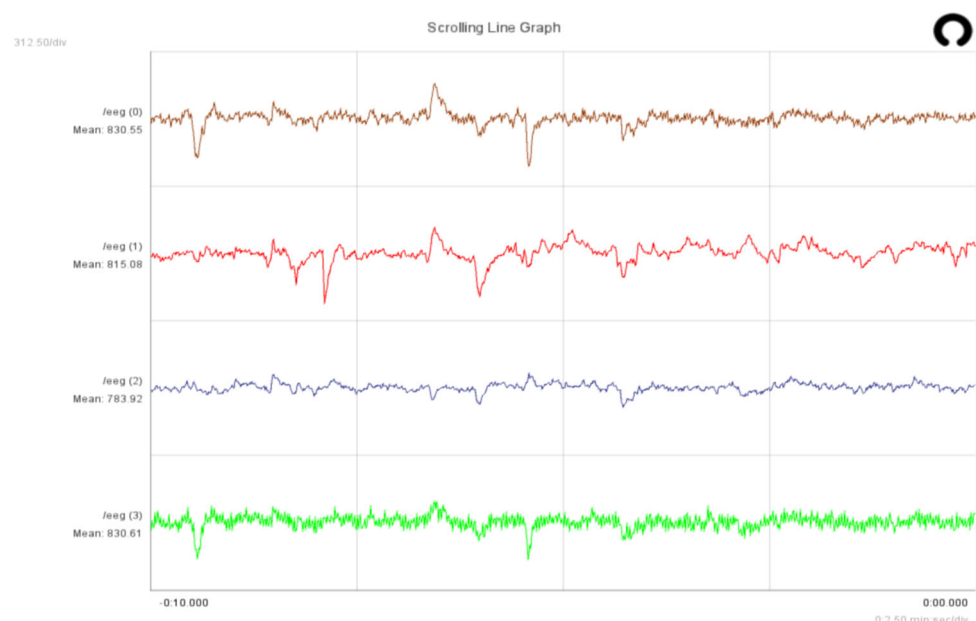


Figure 3. Live four-channel traces used for quality control during acquisition.

Raw EEG and QC logs were exported as CSV with subject, session, task, and trial identifiers; where applicable, the metadata layout followed EEG-BIDS conventions for events and timing, facilitating reuse and reproducibility (Figure 4) [100,101]. Automated artifact-rejection toolboxes—such as ICA-based identification of ocular/muscle components and Autoreject for bad-segment detection—are documented alternatives [71,102], but were not employed here given the low-channel dry-EEG setup and our preference for conservative, acquisition-time QC to avoid over-correction; instead, muscle and move-

ment contamination were mitigated primarily through prospective monitoring and trial repetition [98,103].

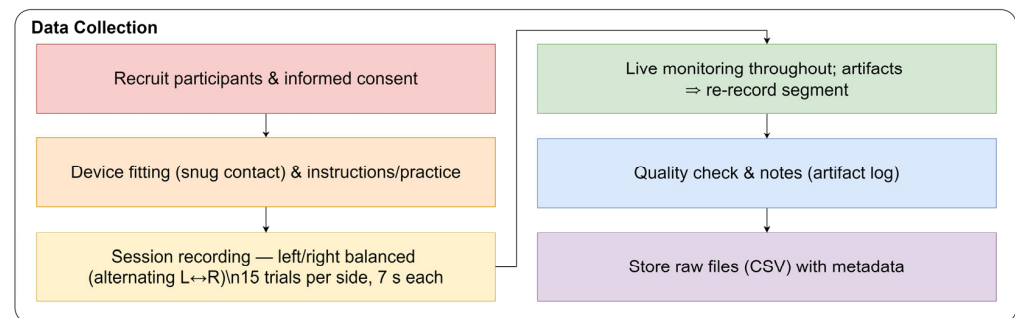


Figure 4. Data-collection workflow from recruitment to archival.

3.2. Preprocessing

Preprocessing was explicitly leakage-safe under subject-wise evaluation: any parameter that requires fitting—including normalization statistics, artifact-rejection thresholds, and any filter with learned or data-dependent state—was estimated only on the training subjects within a fold and then applied unchanged to validation/test subjects, preventing optimistic bias that often inflates performance in small-sample neuroimaging when pipelines “peek” at held-out data [79]. Each EEG channel was band-pass filtered with a 5th-order Butterworth design at 1–50 Hz by default; a narrow 50/60 Hz notch was applied only when line-noise peaks were detected (IIR notch, typical $Q \approx 30$). Offline processing used a zero-phase IIR implementation to avoid phase distortion, whereas the online stream used a phase-matched causal implementation with identical poles/zeros [104–107]. Filter order and transition bands followed electrophysiology guidance to minimize ringing and transient smearing of short-latency ERPs, with high-pass settings kept conservative to avoid artifactual peaks and latency shifts reported for aggressive cutoffs [104,108,109]. Taken together, these choices prioritize reproducibility (full reporting of filter type/order and implementation) and validity (bias-aware evaluation and artifact-aware design).

Reference estimation and line-noise stabilization followed the PREP pipeline in a form adapted to a four-channel montage. Specifically, we used PREP’s robust line-noise characterization to model 50/60-Hz sinusoids and harmonics and applied a narrow notch only when spectral peaks were detected; channels were screened for outliers using PREP’s amplitude/correlation heuristics, but global average re-referencing and PREP’s spatial interpolation were disabled to avoid topology distortion under sparse sampling [70]. Because a low-density montage makes aggressive spatial interpolation or re-referencing liable to smear lateralization (e.g., alpha asymmetry) and to bias topographies, we retained the device reference at FPz and limited re-referencing to diagnostics [48,49,72]. Clear outlier epochs were repaired or rejected using Autoreject (local mode), which fits channel-specific peak-to-peak thresholds by CV and, when possible, interpolates only the flagged sensors in that epoch; to preserve spatial information with 4 channels, we constrained the grid to $n_interpolates = \{1, 2\}$ and $consensus = \{0.2, 0.4, 0.6\}$, rejecting an epoch if $>50\%$ sensors were bad. Interpolation used spherical splines as implemented in the MNE-Python toolbox (version 0.24.0) [110], but we capped interpolation to at most two sensors per epoch because spline-based repair can inflate spatial correlations if over-used [71]. Independent component removal was not part of the default path: with few sensors, ICA decompositions are unstable and poorly constrained; when components were inspected for documentation, ICLabel assisted annotation only (no mandatory removal) [72]. As a result, residual ocular leakage cannot be completely excluded in this four-channel setting. We therefore interpret decoding performance as reflecting practical discriminability under wearable constraints,

and we explicitly discuss ocular confounds and needed future validation (EOG/eye tracking or additional posterior sensors) in the Limitations. Signals were epoched at cue onset and segmented into 2-s windows (512 samples at 256 Hz) with 50% overlap (1-s step), aligning with the feature-extraction settings used downstream.

Window length and overlap were selected on physiological and statistical grounds. A 2-s analysis window yields a Fourier resolution of ≈ 0.5 Hz ($\Delta f = 1/T$), which is adequate to isolate the alpha-band (8–13 Hz) while retaining sensitivity to transient ERD/ERS dynamics [91,111,112]. At 10 Hz, a 2-s window contains ~ 20 cycles, providing enough repetitions to average phase variability without washing out brief attentional modulations; this choice follows standard time–frequency guidance balancing cycle count against temporal precision [113]. A 50% overlap increases the number of effective samples for stable parameter estimation and maintains temporal continuity for downstream smoothing or streaming inference, consistent with classical Welch-type segmenting and common digital signal processing (DSP) practice [114,115]. Figure 5 summarizes the pipeline from raw CSV through leakage-safe preprocessing and windowing to model ingestion.

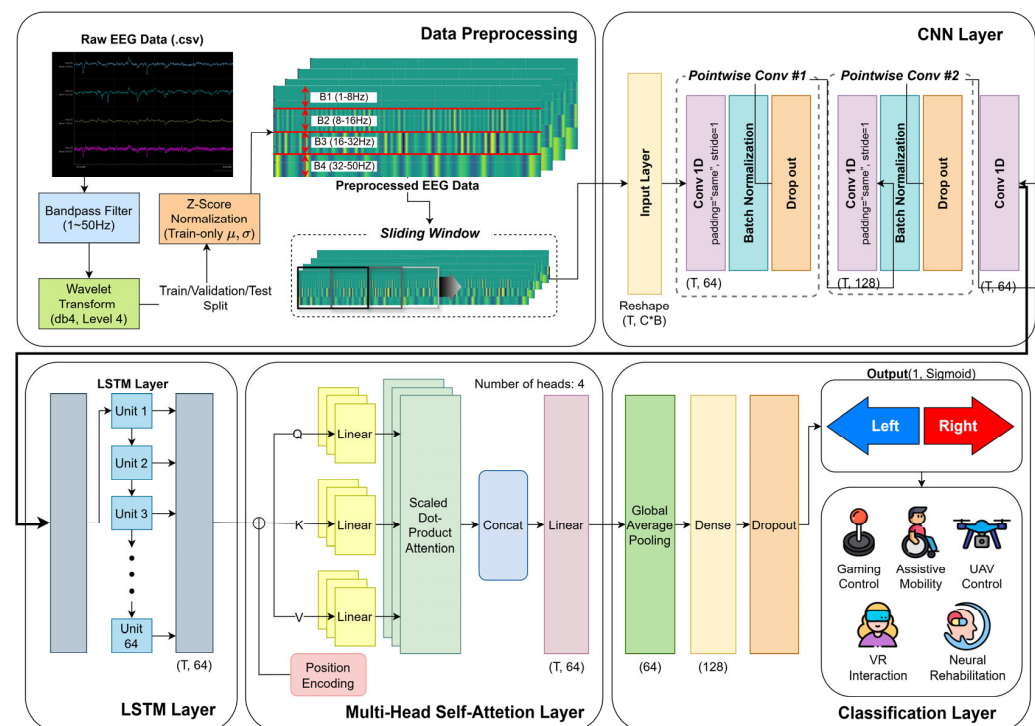


Figure 5. End-to-end pipeline: preprocessing, feature construction, compact CNN/LSTM/MHSA encoder, and classification.

3.3. Features

Alongside the raw windows, we computed a multiresolution representation using a level-4 Daubechies-4 discrete wavelet transform (db4-L4). With a sampling rate of 256 Hz and a 2.00-s analysis window, the resulting sub-bands align with rhythms implicated in CSA; canonical accounts emphasize posterior alpha lateralization, but under our four-channel montage these band mappings should be interpreted as coarse proxies for subject-dependent, band-limited transients rather than direct occipito-parietal topography [116]. Wavelet analysis is well-suited to such nonstationary, transient phenomena because its scale-dependent windows narrow in time for higher frequencies and widen for lower frequencies, enabling short-lived α/β bursts to be captured without sacrificing frequency localization at slower rhythms; by contrast, fixed filter banks impose uniform time support

and do not adapt window length across bands [117,118]. Figure 6 illustrates the stack layout.

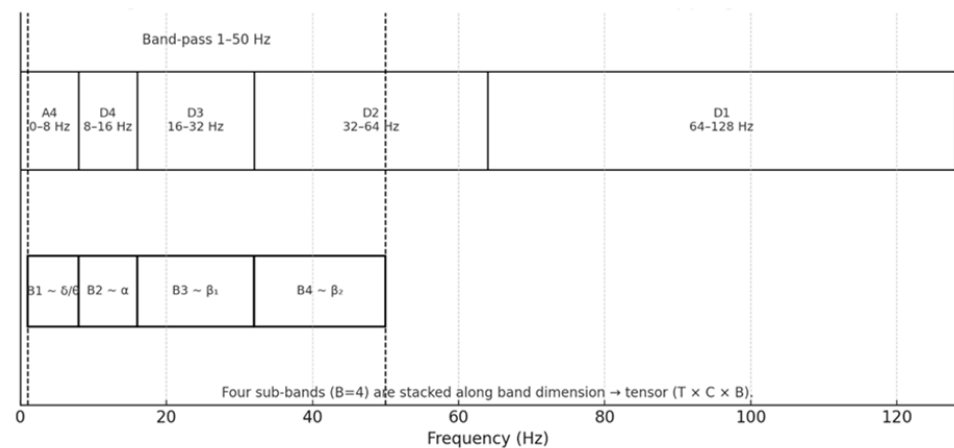


Figure 6. Wavelet sub-band layout (db4-L4) and tensor organization along time (T), channel (C), and band (B).

To preserve task-relevant oscillatory information while keeping the input compact, we first apply a 1–50 Hz band-pass filter and then perform a level-4 discrete wavelet decomposition (db4). The decomposition yields sub-band re-constructions corresponding to A4 (≈ 1 –8 Hz), D4 (≈ 8 –16 Hz; α), D3 (≈ 16 –32 Hz; low- β), and D2 (≈ 32 –64 Hz). Because the signals are pre-filtered at 50 Hz, the effective support of D2 is ≈ 32 –50 Hz (the 50–64 Hz portion is attenuated). The four reconstructed bands are stacked in the fixed order [A4, D4, D3, D2] and provided as input to the backbone.

Signals were band-pass filtered at 1–50 Hz by default using a 5th-order Butterworth design; offline processing used zero-phase filtering, whereas online streaming used a causal implementation [104–107]. Data were sampled at 256 Hz with the Muse-2 headset [81]. Windows were extracted with 2.00-s windows (512 samples) and 1.00-s hops (256 samples; 50% overlap), yielding a per-trial window count for a trial of length N samples (sliding-window scheme) as follows:

$$T = \left\lfloor \frac{N - 512}{256} \right\rfloor + 1 \quad (1)$$

The resulting wavelet volume per window had shape $(T \times C \times B)$, where C is the number of EEG channels and B is the feature depth per channel.

For each channel c and band $b \in \{“A4”, “D4”, “D3”, “D2”\}$, let $w_{b,c}[n]$ denote the db4-L4 reconstruction samples within a window containing N_b samples. We summarized each (b, c) with log-power and first/second-order statistics, where log-power was computed with $\varepsilon = 10^{-12}$ for numerical stability:

$$\begin{aligned} LP_{b,c} &= \log(\varepsilon + 1/N_b \sum_{n=1}^{N_b} w_{b,c}[n]^2) \\ \mu_{b,c} &= 1/N_b \sum_{n=1}^{N_b} w_{b,c}[n] \\ \sigma_{b,c}^2 &= 1/N_b - 1 \sum_{n=1}^{N_b} (w_{b,c}[n] - \mu_{b,c})^2, \end{aligned} \quad (2)$$

Log-energy features in the wavelet domain are widely used in EEG because they compress dynamic range and emphasize multiplicative power changes typical of oscillatory bursts, while mean/variance capture slow offsets and dispersion changes in co-efficient distributions that co-vary with micro-state transitions [119,120]. Concatenating

$\{LP, \mu, \sigma^2\}$ across the four sub-bands yields $B = 4 \times 3 = 12$ features per channel, so the per-window tensor is $(T \times C \times 12)$ [119,121].

Compared with fixed filter-bank methods (e.g., band-pass banks prior to spatial filtering), the discrete wavelet transform provides adaptive time–frequency resolution that improves localization of brief, burst-like CSA signatures in α/β while preserving frequency specificity for slower components [117,121]. This makes the wavelet stack a natural complement to the raw stream: the encoder can exploit precise temporal structure from raw windows or rhythm-aware summaries from wavelets, improving robustness to noise and small timing jitter [119].

Rationale for dB4-L4. Daubechies-4 offers compact support, near-orthogonality, and short filter length, yielding low-cost, leakage-resistant decomposition suited to short windows; level-4 supplies sub-bands that align with $\delta/\theta/\alpha/\beta/\gamma$ while keeping coefficient counts balanced for stable statistics [121]. In particular, α -band activity has been shown to track a two-dimensional spotlight of attention over time during spatial working memory maintenance, supporting the relevance of resolving α dynamics in attention-related decoding settings [122]. Alternative choices (e.g., Morlet continuous wavelets or superlet families) can provide excellent frequency localization or super-resolution when continuous maps are required, but they incur higher computational cost and redundancy; the “optimal” mother/level is known to be data- and goal-dependent [121,123]. We therefore report dB4-L4 as an efficiency-oriented default for CSA while acknowledging that other wavelets/levels could trade temporal versus spectral precision differently [118,121,123].

In addition to the wavelet stack, we used fixed filter-bank features to separate the effect of multiresolution analysis from the classifier. The same 2.00-s windows were decomposed with 4th-order zero-phase Butterworth bands at (1–4, 4–8, 8–13, 13–30) Hz, with an optional individualized alpha band (individual alpha frequency (IAF) ± 2 Hz) estimated from a Welch spectrum over training data. These band-pass volumes (with or without IAF) were either fed directly to the Hybrid encoder as a wavelet-free variant or passed to shallow baselines under identical leakage-safe preprocessing [37].

For FBCSP + LDA, sub-band trials yielded class-wise covariance matrices and common spatial pattern filters; band-wise log-variance features were concatenated and classified with LDA [63]. For the Riemannian TS baseline, 1–30 Hz windows were converted to covariance matrices, projected to the tangent space at the Riemannian mean, and fed to LDA [64]. Together, these FBCSP and TS-LDA models provide competitive MI/ERP/SSVEP-style references on the same four-channel dry-EEG features, enabling direct comparison with the proposed Hybrid encoder.

