Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

EEG Emotion Recognition Based on Masked Generative Adversarial Networks and Neural Architecture Search

Kexuan Zhu
Qufu Normal University
Qian Zhao
zhaoq0620@126.com

Qufu Normal University
Huadi Wang
Qufu Normal University
Yang Zhao
Shanxian Vocational Secondary School of Shandong

Research Article

Keywords: EEG, emotion recognition, generative adversarial network, data augmentation, neural architecture search

Posted Date: April 21st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9400088/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

EEG Emotion Recognition Based on Masked Generative Adversarial Networks and Neural Architecture Search

Kexuan Zhu[a], Qian Zhao[a,*], Huadi Wang[a], Yang Zhao[b]

[a] *College of Engineering, Qufu Normal University, Rizhao, 276800, China*

[b] *Department of Mechatronics Engineering, Shanxian Vocational Secondary School of Shandong, Heze, 274300, China*

[*]Corresponding author: Qian Zhao. Email: zhaoq0620@126.com

Abstract: EEG signals face challenges such as a scarcity of training data, significant individual variability, and difficulties in designing neural network architectures, which severely limit the performance and generalization capabilities of emotion recognition models. This paper proposes an EEG-based emotion recognition method that integrates masked generative adversarial data augmentation with personalized neural architecture search. We designed a data augmentation strategy using a frequency-domain feature-based stochastic electrode mask generative adversarial network, which effectively expands the training dataset and enhances the model's generalization ability. Subsequently, we constructed a neural architecture search space tailored for EEG signals and employed a differentiable architecture search method to automatically discover the optimal network topology. Finally, a personalized architecture search strategy is employed to independently identify the classification network best suited to each subject's unique EEG signal characteristics. Experimental validation on the public EEG emotion dataset DEAP achieved a classification accuracy of 98.64% on the Arousal dimension and 98.59% on the Valence dimension, significantly outperforming traditional machine learning methods and deep learning models with fixed architectures.

Keywords: EEG, emotion recognition, generative adversarial network, data augmentation, neural architecture search

1. Introduction

As a core element of human psychological activity, emotions profoundly influence an individual's cognitive processes, decision-making, and social interactions [1]. Accurate identification and understanding of emotional states demonstrate broad application potential in various fields, including human-computer interaction, mental health monitoring, driving safety alerts, and educational assessment [2]. Traditional emotion recognition methods primarily rely on external behavioral characteristics such as facial expressions, voice intonation, and body movements to infer internal emotional states [3]; however, these methods are susceptible to interference from subjective control and external environmental factors [4], resulting in limitations in recognition accuracy and robustness.

Electroencephalogram (EEG) signals originate directly from neural activity and can more fundamentally reflect the physiological basis of emotion generation. With millisecond-level temporal resolution and resistance to conscious masking, they provide an objective and reliable physiological approach to emotion recognition [5–8]. However, the construction of EEG emotion datasets relies on complex experiments and expensive equipment, resulting in small-scale publicly available datasets that are prone to overfitting in deep learning models [9]; At the same time, significant inter-individual variations in EEG activity patterns pose a challenge to the models' generalization capabilities [10]. Furthermore, classifiers often rely on manually designed architectures, which are not only time-consuming

and labor-intensive but also make it difficult to ensure a globally optimal structure [11]. These factors have hindered the further development of EEG emotion recognition research.

Given the challenges outlined above, this paper presents three key innovations:

1. To address the scarcity of EEG data, after extracting features through frequency-domain decomposition, we randomly mask some electrode channels and use generative adversarial networks to generate high-quality samples, thereby effectively augmenting the training dataset.

2. To address the challenge of network architecture design, we construct a neural architecture search space tailored to EEG signals and employ differentiable architecture search methods to automatically discover optimal network topologies, thereby more effectively extracting discriminative features and improving classification accuracy.

3. To address the issue of individual variability, we propose a personalized architecture search strategy that independently searches for the classification network best suited to each subject's signal characteristics, achieving precise adaptation at the architectural level.

2. Related Work

2.1 Data Augmentation

Research on data augmentation for EEG-based emotion recognition has developed diverse technical approaches, providing crucial support for improving the accuracy and robustness of emotion recognition. Li et al. [12] achieved effective data augmentation through temporal distortion, frequency noise addition, and frequency masking, thereby effectively mitigating the problem of overfitting in small-sample models; Wardoyo et al. [13] proposed a radius-constrained SMOTE sampling strategy, which synthesizes samples in the feature space under radius constraints, enabling more reasonable supplementation of small datasets; Liao et al. [14] performed probabilistic sampling of EEG features based on a Gaussian mixture model, generating synthetic samples that conform to the distribution of the original data.

Generative Adversarial Networks (GANs), with their powerful generative capabilities, have demonstrated significant potential in the field of EEG emotion recognition data augmentation. Luo et al. [15] proposed a Wasserstein-based Generative Adversarial Network Domain Adaptation (WGANDA) framework, which achieves cross-subject EEG emotion recognition through a two-step training process and gradient penalization; its performance significantly outperforms existing domain adaptation methods; Zhang et al. [16] combined GANs with self-supervised learning, using masking transformations and adversarial training to increase sample diversity, and introduced a multi-factor training network to balance sample quality and diversity; Tang et al. [17] proposed a cross-domain adaptation GAN (SDC-Net), incorporating semantic dynamic consistency constraints into the generator to enable synthesized samples to effectively bridge distribution differences across different subjects.

