Journal of Neural Engineering

PURPOSE-LED PUBLISHING™

Multi-head noise regression for single-channel EEG: estimating ocular and muscle contamination to guide artifact removal

To cite this article before publication: Usman Qamar Shaikh *et al* 2026 *J. Neural Eng.* in press https://doi.org/10.1088/1741-2552/ae541d

(Major Revision) source file

Manuscript ID: JNE-109929

Title

Multi-head noise regression for single-channel EEG: estimating ocular and muscle contamination to guide artifact removal.

Authors

Usman Qamar Shaikh (Corresponding Author)
Institute of Biomedical Technologies,
Electrical and Electronic Engineering Department,
Faculty of Engineering, Computer and Mathematical Sciences,
Auckland University of Technology, Auckland, New Zealand
0009-0000-0171-3940
usman.shaikh@autuni.ac.nz

Anubha Manju Kalra
Institute of Biomedical Technologies,
Electrical and Electronic Engineering Department,
Faculty of Engineering, Computer and Mathematical Sciences,
Auckland University of Technology, Auckland, New Zealand
0000-0001-7908-3053

Andrew Lowe
Institute of Biomedical Technologies,
Mechanical Engineering Department,
Faculty of Engineering, Computer and Mathematical Sciences,
Auckland University of Technology, Auckland, New Zealand
0000-0002-0612-1036

Imran Khan Niazi
1) Centre for Chiropractic Research,
New Zealand College of Chiropractic,
Auckland, New Zealand,
2) Department of Health Science and Technology,
Aalborg University, Aalborg, Denmark
0000-0001-8752-7224

Abstract

Objective. EEG is often contaminated by ocular (EOG) and muscle (EMG) artifacts, yet many pipelines apply uniform denoising, risking distortion of clean neural activity. We propose a two-head, single-channel regressor that estimates EOG and EMG noise-to-signal ratio (NSR, dB) from short segments and test whether it can guide selective artifact reduction, including downstream BCI decoding.

Approach. Using EEGdenoiseNet clean EEG and artifact exemplars, we synthesised 2-s single-channel mixtures with known EOG/EMG NSR spanning –10 to +10 dB and trained several model families to jointly regress both NSRs. Generalisation was evaluated on an independent eyeblink dataset via agreement with regression-based ocular-reference topographies, and in two applications: (i) gating stationary wavelet blink removal on a P3 ERP dataset and (ii) gating the same denoiser on a 55-subject RSVP P300 speller dataset (FP1/FP2).

Main results. A dilated temporal convolutional network (TCN) performed best (EOG: MAE ≈ 1.8 dB, R^2 ≈ 0.82; EMG: MAE ≈ 1.0 dB, R^2 ≈ 0.94) with low bias across NSR. The EOG head recovered blink topographies (median spatial correlation ≈ 0.91). On the P3 dataset, indiscriminate wavelet denoising reduced significant ERP channels, whereas TCN-guided gating preserved 22–23 of 24 while processing ~9–20% of segments. On the speller dataset, denoising all epochs reduced decoding, while selective denoising improved AUC (θ = 9 dB: ΔAUC = 0.327, p = 0.0040) while denoising 12.45 ± 9.29% of test segments.

Significance. Multi-head noise regression provides interpretable, continuous ocular and muscle contamination estimates that can act as control signals for conservative, noise-aware artifact handling under constrained-lead conditions.

Section 1: Introduction

Electroencephalography (EEG) is a widely used tool for studying human brain function in clinical neuroscience, brain–computer interfaces (BCIs) and neurotechnology applications [1–4]. Its millisecond temporal resolution and relatively low hardware cost make EEG attractive for monitoring cognitive workload [5], sleep [6,7] and pathology [8,9] over extended periods, including outside specialised laboratories. In recent years there has been growing interest in long-term and wearable EEG optimised for comfort and ease of use. This has led to a rise in wireless, wearable, limited-lead EEG systems built with lightweight instrumentation for real-world settings, including headbands, ear-EEG, EEG integrated into VR headsets and other mobile form factors [10–14].

Non-neural physiological artifacts remain a central bottleneck for EEG interpretation [15]. Electrooculographic (EOG) artifacts arise from blinks, saccades and slow eye movements and manifest as sharp deflections, predominantly in frontal electrodes, contaminating the delta (0–4 Hz), theta (4–8 Hz) and alpha (8–12 Hz) bands. Electromyographic (EMG) artifacts arise from cranial muscle activity in the face, jaw and neck, appear as high-frequency noise, and contaminate the alpha (8–12 Hz), beta (12–30 Hz) and gamma (30–100 Hz) bands. Together, EOG and EMG cover much of the characteristic EEG spectrum (0–40 Hz) and can reach amplitudes comparable to or larger than the underlying neural signals [16,17]. These contaminants can distort event-related potentials (ERPs), bias spectral and connectivity estimates, and degrade the performance of automated decoders and clinical decision-support tools [17–19]. The problem is amplified in long-term and free-living recordings, where people move, blink naturally, speak and interact with their surroundings, leading to frequent EOG and EMG onsets. At the same time, many wearable and consumer EEG systems use only a few electrodes and often lack dedicated reference channels [10], which makes separating neural and non-neural activity more difficult.

A wide range of artifact-removal methods has been developed for conventional high-density EEG [20]. Artifact Subspace Reconstruction (ASR) and Independent Component Analysis (ICA) are among the most widely used. ASR and its Riemannian variant learn a “clean” subspace from relatively artifact-free segments of the recording and project noisy segments back into that subspace to suppress artifacts [21,22]. ICA performs blind source separation by decomposing multichannel EEG into statistically independent components so that EOG- and EMG-dominated components can be identified and removed [23–25]. Canonical Correlation Analysis (CCA) provides another blind source separation approach, finding linear combinations of channels that are maximally correlated [26]. In addition, MNE-Python implements Signal Space Projection (SSP), a subspace-projection method that constructs artifact-related projectors (e.g., from EOG/ECG components) and suppresses artifacts by projecting the data away from that subspace [27,28].

These methods work well because EOG and EMG artifacts are high-energy events that appear across many electrodes, providing spatial structure that helps separate artifacts from brain activity [29]. However, source separation techniques generally assume that the number of recorded channels is at least as large as the number of underlying sources [30]. In practice, they require a relatively high channel count and stable montages, which makes them poorly suited for denoising ambulatory EEG recorded with only one or a few electrodes. Artifact removal in single-lead EEG is therefore substantially more challenging [31].

For channel-wise denoising, regression-based algorithms are among the simplest and most widely used approaches [32–34]. They rely on one or more dedicated reference

channels that primarily capture the artifact of interest (for example, EOG channels for eye movements). The basic idea is to model how the artifact propagates into each EEG channel, estimate propagation coefficients from the reference signals, and then subtract a scaled version of the reference from the contaminated EEG to suppress the artifact [35]. A widely used example is the EOG regression approach implemented in MNE-Python, which estimates per-channel propagation coefficients from the data and subtracts the scaled EOG reference to attenuate ocular contamination [36]. In this way, the artifact contribution is linearly regressed out using the measured reference signal for that specific source. More recently, artifact-reference multivariate backward regression (ARMBR) has been proposed as a regression-based blink handling approach that uses artifact references (and blink detection) to estimate and subtract ocular contributions with minimal calibration. Notably, these approaches still assume either dedicated artifact reference channels (EOG regression) and/or sufficient multichannel structure to reliably estimate artifact subspaces/projections (SSP/ARMBR), which can be difficult to satisfy in limited-lead wearable EEG[37].

To address single- or few-channel settings without full BSS, a variety of signal decomposition methods have been proposed. These include wavelet-based approaches [38–41], surface Laplacian transforms [42], morphological component analysis [43], empirical mode decomposition (EMD) and its variants such as ensemble EMD (EEMD) and complete EMD with adaptive noise (CEEMDAN) [44–49], variational mode decomposition (VMD) [50–52], and singular spectrum analysis (SSA) [53,54]. In most of these approaches, a single-channel signal is first decomposed into a set of basis components or modes (e.g. wavelet subbands, intrinsic mode functions, VMD modes, Laplacian projections or SSA components) according to predefined rules. The decomposed components are then evaluated using statistical or spectral criteria to decide which are more likely to reflect neural activity (e.g. plausible EEG band structure, low kurtosis) and which are more characteristic of artifacts (e.g. large low-frequency deflections or broadband EMG). The signal is reconstructed from the subset of components judged to be neural, while components judged to be artifactual are discarded. However, the effectiveness of these methods depends strongly on decomposition settings and decision rules. In many studies, these hyperparameters are tuned heuristically for a specific dataset or artifact type, which can limit robustness across recording setups, tasks and levels of contamination [52].

More recently, several studies have explored end-to-end deep learning for single-channel EEG denoising. Examples include GAN-guided CNN–transformer hybrids, decision-guided networks that route short- and long-term artifacts through specialised denoisers, and transformer-based models that fuse local and non-local structure or combine multiscale convolutions with attention[55–58]. These networks learn nonlinear mappings from noisy to clean EEG, typically using semi-simulated mixtures such as EEGdenoiseNet with added EOG, EMG or mixed artifacts, and they report state-of-the-

art signal-level metrics (e.g. RRMSE, correlation, SNR) on those benchmarks [59]. Across designs, however, they are used as always-on black-box denoisers, generally target specific artifact types, and produce cleaned waveforms but no explicit, per-channel estimates of ocular and muscle noise load that could be used to govern downstream cleaning.

Classical decomposition-based methods and deep-learning denoisers thus offer complementary strengths and weaknesses. Wavelets, EMD/VMD, SSA and related approaches are relatively transparent: they expose intermediate components that can be inspected, and their effects on the signal are often easier to reason about. At the same time, their performance hinges on hand-tuned parameters and rule-based thresholds and can degrade when artifact structure deviates from those assumptions [52]. In contrast, modern deep networks learn powerful nonlinear mappings and often achieve superior signal-level performance on semi-simulated mixtures, but they behave as black boxes and have mostly been evaluated on short segments and signal-level metrics, with fewer studies examining downstream effects such as event-related potential preservation, decoding accuracy or connectivity estimates in real EEG.

Importantly, there is no universally optimal denoising technique across artifact types. Low-frequency EOG and broadband cranial EMG have very different temporal and spectral profiles, and algorithms typically need to be tuned with different settings for each [15,16,29]. In practice, real EEG often contains mixtures of these contaminants whose relative strength varies over time and across channels, so a fixed, one-size-fits-all denoiser is unlikely to be optimal. There is also an inherent trade-off between removing artifacts and preserving underlying brain activity. As highlighted by Delorme et al. in their “EEG is better left alone” study, aggressive preprocessing can attenuate or distort neural signals [60]. This suggests that when noise levels are low, it may be preferable to leave the signal largely untouched and reserve stronger interventions for segments with clearly elevated contamination.

These considerations motivate explicit noise detection and noise quantification [61,62]. Fickling et al. proposed the EEG Quality Index (EQI), a general-purpose framework that uses a sliding window and six simple signal metrics (band-limited power, line noise, RMS amplitude, maximum gradient, zero-crossing rate and kurtosis) to quantify signal quality relative to a normative database and to classify clean versus artifact-contaminated segments [63]. However, EQI relies on hand-crafted features, simulated data and binary clean/noisy classification, and does not provide continuous estimates of specific artifact loads (e.g. ocular vs muscle). Mohamed et al. proposed an automated EEG quality assessment framework that generates six scores based on amplitude histograms, alpha-band topography and inter-hemispheric symmetry, and beta/theta amplitude and sinusoidality to quantify signal quality in multi-channel recordings [61]. These scores rely on fixed normative thresholds (e.g. beta ≤ 20 μV, theta

≤ 30 μV) and multi-channel spatial patterns and provide only global measures of signal integrity rather than source-specific artifact loads.