3.4. Models and Learning

Three compact encoders were trained under identical preprocessing, segmentation, augmentation, and subject-wise splits. EEGNet [48] employs depthwise-separable temporal and spatial convolutions to encode EEG priors with very few parameters, a property that is advantageous for four-channel dry EEG where limited data heighten overfitting risk. ShallowConvNet [65] performs temporal filtering followed by spatial filtering and a squaring/log nonlinearity, providing a transparent baseline that often competes with deeper models on oscillatory tasks.

Design principles are explicitly distinguished. EEGNet [48] operationalizes EEG-specific priors by separating temporal band-pass-like filtering from spatial depthwise reweighting, yielding compactness without sacrificing interpretability. ShallowConvNet [65] instead implements shallow temporal filters, spatial filters, and analytic power extraction via squaring and log transforms, producing features aligned with frequency-specific oscillatory dynamics and facilitating visualization.

As described in Figure 5, the Hybrid encoder that we designed accepts either raw windows alone or a concatenation of raw and wavelet stacks. A depthwise temporal convolution per channel (kernel length 15 samples, depth multiplier 4, stride 1) captures band-limited edges; a pointwise 1×1 convolution mixes channels to 32 feature maps; and a light spatial depthwise stage across the four sensors reweights electrodes (kernel size 4×1 , depth multiplier 1). Nonlinearities are exponential linear unit (ELU) with BatchNorm after each convolutional block, followed by dropout $p = 0.25$. This arrangement preserves inductive bias for locality while keeping low computational cost through depthwise separability [48,124].

Temporal dependencies are modeled with a bidirectional LSTM (64 units per direction; $\tanh/\sigma$ gates; dropout $p = 0.2$ between recurrent layers). Variable latency in endogenous spatial-attention paradigms is addressed by recurrent gating, which retains informative context when alpha-range modulation lags cue onset by hundreds of milliseconds; canonical posterior alpha dynamics and their trial-to-trial timing variability are well documented in visuospatial attention, although our montage can capture them only indirectly [50,63].

Long-range, non-local interactions are captured using a multi-head self-attention (MHSA) block with four heads (embedding dimension 128; head dimension 32), pre-norm residual structure (LayerNorm–MHSA–MLP), MLP hidden size 256, attention dropout $p = 0.1$, MLP dropout $p = 0.1$, and stochastic depth with drop probability 0.1 on the residual path. Self-attention focuses the encoder on informative temporal segments while deemphasizing artifacts, complementing convolutional locality and recurrent order modeling [125]; compact convolution–transformer variants in EEG corroborate gains from global dependency modeling under low SNR [66].

Interpretability is explicitly qualified. While attention maps aid qualitative inspection, attention weights are not guaranteed in explanations of model decisions; competing distributions can yield similar predictions. Interpretability claims are therefore limited and contextualized by prior findings [126,127].

A global average pooling layer aggregates features before a linear classifier with label smoothing $\varepsilon = 0.1$. The combination—convolution for locality and inductive bias, LSTM for order and latency variability, and MHSA for flexible non-local coupling—balances capacity and regularization under four-channel constraints. Model complexity is deliberately restrained to mitigate overfitting in low-SNR, data-limited regimes common to dry EEG; surveys of MI-EEG deep learning emphasize this challenge and motivate compact designs [128].

Optimization uses AdamW with decoupled weight decay [129]. The initial learning rate is 1×10^{-3} with linear warm-up over five epochs [130], followed by cosine annealing to a floor of 1×10^{-5} [131]. Weight decay is 1×10^{-4} ; $\beta_1 = 0.9$, $\beta_2 = 0.999$. Gradient-norm clipping at 1.0 stabilizes recurrent updates [132]. Dropout values are disclosed above; stochastic depth uses drop probability 0.1 within the attention block. Batch size is 64 with class-balanced sampling.

Decision rules and operating points are reported with sensitivity and specificity, with the classification threshold chosen on the validation fold by maximizing Youden's J statistic ($J = \text{sensitivity} + \text{specificity} - 1$) to balance false-positive and false-negative rates [133].

Reproducibility details are consolidated as follows.

- Hybrid encoder: temporal depthwise conv (kernel 15, depth multiplier 4), pointwise conv to 32 channels, spatial depthwise conv over four channels; ELU + BatchNorm after each conv; dropout $p = 0.25$ after the convolutional stack.
- Temporal module: BiLSTM, 64 units per direction, dropout $p = 0.2$ between recurrent layers. Attention module: four heads, embed dim 128, head dim 32, pre-norm, MLP 256, attention and MLP dropout $p = 0.1$, stochastic depth $p = 0.1$.

- Classifier: global average pooling, linear with label smoothing $\varepsilon = 0.1$. Optimization: AdamW, initial learning rate 1×10^{-3} , warm-up five epochs, cosine decay to 1×10^{-5} , weight decay 1×10^{-4} , gradient clip 1.0, batch size 64.

These disclosures mirror the completeness used when specifying EEGNet-style compact architectures [48].

Limitations are acknowledged. Despite compactness, adding LSTM and MHSA increases parameter count and compute; deployment should consider on-device constraints, although depthwise separable convolutions offer favorable efficiency–accuracy trade-offs [124]. Overfitting and interpretability limitations are mitigated by conservative capacity, explicit regularization, and calibrated claims, but they remain intrinsic risks in low-channel, low-SNR EEG.

3.5. Augmentation and Supervised Consistency

To emulate realistic nuisances in dry-sensor recordings, augmentation was applied on-the-fly to training windows only and disabled at validation/test. Each transformation targets common dry-EEG failure modes: small reflective time shifts model latency variability; low-variance Gaussian noise with occasional narrow ripple near 50/60 Hz emulates mains and broadband sensor noise; per-channel gain jitter and low-probability channel dropout/mixout mimic contact-impedance fluctuation and motion-induced transients typical of dry electrodes [1]. Ranges followed physiology-aware bounds summarized in recent EEG augmentation reviews and systematic comparisons [2,3]. The spectro-temporal masking strategy parallels SpecAugment’s masking for sequence models while constraining mask placement to preserve physiologically meaningful rhythms. Figure 7 illustrates representative cases [134].

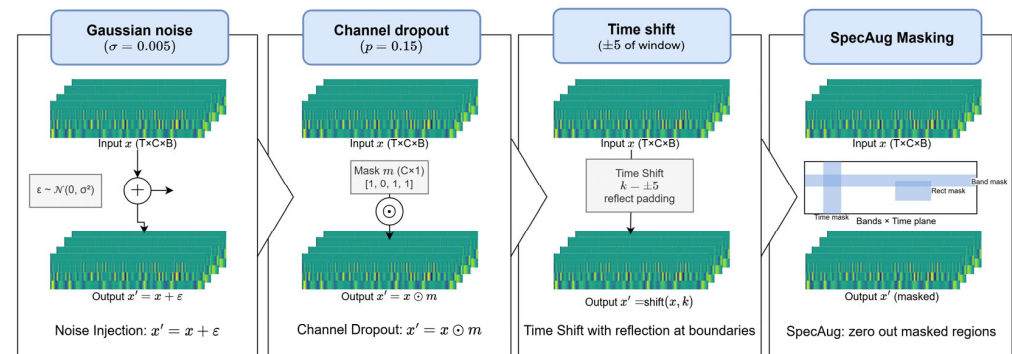


Figure 7. Augmentation families used during training: noise injection, channel dropout, time shift, and spectro-temporal masking.

Parameterization used only label-preserving strengths: time shift $\pm 5\%$ of the window with reflection padding; Gaussian noise $\sigma = 0.005$ for strong views and ≈ 0.0025 for weak views; channel dropout or soft mixout with $p = 0.15$ (strong) and ≈ 0.075 (weak). Short spectro-temporal masks were limited to brief spans and narrow bands to avoid globally suppressing alpha-range content across channels, thereby preserving potential lateralized attention signatures under a low-channel montage [9,10].

A two-view supervised consistency objective enforced invariance to these nuisances using only labeled data. Each window yielded a weak view ($\sigma \approx 0.0025$, shift $\pm 2.5\%$, dropout ≈ 0.075) and a strong view ($\sigma = 0.005$, shift $\pm 5\%$, dropout 0.15); both passed through the shared encoder. The loss combined class-weighted binary cross-entropy with a small penalty on the squared distance between logits, $L = L_{\text{BCE}} + \lambda \|z_{\text{weak}} - z_{\text{strong}}\|_2^2$, with $\lambda = 0.1$. This follows the consistency regularization principle widely used in self-ensembling/student–teacher frameworks (e.g., Mean Teacher), but differs in that

no teacher or EMA targets are used and all samples are labeled. We set $\lambda = 0.1$ to keep the logit-consistency term as a mild auxiliary regularizer (BCE-dominant), consistent with the common practice of using small consistency weights in student–teacher consistency frameworks [135,136]. In parallel, sample-mixing regularizers were used sparingly between compatible pairs: MixUp with $\alpha = 0.2$ and a 1-D CutMix variant to smooth the empirical distribution without violating label semantics [137,138]. Figure 8 depicts the training flow. In our fully labeled, subject-wise CSA setting, this consistency regularizer did not yield a statistically significant improvement ($\Delta \approx 0.003$; $p = 0.98$). Therefore, we treat it as an optional add-on whose utility may depend on the dataset, architecture, and deployment scenario.

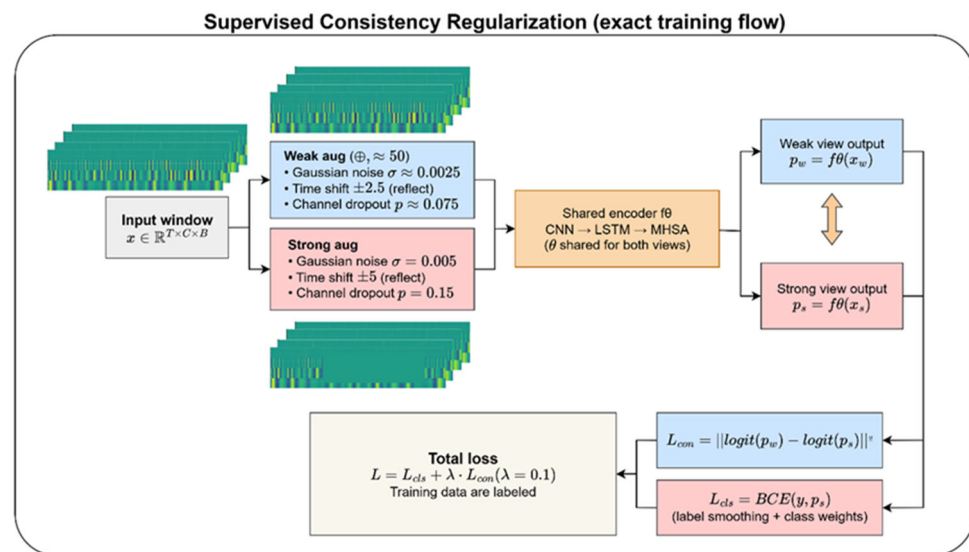


Figure 8. Supervised-consistency training with weak/strong views and combined loss.

3.6. Evaluation and Calibration

Generalization was assessed with leave-one-subject-out (LOSO) and subject-grouped K-fold validation. LOSO directly probes across-subject transfer, whereas grouped K-fold reduces compute by testing on held-out groups while still preventing subject leakage. All fitting steps with learned state (e.g., PREP preprocessing, Autoreject thresholds, scalers) were trained on training subjects only and applied unchanged to validation/test to avoid optimistic bias in small-sample neuroimaging. This protocol follows recommended practice for leakage-free evaluation under limited samples [70,71,79,139].

Bacc (macro recall) was the primary endpoint because it treats false positives and false negatives symmetrically and is robust to incidental class imbalance; overall accuracy, area under the receiver operating characteristic curve (ROC-AUC), sensitivity, specificity, and F1-score were reported secondarily. A single operating threshold per fold was chosen on the validation ROC by maximizing Youden’s J and then fixed on test to decouple threshold selection from the test distribution [23,133,140].

For reporting stability only, we evaluated test-time augmentation (averaging $n = 5$ lightly perturbed weak-view replicas defined in Section 3.5) and a short EMA across consecutive windows within a file, $\hat{p}_t = \alpha p_t + (1 - \alpha)\hat{p}_{t-1}$ with $\alpha = 0.6$ and $\hat{p}_0 = p_0$. In our data, these heuristics did not consistently improve accuracy, so headline metrics are reported without them. They increase inference latency by design—TTA via multiple forward passes and EMA via sequential smoothing—so they are best regarded as reporting stabilizers rather than accuracy boosters for on-device BCI use [38,141].

To obtain reliable probabilities without altering predicted labels, temperature scaling was fit on a held-out subset of training subjects and then applied to test outputs. This

post-hoc calibration improves probability quality and, unlike more complex calibrators, preserves the argmax decisions [142].

Finally, we interpret accuracy in light of finite-sample pitfalls. Chance levels can be exceeded “by chance” when test sets are small or windows are dependent; using subject-wise validation with balanced accuracy mitigates these risks and yields tighter population-level conclusions [22]. Figure 9 consolidates the per-fold pipeline.

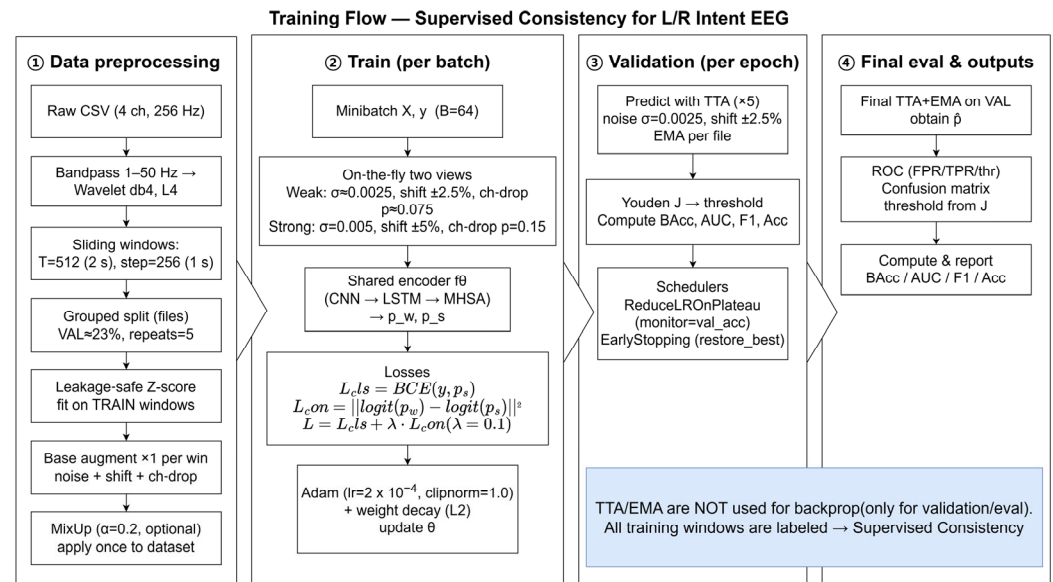


Figure 9. Full per-fold pipeline including leakage-safe preprocessing, two-view training, validation, and final reporting.

To ground cross-task performance differences in the EEG itself, we computed a label-evoked modulation index for each paradigm and canonical frequency band. For each subject and trial window, band power was estimated using Welch’s method and integrated within δ (1–4 Hz), θ (4–8 Hz), α (8–12 Hz), β (13–30 Hz), and γ (30–50 Hz), followed by a log transform and channel-averaging across the four electrodes. Class separability was summarized as an SNR-like effect size, $|d| = |(\mu_1 - \mu_2) / \sigma_{\text{pooled}}|$ (absolute Cohen’s d), computed across trial windows. In addition, to visualize within-window temporal structure in the α band, signals were band-pass filtered (8–12 Hz), the Hilbert envelope was computed, and the mean Δ log-amplitude (class1 – class0) was derived as a function of time and averaged across channels; subject-level curves were resampled to normalized time (0–1) and aggregated as mean $\pm$ SEM across subjects.

4. Results Analyses

4.1. Experimental Setup

Recordings were obtained with a low-density dry electrode headset (Muse-2; four channels, 256 Hz). Sixteen adults completed three binary paradigms: CSA (left vs. right spatial attention with overt eye/body movements suppressed), MI (left vs. right), and affect (Emotion; “cute” vs. “disgust”). Analyses and model selection were conducted at the subject level with repeated subject-wise folds; all fitting steps that can leak information were estimated on training subjects only and applied unchanged to validation/test. Significance testing was performed across subjects (paired contrasts on per-subject metrics) rather than across highly overlapping windows to avoid pseudo-replication and inflated p -values common in small-sample neuroimaging when non-independence is ignored [3,79].

To address concerns on statistical power with $N = 16$, we report for each task the paired permutation p -value on Δ balanced accuracy (BAcc), percentile-bootstrap 95%

confidence intervals from subject-level resampling, and descriptive post-hoc power at $\alpha = 0.05$ computed for the observed paired effect size (CSA $N = 16$, Emotion $N = 12$, MI $N = 13$). This addition follows established guidance that low sample sizes in neuroscience produce wide error bars and unreliable inference unless effect sizes and interval estimates are emphasized; bootstrap resampling provides distribution-free intervals that remain valid under modest deviations from Gaussian assumptions [79,143,144].

To translate classifier accuracy into user-centric terms, we also summarize misclassification rate and ITR, bits/min using the standard discrete-choice BCI channel assumption; these usability readouts are reported alongside ΔBAcc to contextualize practical gains [145].

