

Deep Learning-Based Techniques for Electroencephalogram (EEG) Signal Denoising

Xuanpeng Zhang*

Department of Biomedical Engineering, University of Melbourne, Victoria, Australia

* Corresponding Author: xuanpengz@student.unimelb.edu.au

Abstract. Electroencephalography (EEG) signals denoising is crucial for neural signal interpretation, particularly in complex noise conditions. Traditional methods often fail to address these conditions effectively. In contrast, deep learning-based techniques such as Convolutional Neural Networks (CNNs), Long Short-Term Memory networks (LSTMs), Generative Adversarial Networks (GANs), and Transformers offer enhanced adaptability and robustness by tuning to diverse noise characteristics. This paper presents a systematic review and comparative analysis of updated EEG denoising models, utilizing the EEGdenoiseNet benchmark dataset for a consistent evaluation framework. Performance of each model is assessed using the Relative Root Mean Square Error (RRMSE) and Correlation Coefficient (CC) across time domains. The study aims to elucidate the strengths of each denoising model, providing insights for model selection. Furthermore, it outlines each model's practical implications and limitations, setting the stage for future innovations in EEG signal processing.

Keywords: Deep Learning; Electroencephalogram; Denoising.

1. Introduction

Electroencephalography (EEG) denoising plays an important part in neural signal processing, significantly enhancing the accuracy and clarity of data analysis [1]. Traditional denoising techniques, while grounded in sophisticated mathematical theories, often fall short in adaptability and flexibility under complex noise conditions. In contrast, deep learning (DL) based techniques have revolutionized EEG denoising with their adaptability and robustness, proving effective in complex noise scenarios.

The inception of deep learning can be traced back to 1998 with Lecun's introduction of LeNet-5, the progenitor of convolutional neural networks (CNNs) [2]. The models were initially developed for applications such as image and speech recognition, these models have since been adapted to address challenges in medical fields, particularly EEG denoising.

This paper reviews five DL models, including Convolutional Neural Networks (CNNs), Autoencoders, Long Short-Term Memory networks (LSTMs), Generative Adversarial Networks (GANs), and Transformers, and their variants which have been repurposed for EEG denoising. Utilizing the EEGdenoiseNet benchmark dataset, the denoising models are rigorously evaluated based on the Relative Root Mean Square Error (RRMSE) and Correlation Coefficients (CC) in the time domain. The comparative analysis does not seek to rank these models but rather to highlight their respective denoising capabilities, aiming to provide researchers with insights necessary for selecting appropriate models and to encourage further advancements in EEG signal processing.

2. Artifacts Removal Techniques

2.1. Convolutional Neural Network (CNN):

Convolutional neural network was proposed by Lecun in the late 1980s. He developed LeNet-5, which was applied to the field of handwritten note recognition by 1998 [2]. Over several years as the technology and methodology evolved, the architecture was extended and applied to acoustic modelling and Speech Recognition. In 2020, Sun et al. introduced 1D-ResCNN, which is considered

the first application of a CNN model for EEG denoising [8]. 1D-ResCNN uses a fixed-size sliding window to segment the EEG signal into smaller overlapping segments. The features are extracted in the vertical and horizontal directions by convolving each sub-signal using two convolutional layers and an initial module. The network reconstructs the denoised EEG signal using the learned features. Huang et al. further optimised the model of CNN by proposing DRNet model [4]. They introduced the state segmentation mechanism and skip connection optimizing. The state segmentation mechanism dynamically weights and segments noise and clean signals along the time dimension. And the denoising network is constructed from a series of Inception blocks with skip connections. Additionally, the model employs a Squeeze and Excitation (SE) network after every four inception blocks to enhance amplitude feature recognition. This network calculates channel-wise mean amplitudes, assigns weights through fully connected layers and a sigmoid function, and serves as an attention mechanism to selectively suppress noise and preserve the clean signal segment.