A key conceptual gap is the lack of explicit, artifact-specific noise quantification, especially for single- or few-channel wearable systems. As a result, current single-channel denoisers have no principled way to decide when to intervene, how strongly to intervene, or which type of artifact-reduction method to prioritise. If reliable estimates of ocular and muscle noise loads were available, they could be used to govern downstream cleaning deciding when to engage a denoiser, which family of methods to use (ocular-focused vs muscle-focused) and how aggressively to apply them [15,52]. Such a noise detector would be particularly valuable in single- or few-channel, real-time, wearable settings.

Here we propose a multi-head, single-channel noise regressor that addresses this gap. The model takes 2-s EEG segments from a single channel and jointly estimates ocular (EOG) and muscle (EMG) noise-to-signal ratios in dB. Rather than performing artifact removal directly, it is designed to provide continuous, artifact-specific control signals that can drive other denoising algorithms. We train and benchmark a dilated temporal convolutional network (TCN) against four alternative architectures, a feature-based multilayer perceptron (MLP), a 1D-convolutional (Conv1D) model, a dual-branch Conv1D+STFT model and a Conv1D+BiGRU model on synthetic mixtures derived from the EEGdenoiseNet dataset with known EOG and EMG loads uniformly spanning –10 to +10 dB. We then evaluate whether these models can provide accurate and calibrated estimates of EOG and EMG noise load from short, single-channel segments and whether these estimates generalise to independent data and support downstream artifact management.

This work makes four main contributions. (i) We introduce a two-head, dilated TCN regressor that jointly estimates EOG and EMG noise-to-signal ratios from 2-s single-channel EEG segments. (ii) We demonstrate that the EOG head generalises to an independent eyeblink dataset, accurately recovering blink-related ocular-load topographies. (iii) We show that using the predicted EOG load to gate stationary wavelet blink removal on a clean P3 dataset can prevent over-cleaning. (iv) We further validate the detector-as-control-signal concept on a downstream BCI application by showing that TCN-guided selective denoising improves RSVP P300 speller decoding (55 subjects, FP1/FP2), while intervening on only a small fraction of segments.

The remainder of this paper is organised as follows. In Section 2 we describe the datasets, noise-synthesis procedure and model architectures. Section 3 presents the regression and generalisation results. Section 4 discusses the implications and limitations of our findings and outlines potential applications for noise-aware EEG pipelines, and Section 5 concludes.

Section 2: Methods

Dataset and preprocessing

We used the public EEGdenoiseNet resource as the source of clean EEG segments and artifact exemplars (ocular EOG and cranial EMG) [59]. Each 2-s raw epoch was resampled to 125 Hz. We treated each EEG channel as an independent 2-s training sample.

Clean EEG, EOG and EMG epochs were partitioned into training (70%), validation (15%) and test (15%) sets with anti-leakage control: no raw epoch contributed to more than one split. Before being passed to any model, each 2-s segment was z-scored by subtracting its mean and dividing by its standard deviation.

We z-score each 2-s window to reduce sensitivity to arbitrary amplifier gain, montage scaling, and inter-subject amplitude variability. Although z-scoring removes absolute amplitude, ocular and muscle contamination alters waveform morphology and spectral structure in characteristic ways; the network therefore learns a scale-normalised mapping from these signatures to the NSR targets defined from the underlying mixture components.

Synthetic mixtures and noise labels

Synthetic noisy epochs Y were generated by mixing a clean EEG segment X with EOG and EMG artifacts N_{EOG} and N_{EMG} as given in 1 and shown in Figure 1.

$$Y = X + \lambda_{EOG}.N_{EOG} + \lambda_{EMG}.N_{EMG} \quad (1)$$

The scaling factors λ_{EOG} and λ_{EMG} were set to obtain desired noise-to-signal ratios (NSR) Z_{EOG} and Z_{EMG} in decibels relative to the clean EEG power for EOG and EMG respectively as given in 2 and 3.

$$Z_{EOG} = 10.\log_{10}\left(\lambda_{EOG}^2 \frac{\text{mean}\,(N_{EOG}^2)}{\text{mean}\,(X^2)}\right) \quad (2)$$

$$Z_{EMG} = 10.\log_{10}\left(\lambda_{EMG}^2 \frac{\text{mean}\,(N_{EMG}^2)}{\text{mean}\,(X^2)}\right) \quad (3)$$

Target NSR values were sampled uniformly from –10 to +10 dB during mixture synthesis.

We report contamination as a noise-to-signal ratio (NSR) in dB, defined as the ratio of artifact power to EEG power, so that larger values indicate worse signal quality and can be used directly as a control variable for gating or rejection. This is equivalent to reporting SNR with opposite sign ($\text{NSR}_{\text{dB}} = -\text{SNR}_{\text{dB}}$), and we adopt NSR primarily for interpretability and for thresholding.

We controlled dataset synthesis to obtain a balanced composition of artifact types across this NSR range:

- 25% EOG-only (EMG NSR set to a floor value),
- 25% EMG-only (EOG NSR at floor),
- 5% clean-only ($\lambda_{\mathrm{EOG}} = \lambda_{\mathrm{EMG}} = 0$),
- 45% mixed-noise (both EOG and EMG present).

This composition was maintained approximately uniformly over the NSR spectrum from –10 dB (low contamination) to +10 dB (high contamination), as illustrated in Figure 2. The same proportions were used for the training, validation and test splits. Table 1 summarises the number of epochs per split and per artifact category.

For single-modality mixtures (EOG-only or EMG-only), the non-present head is assigned a floor value of −10 dB. This is a practical numeric proxy for "absent" that stays within the operating range used throughout synthesis (−10 to +10 dB), avoids undefined values during training, and makes the regressor output directly usable in thresholding/gating rules.

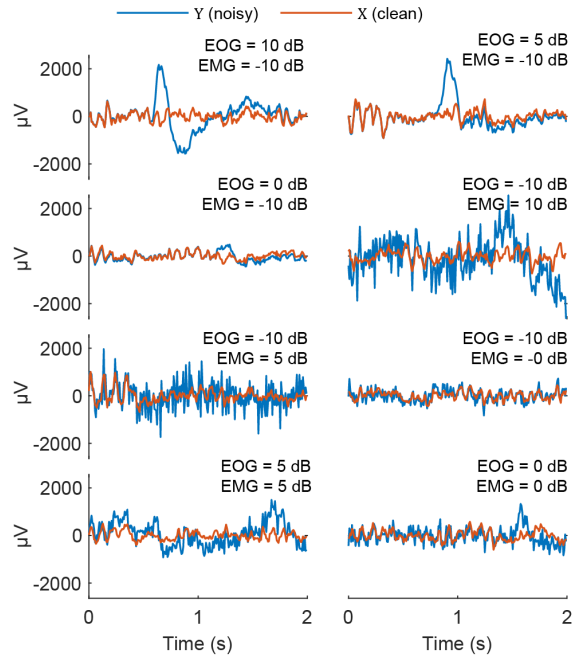

Figure 1: Example of training samples. Each panel shows a 2 s noisy sample Y (blue) overlaid with its clean EEG source X (orange) for different combinations of EOG and EMG artifact loads in dB.

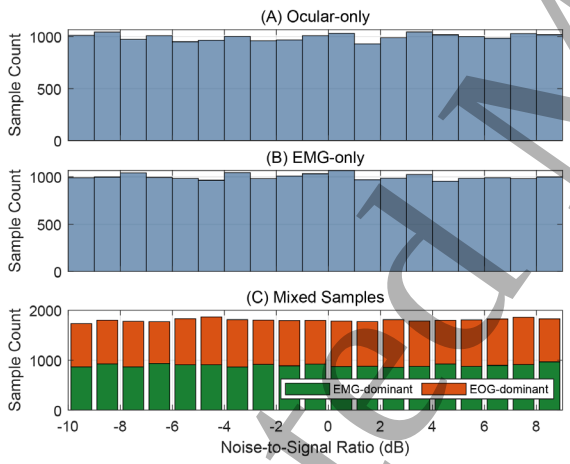

Figure 2: Target noise-to-signal ratio (NSR) distributions for training data synthesis. Histograms show the number of 2-s single-channel segments per target NSR bin (x-axis: NSR, dB; y-axis: sample count). (A) Ocular-only mixtures (EOG). (B) EMG-only mixtures. (C) Mixed EOG+EMG mixtures; stacked bars indicate EOG-dominant vs EMG-dominant samples (dominance defined by the larger target NSR).

TABLE 1: Number of samples for training, validation, and testing.

Data Split	Total Epochs	EOG-only	EMG-only	Clean Only	Mixed
Training	80000	20000	20000	4000	36000
Validation	12000	3000	3000	600	5400
Testing	12000	3000	3000	600	5400

The supervisory task is a two-head regression problem. For each 2-s segment, the model predicts the continuous EOG and EMG NSRs, Z_{EOG} and Z_{EMG}, in dB as defined in (2) and (3). These labels span nearly clean through heavily contaminated conditions and form the targets for the EOG and EMG regression heads.

Model architectures

We evaluated five model families chosen to span distinct inductive biases and practical deployment trade-offs for short-segment, single-channel contamination regression. The families cover hand-crafted feature baseline, standard convolutional networks and architectures designed to capture longer temporal context:

1. Feature-based MLP
2. Conv1D
3. Conv1D + STFT (dual-branch fusion)
4. Conv1D + BiGRU
5. Dilated residual TCN

All models take as input a z-scored 2-s segment sampled at 125 Hz (250 time points) and produce two scalar outputs: Z_{EOG} and Z_{EMG}.

Feature-based MLP

The feature-based multilayer perceptron (MLP) uses a compact set of hand-crafted features as input. Each 2-s segment is reduced to an 18-dimensional feature vector spanning four domains:

- Amplitude and distribution (5 features):
  root-mean-square (RMS), log variance, peak-to-peak range, skewness, kurtosis.
- Spectral power and placement (5 features):
  power-weighted spectral centroid, spectral spread (bandwidth), band power in delta (0–4 Hz) and alpha (8–13 Hz), and the variance of the Hilbert envelope (amplitude modulation).
- Complexity and dynamics (4 features):
  Hjorth mobility, Hjorth complexity, and the coefficient of variation of the instantaneous frequency (mean and variance derived from the analytic signal).
- Temporal and spectral disorder (4 features):
  spectral flatness, normalised spectral entropy, the lag at which the autocorrelation first drops below 0.5, and an Augmented Dickey–Fuller stationarity indicator.

These 18 features are fed to an MLP with two fully connected hidden layers (128 units each) with GELU activation and 0.2 dropout. A final linear layer fan-outs into two regression heads that output Z_{EOG} and Z_{EMG}.

Conv1D

The Conv1D model operates directly on the 250×1 waveform. The backbone comprises three temporal convolutional layers with GELU activations; one layer has stride 2 for light downsampling. The resulting feature map is reduced by global average pooling and passed through a 96-unit dense layer with 0.2 dropout. Two linear output heads then estimate the EOG and EMG NSRs.