Each analysis window was transformed with a discrete wavelet transform (Daubechies-4, level-4) and then standardized by Z-scoring with strict leakage control (fit on training data only; apply to validation/test). In addition to the wavelet stack, two alternative feature families were instantiated for comparison: (i) classical filter-bank (FB) band energies and (ii) FB centered on each participant's IAF. Wavelets were selected a priori because CSA involves non-stationary modulations including posterior alpha bursts and broader spectro-temporal transients. Since these signals are burst-like rather than stationary, multiresolution time–frequency atoms preserve such dynamics more faithfully than static band energies. This physiological and signal-processing rationale aligns with contemporary accounts of alpha lateralization and recommendations to use time–frequency methods for non-stationary EEG; in our dataset, the wavelet feature set yielded more consistent performance than static FB features [146–149].

For the IAF-centered FB variant, IAF was estimated via Welch's method on 8–13 Hz spectra ($f_s = 256$ Hz, $n_{\text{perseg}} = \min(1024, \text{segment length})$, all four channels), averaged across channels with a fallback of 10.0 Hz when no reliable peak was detected [114]. Under our quality criteria, stable peaks were seldom obtained with the four-channel dry electrode montage, yielding an adoption rate near zero and broad/low-prominence candidates—consistent with reports that reliable alpha-peak estimation is sensitive to montage, baseline length, and hardware; low-density/dry systems can reduce reliability in low-frequency ranges or with limited posterior coverage [1,3].

EEG is non-stationary, and discriminative patterns (including α -band modulations in spatial attention) can be transient rather than stationary. Thus, a multi-resolution time–frequency representation is preferred; in our dataset, the level-4 dB4 wavelet decomposition provides four sub-band reconstructions (A4/D4/D3/D2) that align with δ/θ , α , low- β , and high- β /low- γ ranges, respectively.

Future work to recover potential benefits of personalization should combine longer eyes-closed baselines, posterior-focused montages or virtual posterior electrodes, and robust spectral parameterization that explicitly separates aperiodic (1/f) backgrounds from periodic peaks before centering FBs on individualized alpha; complementary burst-detection frameworks can further standardize detection thresholds across frequencies [150,151].

Three encoders were evaluated: a compact EEG-specific CNN, a CNN + LSTM that models temporal dependencies, and a Hybrid encoder that appends MHSA after the recurrent block to reweight informative segments before global pooling. The compact CNN follows the EEGNet design principle of depthwise temporal convolutions per channel coupled with depthwise-separable spatial convolutions, which approximate filter-bank plus spatial filtering while keeping parameters low—well suited to four-channel dry EEG; this architecture encodes EEG priors explicitly and has demonstrated effectiveness across multiple BCI paradigms [48]. The LSTM layer introduces gated memory that preserves long-range temporal structure and mitigates vanishing gradients, allowing the network to integrate evidence across adjacent windows [152]. MHSA then computes context-dependent weights over the recurrent sequence so that transient but informative epochs (e.g., task-relevant

oscillatory transients) are emphasized and noise-dominated intervals are down-weighted before pooling; this yields a soft alignment that complements recurrence [125]. In combination, the Hybrid encoder captures local oscillatory motifs via CNN, sequential dependencies via LSTM, and time-varying saliency via MHSA prior to decision making.

Two training configurations were compared. The Baseline used wavelets and train-only normalization with no data augmentation (Aug), no EMA, and no test-time augmentation (TTA). The All-on-Wav configuration is augmentation-driven, applying per-window Aug—Gaussian noise ($\sigma = 0.005$), channel dropout ($p = 0.15$), and small temporal shifts ($\pm 5\%$ of window). In addition, we evaluated several auxiliary robustness add-ons within the same framework: (i) a supervised two-view consistency/contrastive regularizer (weak view: $\sigma = 0.0025$, $p = 0.075$, $\pm 2.5\%$; strong view: $\sigma = 0.005$, $p = 0.15$, $\pm 5\%$), using a consistency weight $\lambda = 0.1$ and a supervised contrastive term with temperature 0.2 and weight 0.2; (ii) EMA with coefficient $\alpha = 0.6$; and (iii) TTA as an average over 5 weak perturbations ($\sigma = 0.0025$, $\pm 2.5\%$). Training used a maximum of 300 epochs with early stopping on subject-wise validation loss (patience = 60) and ReduceLROnPlateau (patience = 30, factor = 0.8). BAcc (macro-recall) was the primary endpoint; inference used paired Wilcoxon signed-rank tests with Holm's sequential correction, and effect size r is reported where relevant. In the low-channel dry electrode setting, the Baseline omits defenses against data scarcity, low SNR, non-stationarity, and inter-/intra-subject variability that are characteristic of such hardware and protocols—factors that Aug and the evaluated add-ons are intended to mitigate—hence lower robustness is expected [21,23,79,153–155]. All experiments followed the same subject-wise cross-validation, training schedule, feature windowing, and augmentation settings summarized in Table 1.

Table 1. Experimental setup and fixed hyperparameters used throughout Section 4. Subject-wise CV and leak-safe normalization were applied; the Baseline disables Aug/SSL/EMA/TTA, whereas All-on-Wav uses the listed augmentation settings per window.

Block	Parameter	Value
CV	val_fraction	0.25
	test_fraction	0.2
	n_repeats	5
	random_state	42
Training	epochs	300
	batch_size	64
	learning rate	0.0002
	L2 penalty	0.0001
	label smoothing	0.02
	mixup/mixup_alpha	true/0.2
Feature	fs	256
	window/step	512/256
Augment	noise_std	0.005
	drop_p	0.15
	shift_max	0.05

Given the dry, low-channel hardware, we emphasize training-time robustness (leak-safe normalization, augmentation, consistency regularization, EMA) rather than ad-hoc inference-time heuristics, because signal quality in dry systems fluctuates substantially with (i) electrode–skin contact/impedance—air gaps and hair elevate contact impedance and foster low-frequency drift and intermittent “electrode-pop” transients; (ii) motion and sweat—head/lead movement and perspiration induce baseline wander, non-stationary bursts, and EMG contamination that degrade decoding; and (iii) inter-individual skin/hair

physiology that limits repeatable coupling in low-density montages. Validation on Muse-class headsets and recent reviews of dry electrodes document these constraints and motivate robustness-first training protocols that learn invariances to these perturbations—improving practicality and reproducibility—rather than brittle test-time fixes [1,3,13,82,156–160].

4.2. Topline Results and Contribution Analysis

Figure 10 (CSA; 16 subjects) compares the Baseline to All-on-Wav under the Hybrid encoder. The median BAcc increases by +2.9 percentage points ($\Delta = 0.029$; paired Wilcoxon $p = 0.037$; $r \approx 0.56$), with a visible upward shift of both the median and mean markers at comparable interquartile range (IQR), indicating a genuine location shift rather than variance inflation. To translate this gain into user-centric terms, the same subjects show a median reduction in misclassification rate from 0.272 to 0.268 ($\downarrow 1.5\%$) and a median ITR increase from 3.120 to 3.240 bits·min^{−1} ($\uparrow 3.8\%$), implying fewer unintended selections per minute and more efficient command throughput in online use [5,38]. Conceptually, this improvement aligns with the mitigation of small-sample brittleness—overfitting and high variance arising from limited subject-level data—via training-time augmentation that exposes the model to amplitude noise, partial channel loss, and slight temporal misalignment, thereby learning invariances to nuisance processes characteristic of dry electrode recordings [1,79,153].

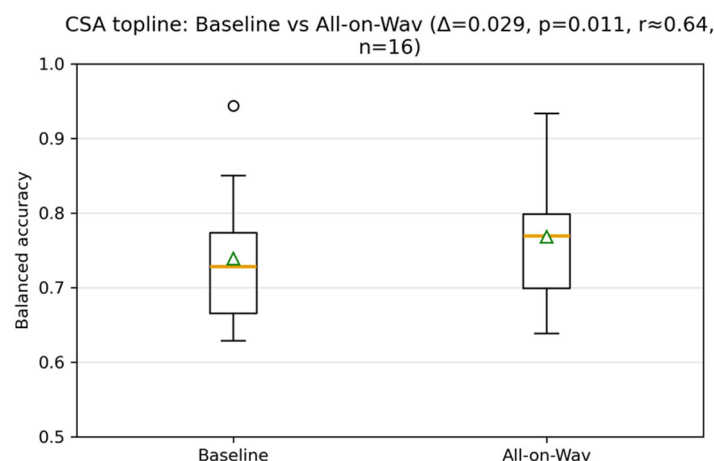


Figure 10. CSA topline: Baseline vs. All-on-Wav under the Hybrid encoder (wavelet features). Boxplots across 16 subjects (triangle = mean, line = median, circles = outliers).

Drop-one ablations around All-on-Wav (Figure 10) isolate which components matter. Removing Aug yields the largest and most consistent drop below zero, establishing augmentation as the primary driver of robustness in the dry, low-channel regime. Our SSL instantiation is a two-view supervised consistency objective that aligns logits between a weakly perturbed view (Gaussian noise $\sigma = 0.0025$, channel-dropout $p = 0.075$, time-shift $\pm 2.5\%$) and a strongly perturbed view ($\sigma = 0.005$, $p = 0.15$, $\pm 5\%$), with a consistency weight $\lambda = 0.1$; labels are fully supervised (no unlabeled pretraining) and the consistency loss is added to cross-entropy on the weak view. Under this setting, removing SSL changes the median by ≈ 0 (Figure 11; $\Delta \approx 0.003$, $p \approx 0.98$, $r \approx 0.01$), indicating no measurable median benefit for this specific consistency regularizer on this dataset. This differs from SSL approaches that rely on large unlabeled corpora or pretraining to learn subject-invariant structure before fine-tuning [5,38,79]. Removing TTA or EMA does not degrade—and occasionally improves—performance. Test-time averaging can reduce confidence by aggregating predictions over noisy transforms, and EMA smooths weight trajectories; both effects may blur decision boundaries without addressing between-subject heterogeneity (anatomy,

contact/impedance, and idiosyncratic alpha dynamics) that dominates generalization in small-N, dry electrode EEG [1,136,153,161]. These results suggest that, in our configuration, training-time augmentation contributes most because it synthetically exposes the model to amplitude noise, partial channel loss, and small timing jitter that mimic real nuisance processes in dry recordings, thereby reducing small-sample brittleness and improving location (median) without inflating variance [79,153]. To isolate the contribution of the consistency term itself, we also evaluated SSL alone on top of the Baseline; the change in CSA balanced accuracy was negligible (Figure 12; $\Delta \approx 0.003$, $p = 0.98$, $r \approx 0.01$; $N = 16$). Therefore, we do not position supervised consistency as a key driver of performance in this work; the observed gains are driven predominantly by augmentation, and the consistency term is reported as an optional, context-dependent regularizer.

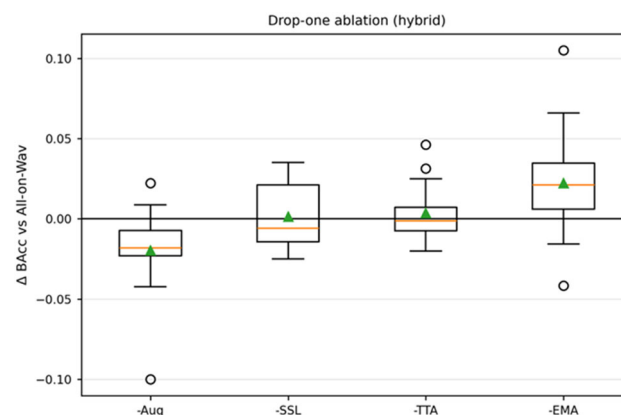


Figure 11. Drop-one ablation relative to All-on-Wav (Hybrid, CSA): removing augmentation hurts most; removing the consistency regularizer has near-zero effect; TTA and EMA are also negligible. Boxplots across 16 subjects (triangle = mean, line = median, circles = outliers).

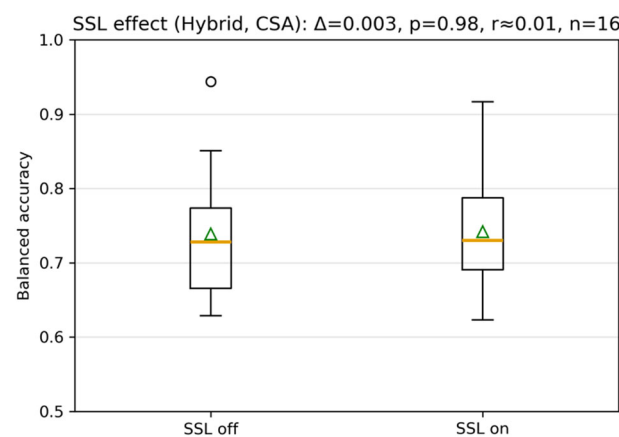


Figure 12. Effect of the supervised consistency regularizer (Hybrid, CSA): $\Delta \approx 0.003$, $p = 0.98$, $r \approx 0.01$. Boxplots across 16 subjects (triangle = mean, line = median, circles = outliers).

4.3. Feature Families Under a Fixed Encoder

Holding the Hybrid encoder and All-on-Wav training constant, Figure 13 compares wavelet features with classical FB band energies and FB centered at each participant's individual alpha frequency (FB + IAF). Wavelets attain the highest median and upper-whisker BAcc, whereas FB trails by several points and FB + IAF shows no reliable advantage over FB. Mechanistically, CSA is expressed through brief, lateralized alpha modulations; a multiresolution wavelet basis preserves such transient time–frequency structure, while FB compresses each window into coarse band energies that can smear short-lived but informative events [9,162]. For FB + IAF, IAF was estimated with Welch spectra in 8–13 Hz

(fs = 256 Hz; nperseg = min(1024, segment length); all four channels averaged) with a 10.0 Hz fallback when no reliable peak was found. Under these settings, stable peaks were rarely detected in the four-channel dry montage (adoption rate $\approx 0\%$ across $N = 16$; median full width at half maximum (FWHM) = 3.50 Hz, CV = 0.067, prominence = 1.93 dB), consistent with known sensitivity of IAF estimation to montage, contact/impedance, and baseline duration in low-density dry systems [1,3–5,114]. This explains the absence of a net FB + IAF benefit in our cohort. While the present evidence favors wavelets, we note remaining design choices and trade-offs: the optimal mother/level can depend on task and subject cohort, and wavelet stacks add computational cost relative to FB (though still compatible with 256 Hz online operation in our pipeline) [63,163].

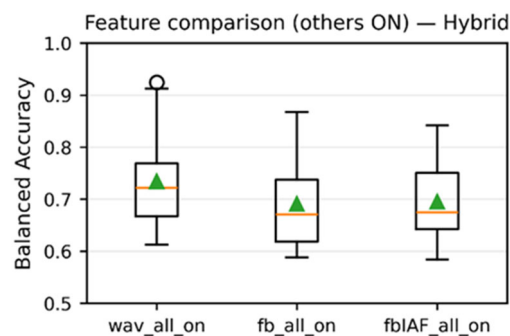


Figure 13. Feature families under the Hybrid encoder (All-on-Wav training): Wavelet vs. FB vs. FB + IAF. Boxplots across 16 subjects (triangle = mean, line = median, circles = outliers).

4.4. Model Sweep and Subject-Wise Distributions

Figure 14 summarizes encoders across subjects and Figure 15 detail per-subject fold distributions (A–P). The ranking is Hybrid $\geq$ CNN + LSTM $>$ CNN, with modest but consistent margins. The Hybrid chiefly raises the upper tail: several participants reach BAcc in the 0.85–0.95 range while the lower quartile shifts only slightly. This pattern is consistent with the attention block computing context-dependent weights over the LSTM sequence and selectively amplifying time segments where task-relevant oscillatory contrasts are expressed (often in $\alpha/\beta/\gamma$ bands, with subject-dependent emphasis); after recurrent aggregation, MHSA performs a soft alignment that highlights transient, informative epochs and suppresses noise-dominated intervals [125,147,164]. For low-SNR participants, lateralization is weak or inconsistent, attention weights become diffuse, and the leverage of the mechanism diminishes, yielding a small median change—an observation consistent with known variability in dry, low-density recordings [1,3].

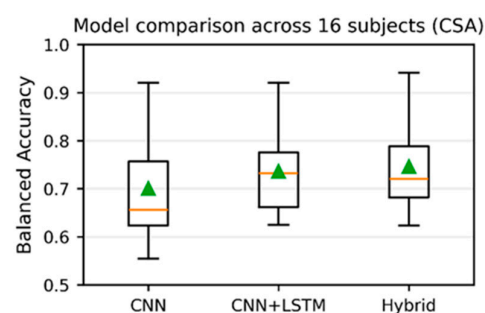


Figure 14. Model comparison across 16 subjects (CSA): CNN, CNN + LSTM, and Hybrid (CNN + LSTM + MHSA). Boxplots across 16 subjects (triangle = mean, line = median, circles = outliers).

