Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

GRAN: An Explainable Generative–Refinement–Annotation Framework for EEG Signal Restoration

Mohammed Roshan Ibrahim[1], Prem Shivanand Jogdhankar[2], Jawahar MV[3], Dr. M Krishnamurthy[4]

[1,2,3,4]Department of Artificial Intelligence and Data science
Kcg college of technology
Chennai,india

Abstract
The use of electroencephalogram (EEG) signals is essential in clinical diagnostics and brain-computer interface (BCI) systems, but is often undermined by physiological and environmental artifacts. Conventional artifact removal methods, such as Independent Component Analysis (ICA) and wavelet-based filtering, can usually involve manual effort and can also potentially corrupt underlying neural patterns. In this paper, the authors present a new three-step pipeline, Generative-Refinement-Annotation (GRAN), a three-step automated EEG signal restoration framework that includes explainability. The Generative module trains a control artifact, i.e. eye blinks, muscle contamination, baseline drift to augment training data. The Refinement module uses a one-dimensional convolutional denoising auto-encoder with skip connectivity to reconstructions of clean signals, using corrupted inputs. The Annotation module generates comprehensible difference maps that indicate the time and space areas where corrections were made, thus making it more transparent to clinical validation. The experimental analysis on BCI Competition IV Dataset 2a shows significant improvements whereby the average SNR improvement of 8.47 dB, Pearson correlation of 0.891 and RMSE reduction of 76.1 percent are achieved over the conventional baselines, which provides an effective and understandable solution to EEG preprocessing.

Keywords—EEG signal processing; artifact removal; denoising autoencoder; explainable AI; brain-computer interface; convolutional neural network

1. Introduction

Electroencephalography (EEG) is a technology that heralds the era of neuroscience and clinical practice and offers non-invasive monitoring of real-time electrical activity in the brain with a temporal resolution of milliseconds [1]. The method has been extensively used in various fields, such as diagnosis of epilepsy, sleep disorders, cognitive neuroscience studies, and designing brain-computer interfaces (BCIs) [2]. Nevertheless, the low amplitude nature of cortical signals (usually 10-100 microvolts) predisposes EEG recordings to contamination by different artifact sources that can seriously impair signal quality [3].

The main challenge in EEG signal processing is physiological artifacts. High-amplitude deflections can be much larger than neural signals and are produced by ocular artifacts, which are produced by eye movements, blinks, and so on [4]. The facial and scalp electromyographic (EMG) contaminant adds broadband high-frequency noise at gamma-band neural activity [5]. Also, low-frequency distortions caused by the motion of electrode impedance give rise to baseline drift artifacts that cause distortions in the morphology of the signal [3].

To identify traditional methods of EEG artifact removal, mostly linear decomposition methods have been used. ICA has become a popular algorithm, and it has been used to separate mixed signals into statistically independent sources to selectively reject artifacts [6]. Although it is effective, ICA needs a lot of user expertise to identify components. Adaptive filtering techniques need reference channels [7]. Wavelet denoising provides multi-resolution analysis at the cost of subjective threshold choice [8].

The introduction of deep learning has triggered the emergence of major innovations in biomedical signal processing [9]. Autoencoders have demonstrated impressive skills in denoising tasks by training compressed representations that maintain critical signal properties and reduce noise [10]. Nevertheless, the current deep learning methods used to denoise EEGs are usually opaque, a major drawback when it comes to clinical implementation where interpretation skills are of utmost importance [11].

The paper introduces the Generative-Refinement-Annotation (GRAN) framework, which deals with the effective artifact removal and transparent processing. The architecture consists of three synergistic modules: (1) a Generative component that generates realistic artifacts to augment data, (2) a Refinement component where a one-dimensional convolutional autoencoder with skip connections is used, and (3) an Annotation component that generates interpretable correction maps that show the spatial-temporal distribution of the changes.

2. Literature Review

A. Traditional EEG Artifact Removal Techniques

Since the inception of clinical electroencephalography, the elimination of artifacts of EEG records has been a constant challenge. Initial methods were based on band-pass filtering and regression-based methods [7]. Band-pass filtering gets rid of the powerline interference and drift, but cannot do anything about artifacts whose spectrals overlap the neural activity. Regression-based techniques use reference channels to estimate artifact contributions. ICA came as a revolutionary method of blind source separation. Makeig et al. [6] have shown that multichannel EEG could be broken down into independent components using ICA to reject artifacts. ICA-based preprocessing became popular in the EEGLAB environment [12]. ICA is effective but necessitates subjective classification of components, and might not work when artifact and neural sources are not statistically independent.