Conv1D + STFT (dual-branch fusion)

The Conv1D + STFT model combines time-domain and time–frequency information.

- The time branch is identical to the Conv1D backbone described above.
- The spectral branch computes a short-time Fourier transform (STFT) using a Hann window with 64-sample frames and 16-sample hop, yielding a time–frequency representation. The log-magnitude STFT is processed by two 3×3 Conv2D layers followed by layer normalisation.

The embeddings from both branches are concatenated and passed through a 96-unit dense layer. Two linear heads then output Z_{EOG} and Z_{EMG}.

Conv1D + BiGRU

The Conv1D + BiGRU model combines local convolutional features with recurrent temporal modelling. Two initial Conv1D layers with GELU activations and 0.1 dropout extract local patterns. Their output is fed to a bidirectional GRU with 32 units per direction, providing access to both past and future context within the 2-s window. The BiGRU output is passed through a 64-unit dense layer and then to the two linear regression heads.

Dilated residual TCN

The dilated temporal convolutional network (TCN) is designed to model long temporal context without recurrence. It starts with a Conv1D stem layer, followed by six residual blocks with exponentially increasing dilation factors (1, 2, 4, 8, 16, 32). Each block contains:

- a Conv1D layer,
- layer normalisation,
- GELU activation,
- spatial dropout,

with a residual (skip) connection around the block. With this dilation pattern, the effective receptive field covers essentially the entire 2-s window. The final feature map is reduced by global average pooling (GAP) and passed through a 96-unit dense layer that feeds the two regression heads. The architecture of the dilated TCN is given in Figure 3.

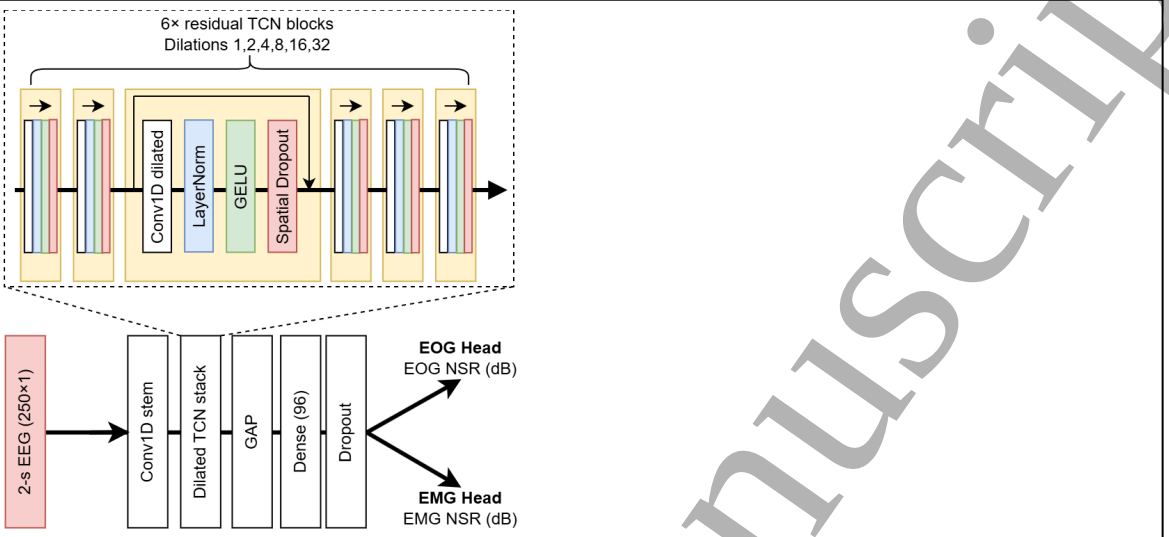

Figure 3: Two-head dilated temporal convolutional network (TCN) regressor for noise-load estimation. A 2-s single-channel EEG segment (250×1 samples) is passed through a 1D convolutional stem and a stack of six residual dilated TCN blocks (dilations 1, 2, 4, 8, 16, 32). Global average pooling (GAP) and a 96-unit dense layer feed two output heads that regress ocular and muscle noise-to-signal ratio (NSR) in decibels: EOG head and EMG head.

Training procedure and evaluation metrics

All models were trained using the Adam optimiser with an initial learning rate of 0.001 and a batch size of 256. The training objective was the mean Huber loss (δ = 1.0) across the two heads (EOG and EMG), computed on the dB targets Z_{EOG} and Z_{EMG}.

Each model was trained for up to 60 epochs and repeated 10 times with independent random seeds. Early stopping was applied if the validation loss did not improve for 10 consecutive epochs. If validation loss plateaued for 4 epochs, the learning rate was reduced by a factor of 0.5.

For each regression head (EOG, EMG), performance on the held-out test set was quantified using:

- Mean absolute error (MAE) in dB, directly matching the NSR definition and capturing average absolute deviation from the true noise level.
- Coefficient of determination (R^2), measuring how well the predictions track the dynamic range of artifact loads across the test set ($R^2 \approx 1$ indicates close agreement; $R^2 < 0$ indicates performance worse than predicting the test-set mean).

To compare models statistically, we applied paired Wilcoxon signed-rank tests on run-wise summaries (10 paired observations per model pair) separately for each head and metric.

External eyeblink dataset and ocular-load validation

To assess whether the EOG head generalises to independent data and reproduces realistic blink topographies, we used a publicly available eyeblink dataset comprising 20 healthy participants performing voluntary and involuntary blink tasks [64]. Recordings were obtained from 15 channels (1 EOG + 14 EEG: Fp1, Fp2, F3, F4, T3, C3, Cz, C4, T4, P3, Pz, P4, O1, O2) at 256 Hz. Continuous recordings were provided for eyes-closed, voluntary and involuntary conditions, with three trials per subject. For the present analysis we used only the voluntary and involuntary conditions.

All channels were downsampled to 125 Hz to match the temporal resolution used in model training, and each 2-s segment was z-scored in the same way as the training data.

Blink detection and epoching

Blinks were detected from the EOG channel using a derivative-based detector targeting the characteristic biphasic slope of blink waveforms. The EOG signal was median-detrended and band-pass filtered between 0.5 and 10 Hz with a 2nd-order zero-phase Butterworth filter to emphasise blink transients while suppressing slow drift and high-frequency noise. The temporal derivative of the filtered EOG was then computed and z-scored.

Candidate blink events were defined as pairs of large, opposite-polarity derivative peaks (positive followed by negative, or vice versa) occurring within 250 ms, with both peaks exceeding $|z| \geq 2$ and separated from neighbouring events by at least 0.4 s. For each candidate pair, the blink time was taken as the point of maximum $|z|$-normalised EOG amplitude between the two derivative peaks.

Around each detected blink, we extracted a 2-s epoch (±1 s) from the downsampled multichannel data. This yielded 1189 voluntary and 4049 involuntary blink epochs.

Ocular-load ground truth and spatial correlation

To obtain a channel-wise ocular ground truth for each blink epoch, we regressed each EEG channel on the EOG channel. Within an epoch, the EOG signal x was mean-centred and used as a single predictor. For each EEG channel y, we fitted a least-squares model. The predicted component $\hat{y} = \beta x$ was interpreted as the ocular contribution, and the residual $r = y - \hat{y}$ as the non-ocular component.

We defined the ground-truth ocular load for that channel and epoch as a relative power ratio in dB as given in (4).

$$OC_{GT} = 10\log_{10}\frac{var\ (\hat{y})}{var\ (r)} \quad (4)$$

This yielded a 14-dimensional ground-truth ocular-load vector (one value per EEG channel) for each blink epoch. The regression-derived ocular component provides a practical proxy for the spatial distribution of blink contamination when only one EOG reference channel is available. Because the same EOG reference drives all channels within an epoch, this approach is well suited to benchmarking whether the model recovers the expected anterior–posterior pattern of blink propagation. We emphasise that this reference is not a perfect ground truth and may include estimation error. Accordingly, we interpret the spatial-correlation analysis as a validation of topographic plausibility rather than an absolute measure of ocular artifact magnitude. An additional independent benchmark (e.g., manual artifact annotation by multiple expert raters) would be valuable; however, such labels were not available for this dataset and were beyond the scope of the present study.

The trained TCN was then applied to each channel of each blink epoch (using the same 2-s, 125 Hz, z-scored format) to obtain a 14-dimensional vector of EOG-head outputs. For each epoch, we computed the spatial Pearson correlation between the 14-channel ground-truth vector and the 14-channel EOG-head outputs, giving a blink-wise spatial correlation coefficient. Correlations were summarised separately for voluntary and involuntary blinks. We then aggregated at the subject level by taking the median blink-wise correlation per subject, and summarised these subject medians as median ± standard deviation across participants.

As a negative control, we also examined whether the EMG head inadvertently tracked ocular contamination by computing the Pearson correlation between EMG-head outputs and the ground-truth ocular load values pooled across channels and epochs. No independent EMG reference channel is available in this dataset; therefore, EMG-head evaluation here is limited to negative-control analyses (i.e., ensuring it does not trivially track the ocular proxy).

EEGLAB P3 dataset and noise-aware blink removal

The proposed model performs artifact detection (noise-load estimation), not artifact removal. We therefore use a standard blink-removal routine only as a downstream module to demonstrate how continuous EOG-load estimates can govern when and where cleaning is applied. We intentionally include a clean-reference setting to test a “do-no-harm” requirement: when residual contamination is low, uniform denoising is expected to be unnecessary and may attenuate ERP features, whereas noise-aware gating should avoid such over-cleaning by restricting processing to the most contaminated segments.

To evaluate the impact of noise-aware blink removal on an ERP, we used the EEGLAB sample dataset from a covert spatial attention task, recorded with 30 scalp EEG channels and two EOG channels at 128 Hz [65]. The “square” event marks visual stimulus onset and “rt” marks the button press.

Cleaned reference dataset

To obtain a clean reference, the continuous data were band-pass filtered between 0.5 and 40 Hz, decomposed using ICA, and components were automatically labelled with ICLabel [66]. Components classified as EOG or EMG with probability > 0.8 were removed. The data were then average-re-referenced and visually inspected to confirm that eye and muscle artifacts had been largely removed.

The cleaned data were epoched from –1 s to +1 s around the 80 “square” events, yielding 30 EEG channels × 80 epochs = 2400 channel×epoch segments. We restricted analyses to the scalp EEG channels and focused on the single-condition P3 ERP.

Noise-load estimation with the TCN

For each 2-s epoch, we downsampled from 128 Hz to 125 Hz, z-scored each channel, and applied the trained two-head TCN noise regressor to obtain per-channel EOG and EMG noise loads in dB. This produced two 30×80 matrices (EOG and EMG load), each with one scalar estimate per channel and epoch.

Wavelet blink removal and gating strategy

Blink removal was implemented using a stationary wavelet transform (SWT) applied channel-wise to the epoched EEG. We used a Coiflet (coif3) mother wavelet with six decomposition levels. Blink-related activity is expected to dominate in the low-to-mid frequency range, so we applied soft thresholding to the detail coefficients at levels 3–6, leaving the coarsest scale (slow ERP components) and the finest scales (higher-frequency activity) unchanged. Thresholds at each level were set using a median-absolute-deviation–based universal rule [38–40,67]. The inverse SWT yielded a cleaned time series for each channel and epoch.