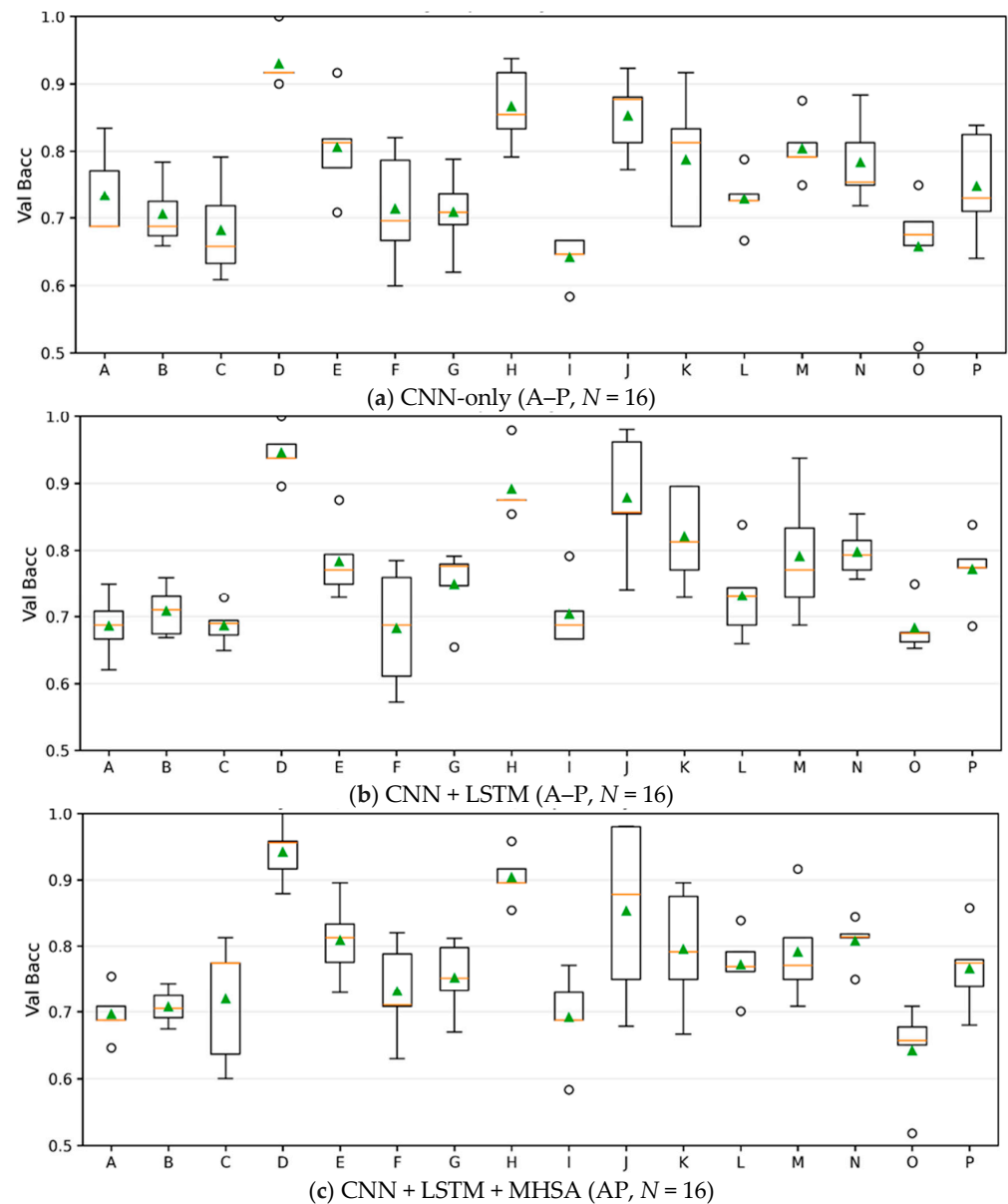


Figure 15. (a) Per-subject fold distributions: CNN-only (subjects A–P); (b) Per-subject fold distributions: CNN + LSTM (subjects A–P); (c) Per-subject fold distributions: Hybrid (CNN + LSTM + MHSA; subjects A–P). Boxplots across 16 subjects (triangle = mean, line = median, circles = outliers).

On the same leakage-safe windows, classical baselines stayed in the mid-0.5 range. Figure 16a shows per-subject fold distributions for FBCSP + LDA: mean CSA balanced accuracy was 0.536 ± 0.108 across subjects and folds, with a few participants (e.g., A, B, E, H, M, P) reaching ~ 0.65 – 0.70 while others hovered near chance. Figure 16b shows the Riemannian tangent-space baseline (TS-LDA) with a similar profile (0.525 ± 0.086), again with substantial inter-subject spread. Figure 16c additionally reports the canonical Riemannian Minimum Distance to Mean (MDM; AIRM), which likewise remains in the mid-0.5 range with pronounced inter-subject variability. Figure 17 plots class distributions along the most discriminant LDA direction for a high-performing subject (E) and a low-performing subject (N): for E, FBCSP + LDA and TS-LDA yield well-separated modes, whereas for N the densities overlap strongly. Together, these results confirm that, under identical four-channel dry-EEG constraints, shallow FBCSP and Riemannian baselines (TS-LDA) pipelines [63,64] provide non-trivial but clearly weaker CSA baselines than the proposed Hybrid encoder.

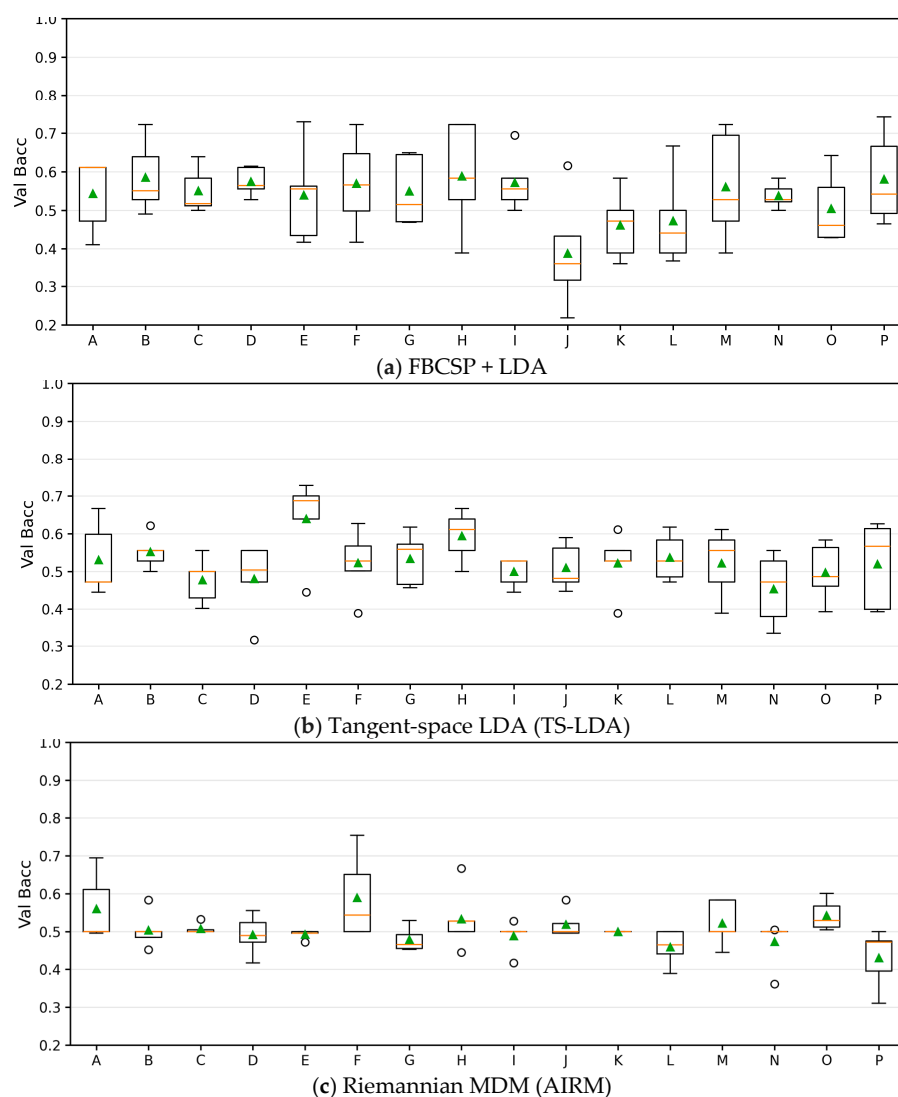


Figure 16. Classical CSA baselines with four-channel dry EEG. (a) FBCSP + LDA—per-subject cross-validated balanced accuracy; boxplots show fold-wise distributions per subject (A–P; $N = 16$), and green triangles mark the fold mean; (b) Tangent-space LDA (TS-LDA)—per-subject cross-validated balanced accuracy with the same format; (c) Riemannian MDM (AIRM)—per-subject cross-validated balanced accuracy with the same format. Boxplots across 16 subjects (triangle = mean, line = median, circles = outliers).

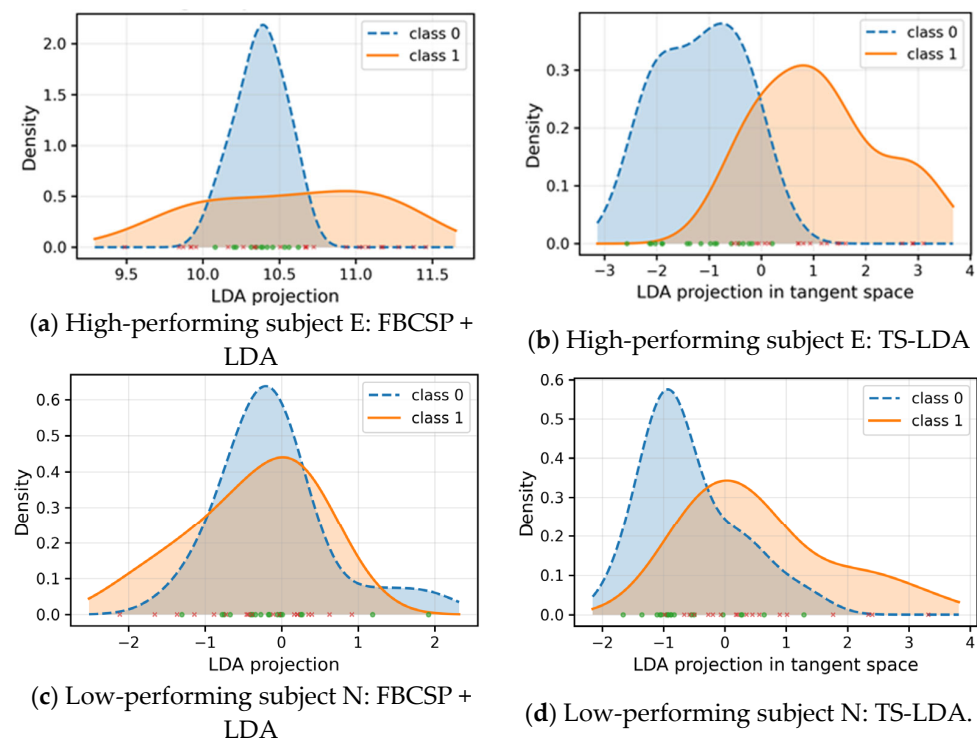


Figure 17. Class distributions along the most discriminant LDA direction for CSA. (a) High-performing subject E: FBCSP + LDA; (b) high-performing subject E: TS-LDA; (c) low-performing subject N: FBCSP + LDA; (d) low-performing subject N: TS-LDA.

As a result, the error profile remains balanced: the micro-averaged confusion matrix in Figure 18 ($N = 60,912$ windows) shows recall of 0.738 (Left) and 0.741 (Right) with precision of 0.746 and 0.733, respectively; macro-F1-score ≈ 0.740 and BAcc ≈ 0.740 indicate no side bias, supporting balanced accuracy as the headline metric for this symmetric task [23]. Looking ahead, gains for low-SNR users may require personalization and robustness beyond generic attention: subject-adaptive or transfer-learning strategies, stronger artifact mitigation and augmentation tailored to dry electrode nuisances, and training curricula that emphasize subject-specific temporal signatures are promising directions [157,161,165].

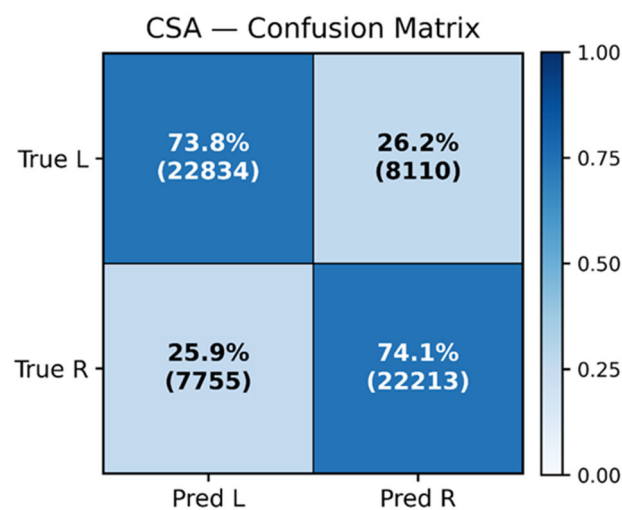


Figure 18. CSA micro-averaged confusion matrix under All-on-Wav (Hybrid).

4.5. Cross-Task Comparison Under an Identical Pipeline

Applying the same pipeline (Hybrid + All-on-Wav) to MI and Emotion yields markedly higher medians than CSA (Figure 19): MI and Emotion cluster in the high-0.8 to low-0.9 range, whereas CSA concentrates in the mid-0.7 s. The gap is neurophysiologically plausible. MI elicits robust contralateral μ (8–13 Hz) and β (13–30 Hz) ERD/ERS over sensorimotor cortices that remain detectable even with few channels [91,166]. Emotion paradigms often present salient patterns such as frontal-alpha asymmetry (valence-related power differences over left/right prefrontal sites) and task-dependent changes in $\theta/\alpha/\gamma$ power and connectivity over fronto-temporal networks, which likewise survive low-density recording [167–169]. By contrast, CSA depends on covert shifts of spatial attention whose posterior-alpha lateralization is subtler, more idiosyncratic across individuals, and more sensitive to dry electrode variability, yielding lower medians under the same hardware constraints [11]. Importantly, this does not diminish CSA's practical value: covert-attention BCIs offer hands-free, gaze-independent control and can complement or replace overt motor/ocular channels, which is attractive for users with severe motor limitations and for hands-busy or eyes-busy human–computer interaction (e.g., AR overlays or silent, non-verbal selection) [145,170].

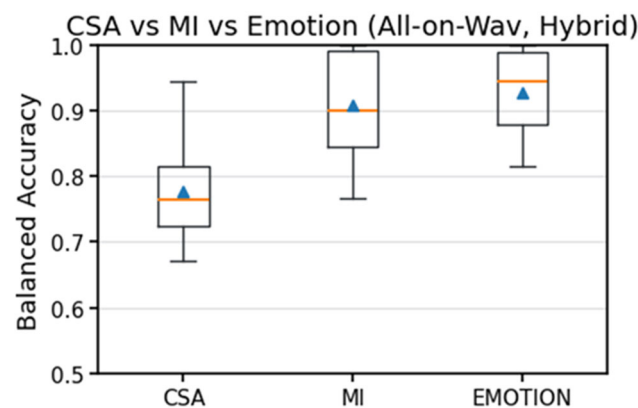


Figure 19. Cross-task comparison (same pipeline): CSA vs. MI vs. Emotion.

Beyond decoder metrics, we quantified the magnitude of label-evoked EEG modulation itself to provide a direct electrophysiological basis for the CSA–MI–Emotion performance gap. Figure 20 summarizes the across-subject distribution of the band-wise modulation index ($|\text{Cohen's } d|$ of log-bandpower between class labels) for each paradigm. CSA shows uniformly small effect sizes across bands, indicating weak label separability in band power under the four-channel dry montage. In contrast, MI and Emotion exhibit substantially stronger modulation, with Emotion showing the largest median effects in β – γ and MI showing larger subject-dependent variability, including high-modulation outliers in higher frequencies. Figure 21 further visualizes the time-resolved Δ log-amplitude of the α -band envelope within the 2.0 s decision window. CSA remains close to zero throughout the window, whereas MI shows a sustained negative deflection suggestive of reduced α -band amplitude, and Emotion shows a sustained positive contrast over a large portion of the window. Together, these results support the interpretation that CSA's lower decoding performance reflects not only task/model difficulty but also intrinsically smaller label-evoked modulation (i.e., lower effective SNR) in the recorded EEG with sparse dry-electrode sampling.

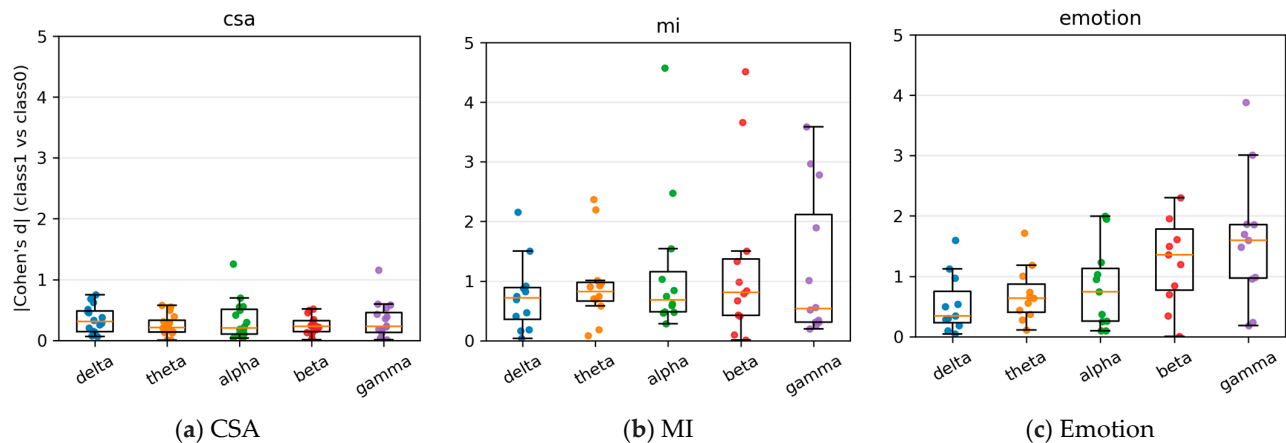


Figure 20. Label-evoked modulation magnitude by band ($|Cohen's d|$). (a) CSA, (b) MI, (c) Emotion. For each subject, $|d|$ was computed between class labels on log-bandpower (Welch; channel-averaged) for $\delta/\theta/\alpha/\beta/\gamma$; boxplots show across-subject distributions.