B. EEG Denoising by Autoencoders.

Autoencoders are trained to encode input signals into latent representations that are compressed and to decode them to recover the original input [10]. Sun et al. [13] were the first to use denoising autoencoders to remove EEG motion artifact, and it performed better than wavelet thresholding. Yang et al. [14] have suggested a one-dimensional convolutional autoencoder to single-channel EEG denoising that showed usefulness in eliminating ocular and muscle artifacts without disrupting the morphology of event-related potentials. Architectures with skip connections in U-Nets maintain fine-grained time information when compressing [15].

C. Generative Models and Explainability.

Realistic models facilitate the synthesis of artifacts in a controlled manner to augment data. GANs have displayed impressive signal generation abilities [16]. The work of Zhang et al. [17] included the EEGDenoiseNet benchmark dataset that consists of generative artifact models. Alternative generative structures are offered by VAEs [18]. Explainability mechanisms deal with the black box nature of neural networks [11]. Saliency mapping determines powerful input features [19]. Attention systems bias the input features based on a task [20].

TABLE I. COMPARISON OF EEG ARTIFACT REMOVAL METHODS

Method	Auto	Multi	Interp.	SNR
Bandpass Filter	Yes	No	Yes	1.23
Wavelet (db4)	Part	Part	Part	3.47
ICA (JADE)	No	Yes	Part	5.21
EEGDenoiseNet	Yes	Yes	No	7.23
GRAN (Ours)	Yes	Yes	Yes	8.47

3. Problem Definition

Considering multi-channel EEG data with artifacting due to physiological factors, design an automated signal restoration system that: (1) can robustly remove a variety of artifacts without human intervention, (2) does not distort the properties of neural activity, (3) can be used to provide an interpretable annotation of where corrections were made, and (4) works across subjects and recording conditions.

Research Objectives: Build generative artifact models that synthesize realistic eyes blink, muscle, and baseline drift artifacts; build a convolutional denoising autoencoder network that is optimized on one-dimensional EEG signals; build an annotation mechanism that generates interpretable difference maps; quantify framework performance by SNR improvement, Pearson correlation, and RMSE metrics.

4. Proposed Methodology

A. Framework Overview

The GRAN system adopts the three-step pipeline of EEG signal restoration as shown in Fig. 1. Raw EEG data are standardized preprocessed in bandpass filtering (0.5-45 Hz) and z-score normalization. The Generative module is used to produce artificial content to supplement training data. The Refinement

module uses a convolutional denoising autoencoder to produce clean signals. The Annotation module calculates difference maps, which display the spatial-temporal distribution of applied corrections.
Fig. 1. GRAN Framework Architecture: Three-stage pipeline comprising Generative (artifact synthesis), Refinement (1D CNN autoencoder), and Annotation (difference mapping) modules.

B. Preprocessing Module

TABLE II. PREPROCESSING PARAMETERS

Parameter	**Value**
Sampling Rate	250 Hz
Window Size	4s (1000 samples)
Overlap	50%
Filter	Butterworth 0.5-45Hz
Normalization	Z-score
Channels	C3, Cz, C4

C. Generative Artifact Module

The Generative module synthesizes three artifact types: Eye blinks modeled as Gaussian pulses (amplitude 5–10σ, duration 200–500ms); muscle artifacts as band-limited noise (20–45 Hz) with a Hanning envelope; baseline drift as sinusoidal oscillation (0.05–0.3 Hz). Combined signal: X_noisy = X_clean + A_blink + A_muscle + A_drift.

TABLE III. ARTIFACT GENERATION PARAMETERS

Artifact	**Parameter**	**Range**
Eye Blink	Amplitude	5-10× σ
Eye Blink	Duration	200-500 ms
Muscle	Frequency	20-45 Hz
Muscle	Duration	100-300 ms
Drift	Frequency	0.05-0.3 Hz
Combined	Probability	95%/window

D. Refinement Module Architecture

The Refinement module implements a U-Net-inspired 1D convolutional autoencoder. Encoder: four blocks with channel progression 3→32→64→128→256, strided convolutions for downsampling. Decoder: mirrors the encoder using transposed convolutions. Skip connections concatenate encoder feature maps with corresponding decoder activations. Model contains 892,419 parameters. Training: combined loss L = MSE + 0.1×L1, Adam optimizer, learning rate 10^{-3}.

