Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.056016

ARTICLE

# Stroke Electroencephalogram Data Synthesizing through Progressive Efficient Self-Attention Generative Adversarial Network

**Suzhe Wang*, Xueying Zhang, Fenglian Li and Zelin Wu**

College of Electronic Information Engineering, Taiyuan University of Technology, Taiyuan, 030024, China

*Corresponding Author: Suzhe Wang. Email: wangsuzhe@tyut.edu.cn

Received: 12 July 2024 Accepted: 09 September 2024 Published: 15 October 2024

## ABSTRACT

Early and timely diagnosis of stroke is critical for effective treatment, and the electroencephalogram (EEG) offers a low-cost, non-invasive solution. However, the shortage of high-quality patient EEG data often hampers the accuracy of diagnostic classification methods based on deep learning. To address this issue, our study designed a deep data amplification model named Progressive Conditional Generative Adversarial Network with Efficient Approximating Self Attention (PCGAN-EASA), which incrementally improves the quality of generated EEG features. This network can yield full-scale, fine-grained EEG features from the low-scale, coarse ones. Specially, to overcome the limitations of traditional generative models that fail to generate features tailored to individual patient characteristics, we developed an encoder with an effective approximating self-attention mechanism. This encoder not only automatically extracts relevant features across different patients but also reduces the computational resource consumption. Furthermore, the adversarial loss and reconstruction loss functions were redesigned to better align with the training characteristics of the network and the spatial correlations among electrodes. Extensive experimental results demonstrate that PCGAN-EASA provides the highest generation quality and the lowest computational resource usage compared to several existing approaches. Additionally, it significantly improves the accuracy of subsequent stroke classification tasks.

## KEYWORDS

Data augmentation; stroke electroencephalogram features; generative adversarial network; efficient approximating self-attention

## 1 Introduction

Stroke is a serious condition resulting from the sudden rupture or obstruction of brain vessels within a few hours, leading to high rates of mortality and disability. It is mainly categorized into two classes: ischemic stroke and hemorrhagic stroke, with ischemic stroke accounting for over 70% of cases [1–3]. Therefore, it is necessary to diagnose stroke early for the patient. Electroencephalography (EEG) is an effective solution which could capture the electrical interactions among brain cells, offering valuable insight into neuronal activity [4]. This signal pattern is non-invasive and cost-effective, providing abundant data on brain functions and diseases.

Researchers have adopted various machine learning methods for this type of medical signal to extract nonlinear and high-dimensional features, among which deep learning methods are particularly famous [5–7]. The deep features of the extensive data could contribute to a more comprehensive understanding and diagnosis of neurological conditions. Therefore, they not only improve stroke diagnosis and classification performance but also advance neuroscience research. However, the collection and preprocessing of EEG data are labor-intensive and time-consuming [8], which hinders the training data provision. Moreover, data sharing between hospitals is another obstacle for data supply. Thirdly, the low proportion of stroke patient EEG data will lead to an unbalanced dataset and a decline in the learning ability of the deep model. Therefore, these issues slow down the development of intelligent-driven aided diagnosis technologies and fail to reduce the workload of doctors effectively.

Over the past decades, several studies have explored methods to generate medical data. One common approach involves geometric transformations [9]. But it is important to note that these transformed EEG features may not be inherently compatible with the original ones, as they can alter the internal time domain characteristics. Another method introduces noise into the training data [10]. However, unlike images, this technique can amplify the noise in the EEG, leading to training instability. A third approach is oversampling the minority class. Yet, the data obtained by resampling is prone to overfitting.

Lately, Generative Adversarial Networks (GANs) have become increasingly more effective for medical data generation. This deep generative model is proposed by Goodfellow et al. [11], creates novel data by data sampling. Regarding data augmentation, some researchers have studied different GANs for biological signal [12]. Although these methods can compensate for the shortage of training data and reduce category imbalance, the generated quality still needs to be improved to support high diagnosis precision. Despite the recent success of attention architecture such as Transformer, which has improved the results of many generative tasks, its optimal configurations for the EEG domain and GANs have not been fully explored. Therefore, the shortcomings of these studies can be summarized as follows:

- Existing literatures lack research on high quality EEG features generation.
- The additional attention mechanism will increase the computational complexity of the deep model when the input feature dimensions increase.
- Limited literatures are presented on the loss function design of across layers and EEG spatial domain.

To address these challenges, we present a progressive GAN framework with efficient approximating self-attention (PCGAN-EASA), designed to incrementally improve generation quality while reducing the overall computational intensity of the model. The primary novelties of this research are summarized as:

- A dual-layer auto-encoding conditional GAN framework is proposed to enhance the representation capability of EEG features across different scales. This structure captures and generates features of stroke EEG from coarse to fine, producing high quality across different scales.
- A lightweight linear effective attention module based on deep sampling approximation is employed in the encoding network. This module automatically extracts style constraint information from each patient's features while simplifying the computational intensity.
- In the design of the loss function, the cross-layer information is considered to compensate for the information degradation during the generation process. In addition, by introducing EEG spatial information reconstruction loss, the values at the electrodes are more natural and consistent.

## 2 Related Works

Unbalanced EEG data often result in overfitting and low prediction accuracy when predicting strokes. In recent years, deep generative models based on GAN have brought fresh perspectives for EEG data augmentation [13–15]. In one of the earliest studies, Abdelfattah et al. [16] utilized GANs to augment the limited EEG Motor imagery dataset, embedding a Recurrent Neural Network (RNN) into the discriminator to capture the time dependencies. Similarly, Pascual et al. [17] employed a GAN-based strategy to generate more realistic epilepsy EEG data, enabling the model to transform ictal seizure samples into comparatively refined signals. To elevate the quality of the generated EEG, Kadri et al. [18] utilized an adapted form of the Deep Convolutional Generative Adversarial Networks (DCGAN) architecture, where fully connected layers and techniques such as batch normalization contributed to generating more realistic Alzheimer's EEG signals. In another study [19], Fahimi et al. adopted the DCGAN framework for brain-computer interface tasks, producing EEG signals that closely resemble the original signals in both features and style.

As GANs have advanced, researchers have increasingly focused on the diversity of generated data styles. One approach is the use of conditional GANs. Zhang et al. [20] used this model to create subject-specific EEG features. Another effective technique is the Wasserstein GAN (WGAN). To address the issue of mode collapse, Bouallegue et al. [21] applied WGAN to gain the classification accuracy of autistic patients by enhancing training datasets. Similarly, Luo et al. [22] introduced the Wasserstein distance GAN (WGAN) with a loss function that integrates spatial and temporal information to reconstruct motor-related EEG data. The EEG data generated by this network significantly improved subsequent classification tasks. Furthermore, to generate high-quality multi-channel EEG data, Panwar et al. [23] employed the Wasserstein GAN with Gradient Penalty (WGAN-GP) architecture with a gradient penalty for data augmentation in affective EEG. Additionally, Zheng et al. [24] introduced three data augmentation techniques rooted in generative approaches to address data scarcity in EEG emotion identification. These methods, namely cWGAN (Conditional Wasserstein GAN), sVAE (Selective VAE), and sWGAN (Selective WGAN), demonstrated superior performance compared to alternatives like cVAE (Conditional VAE) and rotational data augmentation. However, these models need more network layers, which complicates the generation of high-quality EEG signals. Increasing the number of layers in each part of GANs can lead to training instability [25] and produce negligible results.

In recent years, integrating self-attention mechanisms into GAN networks has emerged as an effective approach for enhancing the generation of features with long-distance dependencies [26–28]. Concerning image generation, Tang et al. [29] incorporated a position-aware attention module into the GAN generator to amplify high-frequency features. Similarly, Kuo et al. [30] utilized the self-attention GAN model (SAGAN) to produce more detailed insomnia EEG data samples, significantly improving sleep assessment classification accuracy. In the area of EEG denoising, Yin et al. [31] achieved better performance on error-related metrics by incorporating both Transformer and CNN structures into the GAN generator, capturing both global and local correlations in EEG signals. However, these approaches face the challenge of rapidly increasing module parameters and computational load with growing input dimensions, adding burden to the model.