We compared three pipelines:

1. **Baseline:** no additional wavelet blink removal (post-ICLabel data only).
2. **Wavelet-all:** SWT-based blink denoising applied to all 2400 channel×epoch segments.
3. **Gated wavelet:** the same SWT denoising applied selectively only to segments whose TCN-estimated EOG load exceeded a threshold $\theta \in \{-9, -8, \ldots, 0\}$ dB.

In the gated condition, the EOG load matrix defined a binary mask over channel×epoch pairs, and only those with $Z_{\mathrm{EOG}} > \theta$ were denoised; all others were left unmodified.

ERP computation and significant-channel metric

For each pipeline, we computed ERPs time-locked to the "square" event. To visualise spatial patterns, electrodes were grouped into eight anatomical regions (Frontal, Fronto-central, Central, Centro-parietal, Parietal, Parieto-occipital, Occipital, Temporal) based on standard 10–20 labels, and trial-averaged ERPs were computed per region.

To quantify how denoising affected ERP detectability, we used Delorme's significant-channel metric [60]. For each pipeline and channel, we computed the mean amplitude in a pre-stimulus window (–600 to 0 ms) and a post-stimulus window (0 to 600 ms). For each channel, the per-trial difference (post – pre) was converted to a z-score by normalising by the pre-stimulus standard deviation. We then performed a one-sample t-test against zero on these z-scores across epochs. p-values were corrected across channels using the Benjamini–Hochberg FDR procedure ($q = 0.05$). Channels with $q < 0.05$ were considered significant, and the number of significant channels was used as a summary measure of spatial ERP detectability under each pipeline.

BCI Evaluation on P300 speller decoding

To evaluate whether noise-aware selective artifact removal improves performance in a realistic application scenario, we added a downstream BCI decoding benchmark using the RSVP P300 speller dataset released by Won *et al.* (55 participants) [68]. We used a minimal forehead montage consisting of FP1 and FP2, which represents a constrained-lead setting where ocular contamination is prominent and where spatially driven multichannel artifact methods (e.g., ICA/ASR) are less applicable.

We followed the dataset authors' reference decoding framework and implemented a subject-specific SWLDA pipeline consistent with their MATLAB implementation (details below). For each subject, we compared three denoising strategies:

(i) **raw**: no additional denoising;
(ii) **wav_all**: wavelet denoising applied to all epochs; and
(iii) **TCN-guided selective denoising**: the same wavelet denoiser applied only to channel×epoch segments whose predicted contamination exceeded a threshold θ(dB). Thresholds were swept over $\theta \in \{-5,0,5,7.5,9\}$ dB to quantify the trade-off between denoising aggressiveness and preservation of task-relevant signal structure.

Selective denoising decisions were made at the channel×epoch level using the proposed two-head contamination estimates. For each channel×epoch segment, we defined the contamination index using $Z = Z_{EOG}$, where Z_{EOG} is the predicted ocular noise load (dB). The wavelet denoiser was applied only when $Z \geq \theta$, otherwise the raw segment was left unchanged. This design intentionally restricts processing to highly contaminated segments while preserving clean data.

SWLDA decoding and performance metrics

Letter decoding was performed with a subject-specific stepwise linear discriminant analysis (SWLDA) pipeline aligned with the dataset's reference implementation [68]. Briefly, epochs were band-pass filtered (0.5–10 Hz, 4th-order Butterworth, zero-phase), baseline-corrected using the pre-stimulus interval, and time-windowed to 0–600 ms post-stimulus. Features were computed using average-decimation in 24-sample blocks and then flattened across time bins and channels. Stepwise feature selection used $p_{\text{enter}} = 0.08$ with a maximum of 60 retained features. The resulting linear score was accumulated across row/column flashes over repetitions to decode each character.

Decoding performance was summarized as letter accuracy vs. repetitions. We used the area under the letter-accuracy curve from repetitions 1–10 (AUC) as the primary summary metric. Secondary analyses included mean letter accuracy at 5 and 10 repetitions (Letter@5, Letter@10) and per-repetition paired comparisons between raw and $\theta = 9$dB.

Finally, we quantified the degree of intervention induced by selective denoising by reporting the channel-wise fraction of segments denoised, defined as the percentage of segments satisfying $Z \geq \theta$, averaged across subjects.

Section 3: Results

Multi-head noise regression on synthetic mixtures

Figure 4(a-b) summarises the reliability of the five model families (Feature-MLP, Conv1D, Conv1D+STFT, Conv1D+BiGRU and Dilated TCN). For both regression heads, the median predicted noise load increases monotonically with the true noise level across the full –10 to +10 dB range. For the EOG head, the reliability curves lie slightly below the identity line in the near-clean regime (true noise ≈ –2 to 0 dB), indicating a mild tendency to underestimate ocular noise when contamination is very low. In contrast, the EMG head reliability in Figure 4(b) closely follows the ideal line over the entire range, indicating that muscle-noise levels are recovered with high fidelity. (Reliability curves were obtained by binning test samples by true NSR and plotting the median predicted NSR within each bin.)

The bias curves in Figure 4(c–d) further quantify these effects. For the EOG head, very low-noise segments (true < –5 dB) exhibit a small positive bias (Pred–True > 0), i.e. the models tend to predict slightly higher ocular contamination than is present, whereas the bias moves closer to zero around -5 dB. For EMG, the bias remains close to zero over most of the range, with only a slight negative bias for heavily contaminated segments (true > 5 dB), where the models slightly underestimate the muscle noise. Across all four panels, the Dilated TCN curves sit closest to the ideal reference (identity / zero-bias)

lines, indicating that it is the most reliable and least biased estimator among the tested architectures.

Across the five model families, the Dilated TCN also achieved the best average performance on the held-out test data for both heads. On the EOG head, its MAE was 1.809 ± 0.042dB (mean ± SD across 10 runs), outperforming Conv1D+BiGRU (1.917 ± 0.114 dB), Conv1D+STFT (2.271 ± 0.078 dB), Conv1D (2.446 ± 0.044 dB) and the Feature-MLP (2.669 ± 0.023 dB). On the EMG head, the TCN reached 1.026 ± 0.008dB, ahead of Conv1D+BiGRU (1.141 ± 0.049 dB), Conv1D+STFT (1.178 ± 0.037 dB), Conv1D (1.239 ± 0.032 dB) and the Feature-MLP (1.326 ± 0.011 dB). The MAE distributions across the 10 training runs are shown in Figure 4(e–f).

In terms of R^2, the TCN again ranked first with 0.817 ± 0.008on the EOG head and 0.940 ± 0.001on the EMG head. The next-best model was Conv1D+BiGRU (EOG: 0.799 ± 0.024; EMG: 0.932 ± 0.003), followed by Conv1D+STFT (EOG: 0.727 ± 0.019; EMG: 0.929 ± 0.002), Conv1D (EOG: 0.690 ± 0.011; EMG: 0.924 ± 0.002) and the Feature-MLP (EOG: 0.654 ± 0.006; EMG: 0.906 ± 0.001). Relative to Conv1D+BiGRU, the TCN reduced EOG MAE by 0.107 dB and EMG MAE by 0.115 dB, and improved EOG $R^2$by +0.018 and EMG $R^2$by +0.008. The R^2distributions across the 10 runs are shown in Figure 4(g–h). Pairwise statistical significance across runs was assessed using Wilcoxon signed-rank tests; the resulting p-values are visualised as a heatmap in Figure 5.

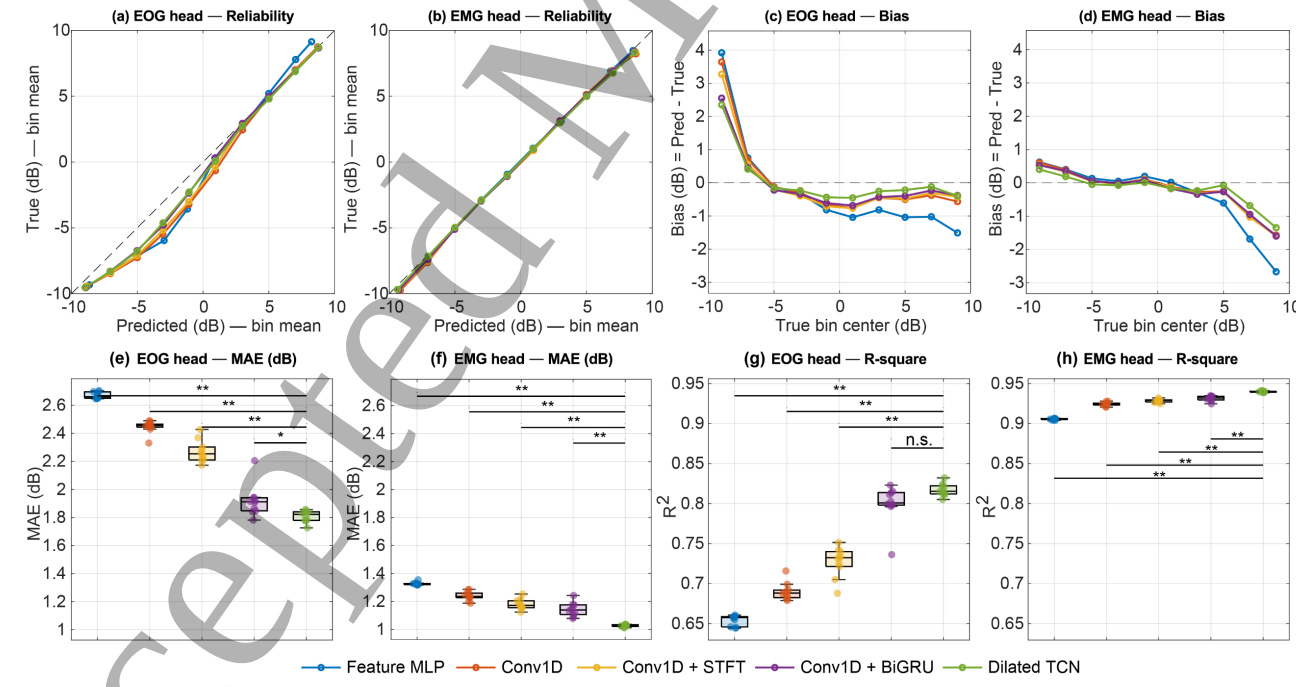

Figure 4: *Performance overview across models on test samples. (a–b) Reliability (Predicted-binned), (c–d) Bias (True-binned), (e–f) MAE (dB), and (g–h) R-square distributions for EOG and EMG regression heads respectively. * represents FDR-BH corrected statistical significance (paired Wilcoxon signed-rank, *: p < 0.05, **: p < 0.01, n.s: p > 0.05).*

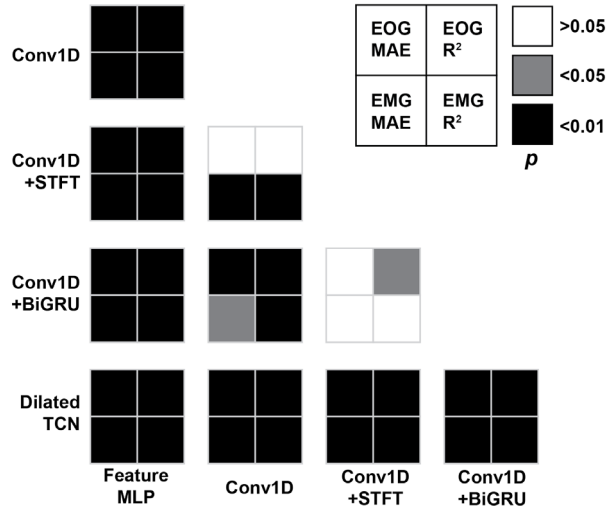

Figure 5: Pairwise significance of model performance differences. Heatmaps show BH-FDR adjusted p-values from two-sided Wilcoxon signed-rank tests for pairwise comparisons between model families (rows vs columns) on four metrics: EOG MAE, EMG MAE (mean absolute error, dB) and EOG R^2, EMG R^2 (coefficient of determination). Cell shading indicates adjusted significance level (white: p>0.05; gray: p≤0.05; black: p<0.01).