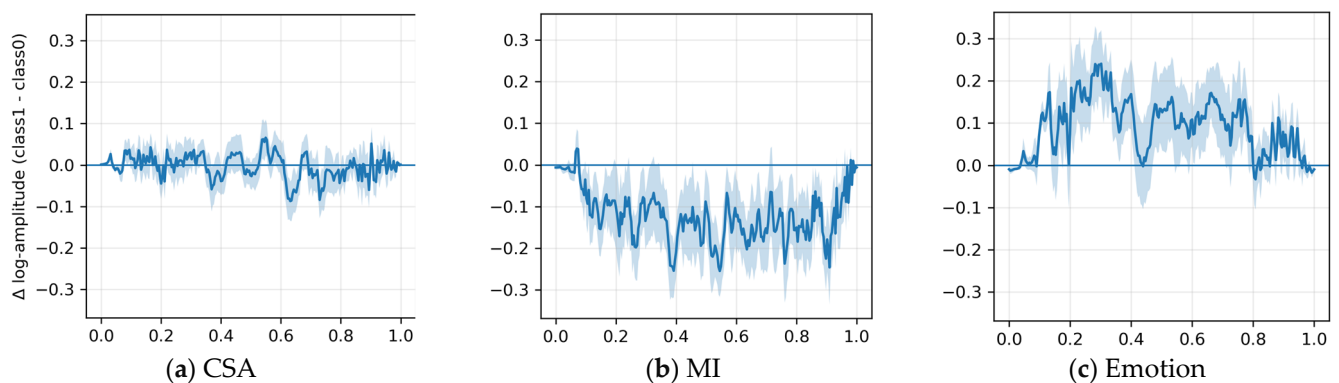
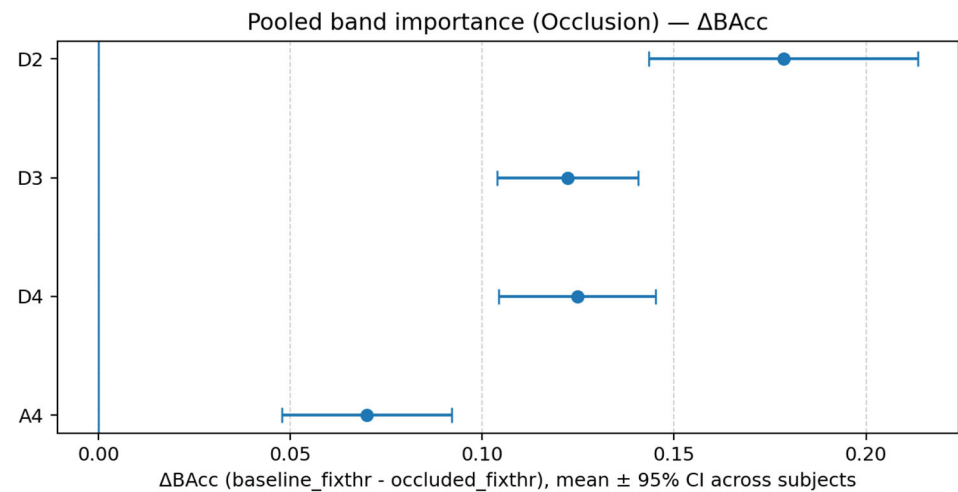
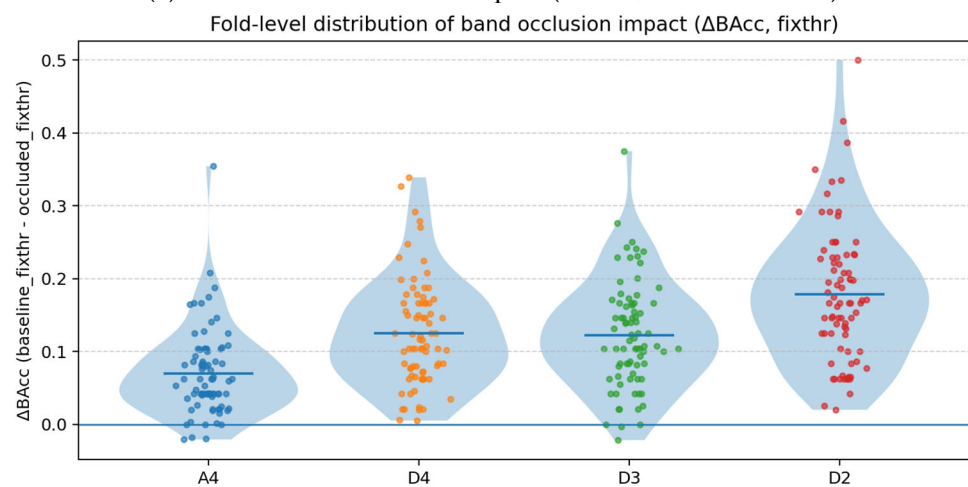


Figure 21. Time-resolved α -band envelope modulation ($\Delta \log$ -amplitude). (a) CSA, (b) MI, (c) Emotion. Curves show mean $\pm$ SEM of subject-level α -envelope differences (class1 – class0) across normalized time (0–1 s) within the decision window.

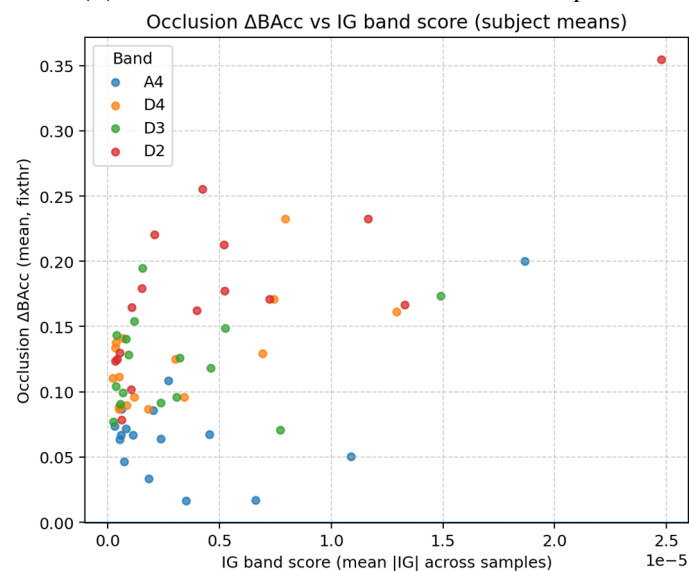
4.6. Explainability Analysis (Temporal, Spatial, and Spectral Relevancy)

To address interpretability, we visualize what the attention-based model learned along three axes: frequency (spectral), space (channels), and time (temporal). The analyses are computed on the CSA/all_on_wav setting using the four-band wavelet input [A4, D4, D3, D2] defined in the Methods.

We quantify each band's contribution by band occlusion: for each validation fold, we mask one sub-band (set it to zero after z-normalization) and measure the performance drop using the baseline Youden threshold fixed from the unoccluded model. Figure 22 shows that occluding D2 (≈ 32 – 50 Hz; high- β /low- γ) yields the largest mean decrease in balanced accuracy, indicating a strong contribution from high-frequency transients, while D4 (≈ 8 – 16 Hz; α) also causes a non-negligible drop. Importantly, the fold-wise spread and integrated-gradients attributions highlight substantial inter-subject variability, consistent with subject-specific spectral signatures.

(a) Pooled mean occlusion impact (ΔBAcc , fixed threshold)

(b) Fold-level distribution of occlusion impact



(c) Relationship between integrated gradients band attribution and occlusion impact

Figure 22. Spectral relevancy analysis. (a) Pooled mean occlusion impact (ΔBAcc , fixed threshold); (b) Fold-level distribution of occlusion impact; (c) Relationship between integrated gradients band attribution and occlusion impact.

Spatial relevancy is summarized as an attention-weighted channel saliency map, computed by aggregating attention weights over time and weighting the absolute band-power change ($|\Delta\text{power}|$) per channel. Figure 23a,b contrasts the average saliency patterns for the top-3 and bottom-3 subjects (ranked by BAcc), illustrating that informative channels are not identical across individuals. In both groups, saliency tends to concentrate on the temporal-parietal channels (TP9/TP10), while the frontal channels (AF7/AF8) exhibit more variable contributions across subjects. Notably, the top-3 group shows a more pronounced and structured saliency pattern—particularly with stronger emphasis on TP9/TP10—whereas the bottom-3 group presents a flatter and less consistent distribution across channels. The time $\times$ channel heatmaps in Figure 23c,d further support this observation by showing when each channel becomes salient within the 2.0 s window. In the top-3 group (Figure 23c), saliency appears as sharper, recurring high-intensity streaks that are predominantly confined to TP9/TP10, indicating more consistent reliance on the temporal-parietal sensors across time. In contrast, the bottom-3 group (Figure 23d) exhibits weaker channel specificity, with salient time points appearing more diffusely across channels and a reduced dominance of TP9/TP10.

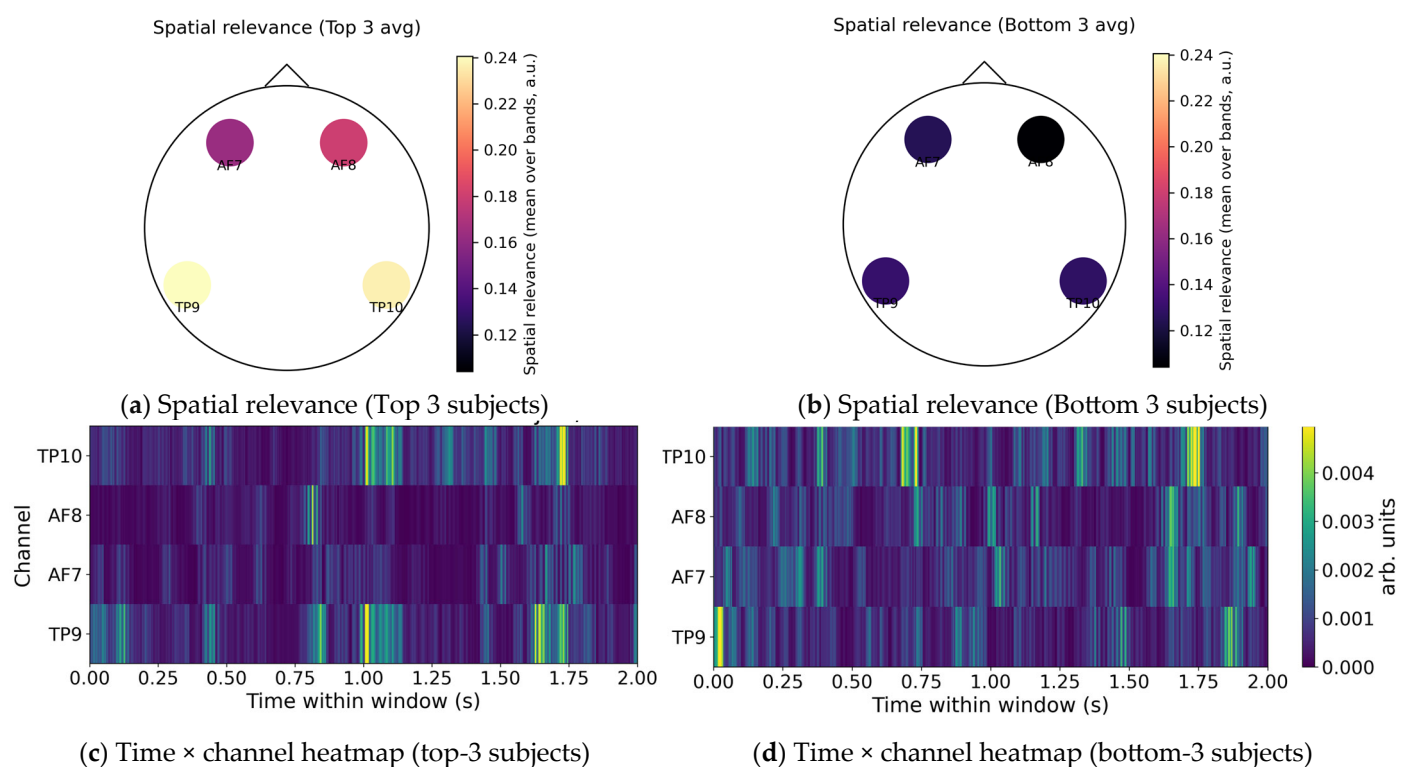


Figure 23. Spatial relevancy analysis. (a) Spatial relevance (Top 3 subjects); (b) Spatial relevance (Bottom 3 subjects); (c) Time $\times$ channel heatmap (top-3 subjects): time-resolved attention-weighted channel saliency; (d) Time $\times$ channel heatmap (bottom-3 subjects).

To probe how the model allocates temporal emphasis within each 2.0-s window, we visualize normalized attention weights over time for all subjects. Figure 24 indicates that attention is non-uniform and tends to concentrate on specific sub-intervals within the window. This pattern suggests that the model assigns higher weights to comparatively informative temporal portions of the input.

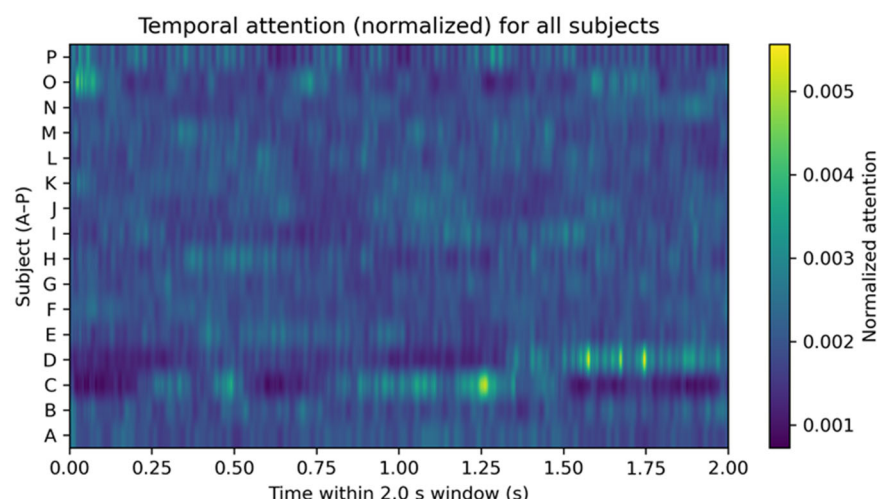


Figure 24. Temporal attention heatmap (normalized) across all subjects.

4.7. Online System Evaluation

We evaluated online accuracy and end-to-end responsiveness per 2.0 s decision window across four configurations: a compact CNN, a CNN + LSTM, a Hybrid (CNN + LSTM + MHSA) with all defenses off (“All-off-Wav”), and the final Hybrid with augmentation + SSL enabled (“All-on-Wav”). Details of the wavelet (“Wav”) front end are provided in Appendix A. The Hybrid All-on-Wav achieved the highest online accuracy (0.695), followed by the same backbone without Aug/SSL (0.673), the CNN + LSTM (0.612), and the CNN (0.578). Thus, adding training-time defenses on the Hybrid backbone provides a +0.022 absolute lift over the clean Hybrid while remaining compact (<0.1 M parameters) and EEG-specific in design [48,125,152]. Table 2 summarizes the trade-offs between accuracy, parameter count, and Total est. latency.

Table 2. Online evaluation (per 2.0 s window). Latency components are measured per window and combined into the “Total est. latency” a user experiences.

Model (Decoder)	Trainable Params	CPU Forward (ms)	GPU Forward (ms)	Total est. Latency (ms)	Online Accuracy
CNN (EEG-specific compact CNN)	34,689	4.873	3.857	2028.9	0.578
CNN + LSTM	67,713	170.126	8.148	2033.1	0.612
Hybrid (CNN + LSTM + MHSA), All-off-Wav	84,353	184.705	10.863	2035.9	0.673
Hybrid (CNN + LSTM + MHSA), All-on-Wav	84,353	184.705	10.863	2035.9	0.695

Notes. (i) “Total est. latency” = 2000 ms evidence buffer + ~20 ms preprocessing + model inference (forward-pass) + ~5 ms postprocessing; under this setup, “Est. reaction” is effectively identical to total latency (ii) All-on-Wav and All-off-Wav share the same Hybrid backbone at inference; only training-time defenses differ, so their runtime/parameter entries coincide. (iii) Accuracy is the online score aggregated across subjects; see dispersion comments below for inter-individual variability [23,145].