## 3 Methodology

### *3.1 PCGAN-EASA Architecture Design*

To improve the quality of synthetic EEG features, this study presents a progressive data generation framework named PCGAN-EASA. The architecture of PCGAN-EASA is shown in Fig. 1. This model

comprises a dual-layer generative network similar to CGAN [32]. However, it differs in that it does not require a real patient labelfor conditional mode information. Instead, the characteristics of each patient are autonomously extracted via an encoding module. The generating process of the PCGAN-EASA can be split into two main steps, each step contains three sub-steps. We will detail it as follows:

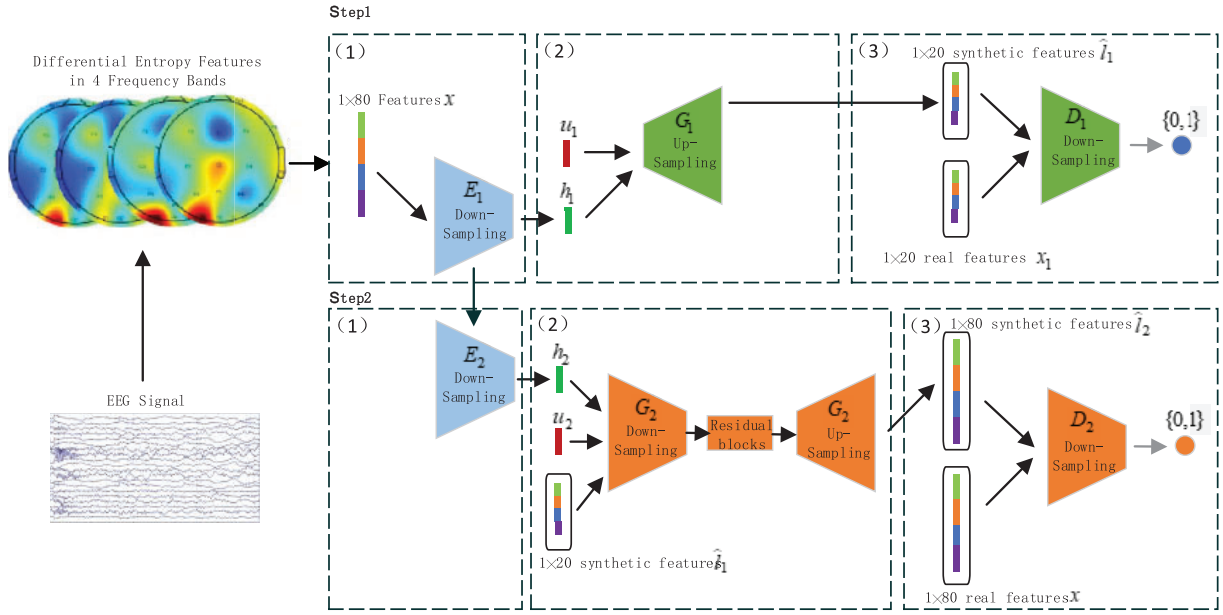


**Figure 1:** Overview of the PCGAN-EASA method

**Step 1:** Generation of Skeleton EEG Feature Data.

The first step involves parts $E_1$, $G_1$ and $D_1$. They are primarily responsible for generating preliminary stroke feature vectors.

**Sub-step 1:** Case-specific conditional feature extraction. The EASA encoder $E_1$ transforms the stroke EEG features $x$ into a compressed conditional feature vector $h_1$, adding extra constraint information to the generator $G_1$.

**Sub-step 2:** Generation of fake features. The generator $G_1$ takes the conditional vector $h_1$ and random noise $u_1$ as input, to produce a low-resolution latent feature $\widehat{l}_1$.

**Sub-step 3:** Discrimination of sample authenticity. The discriminator $D_1$ compares $\widehat{l}_1$ with the genuine feature vectors $x$ and provides feedback to the $G_1$ on the authenticity of the fake samples, thereby improving the generation refinement.

**Step 2:** Improvement of Generated Stroke Data.

This step uses components $E_2$, $G_2$ and $D_2$ with three sub-steps similar to those in the first step. The differences are: In Sub-step 1, the ESEA encoder $E_2$ continues to input the feature vector $h_1$ from the previous step to learn higher-level abstract features $h_2$. In Sub-step 2, the generator $G_2$ uses the latent features $\widehat{l}_1$ generated in the first step, along with random noise $u_2$ to jointly produce the high quantity samples $\widehat{l}_2$.

### *3.2 PCGAN-EASA Architecture Design*

The PCGAN-EASA model employs two encoders to extract latent style features that closely match the minority class distribution. The extracted conditional vector is smoother and more continuous than traditional one-hot encoding. The construction of the EASA encoder is shown in Fig. 2.

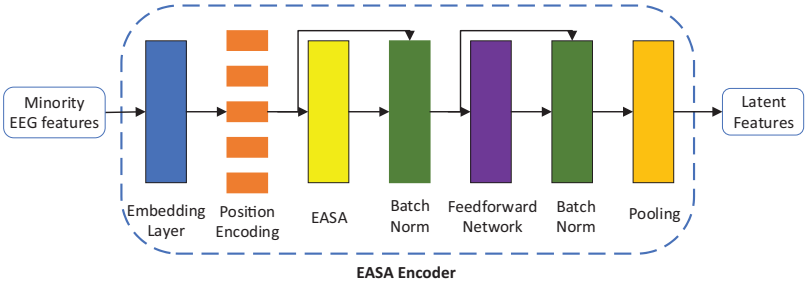


**Figure 2:** The architecture of the EASA encoder

The embedding layer firstly maps the minority class sample $x$ into a codable vector $x_e$ with dimension $d_m$, and the positional encoding layer constructs a fixed-length position vector for each element of $x$, providing efficient positional information for the linearly efficient self-attention mechanism [33]. The positional encoding formula is expressed as follows:

$$PE(i) = \begin{cases} \sin\left(\dfrac{pos}{10000^{2i}/d_{\mathrm{m}}}\right), if\ i = 2k \\ \cos\left(\dfrac{pos}{10000^{2i+1}/d_{\mathrm{m}}}\right), if\ i = 2k+1 \end{cases} \tag{1}$$

In the formula, $pos$ is the position number, the dimension of the embedding vector is $d_m$, and $i$ is the vector position.

Next, the vector $x_e$ is passed through the EASA encoder. The core structure of EASA is inspired by the Transformer self-attention mechanism [34], where each vector is split into three components: the query vector $Q$, key vector $K$, and value vector $V$, $Q$, $K$ and $V$ are obtained by passing $x_e$ through a linear layer, and the efficient approximating self-attention mechanism (EASA) is used to calculate the sample similarity, resulting in an attention matrix $x_A$.

$$x_A = EASA(Q, K, V) \tag{2}$$

Next, we apply a residual connection and layer normalization for $x_A$. The residual connection facilitates easier training of the encoder and enhances the generator's performance. Layer normalization further stabilizes the input distribution and improves the model's training convergence by learning parameters. Finally, a one-dimensional max pooling is used to generate the conditional latent features $h_1$.

#### *3.2.1 Efficient Approximating Self-Attention Mechanism*

This section elaborates on the Efficient Approximating Self-Attention mechanism. Traditional attention models divide input features into multiple groups, calculate the weight for each group, and average these weights to synthesize an aggregate feature representation. Scaled Dot-Product Attention (SDA) is a well-known implementation of this technique, as detailed in the Reference [35]. SDA

calculates the similarity between pairs of feature vectors using the dot product, then uses this similarity as the weight for averaging. The fundamental formula is presented as (3).

$$SDA_{att}(Q,K,V) = soft\max\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

$$q_i = \omega_q^T x_i, k_i = \omega_k^T x_i, v_i = \omega_v^T x_i \tag{3}$$

In Eq. (3), the extent of the series $x$ is denoted as $N$. The dimensions of vectors $Q$, $K$ and $V$ are $n\times d_q, n\times d_k, n\times d_v$, respectively, and the $Q=\{q_i\}, K=\{k_i\}, V=\{v_i\}$ is derived from $x_e$ through a linear layer. A softmax normalization is then applied across both rows and columns. From this equation, it is clear that the computational complexity of $SDA_{att}(Q,K,V)$ is $O\left(d_k N^2\right)$, which increases rapidly with growing input dimensions $x$.