On an identical computer setup (Python 3.12.3, NVIDIA GeForce RTX 3060), we also measured inference latency using a batch size of 256 and 50 timed iterations per model. We report mean wall-clock time per 2-s segment. The plain Conv1D model was fastest at 0.190 ms per segment, followed by Conv1D+STFT at 0.202 ms and the Dilated TCN at 0.236 ms. The recurrent Conv1D+BiGRU incurred 0.487 ms per segment. The Feature-MLP required 3.006 ms per segment in total (2.794 ms for feature extraction and 0.213 ms for model inference). The accuracy–latency trade-off is shown in Figure 6.

Figure 6: Accuracy–latency trade-off. Average test R^2 across the two heads (EOG and EMG) versus end-to-end inference latency (ms/sample) for each model family.

Overall, the Dilated TCN delivered the strongest accuracy on both heads (lowest MAE and highest R^2), with pairwise tests confirming significant gains over alternatives in the majority of comparisons (Figure 5). This accuracy is achieved with competitive latency and a purely convolutional architecture (no recurrence, no external feature extraction), which simplifies deployment. Although the plain Conv1D is marginally faster, it trails in both MAE and R^2. Conv1D+STFT adds a second branch, and the Feature-MLP's end-to-end time is dominated by feature computation. The higher latency of Conv1D+BiGRU, due to explicit temporal memory, did not translate into higher accuracy than the fully convolutional Dilated TCN. We therefore adopt the Dilated TCN as our primary model for subsequent analyses.

External eyeblink dataset: preservation of blink topographies

We next evaluated whether the EOG head generalises to independent recordings and reproduces realistic blink topographies (Methods, eyeblink dataset). To do so, we compared the model's outputs against regression-based ocular ground truth on a separate eyeblink dataset comprising voluntary and involuntary blinks [64].

Across all epochs, the blink-wise spatial correlation between the 14-channel ocular ground truth and the TCN EOG-head outputs was high for both conditions (voluntary: median $r = 0.906$, SD = 0.229, $N = 1189$ blinks; involuntary: median $r = 0.917$, SD = 0.244, $N = 4049$ blinks). This indicates that, for individual blink instances, the model closely reproduces the ground-truth spatial distribution of ocular load. Aggregating at the subject level, the median blink-wise spatial correlation per participant also remained high (voluntary: group median = 0.911, SD across subjects = 0.111, $N = 20$; involuntary: group median = 0.916, SD = 0.112, $N = 20$), demonstrating consistently strong agreement between predicted and ground-truth ocular-load topographies across participants. The subject-wise r distributions for the voluntary and involuntary datasets are shown as boxplots in Figure 7(a–b).

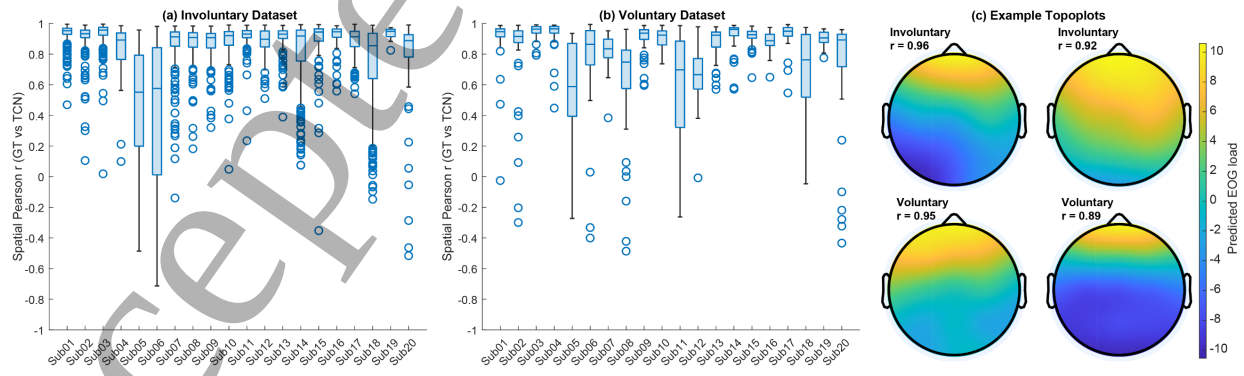

Figure 7: Subject-wise spatial correspondence between ground-truth (GT) and TCN EOG-head predictions on the 14-channel blink dataset. Each box shows the distribution of blink-wise spatial Pearson correlation coefficients (r) between the GT- load and the TCN EOG-head output, aggregated over all blinks for a given subject. (a): involuntary blink epochs; (b): voluntary blink epochs; (c) randomly selected example single-blink EOG artifact topographies estimated by the TCN EOG head.

Figure 7(c) illustrates example single-blink topographies. Each scalp map depicts the TCN-predicted EOG load for a randomly selected blink epoch from the voluntary and involuntary datasets. The expected anterior–posterior gradient is clearly recovered, with the largest predicted ocular contamination over frontal electrodes and a gradual decay toward parietal and occipital regions.

As a negative control, we examined the relationship between the EMG head and the ocular ground truth. When all channels and epochs were pooled, EMG-head outputs were negatively correlated with the EOG-regression ocular load (voluntary: $r = -0.385$; involuntary: $r = -0.319$), indicating that the EMG head does not simply track ocular contamination and instead captures a distinct, non-ocular noise component.

EEGLAB P3 dataset: noise-aware blink removal

Finally, we assessed whether the noise detector can help preserve ERP amplitudes when used to gate a conventional blink-removal method (Methods, EEGLAB P3 dataset). We compared three pipelines on the EEGLAB sample dataset: (i) a clean Baseline (post-ICLabel data, no further blink removal), (ii) Wavelet-all, where all channels and epochs were denoised using stationary wavelet thresholding, and (iii) Gated wavelet, where the same wavelet procedure was applied only to channel×epoch segments whose EOG noise load exceeded a threshold θ.

As expected for this heavily preprocessed dataset, the Baseline pipeline exhibited a strong P3 ERP across channels. For most segments, the TCN predicted low residual artifact (median EOG load = -6.121 ± 3.408 dB; median EMG load = -8.944 ± 0.474 dB), consistent with the notion that the data are largely clean.

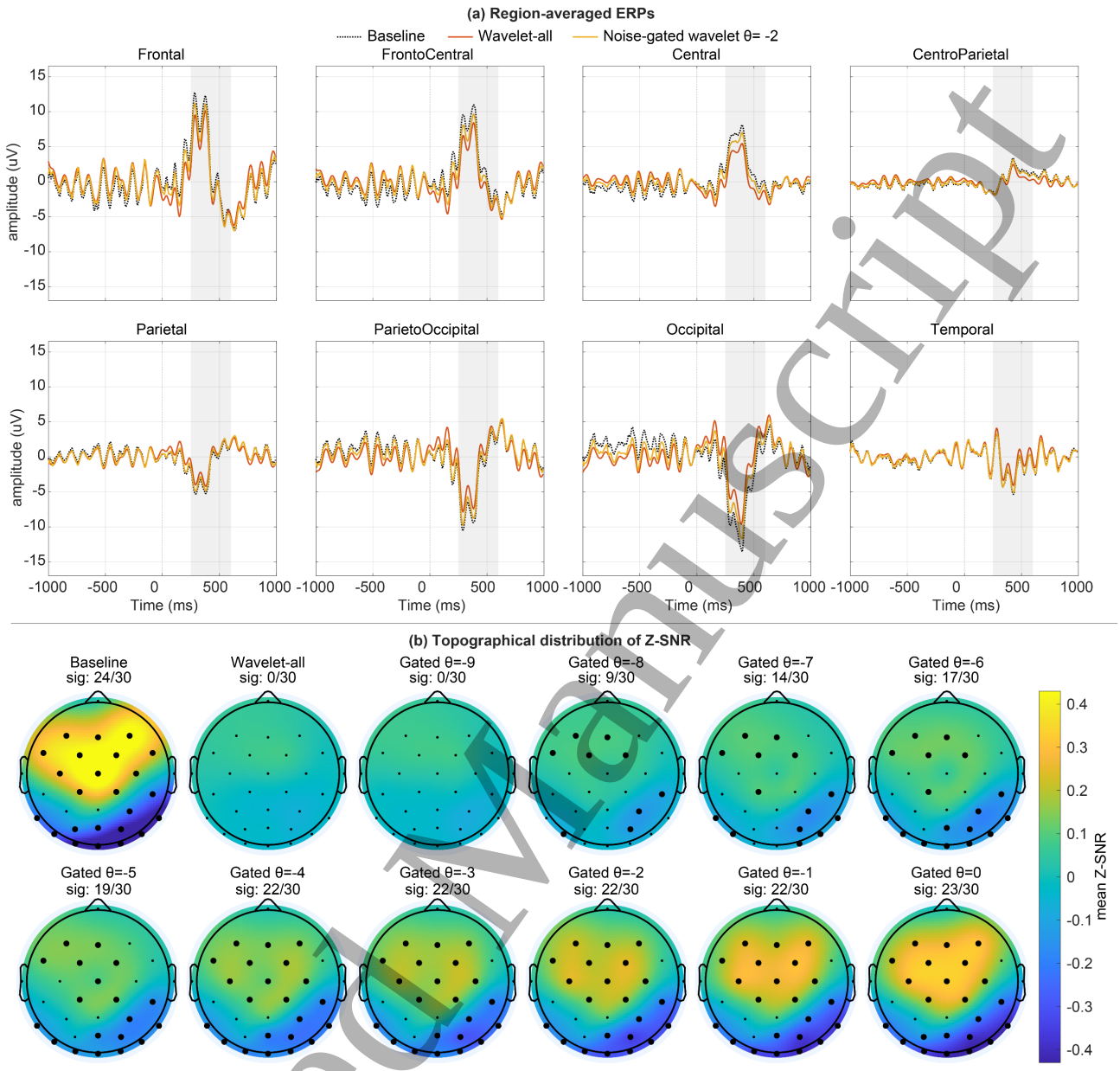

Figure 8: (a) Regional ERPs (channel averaged) for eight scalp regions, comparing baseline pipeline, unguided wavelet denoising applied to all channel–epoch segments ("wavelet-all"), and noise-gated wavelet cleaning using the EOG noise estimator with threshold θ = –2 dB. (b) Topographies of mean Z-SNR with FDR-significant channels (bold dots) for baseline, wavelet-all, and gated wavelet methods with threshold ranging between -9 dB and 0 dB (TCN EOG-head prediction).