2.2. Long Short-Term Memory (LSTM)

Long Short-Term Memory is a variant of RNN. Hochreiter and Schmidhuber developed the model of LSTM based on RNN, which they called the "remedy" of RNN [9]. RNN uses "Back-Propagation Through Time (BPTT)" algorithm. In conventional BPTT method, as the name suggests, the errors need to be propagated backwards through multiple time steps of the network. The weight is multiplied repeatedly in each iteration; hence the error is exponentially related to the weight in backward propagation. If the network weight is small (less than 1), the gradient will gradually decrease to near zero as the time step increases. This phenomenon is known as 'gradient vanishing,' which causes the network to fail at learning dependencies. Conversely, if the network weights are large (greater than 1), the gradient may gradually increase with time steps. Eventually it becomes very large, causing the training process to be unstable (gradient explosion). LSTM introduces a special network structure design that includes memory cells and three types of gating mechanisms (input gates, forgetting gates, and output gates). The improvement makes LSTM more adept at maintaining long-term dependencies in long sequences and learning more efficiently. On the basis of RNN, it improves the performance and application scope of processing sequence data.

Zhang et al. adapted Long Short-Term Memory (LSTM) network based on the research of Hochreiter and Schmidhuber and applied on removing blinking artefacts from EOG contaminated EEG signals [3]. The model starts from an initialised hidden state and processes each EEG sample sequentially through LSTM cells. Outputs are then collected from the states of each cell via a fully connected network. The output layer consists of a three-layer fully connected network with 512 or 1024 neurons per layer and utilises the ReLU activation function and dropout regularisation to improve the generalisation of the model. Building on Zhang et al.'s work, MathWorks successfully reproduced the code using MATLAB's built-in trainingOptions function and the "Adam" optimiser [10].

2.3. Autoencoders

The basic autoencoder proposed by Bengio et al. in 2006 is a neural network model for unsupervised learning aimed at dimensionality reduction and feature extraction [11]. It effectively ignores extraneous "noise" in the signal by learning to encode data into a compressed latent space representation while training to reconstruct the original data from the encoding. An autoencoder consists of both an encoder and a decoder: the encoder compresses the input data into a latent space representation; the decoder attempts to reconstruct the input from this representation to be as close as possible to the original data.

Based on the basic autoencoder, Vincent et al. proposed denoising autoencoders (DAEs). It extends the basic autoencoder model by creating a corrupted version of the input data [12]. The encoder part of a DAE learns a representation of the corrupted data and the decoder learns to reconstruct the original data from the corrupted representation. The training process of the denoising autoencoder (DAE) consists of two main steps. First, a partially corrupted version is created by deliberately corrupting the original input data using a stochastic process. In a DAE, the encoder learns a

representation of the corrupted data, and the decoder reconstructs the original data from this representation.

Based on DAE, Li et al. developed a method to optimise EEG denoising results [13]. Although the method is also based on deep learning for EEG denoising, it is different from other deep learning denoising methods in this paper. Instead of actively removing artefacts, this deep learning model focuses on reconstructing incomplete EEG segments. It is mentioned because it is representative of autoencoder methods and offers a different approach. The researchers used Lomb-Scargle periodograms to analyse the denoised segments and combined them with a denoising autoencoder (DAE) to learn from this incomplete data and generate clean EEG signals.

2.4. GANs

Ian Goodfellow et al. introduced the concept of Generative Adversarial Networks (GANs) [14]. The core of the model is an adversarial conflict between two neural networks, Generator, G and Discriminator, D. The Generator is responsible for generating as much data as possible that can deceive the discriminator model, while the Discriminator is responsible for distinguishing the authenticity of the input data. The Generator gradually learns how to generate more and more authentic data. At the same time, the Discriminator is learning how to better recognise the data. Ideally, the quality of the data produced by the Generator will be infinitely close to the real data.

Eoin Brophy et al. applied the GANs models to EEG signals denoising [15]. The model learns the relationship between contaminated data and clean data in the conflict training. The generator (G) attempts to produce an output that is close to the real EEG signal, while the discriminator (D) tries to distinguish between the real EEG signal from generated ones.