By incorporating the explicit equation of the softmax function into Eq. (3), we can reformulate the equation as follows:

$$SDA_{att}(Q,K,V) = \frac{\sum_{j=1}^{n} e^{q_i^T k_j} v_j}{\sum_{j=1}^{n} e^{q_i^T k_j}} \tag{4}$$

Here, the term $e^{q_i^T k_j}$ represents the nonlinear transformation of $q$ and $k$ within a kernel function space.

We aim to improve SDA by introducing an Efficient Approximating Self-Attention Mechanism (EASA), which reduces computational complexity to $O(N)$, and enhances performance in experimental validations. The process of the EASA module is shown in Fig. 3. Next, we derive the equivalence between EASA and SDA.

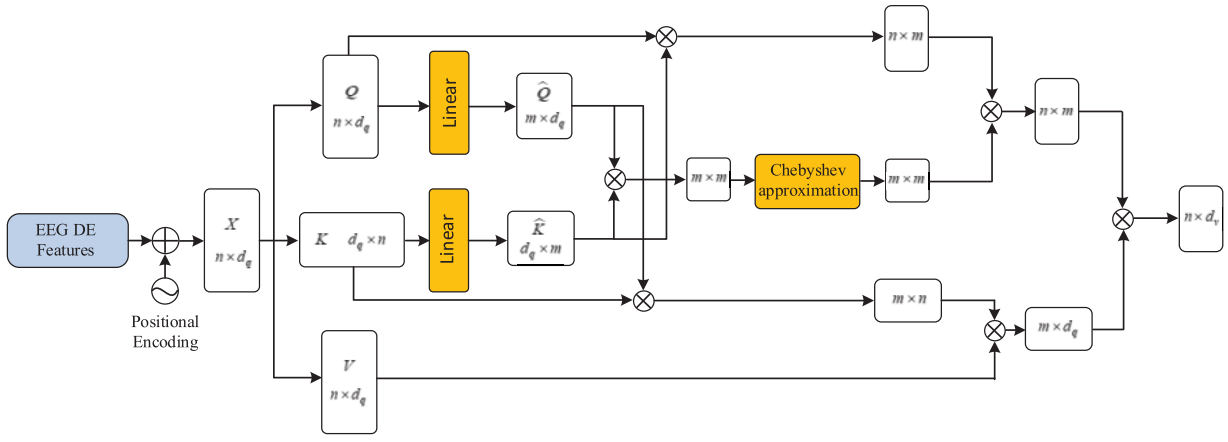


**Figure 3:** Illustration of the efficient approximating self-attention (EASA) mechanism

Initially, the EASA module adopts the Nyström approximation concept. The classic Nyström approximation algorithm [36] commonly treats the term $e^{q_i^T k_j}$ as a unified matrix $M_s$, which is then

segmented into blocks as follows:

$$M_s = \begin{bmatrix} A_s & B_s \\ B_s^T & C_s \end{bmatrix} \tag{5}$$

In the formula, the $A_s$ is the landmark submatrix obtained by randomly sampling each row and column from $M_s$.

Then matrix $M_s$ will then be Nyström decomposed:

$$M_s = \begin{bmatrix} A_s & B_s \\ B_s^T & C_s \end{bmatrix} = \begin{bmatrix} A_s \\ B_s^T \end{bmatrix} A^{\dagger} \begin{bmatrix} A_s & B_s \end{bmatrix} \tag{6}$$

In the equation, $A^{\dagger}$ represents the Moore-Penrose pseudoinverse of matrix $A_s$.

Based on this decomposition approach, to ensure that the dimensions of the decomposed $Q \cdot K^T$ during the softmax operation align with the block matrix $M_s$, we first sample the input vectors $Q$ and $K$. That is, $m$ rows are chosen from the $n$ rows to serve as the landmark matrices. This process is detailed as follows:

$$\widehat{Q} = \mathrm{FC}\,(Q), \widehat{K} = \mathrm{FC}\,(\mathrm{K}) \tag{7}$$

In the expression, FC stands for a fully connection process. The acquisition of the landmark matrix $\widehat{Q}$ and $\widehat{K}$ in the standard Nyström method typically involves random selection of $m$ elements across each row and column of the matrix $Q$ and $K$. However, the similarity between the sampled matrix and the original matrix generally depends on the value of $m$. To avoid potential issues with random sampling, an alternative strategy is adopted. The vectors are first processed through a fully connected layer with an input dimension of $n$ and an output dimension of $m$. Subsequently, a regularized dropout layer is applied to the network to reduce the risk of overfitting.

According to the block matrix $M_s$, the attention matrix $EASA\,(Q, K, V)$ is then rewritten as:

$$EASA\,(Q, K, V) = soft\max\left(\frac{Q\widehat{K}^T}{\sqrt{d_k}}\right) soft\max\left(\frac{\widehat{Q}\widehat{K}^T}{\sqrt{d_k}}\right)^{\dagger} soft\max\left(\frac{\widehat{Q}K^T}{\sqrt{d_k}}\right) \tag{8}$$

To compute the Moore-Penrose pseudoinverse † in Eq. (8), singular value decomposition (SVD) is typically employed. However, due to the high GPU resource consumption required by SVD, we utilize an iterative Chebyshev method [37] as an approximation. This method is represented by the following formula:

$$P_{k+1} = P_k\,(3I - AP_k\,(3I - AP_k)) \tag{9}$$

Subsequently, we will prove that $P_k$ will ultimately converge to the pseudoinverse of $A_{m\times m}$.

**Theorem 1.** For the matrix $A \in R^{m\times m}$, the series $\{A\}_{j=0}^{j=\infty}$ generated by the iterative method $P_{k+1} = P_k\,(3I - AP_k\,(3I - AP_k))$ will converge to the Moore-Penrose pseudoinverse with the third order convergence, given that the initial approximation $P_0$ satisfies condition $||AA^{\dagger} - AP_0|| < 1$.

**Proof of Theorem 1.**

For any non-singular matrix $P_0$ with dimension $n \times n$, we can select $P_0 = \dfrac{A^T}{||A||_1 ||A||_\infty}$ as the initial approximation matrix:

$$||A||_1 = \max_j \left\{ \sum_{i=1}^{n} |a_{ij}| \right\} \text{ and } ||\mathrm{A}||_\infty = \max_{\mathrm{i}} \left\{ \sum_{\mathrm{j}=1}^{\mathrm{n}} |\mathrm{a_{ij}}| \right\} \tag{10}$$

Subsequently, the difference $||E_0|| = ||AA^\dagger - AP_0||$ can be directly obtained:

$$||E_0|| = ||AA^\dagger - AP_0|| = ||I - AP_0|| < 1 \tag{11}$$

Subsequently, the $k+1$ difference term $||E_{k+1}||$ can be derived as:

$$\begin{aligned} ||E_{k+1}|| &= ||I - AP_{k+1}|| = I - A\left[P_k\left(3I - AP_k\left(3I - AP_k\right)\right)\right] \\ &= I - AP_k\left[3I - 3AP_k + (AP_k)^2\right] = (I - AP_k)^3 = (E_k)^3 \end{aligned} \tag{12}$$

Because $||E_0|| < 1$:

$$||E_{k+1}|| \le ||E_k||^3 \le \cdots \le ||E_0||^{3k+1} \to 0, \text{ as k} \to \infty \tag{13}$$

Finally, the proof is completed.

By approximating $soft\max\left(\dfrac{\widehat{Q}\widehat{K}^T}{\sqrt{d_k}}\right)^\dagger$ with $P^*$, the Nyström based approximation of the attention mechanism can be expressed as follows:

$$EASA\left(Q, K, V\right) = soft\max\left(\frac{Q\widehat{K}^T}{\sqrt{d_k}}\right) P^* soft\max\left(\frac{\widehat{Q}K^T}{\sqrt{d_k}}\right) \tag{14}$$

We will now analyze the time complexity of the EASA mechanism. The algorithm's total complexity primarily includes the computation of the pseudoinverse and matrix multiplication. The peak complexity for the iterative approximation of the Moore-Penrose pseudoinverse is $O\left(m^3\right)$, and the complexity of the other multiplication term in Eq. (14) is $O\left(nm^2 + 2mn\right)$. So the time complexity of the EASA is $O\left(nm^2 + 2mn + m^3\right)$.