Fig. 2. 1D CNN Denoising Autoencoder Architecture with skip connections. Encoder compresses 1000×3 input to 62×256 latent representation; decoder reconstructs output with preserved temporal details.

TABLE IV. AUTOENCODER ARCHITECTURE

Layer	**Channels**	**Output**
Input	3	1000×3
Enc1	32	500×32
Enc2	64	250×64
Enc3	128	125×128
Enc4	256	62×256
Dec1+Skip	128	125×128
Dec2+Skip	64	250×64
Dec3+Skip	32	500×32
Dec4	3	1000×3

E. Annotation Module

The Annotation module computes interpretable correction maps: A_map = |X_noisy − X_restored|. This quantifies the correction magnitude at each time point and channel. Temporal profiles reveal when corrections occurred; channel profiles identify electrodes with greater contamination; high-correction regions (>90th percentile) highlight significant artifact locations.

5. Datasets

TABLE V. DATASET CHARACTERISTICS

Dataset	**Subj**	**Ch**	**Fs**	**Task**
BCI-IV 2a	9	22	250	4-class MI

BCI-IV 2b	9	3	250	2-class MI
PhysioNet	109	64	160	MI/ME
Sleep-EDF	197	2	100	Sleep

BCI Competition IV Dataset 2a serves as primary evaluation benchmark, comprising motor imagery EEG from nine subjects performing four-class imagery tasks. Channels C3, Cz, C4 were extracted, yielding ~27,000 training samples after segmentation.

6. ALGORITHM

TABLE VI. GRAN ALGORITHM

Step	**Operation**
1	Preprocess: Bandpass + Segment + Normalize
2	Generate: Synthesize blink/muscle/drift
3	Train: Minimize L = MSE + 0.1×L1
4	Restore: X_out = Model(X_noisy)
5	Annotate: A_map = |X_noisy - X_out|
6	Evaluate: SNR, Correlation, RMSE

7. EXPERIMENTAL RESULTS

A. Experimental Setup

Experiments employed BCI-IV 2a with subjects A01T-A04T for training, A05T for validation. Training: 20 epochs, batch size 64, learning rate 10^{-3} with step decay. Artifact probability 95%. Implementation: PyTorch with NVIDIA GPU acceleration.

B. Quantitative Performance

TABLE VII. PERFORMANCE METRICS

Metric	**Noisy**	**GRAN**	**Δ**
SNR (dB)	-2.34	6.13	+8.47
Correlation	0.412	0.891	+116%

RMSE	1.247	0.298	-76.1%
PSNR (dB)	12.34	24.67	+12.33

Fig. 3. Comparison of GRAN with baseline methods: (a) SNR improvement, (b) correlation with clean signal, (c) RMSE. GRAN achieves best performance across all metrics.

C. Per-Channel Analysis

TABLE VIII. PER-CHANNEL SNR (dB)

Channel	**Location**	**SNR Δ**	**Corr**
C3	Left Motor	8.23	0.887
Cz	Central	8.71	0.894
C4	Right Motor	8.48	0.891
Average	-	8.47	0.891

D. Artifact-Type Analysis

TABLE IX. ARTIFACT-SPECIFIC PERFORMANCE

Artifact	**SNR (dB)**	**Corr**	**RMSE Δ**
Eye Blink	9.82	0.912	-79.3%
Muscle	6.93	0.863	-71.2%
Drift	10.41	0.924	-82.7%
Combined	8.47	0.891	-76.1%

Fig. 4. GRAN performance by artifact type. Baseline drift achieves the highest SNR improvement; muscle artifacts show the lowest due to spectral overlap with neural activity.

E. Signal Restoration Example

Fig. 5. Signal restoration example: (a) clean signal, (b) noisy signal with artifacts, (c) GRAN restored signal, (d) annotation map showing correction magnitude.

F. Method Comparison

TABLE X. COMPARISON WITH BASELINES

Method	**SNR**	**Corr**	**Auto**	**Interp**
Bandpass	1.23	0.534	✓	✓
Wavelet	3.47	0.647	~	~
ICA	5.21	0.756	✗	~
Simple AE	6.12	0.812	✓	✗
EEGDenoiseNet	7.23	0.854	✓	✗
GRAN	8.47	0.891	✓	✓

G. Ablation Study

TABLE XI. ABLATION STUDY

Config	**SNR**	**Δ**	**Impact**
Full GRAN	8.47	-	Base
w/o Generative	5.23	-3.24	High
w/o Skip Conn	6.91	-1.56	Med
MSE Only	7.83	-0.64	Low
Shallow (2 blk)	6.12	-2.35	High

Fig. 6. Ablation study results. Generative module contributes most significantly to performance (+3.24 dB).