Latency analysis indicates the system is buffer-limited rather than compute-limited: even the largest model executes a per-window forward pass in ≈ 185 ms on a laptop-class CPU and in ≈ 11 ms on a GPU. Consequently, end-to-end command latency clusters near 2.03 s across models, dominated by the 2.0-s evidence window. With a modest increase in CPU forward time relative to the CNN, the Hybrid All-on-Wav yields the best online

accuracy; the CNN remains the fastest but lags markedly in accuracy. These results define a clean Pareto trade-off: the CNN has the lowest compute and the lowest accuracy, whereas the Hybrid All-on-Wav achieves the highest accuracy with a small, acceptable compute overhead, while staying well under 0.1 M parameters—aligning with compact, EEG-prior designs intended for wearable/edge deployment [48,125,152]. For user-centric context, online accuracy translates directly to error rate and to throughput via the conventional BCI channel model (i.e., ITR) when needed for application comparisons [145]. All latency benchmarks were obtained on an AMD Ryzen 7 3700X CPU and an NVIDIA GeForce RTX 4070 GPU (12 GB VRAM) using TensorFlow 2.10.0 under Python 3.10.16.

Inter-subject variability is substantial: while the table reports cohort-level means, per-subject distributions span a wide range typical of dry electrode, few-channel EEG. Qualitatively, the Hybrid All-on-Wav raises the upper tail (best participants become markedly more accurate) and suppresses extreme low-end outliers relative to CNN/CNN + LSTM, indicating that attention-based reweighting after recurrent aggregation helps when task-relevant transients (specifically oscillatory bursts in alpha and high-beta bands) are present, while subjects with weak or inconsistent signatures remain challenging [11,23,125,152]. Practically, this pattern argues for personalization on top of the robust default, while keeping the default compact enough for on-device inference.

From a deployment standpoint, reducing total reaction time will primarily require shorter decision windows and/or greater overlap, since model-side compute is already a minority of the budget. Because overlap increases update rate but also inter-prediction correlation, window/stride should be tuned against online accuracy and user workload; the present results show that, given a 2.0 s window, Hybrid All-on-Wav provides the most favorable accuracy–latency–compactness balance among tested options [23,145].

Because the system is buffer-limited, shortening the decision window is the most direct way to reduce reaction time. For example, reducing the window from 2.0 s to 1.5 s would lower the expected end-to-end latency by ~0.5 s (~1.5 s buffer + preprocessing + inference + postprocessing), while keeping model-side compute nearly unchanged. However, a shorter window reduces the amount of spectral evidence available per decision, which may degrade accuracy and increase variability across users. In practice, this can be partially mitigated by increasing overlap (smaller stride) to raise the update rate, but overlap also increases correlation between successive predictions; thus, window/stride should be tuned to balance responsiveness, accuracy, and user workload.

5. Limitations

Although the cohort size ($N = 16$) is comparable to prior exploratory dry-EEG BCI studies, it remains a limitation for strong population-level generalization. In our CSA analysis, the observed statistical power (0.610) falls below the conventional target (0.8), indicating that effect estimates may be less stable and should be interpreted with caution. Therefore, the present findings should be viewed as evidence of feasibility rather than definitive generalizable performance. Validation on a larger, independent cohort (ideally with a pre-planned sample size rationale) is an important next step.

The Muse-2 montage (TP9/AF7/AF8/TP10) does not sample occipito-parietal cortex where canonical CSA markers (retinotopic posterior alpha lateralization; N2pc) are maximal [90,171]. Therefore, our system cannot be interpreted as directly measuring PO7/PO8-like posterior alpha generators; any CSA sensitivity is expected to be attenuated and indirect under sparse sampling [90].

Covert attention can co-occur with subtle, direction-biased microsaccades even when participants attempt to fixate, and such microsaccades can covary with or modulate lateralized EEG markers [172]. Because AF7/AF8 are near the eyes and ICA is not reliable with

four channels, we cannot fully rule out that part of the decodability reflects EOG leakage rather than purely cortical dynamics [173]. However, we emphasize that the exclusion of the post-cue settling period mitigates the impact of initial orienting saccades, which are typically the most prominent directional ocular artifacts in CSA tasks [174,175]. Future work should include simultaneous eye tracking and/or dedicated EOG, and/or add posterior electrodes to directly test whether performance persists when ocular contributions are regressed out [173].

Given the pronounced inter-subject variability inherent in four-channel dry EEG, a practical next step is to incorporate a brief per-user calibration phase for lightweight subject adaptation. We will evaluate fine-tuning strategies using a minimal amount of labeled subject-specific data (e.g., $\approx 1\text{--}2$ min; 10–20 trials per class). Specifically, we will freeze the feature extractor while updating only a limited parameter subset, such as the final classifier layer or a compact affine feature re-scaling layer. This design keeps calibration time short, minimizes overfitting risks, and preserves the computational footprint for deployment [176–180].

In parallel, we will investigate label-free “baseline” adaptation using eyes-closed resting segments to update normalization statistics (e.g., running mean/variance in normalization layers), thereby mitigating impedance-driven amplitude shifts without requiring task-specific labels [175,178]. Furthermore, the calibration data and online QC signals can be used to identify “low-SNR” users via simple biomarkers already available in our pipeline, such as contact instability events, residual line-noise peaks, and the label-evoked modulation index (e.g., $|\text{Cohen's } d|$ across canonical bands). If these indicators fall below thresholds—suggesting uniformly weak modulation or diffuse attention patterns—the online system would trigger an adaptive operating strategy, such as extending decision windows, applying stronger temporal averaging across consecutive windows, increasing confidence thresholds with abstention under uncertainty, or prompting electrode re-adjustment [181,182]. These steps provide a concrete, deployment-oriented roadmap to transform inter-subject variability from an uncontrolled limitation into a manageable operating condition with targeted mitigation.

While the consistency regularizer showed limited benefit for CSA under our labeled, low-channel dry-EEG protocol, its effectiveness may vary with model capacity, data scale, domain shift, and label availability. A systematic analysis across such conditions is an important direction for future work, and practitioners may enable it depending on the target system constraints.

6. Conclusions

We demonstrate the feasibility of above-chance online left–right CSA decoding with a low-density dry headset (Muse-2, four channels), under subject-wise evaluation and robustness-oriented training. Across 16 participants, the proposed training-time defenses (augmentation + supervised consistency with EMA; All-on-Wav) delivered the best online performance (accuracy = 0.695) versus the same backbone without defenses (0.673), a CNN + LSTM (0.612), and a compact CNN (0.578), while keeping the model lightweight (<0.1 M parameters). Under the same leakage-safe preprocessing and subject-wise evaluation, classical FBCSP + LDA and tangent-space LDA (TS-LDA) baselines remained around mid-0.5 balanced accuracy, highlighting the difficulty of CSA with four-channel dry EEG and the added value of the Hybrid encoder. End-to-end reaction times clustered near ~ 2.03 s per 2.0 s decision window, confirming the system is buffer-limited rather than compute-limited (model inference on CPU ≈ 185 ms; GPU ≈ 11 ms), and therefore deployable on laptop-class CPUs without sacrificing responsiveness.

We further connected these online results to subject-level offline findings. In Section 4, the hybrid All-on-Wav configuration increased balanced accuracy at the paired, subject level for CSA by +2.9%p (95% CI [+0.5, +3.7], permutation $p = 0.037$, power = 0.610, $N = 16$), for Emotion by +2.5%p ($p = 0.008$, $N = 12$), and showed ~0.0%p for MI ($p = 0.906$, $N = 13$). Usability conversions echoed these trends (CSA Err 0.272 $\rightarrow$ 0.268; ITR 3.120 $\rightarrow$ 3.240; Emotion Err 0.124 $\rightarrow$ 0.072; ITR 9.191 $\rightarrow$ 12.586; MI Err 0.114 $\rightarrow$ 0.085; ITR 9.787 $\rightarrow$ 11.638), indicating that gains are meaningful at the interface level. Consistent with dry, few-channel EEG, inter-subject dispersion remained large; however, All-on-Wav raised the upper tail and curtailed extreme low-end outliers, suggesting attention-based reweighting helps when task-relevant features (specifically, oscillatory bursts in alpha and high-beta bands) are present, while not over-amplifying noise.

We analyzed alternative feature families and personalization attempts to clarify design choices. Static filter-bank features and IAF-centered variants were less consistent under this hardware, with IAF adoption near zero under our criteria ($f_s = 256$ Hz, Welch 8–13 Hz, fallback = 10.0 Hz), reinforcing the choice of multiresolution wavelets for non-stationary CSA dynamics and highlighting the practical limits of individualized spectral centering with low-density dry montages.

In addition, we provide a post-hoc interpretability analysis of the attention-based model, including temporal attention heatmaps, spatial relevance over channels, and spectral relevance via band occlusion and integrated gradients (Figures 20 and 21). Specifically, occlusion analyses revealed that high-frequency transients (D2, ≈ 32 –50 Hz) contributed most to model performance, followed by the alpha band (D4, ≈ 8 –16 Hz), with substantial inter-subject variability in spectral reliance.

Limitations include the modest cohort size, four-channel coverage with limited posterior sampling, the 2.0 s decision window that dominates latency, and residual variance across users. Future work will target window/stride optimization to lower reaction time, personalization and online adaptation on top of the robust default (e.g., calibration-efficient fine-tuning), and burst-aware spectral modeling that separates aperiodic and periodic components before any individualized centering. The present study shows that, even with simple hardware, accurate and responsive online CSA control is achievable with compact hybrid models and training-time robustness, enabling practical BCI interaction without heavy computing.

Author Contributions: Conceptualization, J.L.; methodology, D.K. and J.L.; software, D.K.; validation, D.K. and J.L.; formal analysis, D.K. and J.L.; investigation, D.K.; resources, D.K.; data curation, D.K.; writing—original draft preparation, D.K.; writing—review and editing, J.L.; visualization, D.K.; supervision, J.L.; project administration, J.L.; funding acquisition, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Duksung Women's University, Seoul, Republic of Korea (IRB approval No. 2025-011-020-B, approval date: 24 November 2025).

Informed Consent Statement: Written informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

The detailed software and hardware configuration is summarized in Table A1.

Table A1. Software and hardware environment used for training and inference.

Category	Specification
CPU	AMD Ryzen 7 3700X 8-Core Processor @ 3.59 GHz
GPU	NVIDIA GeForce RTX 4070 Laptop GPU (12 GB VRAM)
RAM	32 GB DDR4
Operating System	Windows 11 Pro 22H2 (64-bit, Build 22,621.4317)
Driver/CUDA	NVIDIA Driver 560.94·CUDA 12.6
Python	3.10.16 (conda-forge distribution)
TensorFlow	2.10.0
NumPy	1.26.4
PyWavelets	1.7.0

References

1. Lopez-Gordo, M.A.; Sanchez-Morillo, D.; Valle, F.P. Dry EEG electrodes. *Sensors* **2014**, *14*, 12847–12870. [\[CrossRef\]](#)
2. Huang, Z.; Zhou, Z.; Zeng, J.; Lin, S.; Wu, H. Flexible Electrodes for Non-Invasive Brain–Computer Interfaces: A Perspective. *APL Mater.* **2022**, *10*, 090901. [\[CrossRef\]](#)
3. Krigolson, O.E.; Williams, C.C.; Norton, A.; Hassall, C.D.; Colino, F.L. Choosing MUSE: Validation of a low-cost, portable EEG system for ERP research. *Front. Neurosci.* **2017**, *11*, 109. [\[CrossRef\]](#)
4. Hayes, H.B.; Magne, C. Exploring the utility of the muse headset for capturing the N400: Dependability and single-trial analysis. *Sensors* **2024**, *24*, 7961. [\[CrossRef\]](#)
5. Sidelinger, L.; Zhang, M.; Frohlich, F.; Daughters, S.B. Day-to-day individual alpha frequency variability measured by a mobile EEG device relates to anxiety. *Eur. J. Neurosci.* **2023**, *57*, 1815–1833. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Ratti, E.; Waninger, S.; Berka, C.; Ruffini, G.; Verma, A. Comparison of medical and consumer wireless EEG systems for use in clinical trials. *Front. Hum. Neurosci.* **2017**, *11*, 398. [\[CrossRef\]](#)
7. Ikkai, A.; Dandekar, S.; E Curtis, C. Lateralization in alpha-band oscillations predicts the locus and spatial distribution of attention. *PLoS ONE* **2016**, *11*, e0154796. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Boncompagni, G.; Villena-González, M.; Cosmelli, D.; López, V. Spontaneous alpha power lateralization predicts detection performance in an un-cued signal detection task. *PLoS ONE* **2016**, *11*, e0160347. [\[CrossRef\]](#)
9. Van Diepen, R.M.; Miller, L.M.; Mazaheri, A.; Geng, J.J. The role of alpha activity in spatial and feature-based attention. *ENeuro* **2016**, *3*, ENEURO.0204-16.2016. [\[CrossRef\]](#)
10. Desantis, A.; Chan-Hon-Tong, A.; Collins, T.; Hogendoorn, H.; Cavanagh, P. Decoding the temporal dynamics of covert spatial attention using multivariate EEG analysis: Contributions of raw amplitude and alpha power. *Front. Hum. Neurosci.* **2020**, *14*, 570419. [\[CrossRef\]](#)
11. Thiery, T.; Lajnef, T.; Jerbi, K.; Arguin, M.; Aubin, M.; Jolicoeur, P. Decoding the locus of covert visuospatial attention from EEG signals. *PLoS ONE* **2016**, *11*, e0160304. [\[CrossRef\]](#)
12. Shad, E.H.T.; Molinas, M.; Ytterdal, T. Impedance and noise of passive and active dry EEG electrodes: A review. *IEEE Sens. J.* **2020**, *20*, 14565–14577.
13. Kam, J.W.; Griffin, S.; Shen, A.; Patel, S.; Hinrichs, H.; Heinze, H.-J.; Deouell, L.Y.; Knight, R.T. Systematic comparison between a wireless EEG system with dry electrodes and a wired EEG system with wet electrodes. *NeuroImage* **2019**, *184*, 119–129. [\[CrossRef\]](#)
14. Spüler, M. A high-speed brain-computer interface (BCI) using dry EEG electrodes. *PLoS ONE* **2017**, *12*, e0172400.
15. Riccio, A.; Mattia, D.; Simione, L.; Olivetti, M.; Cincotti, F. Eye-gaze independent EEG-based brain–computer interfaces for communication. *J. Neural Eng.* **2012**, *9*, 045001. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Treder, M.S.; Schmidt, N.M.; Blankertz, B. Gaze-independent brain–computer interfaces based on covert attention and feature attention. *J. Neural Eng.* **2011**, *8*, 066003. [\[CrossRef\]](#)
17. Hwang, H.-J.; Ferreria, V.Y.; Ulrich, D.; Kilic, T.; Chatziliadis, X.; Blankertz, B.; Treder, M. A gaze independent brain-computer interface based on visual stimulation through closed eyelids. *Sci. Rep.* **2015**, *5*, 15890. [\[CrossRef\]](#)
18. Hjortkjær, J.; Wong, D.D.; Catania, A.; Märcher-Rørsted, J.; Ceolini, E.; Fuglsang, S.A.; Kiselev, I.; Di Liberto, G.; Liu, S.-C.; Dau, T. Real-time control of a hearing instrument with EEG-based attention decoding. *J. Neural Eng.* **2025**, *22*, 016027. [\[CrossRef\]](#)
19. White, J.; Power, S.D. k-fold cross-validation can significantly over-estimate true classification accuracy in common EEG-based passive BCI experimental designs: An empirical investigation. *Sensors* **2023**, *23*, 6077. [\[CrossRef\]](#)
20. Brookshire, G.; Kasper, J.; Blauch, N.M.; Wu, Y.C.; Glatt, R.; Merrill, D.A.; Gerrol, S.; Yoder, K.J.; Quirk, C.; Lucero, C. Data leakage in deep learning studies of translational EEG. *Front. Neurosci.* **2024**, *18*, 1373515. [\[CrossRef\]](#)