Apart from the EASA attention module, the computations in PCGAN-EASA model occur in fully connected and convolutional layers. For the $i$-th fully connected layer, it is assumed that the size of the input features is $n$, and each layer's dimension is $d_i$. So the complexity for the fully connected layer is:

$$O\left(Ful\right) = \sum_{i=1}^{I} O\left(n \times d_i\right) \tag{15}$$

For the $j$-th convolutional layer, with a kernel size of $k_j \times k_j$, an input channel of $c_{inj}$ and an output channel of $c_{outj}$, the complexity for the convolutional layer is:

$$O\left(Con\right) = \sum_{j=1}^{J} O\left(h_j \times n \times c_{inj} \times c_{outj} \times k_j^2\right) \tag{16}$$

Finally, the time complexity of the PCGAN-EASA is:

$$O\left(PCGAN-EASA\right)=O\left(nm^2+2mn+m^3\right)+\sum_{i=1}^{I}O\left(nd_i\right)+\sum_{j=1}^{J}O\left(h_jnc_{inj}c_{outj}k_j^2\right) \tag{17}$$

### 3.3 Loss Function Design

The loss function for the PCGAN-EASA model $L_{Total}$ is bifurcated into adversarial loss $L_a$ and reconstruction loss $L_r$, presented compositely in Eq. (18):

$$L_{Total}=L_a+L_r=\lambda_1L_a\left(G_1,D_1\right)+\lambda_2L_a\left(G_2,D_2|G_1\right)+\lambda_3L_r\left(G_1\right)+\lambda_4L_r\left(G_2|G_1\right) \tag{18}$$

In this equation, $\lambda_1$, $\lambda_2$, $\lambda_3$ and $\lambda_4$ are the weighting coefficients.

The adversarial loss part $L_a$ represents the alternating training between the two GAN components, where the generator creates EEG features that mimic the real data distribution, and the discriminator $D$ attempts to distinguish the generated samples. Thus, the adversarial loss function $L_a$ is an optimization formula, as expressed in Eq. (19):

$$\min_G\max_D L_a\left(D,G\right)=\min_G\max_D\left\{E_{x\sim P_{data(x)}}\left[\log D\left(x\right)\right]+E_{z\sim p_{t(T)}}\left[\log\left(1-D\left(G\left(t\right)\right)\right)\right]\right\} \tag{19}$$

Here, $x$ symbolizes the authentic EEG feature signal while $t$ denotes a random noise with Gaussian distribution $p_t$. The adversarial training in PCGAN-EASA spans the entire two-stage process. In the front stage, the generator $G_1$ and discriminator $D_1$ interact using the random variable $u$ and conditional vector $h_1$ to refine the hidden layer features $\widehat{l}_1$ towards the real feature $x$. Consequently, the adversarial loss function $L_a(G_1,D_1)$ is divided into discriminator and generator components, as indicated in Eq. (20):

$$\begin{aligned}L_{D_1}&=\max_{D_1}E_{\left(x,h_1\right)\sim p_{data}}\left[\log D_1\left(x|h_1\right)\right]+E_{t\sim p_T(t)}\left[\log\left(1-D_1\left(G_1\left(t|h_1\right)\right)\right)\right]\\L_{G_1}&=\min_{G_1}E_{u\sim p_T(t)}\left[\log\left(1-D_1\left(G_1\left(t|h_1\right)\right)\right)\right]=\min_{G_1}E_{t\sim p_T(t)}\left[-\log D_1\left(G_1\left(t|h_1\right)\right)\right]\end{aligned} \tag{20}$$

In the second step, the coarse-grained stroke features $\widehat{x}$ are further refined with the condition vector $h_2$ and the hidden layer features $\widehat{l}_1$, leading to an adversarial loss function $L_a\left(G_2,D_2|G_1\right)$, that can be similarly expressed in Eq. (21):

$$\begin{aligned}L_{D_2}&=\max_{D_2}E_{\left(\widehat{l}_1,h_2\right)\sim p_{data}}\left[\log D_2\left(x|\widehat{x}\right)\right]+E_{t\sim p_T(t)}\left[\log\left(1-D_2\left(G_2\left(t|\widehat{l}_1,h_2\right)\right)\right)\right]\\L_{G_2}&=\min_{G_2}E_{u\sim p_T(t)}\left[-\log D_2\left(G_2\left(t|\widehat{l}_1,h_2\right)\right)\right]\end{aligned} \tag{21}$$

Moreover, to enhance the detail in the EEG feature generation, a reconstruction loss term $L_r$ is also included. First, we introduce and modify the Charbonnier loss function, whose standard formula [38] is as follows:

$$L_{Cha}=\frac{1}{N}\sum_{s=1}^{N}\sqrt{\left(\widehat{x}_s-x_s\right)^2+\varepsilon^2} \tag{22}$$

where $\widehat{x}_s$ represents the reconstructed features, $x_s$ stands for the label features, and $\varepsilon$ is a small constant typically set to $1\times10^{-3}$.

Given the inherent correlation among EEG signals from adjacent electrodes, we divide the generated and real data within the training batch into four segments to calculate the Pearson correlation coefficient. Four segments were found to be optimal through experimentation. The improved Charbonnier loss function is then obtained as follows:

$$L_{PCha} = \frac{1}{4}\sum_{s=0}^{3}\sqrt{\left(1-\rho\left(\widehat{L}_s, X_s\right)\right)+\varepsilon^2}$$

$$\widehat{L}_s = \left\{\widehat{l}_{s\cdot\frac{N}{4}+1}, \widehat{l}_{s\cdot\frac{N}{4}+2}, \cdots, \widehat{l}_{s\cdot\frac{N}{4}+\frac{N}{4}}\right\} \mathrm{s} = 0,1,2,3$$

$$X_s = \left\{x_{s\cdot\frac{N}{4}+1}, x_{s\cdot\frac{N}{4}+2}, \cdots, x_{s\cdot\frac{N}{4}+\frac{N}{4}}\right\} \mathrm{s} = 0,1,2,3 \tag{23}$$

where $\rho$ denotes the operation of computing the Pearson correlation coefficient.

Additionally, to ensure smoother generated feature signals, we combine the modified Charbonnier loss with the MSE loss function, and describe the reconstruction part $L_r$ under the generators $G_1$ and $G_2$ in the following mixed form:

$$L_r(G_1) = L_{MSE}(G_1) + L_{PCha}(G_1) = \frac{1}{N}\sum_{i=1}^{N}\left(\widehat{l}_i - x_i\right)^2 + \frac{1}{4}\sum_{s=0}^{3}\sqrt{\left(1-\rho\left(\widehat{L}_s, X_s\right)\right)+\varepsilon^2}$$

$$\begin{aligned} L_r(G_2|G_1) &= L_{MSE}(G_2|G_1) + L_{PCha}(G_2|G_1) \\ &= \frac{1}{N}\sum_{i=1}^{N}\left(\widehat{x}_i - x_i\right)^2 + \frac{1}{4}\sum_{s=0}^{3}\sqrt{\left(1-\rho\left(\widehat{X}_s, X_s\right)\right)+\varepsilon^2} \end{aligned} \tag{24}$$

Here, $N$ denotes the batch size, $\widehat{l}_i$ and $\widehat{x}_i$ are the reconstruction vectors from generators $G_1$ and $G_2$, and $x_i$ represents the authentic input vectors. Furthermore, the pseudocode for the PCGAN-EASA can be summarized as follows:

**Algorithm 1:** The PCGAN-EASA algorithm

**Input:** Actual stroke EEG features $x$; random noise $u_1$, $u_2$
**Output:** Fake stroke EEG features $\widehat{l}_2$;
1: **for** epoch=0 **do**
2: **for** number of discriminator iterations **do**
3: Fetch **m** samples from real EEG feature distribution $p_{data}$;
4: Fetch **m** samples from synthetic feature $\widehat{l}_1$, $\widehat{l}_2$;
5: Compute discriminator loss $L_{D1}$, $L_{D2}$;
6: Update discriminator parameters;
7: **end**
8: **for** number of generator iterations **do**
9: Fetch **m** samples from real EEG feature distribution $p_{data}$;
10: Compute conditional features $h_1$, $h_2$;
11: Fetch **m** samples from noise distribution $p_z$;
12: Update generator by $L_{G1}$ and $L_{G2}$;
13: **end**
14: **end**

## 4 Results

The validation of the PCGAN-EASA model involved two primary stages. First, during the ablation study phase, we fine-tuned the EASA layers number in the encoder to determine the optimal structural configuration. Following this, the efficacy of the EASA encoder was benchmarked against other established attention mechanisms, highlighting its unique advantages and efficiency. To address the issue of data insufficiency, particularly in minority stroke case classification tasks, we employed a multi-faceted evaluation approach. This approach included generating additional training instances using various generative adversarial network (GAN) architectures and testing with different quantities and proportions of synthetic data to assess performance improvement after data augmentation comprehensively.