Figure 8(a) shows regional ERPs for eight anatomical regions, comparing Baseline, Wavelet-all and a representative Gated-wavelet configuration with $\theta = -2$dB. The Frontal region comprised FPz, F3, Fz and F4; the Fronto-central region included FC5, FC1, FC2 and FC6; the Central region included C3, Cz and C4; and the Centro-parietal region comprised CP5, CP1, CP2 and CP6. The Parietal region was defined by P7, P3, Pz, P4 and P8, while the Parieto-occipital region included PO7, PO3, POz, PO4 and PO8. The Occipital region comprised O1, Oz and O2, and the Temporal region was defined by T7 and T8. For each region, channels were averaged per trial and then across trials to obtain regional ERPs.

The Baseline pipeline exhibits the expected P3, with a clear positive deflection around 300–400 ms. In the Wavelet-all condition, where wavelet denoising was applied to all 2400 channel×epoch segments, the P3 amplitude was markedly attenuated across regions, illustrating the effect of applying additional denoising to data that are already largely clean. In contrast, the Gated-wavelet pipeline with $\theta = -2$dB, which only denoised channel×epoch segments with a relatively high predicted EOG load, produced ERPs that closely matched the Baseline in both morphology and latency.

These qualitative differences were quantified using channel-wise post–pre z-scores (Z-SNR) and FDR-corrected significance tests Figure 8(b). The Baseline pipeline yielded 24 out of 30 channels that were FDR-significant. Applying Wavelet-all substantially reduced the statistical evidence for the P3: 0/30 channels remained FDR-significant, and the topography of mean z-scores was strongly flattened. This shows that applying the wavelet procedure uniformly to all segments in an already-clean dataset can substantially reduce ERP sensitivity.

When wavelet denoising was gated using the EOG noise estimates, the number of significant channels increased monotonically as the threshold θ was relaxed from –9 to 0 dB: 0/30 at $\theta = -9$ dB, then 9/30, 14/30, 17/30, 19/30, 22/30, 22/30, 22/30, 22/30 and finally 23/30 channels at $\theta = 0$ dB. For $\theta \geq -4$ dB, the Gated-wavelet pipelines recovered 22–23 of the 24 Baseline significant channels, and the spatial distribution of z-scores closely resembled the Baseline P3 topography. In other words, detector-guided cleaning recovered most of the ERP sensitivity lost in the Wavelet-all condition.

To summarise the trade-off between “how much data is modified” and “how much ERP sensitivity is preserved”, Figure 9 plots the percentage of segments denoised against the percentage of Baseline significant channels preserved. The Baseline pipeline sits at (0%, 100%), representing maximal ERP sensitivity with no additional cleaning. Wavelet-all lies at (100%, 0%), representing maximal data modification with complete loss of FDR-significant P3 channels. In contrast, the Gated-wavelet pipelines trace a curve from (88%, 0%) at $\theta = -9$dB to approximately (9–20%, 92–96%) for thresholds between –2 and 0 dB. In these higher-threshold configurations, only ~9–20% of the most contaminated channel×epoch segments are denoised, and nearly all of the Baseline significant channels are preserved.

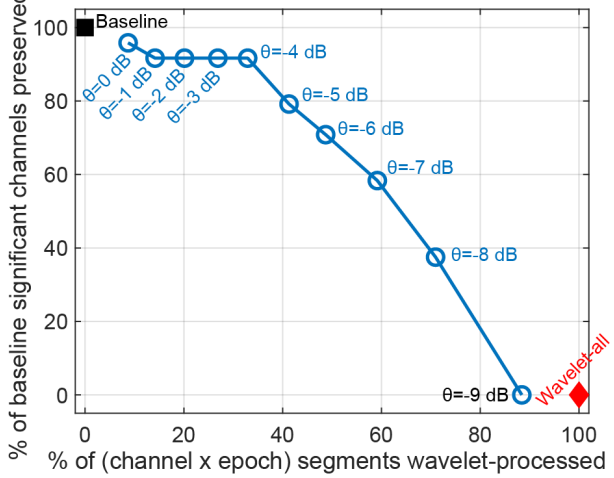

Figure 9: The relationship between the percentage of segments subjected to wavelet denoising (x-axis) and the percentage of baseline significant channels preserved (y-axis, relative to 24/30 FDR-significant channels at baseline).

Overall, this case study on an already-clean ERP dataset shows that the TCN-based EOG noise estimates can be used to control a conventional wavelet blink-removal procedure in a way that avoids over-cleaning. By gating the wavelet denoiser using per-channel noise estimates, we maintain the statistical detectability and spatial extent of the P3 while still targeting the relatively few segments where residual blink contamination is present.

Speller P300 dataset

We evaluated the proposed noise-aware selective denoising on the P300 speller dataset (55 subjects) using a forehead-only montage (FP1/FP2). Applying wavelet denoising to all epochs (wav_all) consistently reduced decoding performance relative to raw, consistent with over-attenuation of task-relevant ERP structure. In contrast, TCN-guided selective denoising improved decoding, and performance increased monotonically as the threshold θ(dB) increased, i.e., as denoising was restricted to progressively more contaminated segments (Figure 10 (a)).

Summary decoding metrics are reported in Table 2. At 10 repetitions, mean±SD letter accuracy improved from 0.183±0.165 (raw) to 0.232±0.162 at $\theta = 9$dB, while wav_all yielded the lowest performance (0.129±0.131). Early decoding also improved (Letter@5: 0.149±0.121 raw vs 0.182±0.122 at $\theta = 9$dB; Table 2).

Using the primary summary metric AUC(1..10), $\theta = 9$dB significantly outperformed raw (paired across subjects): ΔAUC(1..10)=0.327, $p = 0.0040$, $d_z = 0.405$(Figure 10 (b)). Per-repetition paired tests (repetitions 1–10) indicated that gains emerge early and persist across mid-range repetitions; after BH-FDR correction, significant improvements were observed at repetitions 3 and 5–10.

Importantly, gains were achieved while modifying only a small fraction of the data. The test-set channel-wise fraction of denoised segments decreased sharply with increasing θ: 93.13±6.92% ($\theta = -5$), 70.76±13.30% ($\theta = 0$), 35.27±15.30% ($\theta = 5$), 21.23±12.26% ($\theta = 7.5$), and 12.45±9.29% ($\theta = 9$) (Figure 10 (c); Table 2), indicating that $\theta = 9$ improves decoding while intervening on only ~12% of segments on average.

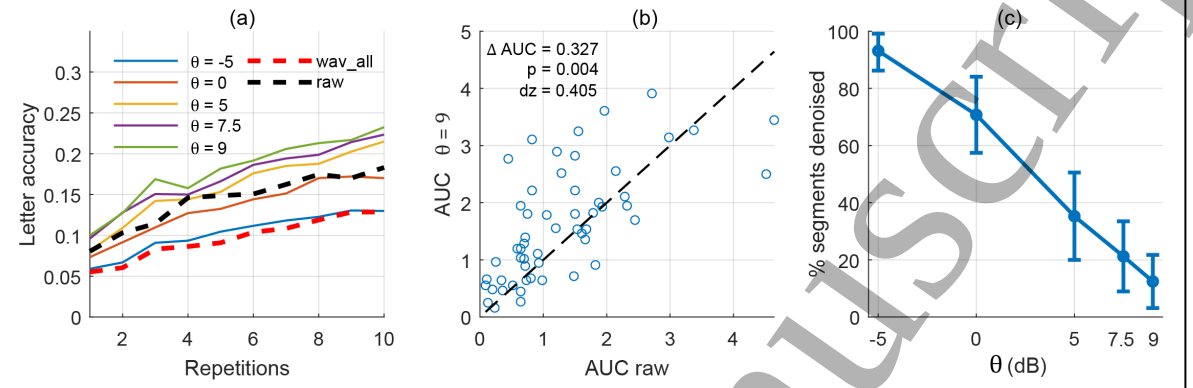

Figure 10: Downstream BCI evaluation on RSVP P300 speller decoding with FP1/FP2. (a) Group mean letter decoding accuracy versus repetitions (1–10) for raw, wavelet denoising applied to all epochs (wav_all), and TCN-guided selective denoising. (b) Paired comparison of AUC between $\theta = 9$dB and raw across 55 subjects. Selective denoising improves AUC. (c) Channel-wise fraction of denoised segments on the test set (mean±SD across subjects).

Table 2: Downstream P300 speller decoding with a forehead-only montage (FP1/FP2; N=55). For each denoising condition (raw, wav_all, and TCN-guided selective denoising at thresholds $\theta \in \{-5,0,5,7.5,9\}$dB), we report letter decoding accuracy after 5 and 10 repetitions (Letter@5, Letter@10; mean±SD across subjects), AUC (area under the letter-accuracy curve across repetitions 1–10; mean±SD), and the channel-wise fraction of test segments denoised (percentage of segments processed; mean±SD).

Condition	Letter@5 (mean±SD)	Letter@10 (mean±SD)	% segments denoised (test, mean±SD)	AUC mean±SD)
raw	0.149 ± 0.121	0.183 ± 0.165	0 ± 0	1.302 ± 0.998
wav_all	0.091 ± 0.099	0.129 ± 0.131	100 ± 0	0.872 ± 0.796
$\theta = -5$	0.105 ± 0.098	0.13 ± 0.123	93.13 ± 6.92	0.933 ± 0.783
$\theta = 0$	0.132 ± 0.112	0.17 ± 0.15	70.76 ± 13.3	1.22 ± 0.935
$\theta = 5$	0.153 ± 0.126	0.215 ± 0.167	35.27 ± 15.3	1.448 ± 1.018
$\theta = 7.5$	0.166 ± 0.132	0.223 ± 0.166	21.23 ± 12.26	1.548 ± 1.06
$\theta = 9$	0.182 ± 0.122	0.232 ± 0.162	12.45 ± 9.29	1.629 ± 0.981

Finally, the optimal θ varied across subjects but favored higher thresholds. By Letter@10, $\theta \geq 5$dB was optimal for 76.4% of subjects (with $\theta = 9$dB most frequent, 30.9%), and similar trends were observed for Letter@5 (72.7% with $\theta \geq 5$dB).

Section 4: Discussion

In this work we introduced a two-head, single-channel noise regressor that estimates ocular and muscle noise loads in dB, and evaluated it on synthetic mixtures, an independent eyeblink dataset and a clean ERP paradigm. On synthetic data, the model provided accurate and calibrated estimates of EOG and EMG noise levels over a –10 to +10 dB range. On the eyeblink dataset, the EOG head reproduced blink-related topographies with high spatial correlation, and in the ERP case study, the EOG load estimates were sufficient to gate a conventional wavelet blink-removal method in a way that preserved P3 amplitudes. We further validated this "detector-as-control-signal" concept on a downstream BCI task using a 55-subject P300 speller dataset with a minimal forehead montage (FP1/FP2). In this setting, indiscriminate wavelet denoising applied to all epochs degraded letter decoding accuracy, whereas detector-guided selective denoising improved decoding in a threshold-dependent manner. For example, at $\theta = 9$dB the selective pipeline increased Letter@10 from 0.183±0.165 (raw) to 0.232±0.162 and improved AUC by 0.327 ($p = 0.0040$, $d_z = 0.405$) while denoising only 12.45±9.29% of test segments on average. Together, these results suggest that continuous noise-load estimates can serve not only as a quality indicator but also as a control signal for deciding when and where to apply artifact-removal techniques.