21. Kapoor, S.; Narayanan, A. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns* **2023**, *4*, 100804. [CrossRef]
22. Combrisson, E.; Jerbi, K. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *J. Neurosci. Methods* **2015**, *250*, 126–136. [CrossRef]
23. Brodersen, K.H.; Ong, C.S.; Stephan, K.E.; Buhmann, J.M. The balanced accuracy and its posterior distribution. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 3121–3124.
24. Scikit-Learn Developers. Balanced_Accuracy_Score: Compute the Balanced Accuracy. Available online: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.balanced_accuracy_score.html (accessed on 6 November 2025).
25. Gosala, B.; Kapgate, P.D.; Jain, P.; Chaurasia, R.N.; Gupta, M. Wavelet transforms for feature engineering in EEG data processing: An application on Schizophrenia. *Biomed. Signal Process. Control* **2023**, *85*, 104811.
26. Lashgari, E.; Liang, D.; Maoz, U. Data augmentation for deep-learning-based electroencephalography. *J. Neurosci. Methods* **2020**, *346*, 108885. [PubMed]
27. George, O.; Smith, R.; Madiraju, P.; Yahyasoltani, N.; Ahamed, S.I. Data augmentation strategies for EEG-based motor imagery decoding. *Heliyon* **2022**, *8*, e10240. [CrossRef] [PubMed]
28. Banville, H.; Wood, S.U.; Aimone, C.; Engemann, D.-A.; Gramfort, A. Robust learning from corrupted EEG with dynamic spatial filtering. *NeuroImage* **2022**, *251*, 118994. [CrossRef] [PubMed]
29. Chakravarthi, B.; Ng, S.-C.; Ezilarasan, M.; Leung, M.-F. EEG-based emotion recognition using hybrid CNN and LSTM classification. *Front. Comput. Neurosci.* **2022**, *16*, 1019776. [CrossRef]
30. Oka, H.; Ono, K.; Panagiotis, A. Attention-Based PSO-LSTM for Emotion Estimation Using EEG. *Sensors* **2024**, *24*, 8174. [CrossRef]
31. Rommel, C.; Moreau, T.; Gramfort, A. CADDA: Class-wise automatic differentiable data augmentation for EEG signals. *arXiv* **2021**, arXiv:2106.13695.
32. Jin, X.; Li, L.; Dang, F.; Chen, X.; Liu, Y. A survey on edge computing for wearable technology. *Digit. Signal Process.* **2022**, *125*, 103146. [CrossRef]
33. Wilson, J.A.; Mellinger, J.; Schalk, G.; Williams, J. A procedure for measuring latencies in brain–computer interfaces. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 1785–1797. [CrossRef]
34. Arvaneh, M.; Ward, T.E.; Robertson, I.H. Effects of feedback latency on P300-based brain-computer interface. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2015**, *2015*, 2315–2318.
35. Mwata-Velu, T.; Niyonsaba-Sebigunda, E.; Avina-Cervantes, J.G.; Ruiz-Pinales, J.; Velu-A-Gulenga, N.; Alonso-Ramírez, A.A. Motor Imagery Multi-Tasks Classification for BCIs Using the NVIDIA Jetson TX2 Board and the EEGNet Network. *Sensors* **2023**, *23*, 4164. [CrossRef] [PubMed]
36. Van Erp, J.; Lotte, F.; Tangermann, M. Brain-computer interfaces: Beyond medical applications. *Computer* **2012**, *45*, 26–34. [CrossRef]
37. Gomez-Rivera, Y.; Cardona Álvarez, Y.; Gomez-Morales, O.; Alvarez-Meza, A.; Castellanos-Domínguez, G. BCI-based real-time processing for implementing deep learning frameworks using motor imagery paradigms. *J. Appl. Res. Technol.* **2024**, *22*, 646–653. [CrossRef]
38. Nicolas-Alonso, L.F.; Gomez-Gil, J. Brain computer interfaces, a review. *Sensors* **2012**, *12*, 1211–1279. [CrossRef] [PubMed]
39. Farwell, L.; Donchin, E. Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr. Clin. Neurophysiol.* **1988**, *70*, 510–523. [CrossRef]
40. Pan, J.; Chen, X.; Ban, N.; He, J.; Chen, J.; Huang, H. Advances in P300 brain–computer interface spellers: Toward paradigm design and performance evaluation. *Front. Hum. Neurosci.* **2022**, *16*, 1077717. [CrossRef]
41. Lin, Z.; Zhang, C.; Wu, W.; Gao, X. Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 2610–2614. [CrossRef]
42. Chen, X.; Wang, Y.; Gao, S.; Jung, T.-P.; Gao, X. Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain–computer interface. *J. Neural Eng.* **2015**, *12*, 046008. [CrossRef]
43. Nakanishi, M.; Wang, Y.; Chen, X.; Wang, Y.-T.; Gao, X.; Jung, T.-P. Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis. *IEEE Trans. Biomed. Eng.* **2017**, *65*, 104–112. [CrossRef]
44. Li, M.; He, D.; Li, C.; Qi, S. Brain–computer interface speller based on steady-state visual evoked potential: A review focusing on the stimulus paradigm and performance. *Brain Sci.* **2021**, *11*, 450. [CrossRef]
45. Xie, J.; Xu, G.; Wang, J.; Li, M.; Han, C.; Jia, Y. Effects of mental load and fatigue on steady-state evoked potential based brain computer interface tasks: A comparison of periodic flickering and motion-reversal based visual attention. *PLoS ONE* **2016**, *11*, e0163426. [CrossRef]
46. Van Diepen, R.M.; Foxe, J.J.; Mazaheri, A. The functional role of alpha-band activity in attentional processing: The current zeitgeist and future outlook. *Curr. Opin. Psychol.* **2019**, *29*, 229–238. [CrossRef]

47. Van Ede, F.; Quinn, A.J.; Woolrich, M.W.; Nobre, A.C. Neural oscillations: Sustained rhythms or transient burst-events? *Trends Neurosci.* **2018**, *41*, 415–417. [[CrossRef](#)] [[PubMed](#)]
48. Lawhern, V.J.; Solon, A.J.; Waytowich, N.R.; Gordon, S.M.; Hung, C.P.; Lance, B.J. EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces. *J. Neural Eng.* **2018**, *15*, 056013. [[CrossRef](#)]
49. Cisotto, G.; Zanga, A.; Chlebus, J.; Zoppis, I.; Manzoni, S.; Markowska-Kaczmar, U. Comparison of attention-based deep learning models for eeg classification. *arXiv* **2020**, arXiv:2012.01074. [[CrossRef](#)]
50. Kiss, M.; Van Velzen, J.; Eimer, M. The N2pc component and its links to attention shifts and spatially selective visual processing. *Psychophysiology* **2008**, *45*, 240–249. [[CrossRef](#)]
51. Li, C.; Liu, Q.; Hu, Z. Further evidence that N2pc reflects target enhancement rather than distracter suppression. *Front. Psychol.* **2018**, *8*, 2275. [[CrossRef](#)]
52. Foxe, J.J.; Snyder, A.C. The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Front. Psychol.* **2011**, *2*, 154. [[CrossRef](#)]
53. Wöstmann, M.; Alavash, M.; Obleser, J. Alpha oscillations in the human brain implement distractor suppression independent of target selection. *J. Neurosci.* **2019**, *39*, 9797–9805. [[CrossRef](#)]
54. Schneider, D.; Herbst, S.K.; Klatt, L.; Wöstmann, M. Target enhancement or distractor suppression? Functionally distinct alpha oscillations form the basis of attention. *Eur. J. Neurosci.* **2022**, *55*, 3256–3265. [[CrossRef](#)]
55. Banville, H.; Chehab, O.; Hyvärinen, A.; Engemann, D.-A.; Gramfort, A. Uncovering the structure of clinical EEG signals with self-supervised learning. *J. Neural Eng.* **2021**, *18*, 046020. [[CrossRef](#)] [[PubMed](#)]
56. Kostas, D.; Aroca-Ouellette, S.; Rudzicz, F. BENDR: Using transformers and a contrastive self-supervised learning task to learn from massive amounts of EEG data. *Front. Hum. Neurosci.* **2021**, *15*, 653659. [[CrossRef](#)]
57. Reichert, C.; Tellez Ceja, I.F.; Sweeney-Reed, C.M.; Heinze, H.-J.; Hinrichs, H.; Dürschmid, S. Impact of stimulus features on the performance of a gaze-independent brain-computer interface based on covert spatial attention shifts. *Front. Neurosci.* **2020**, *14*, 591777. [[CrossRef](#)]
58. Kim, S.; Lee, S.; Kang, H.; Kim, S.; Ahn, M. P300 brain–computer interface-based drone control in virtual and augmented reality. *Sensors* **2021**, *21*, 5765. [[CrossRef](#)]
59. LaFleur, K.; Cassidy, K.; Doud, A.; Shades, K.; Rogin, E.; He, B. Quadcopter control in three-dimensional space using a noninvasive motor imagery-based brain–computer interface. *J. Neural Eng.* **2013**, *10*, 046003. [[CrossRef](#)] [[PubMed](#)]
60. Prapas, G.; Angelidis, P.; Sarigiannidis, P.; Bibi, S.; Tsiouras, M.G. Connecting the brain with augmented reality: A systematic review of BCI-AR systems. *Appl. Sci.* **2024**, *14*, 9855. [[CrossRef](#)]
61. Mikhaylov, D.; Saeed, M.; Husain Alhosani, M.; F. Al Wahedi, Y. Comparison of EEG Signal Spectral Characteristics Obtained with Consumer-and Research-Grade Devices. *Sensors* **2024**, *24*, 8108. [[CrossRef](#)]
62. Hinrichs, H.; Scholz, M.; Baum, A.K.; Kam, J.W.Y.; Knight, R.T.; Heinze, H.-J. Comparison between a wireless dry electrode EEG system with a conventional wired wet electrode EEG system for clinical applications. *Sci. Rep.* **2020**, *10*, 5218. [[CrossRef](#)]
63. Ang, K.K.; Chin, Z.Y.; Wang, C.; Guan, C.; Zhang, H. Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. *Front. Neurosci.* **2012**, *6*, 39. [[CrossRef](#)] [[PubMed](#)]
64. Barachant, A.; Bonnet, S.; Congedo, M.; Jutten, C. Multiclass brain–computer interface classification by Riemannian geometry. *IEEE Trans. Biomed. Eng.* **2011**, *59*, 920–928. [[CrossRef](#)]
65. Schirrmeister, R.T.; Springenberg, J.T.; Fiederer, L.D.J.; Glasstetter, M.; Eggensperger, K.; Tangermann, M.; Hutter, F.; Burgard, W.; Ball, T. Deep learning with convolutional neural networks for EEG decoding and visualization. *Hum. Brain Mapp.* **2017**, *38*, 5391–5420. [[CrossRef](#)]
66. Zhao, W.; Jiang, X.; Zhang, B.; Xiao, S.; Weng, S. CTNet: A convolutional transformer network for EEG-based motor imagery classification. *Sci. Rep.* **2024**, *14*, 20237. [[CrossRef](#)]
67. Lee, Y.-E.; Lee, S.-H. EEG-transformer: Self-attention from transformer architecture for decoding EEG of imagined speech. In Proceedings of the 2022 10th International Winter Conference on Brain-Computer Interface (BCI), Gangwon-do, Republic of Korea, 21–23 February 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–4.
68. Wan, Z.; Li, M.; Liu, S.; Huang, J.; Tan, H.; Duan, W. EEGformer: A transformer–based brain activity classification method using EEG signal. *Front. Neurosci.* **2023**, *17*, 1148855. [[CrossRef](#)] [[PubMed](#)]
69. Vafaei, E.; Hosseini, M. Transformers in EEG Analysis: A review of architectures and applications in motor imagery, seizure, and emotion classification. *Sensors* **2025**, *25*, 1293. [[CrossRef](#)] [[PubMed](#)]
70. Bigdely-Shamlo, N.; Mullen, T.; Kothe, C.; Su, K.-M.; Robbins, K.A. The PREP pipeline: Standardized preprocessing for large-scale EEG analysis. *Front. Neuroinform.* **2015**, *9*, 16. [[CrossRef](#)] [[PubMed](#)]
71. Jas, M.; Engemann, D.A.; Bekhti, Y.; Raimondo, F.; Gramfort, A. Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage* **2017**, *159*, 417–429. [[CrossRef](#)]
72. Pion-Tonachini, L.; Kreutz-Delgado, K.; Makeig, S. ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage* **2019**, *198*, 181–197. [[CrossRef](#)]

73. Lopez, K.L.; Monachino, A.D.; Morales, S.; Leach, S.C.; Bowers, M.E.; Gabard-Durnam, L.J. HAPPILEE: HAPPE In Low Electrode Electroencephalography, a standardized pre-processing software for lower density recordings. *NeuroImage* **2022**, *260*, 119390. [CrossRef]
74. Hill, A.T.; Enticott, P.G.; Fitzgerald, P.B.; Bailey, N.W. RELAX-Jr: An Automated Pre-Processing Pipeline for Developmental EEG Recordings. *Hum. Brain Mapp.* **2024**, *45*, e70034. [CrossRef]
75. Habashi, A.G.; Azab, A.M.; Eldawlatly, S.; Aly, G.M. Generative adversarial networks in EEG analysis: An overview. *J. Neuroeng. Rehabil.* **2023**, *20*, 40. [CrossRef] [PubMed]
76. Du, X.; Wang, X.; Zhu, L.; Ding, X.; Lv, Y.; Qiu, S.; Liu, Q. Electroencephalographic signal data augmentation based on improved generative adversarial network. *Brain Sci.* **2024**, *14*, 367. [CrossRef]
77. Bao, G.; Yan, B.; Tong, L.; Shu, J.; Wang, L.; Yang, K.; Zeng, Y. Data augmentation for EEG-based emotion recognition using generative adversarial networks. *Front. Comput. Neurosci.* **2021**, *15*, 723843. [CrossRef] [PubMed]
78. Weng, W.; Gu, Y.; Guo, S.; Ma, Y.; Yang, Z.; Liu, Y.; Chen, Y. Self-supervised learning for electroencephalogram: A systematic survey. *ACM Comput. Surv.* **2025**, *57*, 1–38. [CrossRef]
79. Varoquaux, G. Cross-validation failure: Small sample sizes lead to large error bars. *Neuroimage* **2018**, *180*, 68–77. [CrossRef]
80. Jamalabadi, H.; Alizadeh, S.; Schönauer, M.; Leibold, C.; Gais, S. Multivariate classification of neuroimaging data with nested subclasses: Biased accuracy and implications for hypothesis testing. *PLoS Comput. Biol.* **2018**, *14*, e1006486. [CrossRef]
81. InteraXon. Muse 2—EEG Headband Technical Specifications. Available online: <https://choosemuse.com/products/muse-2> (accessed on 6 November 2025).
82. Krigolson, O.E.; Hammerstrom, M.R.; Abimbola, W.; Trska, R.; Wright, B.W.; Hecker, K.G.; Binsted, G. Using Muse: Rapid mobile assessment of brain performance. *Front. Neurosci.* **2021**, *15*, 634147. [CrossRef]
83. Date, T.S. Muse 2 Headband Specifications. *Tecnológico de Monterrey*. Available online: <https://ifeldh.tec.mx/> (accessed on 12 November 2025).
84. Li, G.; Wang, S.; Duan, Y.Y. Towards gel-free electrodes: A systematic study of electrode-skin impedance. *Sens. Actuators B Chem.* **2017**, *241*, 1244–1255. [CrossRef]
85. Tautan, A.-M.; Mihajlovic, V.; Chen, Y.-H.; Grundlehner, B.; Penders, J.; Serdijn, W. Signal quality in dry electrode EEG and the relation to skin-electrode contact impedance magnitude. In Proceedings of the International Conference on Biomedical Electronics and Devices (BIODEVICES 2014), ESEO, Angers, France, 3–6 March 2014; SciTePress: Setúbal, Portugal, 2014; pp. 12–22.
86. Lee, S.; Kim, M.; Ahn, M. Evaluation of consumer-grade wireless EEG systems for brain-computer interface applications. *Biomed. Eng. Lett.* **2024**, *14*, 1433–1443. [CrossRef]
87. World Medical Association. World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA* **2013**, *310*, 2191–2194. [CrossRef]
88. InteraXon Inc. MuseLab (v1.9.5) [Software]. Available online: <https://choosemuse.com/> (accessed on 6 November 2025).
89. Worden, M.S.; Foxe, J.J.; Wang, N.; Simpson, G.V. Anticipatory biasing of visuospatial attention indexed by retinotopically specific alpha-band electroencephalography increases over occipital cortex. *J. Neurosci. Off. J. Soc. Neurosci.* **2000**, *20*, RC63. [CrossRef]
90. Thut, G.; Nietzel, A.; Brandt, S.A.; Pascual-Leone, A. α -Band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *J. Neurosci.* **2006**, *26*, 9494–9502. [CrossRef] [PubMed]
91. Pfurtscheller, G.; Da Silva, F.L. Event-related EEG/MEG synchronization and desynchronization: Basic principles. *Clin. Neurophysiol.* **1999**, *110*, 1842–1857. [CrossRef] [PubMed]
92. Neuper, C.; Scherer, R.; Reiner, M.; Pfurtscheller, G. Imagery of motor actions: Differential effects of kinesthetic and visual-motor mode of imagery in single-trial EEG. *Cogn. Brain Res.* **2005**, *25*, 668–677. [CrossRef] [PubMed]
93. Koelstra, S.; Muhl, C.; Soleymani, M.; Lee, J.-S.; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijholt, A.; Patras, I. Deap: A database for emotion analysis; using physiological signals. *IEEE Trans. Affect. Comput.* **2011**, *3*, 18–31. [CrossRef]
94. Zheng, W.-L.; Lu, B.-L. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Trans. Auton. Ment. Dev.* **2015**, *7*, 162–175. [CrossRef]
95. Allison, B.Z.; Pineda, J.A. Effects of SOA and flash pattern manipulations on ERPs, performance, and preference: Implications for a BCI system. *Int. J. Psychophysiol.* **2006**, *59*, 127–140. [CrossRef]
96. Sugi, M.; Hagimoto, Y.; Nambu, I.; Gonzalez, A.; Takei, Y.; Yano, S.; Hokari, H.; Wada, Y. Improving the performance of an auditory brain-computer interface using virtual sound sources by shortening stimulus onset asynchrony. *Front. Neurosci.* **2018**, *12*, 108. [CrossRef]
97. Lotte, F.; Larrue, F.; Mühl, C. Flaws in current human training protocols for spontaneous brain-computer interfaces: Lessons learned from instructional design. *Front. Hum. Neurosci.* **2013**, *7*, 568. [CrossRef]
98. Keil, A.; Debener, S.; Gratton, G.; Junghöfer, M.; Kappenman, E.S.; Luck, S.J.; Luu, P.; Miller, G.A.; Yee, C.M. Committee report: Publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. *Psychophysiology* **2014**, *51*, 1–21. [CrossRef]