### *4.1 Dataset and Preprocessing*

The study utilized the EEG data collected from the Neurology Department of Shanxi Provincial People's Hospital, China. There are 204 participants, with 106 males and 98 females, aged between 29 and 73 years old.

Under the supervision of the ethics committee, the data were utterly anonymized before training. All participants were informed about the purpose and scope of data usage, and the synthetic data were only used for the diagnosing the stroke-related diseases. During the recording, the psychological states of the participants were closely paid attention to avoid any potential negative impacts.

All participants initially lay supine in a quiet room. After a 5-min relaxation, which established a baseline EEG, the participants were instructed to close their eyes for a 10-min recording. Every 3 min, they were asked to open their eyes and rest for 20 s to maintain focus. Each participant was recorded for at least half an hour.

The EEG signals were sampled at a frequency of 500 Hz using 20 electrodes. The original data was divided into multiple 60-s segments, resulting in a total of 4444 samples, with 1034 from acute ischemic stroke patients and 3410 from healthy individuals. The dataset had a degree of imbalance of 1:3.3. To enhance the quality of the EEG samples, several preprocessing methods were applied, including electrode localization, noise reduction, and artifact removal. Differential entropy features were extracted from the $\delta$, $\theta$, $\alpha$, and $\beta$ frequency bands, resulting in a feature dimension of 80 for each sample.

### *4.2 Analytical Metrics*

To comprehensively evaluate the quality of the synthetic EEG features, we employed five metrics: Mean Correlation Coefficient (MCC), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index Measure (SSIM) [39]. Additionally, the model's complexity and memory demands were quantified using the number of parameters (Params) and Floating Point Operations Per Second (FLOPs) [40]. For assessing the performance of subsequent stroke classification tasks, a series of evaluation metrics were used, including Accuracy (Acc), Sensitivity (Se), Specificity (Sp), F1 score (F1), and Geometric Mean (G-mean, Gm).

### *4.3 Experimental Environment and Parameter Configuration*

During the PCGAN-EASA training process, the dataset was divided into training, validation, and testing sets in a ratio of 8:1:1. The Nadam optimizer was used for parameter updates, with a learning rate of 0.0001. Training was conducted with a batch size of 64 over 200 epochs.

The hardware configuration for the experiment included a computer with a 3.6 GHz Intel 12700 K processor, 32 GB of memory, and a GeForce RTX 3090 graphics card. The software environment consisted of Windows 10 OS, with Python 3.8 and TensorFlow used for model development. To ensure consistency in the results, each experimental method was executed five times, and the average outcome of these repetitions was recorded as the final result.

### *4.4 Experimental Results*

#### *4.4.1 The Layers Number in EASA Encoding*

In the PCGAN-EASA architecture, the encoder plays a crucial role in capturing the case-specific style from the input EEG data. The depth of the EASA affects the complexity and detail of the information it captures. Therefore, the study focused on the impact of varying the number of EASA layers in the encoder on the quality of synthetic EEG features. The number of layers was adjusted from 1 to 5, and the results of these adjustments are quantitatively presented in Table 1.

**Table 1:** The effect of different layers of encoding in EASA on the generation performance

| Numbers of layer | MCC/% | RMSE | MAE | PSNR | SSIM/% |
|---|---|---|---|---|---|
| 1 | 86.31 | 0.3071 | 0.2513 | 62.09 | 88.68 |
| 2 | 89.93 | 0.2604 | 0.2111 | 63.65 | 88.37 |
| **3** | **91.64** | **0.2464** | **0.1909** | **65.48** | **91.26** |
| 4 | 87.41 | 0.3348 | 0.2626 | 62.77 | 87.15 |
| 5 | 75.76 | 0.5156 | 0.4395 | 59.82 | 64.17 |

It was observed that the metrics for feature generation consistently improved as the number of EASA layers increased from one. The optimal generation quality was achieved with a three-layer EASA, where MCC and SSIM peaked at 91.64 and 91.26, respectively. However, further increasing the number of layers resulted in a noticeable decline in the quality of EEG feature generation. This decline is attributed to overfitting, which arises from the increased complexity and computational cost of the model. In summary, appropriately selecting the number of attention encoding layers increases PCGAN-EASA's learning and representation capabilities, allowing it to extract more representative style information from EEG features. This style of information, through the interaction between the generator and discriminator, helps make the generated EEG features more realistic.

#### *4.4.2 Comparison of Experimental Results Generated by Different Encoders*

To further examine the competence of the EASA module, we compared it against the lightweight deep neural network structure MobileNet, the Transformer module incorporating conventional Self-Dot Attention (SDA), and the Linformer module, which implements lightweight, low-rank factorized self-attention. The evaluation metrics included the previously mentioned generation quality indicators, as well as the number of Params and FLOPs of the overall model under different modules to

assess computational efficiency. The results, presented in Table 2, show that incorporating attention mechanism modules significantly improved the similarity index of generation effects. Notably, the model with the EASA module achieved an average improvement of approximately 6% over the Transformer and Linformer modules, yielding the best generative performance. Additionally, the EASA module demonstrated a similar trend in improving computational efficiency. For example, the total number of network parameters with EASA is only 48% and 85% of those with SDA and Linformer, respectively, and the computational performance improved by at least 10% compared to the Transformer and Linformer modules.

**Table 2:** Comparison of the generation effects of different encoding modules in PCGAN-EASA

| Encoding module | MCC/% | RMSE | MAE | PSNR | SSIM/% | Params | FLOPs |
|---|---|---|---|---|---|---|---|
| +MobileNet | 50.67 | 0.9097 | 0.9281 | 29.56 | 34.51 | 17 M | 3.0 G |
| +Transformer | 74.21 | 0.3415 | 0.3071 | 60.47 | 72.24 | 23 M | 4.8 G |
| +Linformer | 85.58 | 0.2924 | 0.2542 | 61.99 | 75.56 | 14 M | 2.2 G |
| **+EASA** | **91.64** | **0.2464** | **0.1909** | **65.48** | **91.26** | **11 M** | **1.8 G** |

To further validate the computational efficiency gains provided by the EASA module, we compared the training times across generative models that utilized various encoding modules, under increasing dimensions of EEG feature inputs, ranging from a single $\delta$ band feature to all four bands: $\delta$, $\theta$, $\alpha$ and $\beta$. As depicted in Fig. 4, a clear pattern emerged: training times inevitably increased with the rise in input feature dimensions. Notably, the model equipped with the EASA module exhibited the most gradual increase in training times, outperforming both Linformer and MobileNet. In contrast, the Transformer-encoded model showed the most rapid increase in training time. This demonstrates that the EASA module has a distinct advantage in reducing computational demands. However, the EASA's training time is close to Linformer. It does not show a significant advantage, which may be due to the low-rank approximation matrix introduced by Linformer, reducing the algorithm's complexity to $O(n)$.