H. Cross-Dataset Generalization

TABLE XII. CROSS-DATASET RESULTS

Dataset	**SNR**	**Corr**
BCI-IV 2a (Train)	8.47	0.891
BCI-IV 2b	7.82	0.874
PhysioNet	6.94	0.841

Sleep-EDF	5.67	0.789

Fig. 7. Cross-dataset generalization. Performance degrades gradually for different domains, suggesting the benefit of domain-specific fine-tuning.

I. Computational Performance

TABLE XIII. COMPUTATIONAL METRICS

Metric	**GPU**	**CPU**
Inference/Window	2.8ms	18.4ms
Throughput	357/s	54/s
Real-time Factor	714×	108×
Parameters	892K	892K
Model Size	3.4MB	3.4MB

The framework achieves real-time capable inference at 2.8 ms per window on GPU, corresponding to 714× real-time processing, enabling deployment in latency-sensitive BCI applications.

8. CONCLUSION

The GRAN framework of explainable EEG artifact removal was discussed in this paper. Significant advances are: (1) built-in generative-refinement-annotation network that generates high-quality multi-artifact removals; (2) 1D CNN autoencoders with skip connections delivering 8.47 dB SNR gain; (3) annotation system that can produce interpretable correction maps to be used clinically. The framework is superior in performance to other existing approaches and is the only one to be fully automated and interpretable.

Future directions will focus on larger artifact models, sequence-to-sequence architecture with temporal attention, and real-time implementation to embedded platforms. The design philosophy offers a blueprint that can be applied to more general biomedical signal processing situations.

ACKNOWLEDGMENT

The authors would like to thank BCI Competition IV organizers who have made publicly available datasets and the PhysioNet team that has taken care of the EEG Motor Move/Imagery Database.

REFERENCES

1. S. Sanei and J. A. Chambers, EEG Signal Processing, Wiley, 2013.
2. J. R. Wolpaw et al., "Brain-computer interfaces," Clin. Neurophysiol., vol. 113, pp. 767–791, 2002.

3. M. K. Islam et al., "Artifact removal from scalp EEG," Neurophysiol. Clin., vol. 46, pp. 287–305, 2016.
4. R. J. Croft and R. J. Barry, "Removal of ocular artifact," Neurophysiol. Clin., vol. 30, pp. 5–19, 2000.
5. J. A. Urigüen and B. Garcia-Zapirain, "EEG artifact removal," J. Neural Eng., vol. 12, p. 031001, 2015.
6. S. Makeig et al., "ICA of EEG data," NIPS, vol. 8, pp. 145–151, 1996.
7. P. He et al., "Adaptive filtering for EEG," Med. Biol. Eng. Comput., vol. 42, pp. 407–412, 2004.
8. V. Krishnaveni et al., "Wavelet for ocular artifacts," Meas. Sci. Rev., vol. 6, pp. 45–57, 2006.
9. Y. LeCun et al., "Deep learning," Nature, vol. 521, pp. 436–444, 2015.
10. P. Vincent et al., "Denoising autoencoders," ICML, pp. 1096–1103, 2008.
11. A. B. Arrieta et al., "Explainable AI," Inf. Fusion, vol. 58, pp. 82–115, 2020.
12. A. Delorme and S. Makeig, "EEGLAB," J. Neurosci. Methods, vol. 134, pp. 9–21, 2004.
13. W. Sun et al., "1D-ResCNN for EEG," Neurocomputing, vol. 404, pp. 108–121, 2020.
14. B. Yang et al., "Deep learning for EEG," Biomed. Signal Process., vol. 43, pp. 148–158, 2018.
15. O. Ronneberger et al., "U-Net," MICCAI, pp. 234–241, 2015.
16. I. Goodfellow et al., "GANs," NIPS, vol. 27, pp. 2672–2680, 2014.
17. H. Zhang et al., "EEGdenoiseNet," J. Neural Eng., vol. 18, p. 056057, 2021.
18. D. P. Kingma and M. Welling, "VAE," ICLR, 2014.
19. I. Sturm et al., "Interpretable EEG DNNs," J. Neurosci. Methods, vol. 274, pp. 141–145, 2016.
20. A. Vaswani et al., "Attention," NIPS, vol. 30, pp. 5998–6008, 2017.