Building Eoin Brophy's application of GANs, Dong and his team developed AR-WGAN [5]. They introduced a one-dimensional convolutional layer with maximal pooling to reduce the dimensionality and one-dimensional upsampling to recover the dimensionality, thus optimising feature concentration and data integrity. In addition, they replaced the Jensen-Shannon divergence with the Wasserstein distance in the GAN framework, which improves the stability of training and provides a reliable gradient for the generator, thus enhancing the model's ability to denoise EEG signals efficiently.

2.5. Transformer

The Transformer model proposed by Ashish Vaswani et al. completely abandons recursive and convolutional networks and relies comprehensively on an attention mechanism to capture global dependencies between inputs and outputs [16]. The Transformer model consists of an encoder and a decoder, each of which consists of multiple identical layers stacked on top of each other, each of which employs a self-attention mechanism and a peer-to-peer feed-forward network. This structure allows the model to handle sequence conversion tasks in parallel, and significantly reduces the training time and improves the performance of the translation task. Transformer was originally applied to language problems.

Xiaorong Pu et al. developed EEGDnet, a Transformer variant specifically designed for denoising EEG signals [6]. EEGDnet innovatively fuses non-local and local self-similarity to remove noise from EEG signals. By handling non-local self-similarity in the self-attention block and local self-similarity in the feed-forward block, EEGDnet is able to significantly reduce the negative effects caused by noise and outliers [14].

On the basis of EEGDnet, the Denoiseformer model has been further optimised [7]. This model is proposed by Chen et al. It adopts the Transformer architecture and designs the residual variational self-encoder structure on the basis of it, which is capable of obtaining multiple feature representations, distinguishing different contaminated EEG modes, and hence improving denoising performance.

3. Comparative Analysis

3.1. EEGdenoiseNet: a benchmark

Zhang et al. constructed a semi-simulated electroencephalogram (EEG) dataset, pure EEG data contaminated by generated blinking artefacts [3]. The EEGdenoiseNet dataset set a benchmark for EEG denoising studies performance analysis. The dataset was created by preprocessing the EEG, eye oculogram (EOG) and electromyogram (EMG) data by normalising and splitting them into 2-second intervals, which is the appropriate duration to capture the temporal and spectral features of the EEG while avoiding artefacts caused by random blinks or other involuntary movements. When processing the EEG data, band-pass filtering and resampling techniques were used to accurately adjust the signal, followed by independent component analysis using the ICLabel tool to remove artefacts from the data and generate contamination-free EEG segments. The final dataset consists of 4514 clean EEG data segments, 3400 EOG artefact data segments, and 5598 EMG artefact data segments, all of which have been visually inspected by experts to ensure data purity and usability. In addition, the dataset supports the generation of contaminated EEG signals by linearly mixing clean EEG segments with EOG or EMG artefact segments and controlling the signal-to-noise ratio (SNR) of the contaminated signals by adjusting the hyper-parameter λ to mimic varying degrees of noise contamination.

3.2. Quantitative Analysis

The EEGdenoiseNet benchmark employs three objective measurements to evaluate the performance of denoising networks: RRMSE in the time domain (equation (1)), RRMSE in the spectral domain, and the correlation coefficient (CC, equation (2)) [3]:

$$RRMSE_{temporal} = \frac{RMS(f(y)-x)}{RMS(x)} \quad (1)$$

Here, y represents the simulated contaminated EEG signal, $f(y)$ denotes the denoised EEG signal, and x is the ground truth (clean EEG signal).

$$CC = \frac{Cov(f(y),x)}{\sqrt{Var(f(y))*Var(x)}} \quad (2)$$

In this formula, $Cov(f(y), x)$ indicates the covariance between $f(y)$ and x , with $Var(f(y))$ and $Var(x)$ representing their respective variances.