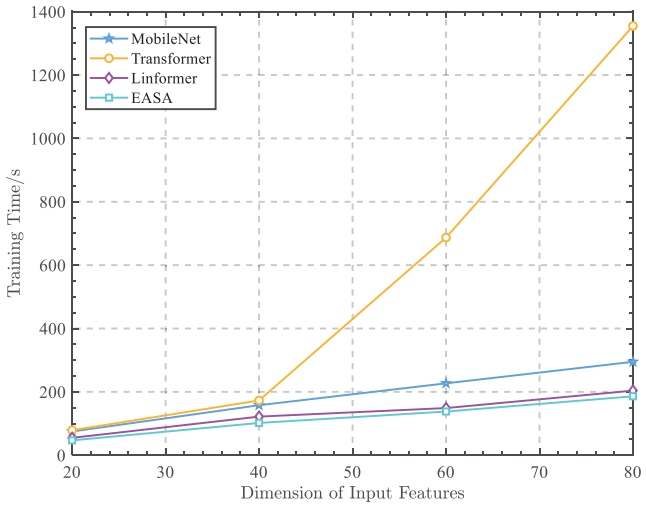


**Figure 4:** The training time of each encoder with EEG feature dimensions increasing

In Fig. 5, the data from Table 2 is visually represented. The first column of images displays topographic maps illustrating the differential entropy characteristics of EEG across four frequency bands for a stroke patient. The subsequent columns, from the second to the fifth, sequentially show the EEG topographic feature maps generated by the PCGAN-EASA network architecture after integrating various encoding modules. Comparative analysis reveals that without the attention mechanism—specifically when using only MobileNet—there are significant deviations in feature values across multiple frequency bands compared to the original images, particularly in the theta and beta frequency bands. Although integrating attention mechanisms like Transformer and Linformer resulted in feature maps that more closely resemble the original data, considerable differences in feature values remain in certain brain regions, which may adversely affect the accuracy of subsequent disease diagnosis by classifiers. Among all comparisons, the proposed EASA module demonstrated the highest similarity in feature values to the original data across different electrode positions and frequency bands. This indicates a more authentic and natural enhancement effect, further confirming the EASA module's superior performance in capturing the details of EEG features.

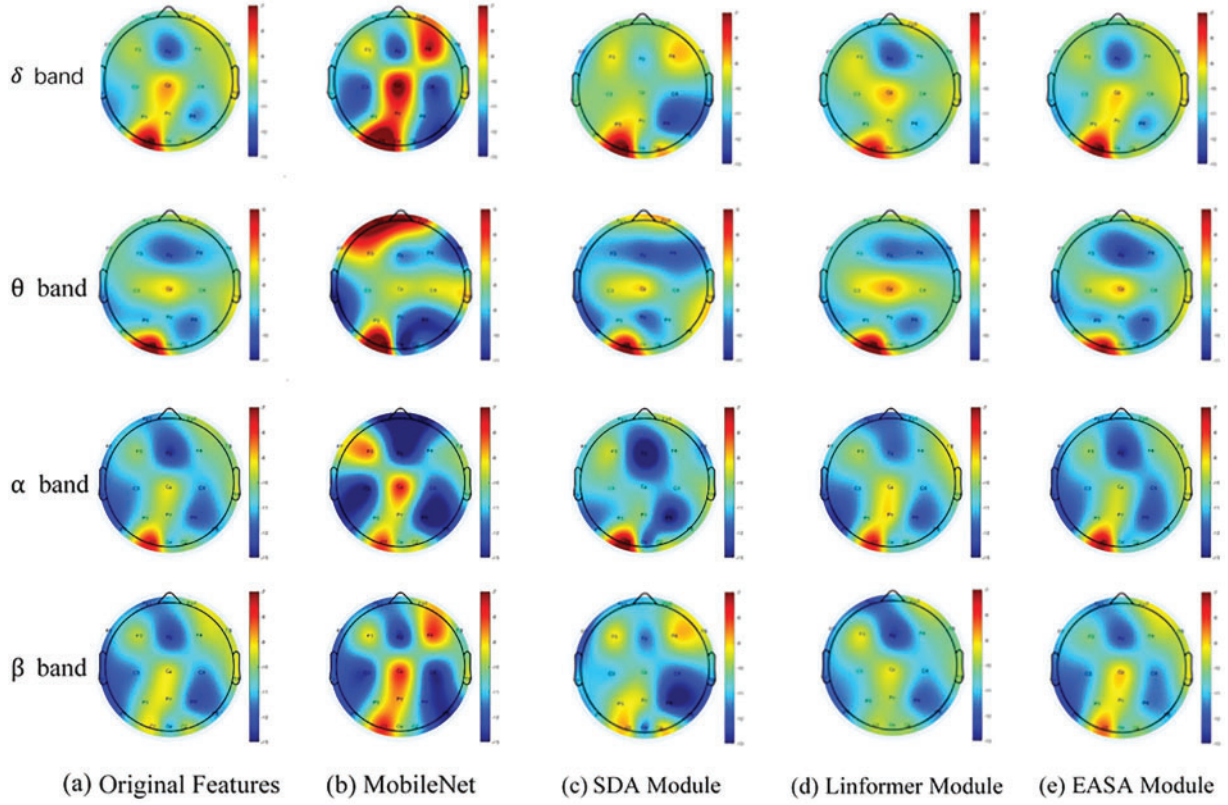


**Figure 5:** Visualization of EEG feature generation across four frequency bands by different encoding modules

### *4.4.3 Comparison of Generative Models for Stroke Classification Effectiveness*

To quantitatively evaluate the impact of the PCGAN-EASA generative model on enhancing stroke classification efficiency, this study employed a five-fold cross-validation strategy to partition the initial dataset. We utilized a range of standard Generative Adversarial Networks, including DCGAN

[41], WGAN [42], E-WACGAN [43], SA-GAN [28], Transformer-CGAN, and PCGAN-EASA, to augment an imbalanced EEG dataset. Subsequently, for the balanced training sets after augmentation, a CNN was used as the classifier. The performance of each model is presented in Table 3. It is clear from the table that PCGAN-EASA outperformed the others in overall performanceh assessment, followed by SA-CGAN. Comparatively, WGAN's performance in data enhancement was not as strong as DCGAN's, but E-WACGAN demonstrated higher sample quality and model robustness due to its improved loss function. Notably, PCGAN-EASA achieved optimal results in sensitivity, geometric mean, and F1 score, with values of 97.79%, 98.04%, and 97.91%, respectively. These results reflect its significant impact on increasing sample diversity and enhancing the quality of minority class samples. PCGAN-EASA was not only successful in generating high-quality minority class samples but also achieved 98.59% accuracy and 97.82% specificity, fully demonstrating its ability to improve the classifier's efficacy. Of course, in the stroke EEG samples, a few training samples are inherently very similar to normal EEG features, or the deep features extracted by EASA are not enough, even after applying data augmentation. This increases the risk of undetected stroke patients.

**Table 3:** Evaluating the stroke classification performance across different generative models

| Methods | Acc/% | Se/% | Gm/% | F1/% | Sp/% |
|---|---|---|---|---|---|
| DCGAN | 86.99 | 64.51 | 77.76 | 69.76 | 93.81 |
| WGAN | 85.89 | 62.00 | 75.94 | 67.09 | 93.13 |
| E-WACGAN | 90.73 | 71.38 | 83.00 | 78.13 | 96.60 |
| SA-GAN | 93.31 | 87.56 | 91.29 | 89.53 | 95.07 |
| Transformer-CGAN | 95.26 | 95.45 | 94.91 | 93.64 | 96.15 |
| **PCGAN-EASA** | **98.59** | **97.79** | **98.04** | **97.91** | **97.82** |

#### *4.4.4 Influence of Different Ratios of Generated Data on Stroke Classification*

Fig. 6 illustrates the performance of stroke classification when the training set is mixed with varying proportions of synthetic samples. The results show that as the proportion of synthetic stroke data increases, all classifiers exhibit improved ability to identify EEG signals. The most significant prediction performance is observed when synthetic data comprises 70% of the total training set. However, when the proportion of synthetic samples becomes excessively high, the data distribution of the training set can become skewed, leading to overfitting to the artificial samples and ultimately impairing the accuracy of real data classification.

## 5 Discussion

In general, the proposed GANs augmentation algorithm for feature datasets has played a significant role in improving the imbalance in stroke datasets.

From the perspective of advantages, based on the experimental results, this algorithm has effectively alleviated the problem of stroke data scarcity in deep diagnostic model, providing the possibility for independent deployment in other hospitals. Meanwhile, with the reduced training time and computational resources, the hardware configuration pressure on future service providers will also decrease, which is conducive to providing more cost-effective solutions. Thirdly, using the automatic

patient feature extraction module may also inspire a future controlled generation of case signals in other GAN architectures.