Absolute speller performance with only FP1/FP2 is expectedly modest given the severely constrained spatial sampling and the fact that row/column P300 decoding typically benefits from broader centro-parietal coverage [68]. Nevertheless, performance remains above theoretical chance for 6×6 letter selection (1/36 ≈ 2.8%). Our intent here is therefore not to maximize speller accuracy, but to test whether noise-aware selective denoising improves signal quality and downstream decoding in a limited-lead regime relevant to long-term wearable monitoring, where only a small number of frontal electrodes is practical and spatial multichannel artifact methods (e.g., ICA/ASR) are less applicable.

The proposed model estimates contamination level; it does not guarantee recovery of neural information that may be distorted or lost due to electrode–skin interface changes or motion-related impedance fluctuations during eye movements. Likewise, cranial EMG overlaps spectrally with EEG, making perfect separation fundamentally challenging. Our contribution is therefore not complete artifact removal, but continuous quantification of ocular and muscle contamination to enable stronger interventions only when warranted and to avoid unnecessary processing of relatively clean data.

Beyond the specific wavelet example, the per-channel, per-epoch noise-load estimates could be used to govern a range of artifact-reduction methods. For example, predicted EOG load can determine when to apply more invasive processing, such as EMD/VMD decompositions or matrix-based approaches (PCA/SVD/low-rank), instead of running these uniformly across all data. Because the model outputs continuous loads in dB, the same signal can also support graded control: the predicted load may scale denoising strength (e.g., shrinkage thresholds, number of removed modes, or mixing weights between raw and processed signals), enabling light correction for mild contamination and stronger intervention only for highly contaminated segments.

Although we focused empirically on blink-related artifacts, the EMG head provides analogous information for muscle contamination and could similarly gate or scale EMG suppression steps (e.g., high-frequency pruning in wavelet/EMD pipelines or component-based removal). Finally, the proposed detector also complements emerging end-to-end deep denoisers that map noisy EEG to a putative clean signal using sequence-to-sequence models (e.g., CNNs, TCNs, U-Nets). Such models are often trained for a dominant artifact type and operate as always-on filters, without an explicit estimate of contamination severity. In contrast, our regressor provides continuous EOG and EMG noise-load estimates (in dB) that can be used to condition denoisers, e.g., as additional inputs, gating variables, or soft mixing weights between raw and denoised outputs, so that the model behaves close to an identity mapping when contamination is low and becomes more aggressive only when warranted. This same principle distinguishes our approach from segmentation–denoising networks, which tightly couple artifact localization and reconstruction in a single end-to-end model: by separating estimation from removal, we obtain a lightweight, modular control signal that can gate or scale a range of denoisers and supports interpretable operating points (e.g., threshold θ).

This study has several limitations. First, the regressor was trained on synthetic mixtures derived from a single public dataset, with noise loads defined as channel-wise power ratios; although we validated generalisation on an independent blink dataset and demonstrated ERP and BCI utility, broader testing across recording systems, tasks, and populations is needed to characterise robustness to domain shifts. Second, the downstream evaluation focused on a single BCI paradigm (RSVP P300 speller) using a minimal FP1/FP2 montage; future work should assess additional paradigms (e.g., motor imagery, SSVEP) and electrode configurations to test generality. Third, we demonstrated detector-guided control using one wavelet-based pipeline, but did not systematically benchmark gating across other artifact-reduction families (e.g., EMD/VMD, component-based methods, SSP-like projections, or end-to-end deep denoisers). Finally, while we compared several model families under a consistent protocol, we did not perform a detailed within-architecture ablation of the best-performing TCN; such ablations may further improve performance. Future work will therefore focus on expanding validation

datasets and tasks, integrating the detector with multiple denoising families, and quantifying impact on broader downstream applications such as clinical event detection and long-term ambulatory monitoring.

Beyond artifact suppression, the noise-load estimates could also be used as covariates in downstream analyses, for example, as nuisance regressors in connectivity or spectral studies, or as quality indicators in large-scale EEG repositories. Because the detector operates at the level of channel–epoch pairs with low latency and a purely convolutional architecture, it is amenable to real-time deployment in wearable or mobile systems. In that setting, it could support adaptive gain control, trigger artifact-sensitive event marking, or provide on-device quality feedback to participants and clinicians.

Section 5: Conclusion

We presented a two-head, single-channel noise regressor that estimates ocular and muscle contamination in dB from short EEG segments and enables noise-aware control of artifact reduction. Trained on synthetic mixtures with known EOG and EMG loads, a dilated TCN outperformed four alternative model families, achieving the best regression accuracy on both heads (EOG: MAE ≈ 1.8 dB, R^2 ≈ 0.82; EMG: MAE ≈ 1.0 dB, R^2 ≈ 0.94). On an independent eyeblink dataset, the EOG head recovered realistic blink topographies (subject-wise median spatial correlation ≈ 0.91).

Across application studies, uniform denoising could degrade downstream outcomes, whereas detector-guided selective denoising was more conservative. On the P3 ERP dataset, indiscriminate wavelet blink removal reduced the number of significant ERP channels, while EOG-load gating preserved most baseline significant channels while denoising only a minority of segments. On a 55-subject P300 speller benchmark using FP1/FP2, denoising all epochs reduced letter decoding performance, whereas selective denoising improved decoding at conservative thresholds (e.g., θ = 9 dB) while modifying only ~12% of test segments on average.

Overall, multi-head noise regression provides interpretable, continuous control signals that can make artifact suppression more targeted under constrained-lead conditions. Future work will extend validation across datasets and paradigms, integrate the detector with additional denoising families beyond wavelets, and evaluate generalisation to clinical event detection and long-term wearable EEG.

Code and data availability.

The trained model weights and inference code will be made publicly available in an open repository upon acceptance of this manuscript.

References

[1] Ofner P, Schwarz A, Pereira J, Wyss D, Wildburger R, Müller-Putz GR. Attempted Arm and Hand Movements can be Decoded from Low-Frequency EEG from Persons with Spinal Cord Injury. Sci Rep 2019;9:7134. https://doi.org/10.1038/s41598-019-43594-9.

[2] Abiri R, Borhani S, Sellers EW, Jiang Y, Zhao X. A comprehensive review of EEG-based brain–computer interface paradigms. J Neural Eng 2019;16:011001. https://doi.org/10.1088/1741-2552/aaf12e.

[3] Noachtar S, Rémi J. The role of EEG in epilepsy: A critical review. Epilepsy & Behavior 2009;15:22–33. https://doi.org/10.1016/j.yebeh.2009.02.035.

[4] Jeong J. EEG dynamics in patients with Alzheimer's disease. Clinical Neurophysiology 2004;115:1490–505. https://doi.org/10.1016/j.clinph.2004.01.001.

[5] Antonenko P, Paas F, Grabner R, van Gog T. Using Electroencephalography to Measure Cognitive Load. Educ Psychol Rev 2010;22:425–38. https://doi.org/10.1007/s10648-010-9130-y.

[6] Šušmáková K. Human Sleep and Sleep EEG. MEASUREMENT SCIENCE REVIEW 2004;4.

[7] Motamedi-Fakhr S, Moshrefi-Torbati M, Hill M, Hill CM, White PR. Signal processing techniques applied to human sleep EEG signals—A review. Biomedical Signal Processing and Control 2014;10:21–33. https://doi.org/10.1016/j.bspc.2013.12.003.

[8] Knyazeva MG, Innocenti GM. EEG coherence studies in the normal brain and after early-onset cortical pathologies. Brain Research Reviews 2001;36:119–28. https://doi.org/10.1016/S0165-0173(01)00087-X.

[9] Alhussein M, Muhammad G, Hossain MS. EEG Pathology Detection Based on Deep Learning. IEEE Access 2019;7:27781–8. https://doi.org/10.1109/ACCESS.2019.2901672.

[10] ChooseMuse. Muse® EEG Mental Fitness & Sleep Headband 2025. https://choosemuse.com/ (accessed October 31, 2025).

[11] IDUN Technologies - The Smartwatch for Your Brain n.d. https://iduntechnologies.com (accessed October 31, 2025).

[12] Emotiv | Brain Data Measuring Hardware and Software Solutions. EMOTIV n.d. https://www.emotiv.com/ (accessed October 31, 2025).

[13] Kaongoen N, Choi J, Woo Choi J, Kwon H, Hwang C, Hwang G, et al. The future of wearable EEG: a review of ear-EEG technology and its applications. Journal of Neural Engineering 2023;20. https://doi.org/10.1088/1741-2552/acfcda.

[14] Choi JW, Kwon H, Choi J, Kaongoen N, Hwang C, Kim M, et al. Neural Applications Using Immersive Virtual Reality: A Review on EEG Studies. IEEE Transactions on Neural Systems and Rehabilitation Engineering 2023;31:1645–58. https://doi.org/10.1109/TNSRE.2023.3254551.

[15] Mumtaz W, Rasheed S, Irfan A. Review of challenges associated with the EEG artifact removal methods. Biomedical Signal Processing and Control 2021;68:102741. https://doi.org/10.1016/j.bspc.2021.102741.
[16] Mannan MMN, Kamran MA, Jeong MY. Identification and Removal of Physiological Artifacts From Electroencephalogram Signals: A Review. IEEE Access 2018;6:30630–52. https://doi.org/10.1109/ACCESS.2018.2842082.
[17] Fatourechi M, Bashashati A, Ward RK, Birch GE. EMG and EOG artifacts in brain computer interface systems: A survey. Clinical Neurophysiology 2007;118:480–94. https://doi.org/10.1016/j.clinph.2006.10.019.
[18] Gratton G. Dealing with artifacts: The EOG contamination of the event-related brain potential. Behavior Research Methods, Instruments, & Computers 1998;30:44–53. https://doi.org/10.3758/BF03209415.
[19] Romero S, Mañanas MA, Barbanoj MJ. A comparative study of automatic techniques for ocular artifact reduction in spontaneous EEG signals based on clinical target variables: A simulation case. Computers in Biology and Medicine 2008;38:348–60. https://doi.org/10.1016/j.compbiomed.2007.12.001.
[20] Sharma R, Meena HK. Emerging Trends in EEG Signal Processing: A Systematic Review. SN COMPUT SCI 2024;5:415. https://doi.org/10.1007/s42979-024-02773-w.
[21] Chang C-Y, Hsu S-H, Pion-Tonachini L, Jung T-P. Evaluation of Artifact Subspace Reconstruction for Automatic Artifact Components Removal in Multi-Channel EEG Recordings. IEEE Transactions on Biomedical Engineering 2020;67:1114–21. https://doi.org/10.1109/TBME.2019.2930186.
[22] Blum S, Jacobsen NSJ, Bleichner MG, Debener S. A Riemannian Modification of Artifact Subspace Reconstruction for EEG Artifact Handling. Front Hum Neurosci 2019;13. https://doi.org/10.3389/fnhum.2019.00141.
[23] Vigario R, Sarela J, Jousmiki V, Hamalainen M, Oja E. Independent component approach to the analysis of EEG and MEG recordings. IEEE Transactions on Biomedical Engineering 2000;47:589–93. https://doi.org/10.1109/10.841330.
[24] Delorme A, Sejnowski T, Makeig S. Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. NeuroImage 2007;34:1443–9. https://doi.org/10.1016/j.neuroimage.2006.11.004.
[25] Jung T-P, Makeig S, Humphries C, Lee T-W, McKEOWN MJ, Iragui V, et al. Removing electroencephalographic artifacts by blind source separation. Psychophysiology 2000;37:163–78. https://doi.org/10.1111/1469-8986.3720163.
[26] Somers B, Bertrand A. Removal of eye blink artifacts in wireless EEG sensor networks using reduced-bandwidth canonical correlation analysis. J Neural Eng 2016;13:066008. https://doi.org/10.1088/1741-2560/13/6/066008.
[27] Uusitalo MA, Ilmoniemi RJ. Signal-space projection method for separating MEG or EEG into components. Med Biol Eng Comput 1997;35:135–40. https://doi.org/10.1007/BF02534144.
[28] Gramfort A. MEG and EEG data analysis with MNE-Python. Front Neurosci 2013;7. https://doi.org/10.3389/fnins.2013.00267.
[29] Barban F, Chiappalone M, Bonassi G, Mantini D, Semprini M. Yet another artefact rejection study: an exploration of cleaning methods for biological and neuromodulatory noise. J Neural Eng 2021;18:0460c2. https://doi.org/10.1088/1741-2552/ac01fe.