99. Kappenman, E.S.; Luck, S.J. Best practices for event-related potential research in clinical populations. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **2016**, *1*, 110–115. [CrossRef]
100. Pernet, C.R.; Appelhoff, S.; Gorgolewski, K.J.; Flandin, G.; Phillips, C.; Delorme, A.; Oostenveld, R. EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Sci. Data* **2019**, *6*, 103. [CrossRef] [PubMed]
101. BIDS Maintainers. The Brain Imaging Data Structure (BIDS) Specification, v1.10+. Available online: <https://bids-specification.readthedocs.io/> (accessed on 6 November 2025).
102. Chaumon, M.; Bishop, D.V.; Busch, N.A. A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *J. Neurosci. Methods* **2015**, *250*, 47–63. [CrossRef]
103. Muthukumaraswamy, S.D. High-frequency brain activity and muscle artifacts in MEG/EEG: A review and recommendations. *Front. Hum. Neurosci.* **2013**, *7*, 138. [CrossRef] [PubMed]
104. Widmann, A.; Schröger, E.; Maess, B. Digital filter design for electrophysiological data—A practical approach. *J. Neurosci. Methods* **2015**, *250*, 34–46. [CrossRef]
105. SciPy Developers. signal.iirnotch—Notch Filter Design. Available online: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.iirnotch.html> (accessed on 6 November 2025).
106. SciPy Developers. signal.lfilter—IIR Filtering. Available online: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.lfilter.html> (accessed on 6 November 2025).
107. SciPy Developers. scipy.signal (Butter, Butterd, Sossfiltfilt). Available online: <https://docs.scipy.org/doc/scipy/reference/signal.html> (accessed on 6 November 2025).
108. Acunzo, D.J.; Mackenzie, G.; Van Rossum, M.C.W. Systematic biases in early ERP and ERF components as a result of high-pass filtering. *J. Neurosci. Methods* **2012**, *209*, 212–218. [CrossRef] [PubMed]
109. Tanner, D.; Morgan-Short, K.; Luck, S.J. How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology* **2015**, *52*, 997–1009. [CrossRef]
110. Gramfort, A.; Luessi, M.; Larson, E.; Engemann, D.A.; Strohmeier, D.; Brodbeck, C.; Parkkonen, L.; Hämäläinen, M.S. MNE software for processing MEG and EEG data. *Neuroimage* **2014**, *86*, 446–460. [CrossRef]
111. Klimesch, W. EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Res. Rev.* **1999**, *29*, 169–195. [CrossRef]
112. Harris, F.J. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proc. IEEE* **2005**, *66*, 51–83. [CrossRef]
113. Cohen, M.X. *Analyzing Neural Time Series Data: Theory and Practice*; MIT Press: Cambridge, MA, USA, 2014.
114. Welch, P. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* **2003**, *15*, 70–73. [CrossRef]
115. SciPy Developers. signal.welch—Welch’s Method for PSD Estimation. Available online: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.welch.html> (accessed on 6 November 2025).
116. Thorpe, S.; D’zMura, M.; Srinivasan, R. Lateralization of frequency-specific networks for covert spatial attention to auditory stimuli. *Brain Topogr.* **2012**, *25*, 39–54. [CrossRef]
117. Allen, D.P.; MacKinnon, C.D. Time–frequency analysis of movement-related spectral power in EEG during repetitive movements: A comparison of methods. *J. Neurosci. Methods* **2010**, *186*, 107–115. [CrossRef]
118. Moca, V.V.; Bârzan, H.; Nagy-Dăbâcan, A.; Mureșan, R.C. Time-frequency super-resolution with superlets. *Nat. Commun.* **2021**, *12*, 337. [CrossRef] [PubMed]
119. Amin, H.U.; Malik, A.S.; Ahmad, R.F.; Badruddin, N.; Kamel, N.; Hussain, M.; Chooi, W.-T. Feature extraction and classification for EEG signals using wavelet transform and machine learning techniques. *Australas. Phys. Eng. Sci. Med.* **2015**, *38*, 139–149. [CrossRef]
120. Wu, T.; Kong, X.; Zhong, Y.; Chen, L. Automatic detection of abnormal EEG signals using multiscale features with ensemble learning. *Front. Hum. Neurosci.* **2022**, *16*, 943258. [CrossRef]
121. Al-Qazzaz, N.K.; Hamid Bin Mohd Ali, S.; Ahmad, S.A.; Islam, M.S.; Escudero, J. Selection of mother wavelet functions for multi-channel EEG signal analysis during a working memory task. *Sensors* **2015**, *15*, 29015–29035. [CrossRef] [PubMed]
122. Sutterer, D.W.; Polyn, S.M.; Woodman, G.F. α -Band activity tracks a two-dimensional spotlight of attention during spatial working memory maintenance. *J. Neurophysiol.* **2021**, *125*, 957–971. [CrossRef]
123. Gandhi, T.; Panigrahi, B.K.; Anand, S. A comparative study of wavelet families for EEG signal classification. *Neurocomputing* **2011**, *74*, 3051–3057. [CrossRef]
124. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861. [CrossRef]
125. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.
126. Jain, S.; Wallace, B.C. Attention is not explanation. *arXiv* **2019**, arXiv:1902.10186.

127. Wiegreffe, S.; Pinter, Y. Attention is not not explanation. *arXiv* **2019**, arXiv:1908.04626.
128. Wang, X.; Liesaputra, V.; Liu, Z.; Wang, Y.; Huang, Z. An in-depth survey on deep learning-based motor imagery electroencephalogram (EEG) classification. *Artif. Intell. Med.* **2024**, *147*, 102738. [[CrossRef](#)]
129. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. *arXiv* **2017**, arXiv:1711.05101.
130. Goyal, P.; Dollár, P.; Girshick, R.; Noordhuis, P.; Wesolowski, L.; Kyrola, A.; Tulloch, A.; Jia, Y.; He, K. Accurate, large minibatch sgd: Training imagenet in 1 hour. *arXiv* **2017**, arXiv:1706.02677.
131. Loshchilov, I.; Hutter, F. Sgdr: Stochastic gradient descent with warm restarts. *arXiv* **2016**, arXiv:1608.03983.
132. Pascanu, R.; Mikolov, T.; Bengio, Y. On the difficulty of training recurrent neural networks. *Int. Conf. Mach. Learn. PMLR* **2013**, *28*, 1310–1318.
133. Youden, W.J. Index for rating diagnostic tests. *Cancer* **1950**, *3*, 32–35. [[CrossRef](#)]
134. Park, D.S.; Chan, W.; Zhang, Y.; Chiu, C.-C.; Zoph, B.; Cubuk, E.D.; Le, Q.V. SpecAugment: A simple data augmentation method for automatic speech recognition. *arXiv* **2019**, arXiv:1904.08779. [[CrossRef](#)]
135. Li, B.; Xu, Y.; Wang, Y.; Li, L.; Zhang, B. The student-teacher framework guided by self-training and consistency regularization for semi-supervised medical image segmentation. *PLoS ONE* **2024**, *19*, E0300039. [[CrossRef](#)] [[PubMed](#)]
136. Tarvainen, A.; Valpola, H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1195–1204.
137. Zhang, H.; Cisse, M.; Dauphin, Y.N.; Lopez-Paz, D. mixup: Beyond empirical risk minimization. *arXiv* **2017**, arXiv:1710.09412.
138. Yun, S.; Han, D.; Oh, S.J.; Chun, S.; Choe, J.; Yoo, Y. Cutmix: Regularization strategy to train strong classifiers with localizable features. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6023–6032. [[CrossRef](#)]
139. Varoquaux, G.; Raamana, P.R.; Engemann, D.A.; Hoyos-Idrobo, A.; Schwartz, Y.; Thirion, B. Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *NeuroImage* **2017**, *145*, 166–179. [[CrossRef](#)]
140. Thölke, P.; Mantilla-Ramos, Y.-J.; Abdelhedi, H.; Maschke, C.; Dehgan, A.; Harel, Y.; Kemtur, A.; Berrada, L.M.; Sahraoui, M.; Young, T. Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data. *NeuroImage* **2023**, *277*, 120253. [[CrossRef](#)]
141. Cha, H.; Kim, D.M.; Gong, T.; Chung, H.W.; Lee, S.-J. SNAP-TTA: Sparse Test-Time Adaptation for Latency-Sensitive Applications. *arXiv* **2025**, arXiv:2511.15276.
142. Guo, C.; Pleiss, G.; Sun, Y.; Weinberger, K.Q. On calibration of modern neural networks. *Int. Conf. Mach. Learn. PMLR* **2017**, *70*, 1321–1330.
143. Button, K.S.; Ioannidis, J.P.; Mokrysz, C.; Nosek, B.A.; Flint, J.; Robinson, E.S.; Munafò, M.R. Power failure: Why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* **2013**, *14*, 365–376. [[CrossRef](#)] [[PubMed](#)]
144. Efron, B.; Tibshirani, R.J. *An Introduction to the Bootstrap*; Chapman and Hall/CRC: Boca Raton, FL, USA, 1994.
145. Wolpaw, J.; Birbaumer, N.; McFarland, D.J.; Pfurtscheller, G.; Vaughan, T. Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* **2002**, *113*, 767–791. [[CrossRef](#)]
146. Nadra, J.G.; Bengson, J.J.; Morales, A.B.; Mangun, G.R. Attention without constraint: Alpha lateralization in uncued willed attention. *Eneuro* **2023**, *10*, ENEURO.0258-22.2023. [[CrossRef](#)]
147. Foster, J.J.; Awh, E. The role of alpha oscillations in spatial attention: Limited evidence for a suppression account. *Curr. Opin. Psychol.* **2019**, *29*, 34–40. [[CrossRef](#)] [[PubMed](#)]
148. Rhif, M.; Ben Abbes, A.; Farah, I.R.; Martínez, B.; Sang, Y. Wavelet transform application for/in non-stationary time-series analysis: A review. *Appl. Sci.* **2019**, *9*, 1345. [[CrossRef](#)]
149. Al-Fahoum, A.S.; Al-Fraihat, A.A. Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains. *Int. Sch. Res. Not.* **2014**, *2014*, 730218. [[CrossRef](#)]
150. Donoghue, T.; Haller, M.; Peterson, E.J.; Varma, P.; Sebastian, P.; Gao, R.; Noto, T.; Lara, A.H.; Wallis, J.D.; Knight, R.T. Parameterizing neural power spectra into periodic and aperiodic components. *Nat. Neurosci.* **2020**, *23*, 1655–1665. [[CrossRef](#)] [[PubMed](#)]
151. Seymour, R.A.; Alexander, N.; Maguire, E.A. Robust estimation of 1/f activity improves oscillatory burst detection. *Eur. J. Neurosci.* **2022**, *56*, 5836–5852. [[CrossRef](#)]
152. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
153. Rommel, C.; Paillard, J.; Moreau, T.; Gramfort, A. Data augmentation for learning predictive models on EEG: A systematic comparison. *J. Neural Eng.* **2022**, *19*, 066020. [[CrossRef](#)]
154. Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; Krishnan, D. Supervised contrastive learning. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 18661–18673.
155. Keras Team. ReduceLROnPlateau (Callbacks API), v2.16.1. 2025. Available online: <https://keras.io/api/callbacks/> (accessed on 13 October 2025).

156. Clements, J.; Sellers, E.; Ryan, D.; Caves, K.; Collins, L.; Throckmorton, C. Applying dynamic data collection to improve dry electrode system performance for a P300-based brain–computer interface. *J. Neural Eng.* **2016**, *13*, 066018. [[CrossRef](#)]
157. Gorjan, D.; Gramann, K.; De Pauw, K.; Marusic, U. Removal of movement-induced EEG artifacts: Current state of the art and guidelines. *J. Neural Eng.* **2022**, *19*, 011004. [[CrossRef](#)]
158. Yang, S.-Y.; Lin, Y.-P. Movement artifact suppression in wearable low-density and dry eeg recordings using active electrodes and artifact subspace reconstruction. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2023**, *31*, 3844–3853. [[CrossRef](#)]
159. Ledwidge, P.S.; McPherson, C.N.; Faulkenberg, L.; Morgan, A.; Baylis, G.C. A comparison of approaches for motion artifact removal from wireless mobile EEG during overground running. *Sensors* **2025**, *25*, 4810. [[CrossRef](#)] [[PubMed](#)]
160. Ravichandran, V.; Ciesielska-Wrobel, I.; Rumon, M.A.A.; Solanki, D.; Mankodiya, K. Characterizing the impedance properties of dry e-textile electrodes based on contact force and perspiration. *Biosensors* **2023**, *13*, 728. [[CrossRef](#)]
161. Lotte, F.; Bougrain, L.; Cichocki, A.; Clerc, M.; Congedo, M.; Rakotomamonjy, A.; Yger, F. A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update. *J. Neural Eng.* **2018**, *15*, 031005. [[CrossRef](#)]
162. Bruns, A. Fourier-, Hilbert- and wavelet-based signal analysis: Are they really different approaches? *J. Neurosci. Methods* **2004**, *137*, 321–332. [[CrossRef](#)] [[PubMed](#)]
163. Frikha, T.; Abdennour, N.; Chaabane, F.; Ghorbel, O.; Ayedi, R.; Shahin, O.R.; Cheikhrouhou, O. Source Localization of EEG Brainwaves Activities via Mother Wavelets Families for SWT Decomposition. *J. Healthc. Eng.* **2021**, *2021*, 9938646. [[CrossRef](#)]
164. Bahdanau, D.; Cho, K.; Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv* **2014**, arXiv:1409.0473.
165. Jayaram, V.; Alamgir, M.; Altun, Y.; Scholkopf, B.; Grosse-Wentrup, M. Transfer learning in brain-computer interfaces. *IEEE Comput. Intell. Mag.* **2016**, *11*, 20–31. [[CrossRef](#)]
166. Neuper, C.; Pfurtscheller, G. Event-related dynamics of cortical rhythms: Frequency-specific features and functional correlates. *Int. J. Psychophysiol.* **2001**, *43*, 41–58. [[CrossRef](#)]
167. Lin, Y.-P.; Wang, C.-H.; Jung, T.-P.; Wu, T.-L.; Jeng, S.-K.; Duann, J.-R.; Chen, J.-H. EEG-based emotion recognition in music listening. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 1798–1806. [[CrossRef](#)]
168. Alarcao, S.M.; Fonseca, M.J. Emotions recognition using EEG signals: A survey. *IEEE Trans. Affect. Comput.* **2017**, *10*, 374–393. [[CrossRef](#)]
169. Li, X.; Song, D.; Zhang, P.; Zhang, Y.; Hou, Y.; Hu, B. Exploring EEG features in cross-subject emotion recognition. *Front. Neurosci.* **2018**, *12*, 162. [[CrossRef](#)]
170. Treder, M.S.; Blankertz, B. (C) overt attention and visual speller design in an ERP-based brain-computer interface. *Behav. Brain Funct.* **2010**, *6*, 28. [[CrossRef](#)]
171. Eimer, M. The N2pc component as an indicator of attentional selectivity. *Electroencephalogr. Clin. Neurophysiol.* **1996**, *99*, 225–234. [[CrossRef](#)] [[PubMed](#)]
172. Yuval-Greenberg, S.; Merriam, E.P.; Heeger, D.J. Spontaneous microsaccades reflect shifts in covert attention. *J. Neurosci.* **2014**, *34*, 13693–13700. [[CrossRef](#)]
173. Thielen, J.; Bosch, S.E.; Van Leeuwen, T.M.; Van Gerven, M.A.J.; Van Lier, R. Evidence for confounding eye movements under attempted fixation and active viewing in cognitive neuroscience. *Sci. Rep.* **2019**, *9*, 17456. [[CrossRef](#)]
174. Hafed, Z.M.; Clark, J.J. Microsaccades as an overt measure of covert attention shifts. *Vis. Res.* **2002**, *42*, 2533–2545. [[CrossRef](#)] [[PubMed](#)]
175. Gu, Q.; Zhang, Q.; Han, Y.; Li, P.; Gao, Z.; Shen, M. Microsaccades reflect attention shifts: A mini review of 20 years of microsaccade research. *Front. Psychol.* **2024**, *15*, 1364939. [[CrossRef](#)]
176. Wu, D.; Xu, Y.; Lu, B.-L. Transfer Learning for EEG-Based Brain-Computer Interfaces: A Review of Progress Made Since 2016. *IEEE Trans. Cogn. Dev. Syst.* **2020**, *14*, 4–19. [[CrossRef](#)]
177. He, H.; Wu, D. Transfer Learning for Brain-Computer Interfaces: A Euclidean Space Data Alignment Approach. *IEEE Trans. Biomed. Eng.* **2020**, *67*, 399–410. [[CrossRef](#)]
178. Zanini, P.; Congedo, M.; Jutten, C.; Said, S.; Berthoumieu, Y. Transfer learning: A Riemannian geometry framework with applications to brain-computer interfaces. *IEEE Trans. Biomed. Eng.* **2018**, *65*, 1107–1116. [[CrossRef](#)]
179. Xu, L.; Ma, Z.; Meng, J.; Xu, M.; Jung, T.-P.; Ming, D. Improving Transfer Performance of Deep Learning with Adaptive Batch Normalization for Brain-computer Interfaces. In Proceedings of the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Virtual, 1–5 November 2021; pp. 5800–5803.
180. Liu, S.; Zhang, J.; Wang, A.; Wu, H.; Zhao, Q.; Long, J. Subject adaptation convolutional neural network for EEG-based motor imagery classification. *J. Neural Eng.* **2022**, *19*, 066003. [[CrossRef](#)] [[PubMed](#)]

181. Chow, C.K. On optimum recognition error and reject tradeoff. *IEEE Trans. Inf. Theory* **1970**, *16*, 41–46. [[CrossRef](#)]
182. El-Yaniv, R.; Wiener, Y. On the Foundations of Noise-free Selective Classification. *J. Mach. Learn. Res.* **2010**, *11*, 1605–1641.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.