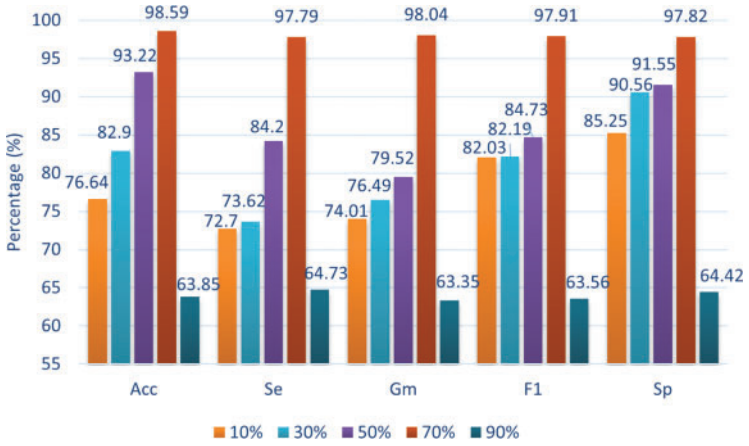


**Figure 6:** Comparison of training times for each encoder across different feature dimensions

In terms of disadvantages, although the PCGAN-EASA can effectively increase accuracy, the diagnostic efficiency is still insufficient because neurologists or radiologists in actual diagnostics need to face multiple possible disease types simultaneously. If the models for all the disease types have to be stored, it would increase the storage space and computation time. Recently proposed general-purpose large models [44] may be one of the effective means to solve these problems. Additionally, although the generative effects of GAN models are relatively sound, the complexity of the training process leads to decreased training stability, making the model prone to issues such as mode collapse and gradient vanishing. Therefore, it is necessary to explore methods such as Wasserstein loss functions [45] or automated architecture search [46] in future research to make adjustments.

## 6 Conclusions

In response to the challenges faced by deep generative models in generating high-quality stroke EEG signals while managing computational resource consumption, this study introduces a novel feature augmentation technique based on an efficient attention mechanism. The model refines the EEG feature generation process by constructing a progressive conditional generative network, addressing the instability issues commonly encountered in traditional deep network training. During the model's encoding phase, an efficient approximating self-attention mechanism based on Nyström approximation is introduced, enabling deeper extraction of case-specific information from stroke EEG features while reducing the computational resources typically required by traditional scaled dot-product attention mechanisms. PCGAN-EASA outperforms traditional generative models across various evaluation metrics, demonstrating its effectiveness in refining EEG signal feature generation and enhancing the accuracy of stroke patient classification.

Future research will focus on addressing the challenges posed by imbalanced stroke EEG data with multiple categories and limited samples per category. Developing deep neural network models that can more effectively extract features from EEG signals will be central to this research. Additionally, integrating other modalities, such as image data, with EEG signal features to improve sample generation quality is a key direction for future efforts. Furthermore, these methodologies could

be extended to super-resolution generation of EEG signals, semi-supervised learning, and domain adaptation, offering significant potential for related applications.

**Acknowledgement:** None.

**Funding Statement:** This work is supported by the General Program under grant funded by the National Natural Science Foundation of China (NSFC) (No. 62171307), and the Basic Research Program of Shanxi Province under grant funded by the Department of Science and Technology of Shanxi Province (China) (No. 202103021224113).

**Author Contributions:** Conceptualization and methodology, Suzhe Wang and Xueying Zhang; formal analysis, Suzhe Wang, Xueying Zhang, Fenglian Li and Zelin Wu; software, validation and writing—original draft, Suzhe Wang and Xueying Zhang; writing—review and editing and data curation, Suzhe Wang, Xueying Zhang and Fenglian Li; supervision and funding acquisition, Suzhe Wang, Xueying Zhang and Fenglian Li. All authors reviewed the results and approved the final version of the manuscript.

**Availability of Data and Materials:** Data not available due to ethical restrictions.

**Ethics Approval:** Ethical approvals for this research were obtained from the Ethics Committees of Taiyuan University of Technology and Shanxi Provincial People's Hospital (approval numbers: TYUT-2023122501 and 1911140805). The study was conducted in accordance with the Declaration of Helsinki. Informed consent was obtained from the participants and informed consent was obtained from Shanxi Provincial People's Hospital.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

[1] K. D. Seo, M. J. Kang, G. S. Kim, J. H. Lee, S. H. Suh and K. Y. Lee, "National trends in clinical outcomes of endovascular therapy for ischemic stroke in South Korea between 2008 and 2016," *J. Stroke*, vol. 22, no. 3, pp. 412–415, Sep. 2020. doi: 10.5853/jos.2020.01928.

[2] J. Y. Kim and H. J. Bae, "Spontaneous intracerebral hemorrhage: Management," *J. Stroke*, vol. 19, no. 1, pp. 28–39, Jan. 2017. doi: 10.5853/jos.2016.01935.

[3] M. P. Kate *et al.*, "Pre-hospital triage of suspected acute stroke patients in a mobile stroke unit in the rural alberta," *Sci. Rep.*, vol. 11, no. 1, pp. 88–93, Mar. 2021. doi: 10.1038/s41598-021-84441-0.

[4] J. I. Doerrfuss, T. Kilic, M. Ahmadi, M. Holtkamp, and J. E. Weber, "Quantitative and qualitative EEG as a prediction tool for outcome and complications in acute stroke patients," *Clin. EEG Neurosci.*, vol. 51, no. 2, pp. 121–129, Sep. 2019. doi: 10.1177/1550059419875916.

[5] D. M. Michels, L. C. Van Dijk, and D. L. J. Tavy, "Perioperative stroke during carotid endarterectomy: Benefits of multimodal neuromonitoring-A case report," *BMC Neurol.*, vol. 22, no. 1, Aug. 2022, Art. no. 325. doi: 10.1186/s12883-022-02835-7.

[6] P. D. Barua *et al.*, "Multilevel hybrid accurate handcrafted model for myocardial infarction classification using ECG signals," *Int. J. Mach. Learn. Cyb.*, vol. 14, no. 5, pp. 1651–1668, 2023. doi: 10.1007/s13042-022-01718-0.

[7] A. Subasi, S. Dogan, and T. Tuncer, "A novel automated tower graph based ECG signal classification method with hexadecimal local adaptive binary pattern and deep learning," *J. Amb. Intel. Hum. Comp.*, vol. 14, no. 2, pp. 711–725, 2023. doi: 10.1007/s12652-021-03324-4.