Relative Root Mean Square Error (RRMSE) quantifies the error magnitude relative to the original signal, with values from 0 to 1. A score of 0 denotes perfect signal reconstruction post-denoising, while a score of 1 indicates poor denoising effectiveness. Lower RRMSE values suggest that the denoising algorithm has effectively minimized noise while preserving the original signal's quality. Higher values indicate significant distortion and inadequate noise removal.

The Correlation Coefficient (CC) assesses the linearity of the relationship between the original and denoised EEG signals. A CC of 1 implies a perfect positive linear relationship, indicating optimal signal integrity preservation. A CC of 0 shows no linear relationship, suggesting unpredictable effects from denoising. A CC of -1, representing a perfect negative linear relationship, is undesirable as it indicates that original signal variations are inverted in the denoised output. Higher CC values are preferable, indicating that the denoising method maintains both the structural and dynamic features of the original signal effectively.

The three criteria are commonly used in the selected articles, while some are not mentioned, such as DRNet and AR-WGAN [4,5]. Due to the lack of information, RRMSE in the spectral domain was not considered.

3.3. Model Performance Comparison

A systematic evaluation of various deep learning models highlights significant differences in their effectiveness for EEG signal denoising. The models assessed include LSTM, AR-WGAN, DRNet, EEGDNet, and the Denoising Deformer. The evaluation metrics utilized are RRMSE_temporal and the correlation coefficient (CC), which measure temporal accuracy and the preservation of linear relationships, respectively (shown in Table 1).

Table 1. Deep learning-based electroencephalography analysis: a systematic review in RRMSE_temporal and Correlation Coefficient.

Denoising DL Models	RRMSE_temporal	Correlation Coefficient
LSTM [3]	0.404	0.895
DRNet [4]	0.611 ± 0.097	0.806 ± 0.075
AR-WGAN [5]	0.176 ± 0.046	0.726 ± 0.033
EEGDNet [6]	0.497	0.868
Denoiseformer [7]	0.444	0.859

The LSTM model demonstrates strong performance with an RRMSE_temporal of 0.404 and a Correlation Coefficient (CC) of 0.895 [3]. This model excels in capturing continuous data patterns, reflected by its high linear correlation with the authentic EEG signals. These results indicate effective noise reduction and excellent preservation of signal integrity.

DRNet recorded an RRMSE_temporal of 0.611 and a CC of 0.806, indicating moderate temporal accuracy and linear correlation capabilities [4]. These characteristics suggest that DRNet is capable of adequately capturing the general dynamics of EEG signals, although with less precision in temporal resolution compared to more focused models. This model may be preferred in scenarios where a balance is sought between computational demands and denoising performance.

The AR-WGAN model, used generative adversarial network techniques, achieved the lowest RRMSE_temporal at 0.176, indicating superior temporal accuracy among the evaluated models [5]. However, with a CC of 0.726, it suggests it is less effective at maintaining linear relationships compared to the LSTM model.

EEGDNet and Denoising Deformer reported RRMSE_temporal scores of 0.497 and 0.444, respectively, with CC values of 0.868 and 0.859 [6, 7]. These models effectively maintain structural integrity and exhibit close correlation with the real EEG signals, indicating their capability to preserve essential signal features. The performance makes them suitable for diverse denoising applications, particularly when both structural integrity and temporal accuracy are prioritized.

The comparative analysis elucidates the varying strengths and trade-offs of each model. AR-WGAN excels in reducing temporal errors but at the expense of lower correlation performance, whereas LSTM maintains a balance between noise reduction and signal integrity [3, 5]. The selection of a denoising model should thus be informed by the specific requirements of the EEG denoising task, considering both error minimization and the preservation of signal relationships.

4. Conclusion

This study presents a comprehensive review and comparative analysis of leading deep learning models for the denoising of EEG signals. Utilizing the EEGdenoiseNet benchmark such as Relative Root Mean Square Error (RRMSE) and Correlation Coefficient (CC), the research evaluates models like DRNet, LSTM, AR-WGAN, EEGDNet, and Denoiseformer across time domains.