[30] Mijović B, De Vos M, Gligorijević I, Taelman J, Van Huffel S. Source Separation From Single-Channel Recordings by Combining Empirical-Mode Decomposition and Independent Component Analysis. IEEE Trans Biomed Eng 2010;57:2188–96. https://doi.org/10.1109/TBME.2010.2051440.

[31] Ulate-Campos A, Loddenkemper T. Review on the current long-term, limited lead electroencephalograms. Epilepsy & Behavior 2024;150:109557. https://doi.org/10.1016/j.yebeh.2023.109557.

[32] Croft RJ, Barry RJ. EOG correction: a new perspective. Electroencephalography and Clinical Neurophysiology 1998;107:387–94. https://doi.org/10.1016/S0013-4694(98)00086-8.

[33] Elbert T, Lutzenberger W, Rockstroh B, Birbaumer N. Removal of ocular artifacts from the EEG — A biophysical approach to the EOG. Electroencephalography and Clinical Neurophysiology 1985;60:455–63. https://doi.org/10.1016/0013-4694(85)91020-X.

[34] He P, Wilson G, Russell C, Gerschutz M. Removal of ocular artifacts from the EEG: a comparison between time-domain regression method and adaptive filtering method using simulated data. Med Bio Eng Comput 2007;45:495–503. https://doi.org/10.1007/s11517-007-0179-9.

[35] van den Berg-Lenssen MMC, van Gisbergen JAM, Jervis BW. Comparison of two methods for correcting ocular artefacts in EEGs. Med Biol Eng Comput 1994;32:501–11. https://doi.org/10.1007/BF02515308.

[36] Gratton G, Coles MGH, Donchin E. A NEW METHOD FOR OFF-LINE REMOVAL OF OCULAR ARTIFACT t n.d.

[37] Alkhoury L, Scanavini G, Louviot S, Radanovic A, Shah SA, Hill NJ. Artifact-reference multivariate backward regression (ARMBR): a novel method for EEG blink artifact removal with minimal data requirements. J Neural Eng 2025;22:036048. https://doi.org/10.1088/1741-2552/ade566.

[38] Zikov T, Bibian S, Dumont GA, Huzmezan M, Ries CR. A wavelet based de-noising technique for ocular artifact correction of the electroencephalogram. Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society] [Engineering in Medicine and Biology, vol. 1, 2002, p. 98–105 vol.1. https://doi.org/10.1109/IEMBS.2002.1134407.

[39] Peng H, Hu B, Shi Q, Ratcliffe M, Zhao Q, Qi Y, et al. Removal of Ocular Artifacts in EEG—An Improved Approach Combining DWT and ANC for Portable Applications. IEEE Journal of Biomedical and Health Informatics 2013;17:600–7. https://doi.org/10.1109/JBHI.2013.2253614.

[40] Krishnaveni V, Jayaraman S, Malmurugan N, Kandasamy A, Ramadoss D. Non adaptive thresholding methods for correcting ocular artifacts in EEG. Academic Open Internet Journal 2004;13.

[41] Khatun S, Mahajan R, Morshed BI. Comparative analysis of wavelet based approaches for reliable removal of ocular artifacts from single channel EEG. 2015 IEEE International Conference on Electro/Information Technology (EIT), Dekalb, IL, USA: IEEE; 2015, p. 335–40. https://doi.org/10.1109/EIT.2015.7293364.

[42] Fitzgibbon SP, Lewis TW, Powers DMW, Whitham EW, Willoughby JO, Pope KJ. Surface Laplacian of Central Scalp Electrical Signals is Insensitive to Muscle Contamination. IEEE Transactions on Biomedical Engineering 2013;60:4–9. https://doi.org/10.1109/TBME.2012.2195662.

[43] Yong X, Ward RK, Birch GE. Artifact removal in EEG using Morphological Component Analysis. 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, 2009, p. 345–8. https://doi.org/10.1109/ICASSP.2009.4959591.

[44] Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc R Soc Lond A 1998;454:903–95. https://doi.org/10.1098/rspa.1998.0193.

[45] Sweeney-Reed CM, Nasuto SJ, Vieira MF, Andrade AO. Empirical Mode Decomposition and its Extensions Applied to EEG Analysis: A Review. Adv Data Sci Adapt Data Anal 2018;10:1840001. https://doi.org/10.1142/S2424922X18400016.

[46] Zhang D-X, Wu X-P, Guo X. The EEG Signal Preprocessing Based on Empirical Mode Decomposition. 2008 2nd International Conference on Bioinformatics and Biomedical Engineering, 2008, p. 2131–4. https://doi.org/10.1109/ICBBE.2008.862.

[47] Sweeney KT, McLoone SF, Ward TE. The Use of Ensemble Empirical Mode Decomposition With Canonical Correlation Analysis as a Novel Artifact Removal Technique. IEEE Transactions on Biomedical Engineering 2013;60:97–105. https://doi.org/10.1109/TBME.2012.2225427.

[48] Bennia F, Moussaoui S, Boutalbi MC, Messaoudi N. Comparative study between EMD, EEMD, and CEEMDAN based on De-Noising Bioelectric Signals. 2024 8th International Conference on Image and Signal Processing and their Applications (ISPA), 2024, p. 1–6. https://doi.org/10.1109/ISPA59904.2024.10536839.

[49] Wu Q, Zhang W, Wang Y, Zhang W, Liu X. Research on removal algorithm of EOG artifacts in single-channel EEG signals based on CEEMDAN-BD. Computer Methods in Biomechanics and Biomedical Engineering 2021;24:1368–79. https://doi.org/10.1080/10255842.2021.1889525.

[50] Dragomiretskiy K, Zosso D. Variational Mode Decomposition. IEEE Transactions on Signal Processing 2014;62:531–44. https://doi.org/10.1109/TSP.2013.2288675.

[51] Saini M, Payal, Satija U. An Effective and Robust Framework for Ocular Artifact Removal From Single-Channel EEG Signal Based on Variational Mode Decomposition. IEEE Sensors Journal 2020;20:369–76. https://doi.org/10.1109/JSEN.2019.2942153.

[52] Massar H, Belhoussine Drissi T, Nsiri B, Miyara M. Advancements in Blind Source Separation for EEG Artifact Removal: A comparative analysis of Variational Mode Decomposition and Discrete Wavelet Transform approaches. Applied Acoustics 2025;228:110300. https://doi.org/10.1016/j.apacoust.2024.110300.

[53] Maddirala AK, Shaik RA. Removal of EMG artifacts from single channel EEG signal using singular spectrum analysis. 2015 IEEE International Circuits and Systems Symposium (ICSyS), 2015, p. 111–5. https://doi.org/10.1109/CircuitsAndSystems.2015.7394075.

[54] Maddirala AK, Veluvolu KC. Eye-blink artifact removal from single channel EEG with k-means and SSA. Sci Rep 2021;11:11043. https://doi.org/10.1038/s41598-021-90437-7.

[55] Yin J, Liu A, Li C, Qian R, Chen X. A GAN Guided Parallel CNN and Transformer Network for EEG Denoising. IEEE Journal of Biomedical and Health Informatics 2025;29:3930–41. https://doi.org/10.1109/JBHI.2023.3277596.

[56] Wu F, Yang J, Liu J, Lin Z, He Y, Zhang L. RLANET: An EEG denoising network for judgemental removal of long- and short-term distribution artefacts. Biomedical Signal Processing and Control 2025;109:107962. https://doi.org/10.1016/j.bspc.2025.107962.

[57] Pu X, Yi P, Chen K, Ma Z, Zhao D, Ren Y. EEGDnet: Fusing non-local and local self-similarity for EEG signal denoising with transformer. Computers in Biology and Medicine 2022;151:106248. https://doi.org/10.1016/j.compbiomed.2022.106248.

[58] Chen J, Pi D, Jiang X, Xu Y, Chen Y, Wang X. Denosieformer: A Transformer-Based Approach for Single-Channel EEG Artifact Removal. IEEE Transactions on Instrumentation and Measurement 2024;73:1–16. https://doi.org/10.1109/TIM.2023.3341114.

[59] Zhang H, Zhao M, Wei C, Mantini D, Li Z, Liu Q. EEGdenoiseNet: a benchmark dataset for deep learning solutions of EEG denoising. J Neural Eng 2021;18:056057. https://doi.org/10.1088/1741-2552/ac2bf8.

[60] Delorme A. EEG is better left alone. Sci Rep 2023;13:2372. https://doi.org/10.1038/s41598-023-27528-0.

[61] Mohamed S, Haggag S, Nahavandi S, Haggag O. Towards automated quality assessment measure for EEG signals. Neurocomputing 2017;237:281–90. https://doi.org/10.1016/j.neucom.2017.01.002.

[62] Wei C-S. Quantitative Quality Assessment for EEG Data: A Mini Review. 2023 IEEE Symposium Series on Computational Intelligence (SSCI), 2023, p. 64–8. https://doi.org/10.1109/SSCI52147.2023.10372021.

[63] Fickling SD, Liu CC, D'Arcy RCN, Ghosh Hajra S, Song X. Good data? The EEG Quality Index for Automated Assessment of Signal Quality. 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2019, p. 0219–29. https://doi.org/10.1109/IEMCON.2019.8936246.

[64] Kanoga S, Nakanishi M, Mitsukura Y. Assessing the effects of voluntary and involuntary eyeblinks in independent components of electroencephalogram. Neurocomputing 2016;193:20–32. https://doi.org/10.1016/j.neucom.2016.01.057.

[65] Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods 2004;134:9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009.

[66] Pion-Tonachini L, Kreutz-Delgado K, Makeig S. ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. NeuroImage 2019;198:181–97. https://doi.org/10.1016/j.neuroimage.2019.05.026.

[67] Khatun S, Mahajan R, Morshed BI. Comparative analysis of wavelet based approaches for reliable removal of ocular artifacts from single channel EEG. 2015 IEEE International Conference on Electro/Information Technology (EIT), Dekalb, IL, USA: IEEE; 2015, p. 335–40. https://doi.org/10.1109/EIT.2015.7293364.

[68] Won K, Kwon M, Ahn M, Jun SC. EEG Dataset for RSVP and P300 Speller Brain-Computer Interfaces. Sci Data 2022;9:388. https://doi.org/10.1038/s41597-022-01509-w.