[8] X. Liu, "Detection and classification of ADHD using deep learning based on EEG signals," *Highlig. Sci., Eng. Technol.*, vol. 91, pp. 191–199, 2024. doi: 10.54097/12cjyf16.
[9] M. Xu *et al.*, "BWGAN-GP: An EEG data generation method for class imbalance problem in RSVP tasks," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 30, pp. 251–263, 2022. doi: 10.1109/TNSRE.2022.3145515.
[10] C. He, J. Liu, Y. Zhu, and W. Du, "Data augmentation for deep neural networks model in EEG classification task: A review," *Front. Hum. Neurosci.*, vol. 15, no. 1, pp. 1–14, Dec. 2021. doi: 10.3389/fnhum.2021.765525.
[11] I. Goodfellow *et al.*, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020. doi: 10.1145/3422622.
[12] A. G. Habashi, A. M. Azab, S. Eldawlatly, and M. A. Gamal, "Generative adversarial networks in EEG analysis: An overview," *J. NeuroEng. Rehabil.*, vol. 20, no. 40, pp. 1–24, 2023. doi: 10.1186/s12984-023-01169-w.
[13] Z. Zhang, S. Zhong, and Y. Liu, "Beyond mimicking under-represented emotions: Deep data augmentation with emotional subspace constraints for EEG-based emotion recognition," presented at AAAI Conf. Artif. Intell., Vancouver, BC, Canada, Feb. 22–25, 2024, pp. 22–25.
[14] C. Wang, L. Liu, W. Zhuo, and Y. Xie, "An epileptic EEG detection method based on data augmentation and lightweight neural network," *IEEE J. Transl. Eng. He.*, vol. 12, no. 8, pp. 22–31, Aug. 2023. doi: 10.1109/JTEHM.2023.3308196.
[15] X. Du *et al.*, "Electroencephalographic signal data augmentation based on improved generative adversarial network," *Brain Sci.*, vol. 14, no. 4, Apr. 2024, Art. no. 367. doi: 10.3390/brainsci14040367.
[16] S. M. Abdelfattah, G. M. Abdelrahman, and M. Wang, "Augmenting the size of EEG datasets using generative adversarial networks," presented at the 2018 Int. Jt. Conf. Neural Networks, Rio de janeiro, Brazil, Jul. 08–13, 2018.
[17] D. Pascual, A. Amirshahi, A. Aminifar, D. Atienza, P. Ryvlin and R. Wattenhofer, "Epilepsygan: Synthetic epileptic brain activities with privacy preservation," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 8, pp. 2435–2446, Dec. 2020. doi: 10.1109/TBME.2020.3042574.
[18] R. Kadri, M. Tmar, B. Bouaziz, and F. Gargouri, "Alzheimer's disease detection using deep ECA-ResNet101 network with DCGAN," presented at the 2021 Int. Jt. Conf. Hybrid Intell. Syst., Ottawa, ON, Canada, Dec. 14–16, 2021, pp. 14–16.
[19] F. Fahimi, Z. Zhang, W. B. Goh, K. K. Ang, and C. Guan, "Towards EEG generation using GANs for BCI applications," presented at the 2019 Int. Jt. Conf. Biomed. Health. Inf., Chicago, IL, USA, May 19–22, 2019.
[20] A. Zhang, L. Su, Y. Zhang, Y. Fu, L. Wu and S. Liang, "EEG data augmentation for emotion recognition with a multiple generator conditional Wasserstein GAN," *Complex Intell. Syst.*, vol. 8, no. 4, pp. 1–13, 2021. doi: 10.1007/s40747-021-00336-7.
[21] G. Bouallegue and R. Djemal, "EEG data augmentation using Wasserstein GAN," presented at the 2020 Int. Conf. Sci. Tech. Autom., Monastir, Tunisia, Dec. 20–22, 2020.
[22] T. Luo, Y. Fan, L. Chen, G. Guo, and C. Zhou, "EEG signal reconstruction using a generative adversarial network with wasserstein distance and temporal-spatial-frequency loss," *Front. Neuroinform.*, vol. 14, Apr. 2020, Art. no. 15. doi: 10.3389/fninf.2020.00015.
[23] S. Panwar, P. Rad, T. -P. Jung, and Y. Huang, "Modeling EEG data distribution with a Wasserstein generative adversarial network to predict RSVP events," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 28, no. 8, pp. 1720–1730, Aug. 2020. doi: 10.1109/TNSRE.2020.3006180.
[24] W. Zheng and H. Zhao, "Cost-sensitive hierarchical classification for imbalance classes," *Appl. Intell.*, vol. 50, no. 8, pp. 2328–2338, Mar. 2020. doi: 10.1007/s10489-019-01624-z.
[25] V. Kushwaha and G. C. Nandi, "Study of prevention of mode collapse in generative adversarial network (GAN)," presented at the 2020 Int. Conf. Inf. Commun. Tech., Changchun, China, May 24–26, 2020.
[26] X. Wang, Z. Cao, R. Wang, Z. Liu, and X. Zhu, "Improving human pose estimation with self-attention generative adversarial networks," *IEEE Access*, vol. 7, pp. 119668–119680, 2019. doi: 10.1109/ACCESS.2019.2936709.

[27] G. Sun, S. Ding, T. Sun, and C. Zhang, "SA-CapsGAN: Using capsule networks with embedded self-attention for generative adversarial networks," *Neurocomputing*, vol. 423, no. 5, pp. 399–406, Jan. 2021. doi: 10.1016/j.neucom.2020.10.092.

[28] P. F. de Araujo-Filho, M. Naili, G. Kaddoum, E. T. Fapi, and Z. Zhu, "Unsupervised GAN-based intrusion detection system using temporal convolutional networks and self-attention," *IEEE Trans. Netw. Serv.*, vol. 20, no. 4, pp. 4951–4963, Dec. 2023. doi: 10.1109/TNSM.2023.3260039.

[29] J. Tang, B. Zou, C. Li, S. Feng, and H. Peng, "Plane-wave image reconstruction via generative adversarial network and attention mechanism," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–15, 2021. doi: 10.1109/TIM.2021.3087819.

[30] C. E. Kuo, T. H. Lu, G. T. Chen, and P. Y. Liao, "Towards precision sleep medicine: Self-attention GAN as an innovative data augmentation technique for developing personalized automatic sleep scoring classification," *Comput. Biol. Med.*, vol. 148, Sep. 2022, Art. no. 105828. doi: 10.1016/j.compbiomed.2022.105828.

[31] J. Yin, A. Liu, C. Li, R. Qian, and X. Chen, "A GAN guided parallel CNN and transformer network for EEG denoising," *IEEE J. Biomed. Health.*, vol. 5, pp. 1–12, 2024. doi: 10.1109/JBHI.2023.3277596.

[32] X. Ding, Y. Wang, Z. Xu, Z. J. Wang, and W. J. Welch, "Distilling and transferring knowledge via cGAN-generated samples for image classification and regression," *Expert Syst. Appl.*, vol. 213, no. 4, 2023, Art. no. 119060. doi: 10.1016/j.eswa.2022.119060.

[33] H. Zhao, J. Jia, and V. Koltun, "Exploring self-attention for image recognition," presented at the 2020 Proc. Int. Conf. Comput. Vis. Pattern Recog., Seattle, WA, USA, Jun. 14–19, 2020.

[34] A. Vaswani *et al.*, "Attention is all you need," presented at Adv. Neural Inf. Process. Syst., Long Beach, CA, USA, Dec. 4–9, 2017.

[35] G. Chen and Y. Maday, "Directed message passing based on attention for prediction of molecular properties," *Comp. Mater. Sci.*, vol. 229, no. 3, Oct. 2023, Art. no. 112443. doi: 10.1016/j.commatsci.2023.112443.

[36] K. Zhang, I. W. Tsang, and J. T. Kwok, "Improved Nyström low-rank approximation and error analysis," presented at the 2020 Proc. Int. Conf. Mach. Learn., Vienna, Austria, Jul. 12–18, 2020.

[37] M. He, Z. Wei, and J. R. Wen, "Convolutional neural networks on graphs with Chebyshev approximation, revisited," *Adv. Neural Inf. Process. Syst.*, vol. 35, no. 3, pp. 7264–7276, 2022. doi: 10.1016/j.commatsci.2023.112443.

[38] J. T. Barron, "A general and adaptive robust loss function," presented at the 2019 Proc. Int. Conf. Comput. Vis. Pattern Recog., Long Beach, CA, USA, Jun. 15–20, 2019.

[39] S. Cao, "Choose a transformer: Fourier or Galerkin," presented at the 2021 Proc. Int. Conf. Adv. Neural Inf. Process. Syst., Dec. 8–14, 2021.

[40] K. Rasheed, J. Qadir, T. J. O'Brien, L. Kuhlmann, and A. Razi, "A generative model to synthesize EEG data for epileptic seizure prediction," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 29, pp. 2322–2332, 2021. doi: 10.1109/TNSRE.2021.3125023.

[41] Q. Wu, Y. Chen, and J. Meng, "DCGAN-based data augmentation for tomato leaf disease identification," *IEEE Access*, vol. 8, pp. 98716–98728, 2022. doi: 10.1109/ACCESS.2020.2997001.

[42] Q. Wang *et al.*, "WGAN-based synthetic minority over-sampling technique: Improving semantic fine-grained classification for lung nodules in CT images," *IEEE Access*, vol. 7, pp. 18450–18463, 2019. doi: 10.1109/ACCESS.2019.2896409.

[43] Q. Jin, R. Lin, and F. Yang, "E-WACGAN: Enhanced generative model of signaling data based on WGAN-GP and ACGAN," *IEEE Syst. J.*, vol. 14, no. 3, pp. 3289–3300, Sep. 2020. doi: 10.1109/JSYST.2019.2935457.

[44] Y. Chen *et al.*, "Generative adversarial networks in medical image augmentation: A review," *Comput. Biol. Med.*, vol. 6, no. 1, Nov. 2023, Art. no. 105382. doi: 10.1038/s41746-023-00958-w.

[45] S. H. Gobato and A. Moradi, "A novel loss function for neural network models exploring stock realized volatility using wasserstein distance," *Dec. Anal. J.*, vol. 10, no. 3, May 2024, Art. no. 100369. doi: 10.1016/j.dajour.2023.100369.

[46] A. Heuillet, A. Nasser, H. Arioui, and H. Tabia, "Efficient automation of neural network design: A survey on differentiable neural architecture search," *ACM Comput. Surv.*, vol. 56, no. 11, pp. 1–36, Jun. 2024. doi: 10.1145/3665138.