The findings show that no single model consistently excels across all metrics, highlighting the importance of selecting a denoising model based on specific noise characteristics and application needs. For example, LSTM is particularly effective at preserving the integrity of temporal features, whereas AR-WGAN is superior in minimizing temporal errors but may compromise some linear relationships.

Future research could focus on developing hybrid models that merge LSTM's temporal fidelity with AR-WGAN's precision to enhance EEG signal denoising. Additional studies are also warranted to test these models in real-world clinical and non-clinical settings, which may present noise variables not fully captured in the benchmark dataset.

Through this detailed comparative analysis, the paper elucidates the diverse capabilities of different deep learning models. It provides insights for tailored model selection to enhance the practical application of EEG in medical diagnostics and neuroscientific research.

References

- [1] Jiang, X., Bian, G. B., & Tian, Z., Removal of Artifacts from EEG Signals: A Review, *Sensors* 19(5) (2019) 987.
- [2] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proceedings of the IEEE* 86(11) (1998) 2278-2324.
- [3] Zhang, H., Zhao, M., Wei, C., Mantini, D., Li, Z., & Liu, Q., EEGdenoiseNet: a benchmark dataset for deep learning solutions of EEG denoising, *Journal of Neural Engineering* 18(5) (2021).
- [4] X. Huang, Z. Hao, Y. Pan, W. Dou, Enhancing Artifact Removal From Scalp EEG Using State-Wise Deep Convolutional Network, 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Istanbul, Turkiye, 2023.
- [5] Dong, Y., Tang, X., Li, Q., Wang, Y., Jiang, N., Tian, L., Zheng, Y., Li, X., Zhao, S., Li, G., & Fang, P., An Approach for EEG Denoising Based on Wasserstein Generative Adversarial Network, *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 31 (2023) 3524-3534.
- [6] Xiaorong Pu, Peng Yi, Kecheng Chen, Zhaoqi Ma, Di Zhao, Yazhou Ren, EEGDnet: Fusing non-local and local self-similarity for EEG signal denoising with Transformer, *Computers in Biology and Medicine* 151 (2022) 106248.
- [7] J. Chen, D. Pi, X. Jiang, Y. Xu, Y. Chen, X. Wang, Denosieformer: A Transformer-Based Approach for Single-Channel EEG Artifact Removal, *IEEE Transactions on Instrumentation and Measurement* 73 (2024) 1-16.
- [8] Weitong Sun, Yuping Su, Xia Wu, Xiaojun Wu, A novel end-to-end 1D-ResCNN model to remove artifact from EEG signals, *Neurocomputing* 404 (2020) 108-121.
- [9] Hochreiter, S., Schmidhuber, J. Long short-term memory, *Neural Comput.* 9 (1997) 1735–1780.
- [10] MathWorks. Long short-term memory networks, Retrieved from <https://au.mathworks.com/help/deeplearning/ug/long-short-term-memory-networks.html>
- [11] Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H. Greedy layer-wise training of deep networks, *Advances in Neural Information Processing Systems*, 19 (2006).
- [12] Vincent, P., Larochelle, H., Lajoie, I., et al. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion, *J. Mach. Learn. Res.* 11 (2010) 12.
- [13] Li, J., Struzik, Z., Zhang, L., et al. Feature learning from incomplete EEG with denoising autoencoder, *Neurocomputing* 165 (2015) 23-31.
- [14] Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al. Generative adversarial nets, *Advances in Neural Information Processing Systems*, 27 (2014).
- [15] Brophy, E., Redmond, P., Fleury, A., et al. Denoising EEG signals for real-world BCI applications using GANs, *Front. Neuroergon.* 2 (2022) 805573.
- [16] Vaswani, A., Shazeer, N., Parmar, N., et al. Attention is all you need, *Advances in Neural Information Processing Systems*, 30 (2017).