These data augmentation techniques optimize EEG data from various dimensions, effectively enhancing the performance of EEG emotion recognition models and laying a solid technical foundation for the development of the field.

2.2 Neural Architecture Search

In the field of EEG emotion recognition, Neural Architecture Search (NAS) has made significant progress by leveraging its automated network design capabilities to effectively

address issues such as the difficulty of traditional manually designed models in adapting to the spatiotemporal characteristics of EEG signals and their limited generalization ability. The team led by Li, Zhang, et al. [11] was the first to systematically apply NAS to EEG emotion recognition, exploring how to automatically construct networks capable of learning discriminative features from multi-channel EEG signals, thereby laying the foundation for subsequent research. In their subsequent work [18], the team further refined their approach by proposing the Transformer-based Neural Architecture Search (TNAS) framework, which employs Multi-Objective Evolutionary Algorithms (MOEA) to discover the optimal model for emotion recognition from a well-trained pool of hypernetworks. Li et al. [19] proposed Spatio-Temporal Neural Architecture Search for EEG-based Emotion Recognition (SSTNAS), which utilizes spatial and temporal EEG features through spiking convolutional neural networks and spiking long-short-term memory networks, respectively, and explores the appropriate SNN architecture for each task via spiking activation patterns. Wu et al. [20] proposed the AutoEER framework, designing a search space specifically tailored for EEG signals to derive the optimal network architecture from the proposed search space. Neural architecture search not only discovers superior domain-specific architectures that surpass human design experience but also adaptively optimizes EEG-based emotion recognition, enhancing classification accuracy and robustness.

However, existing research suffers from the following shortcomings: First, data augmentation methods are often limited to time-domain or simple frequency-domain transformations, failing to fully exploit the multi-channel spatial topology and frequency-domain discriminative features of EEG signals, resulting in insufficient sample diversity. Second, neural architecture search typically employs a uniform network, neglecting individual differences and limiting personalized recognition capabilities. Finally, generative adversarial networks and architecture search are often applied independently at the data and model layers, failing to achieve deep integration between data augmentation and architecture.

3 Proposed Method

The proposed method framework consists of three core modules, as shown in Fig. 1.

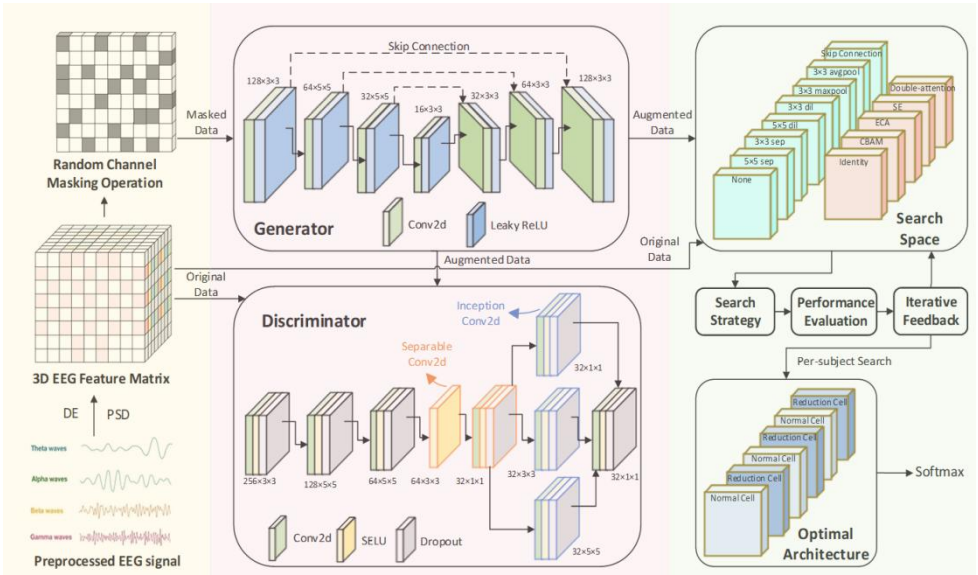


Fig. 1 System Architecture Diagram

3.1 Frequency-Domain and Spatial Feature Representation

3.1.1 Frequency Domain Decomposition and Feature Extraction

As a reflection of the electrophysiological activity of brain neurons, the frequency components of EEG show a clear correlation with specific states of brain function [21]. To accurately extract emotion-related feature information from EEG signals, this study decomposes raw single-channel EEG signals into four major frequency bands—δ (0.5 – 4 Hz), θ (4 – 8 Hz), α (8 – 13 Hz), and β (13 – 30 Hz)—each with distinct physiological significance, in accordance with the standardized frequency band classification rules of international clinical electroencephalography [22].

For each decomposed frequency band, two types of features are extracted to capture the association patterns between the signal and emotional states from different dimensions:

(1) Differential Entropy (DE)

DE is a quantitative measure of the uncertainty in the probability distribution of continuous random variables; it reflects the complexity and variability of a signal and has been widely applied in EEG-based emotion recognition tasks [23]. It is typically assumed that the EEG signals in each frequency band follow a Gaussian distribution, and the definition of differential entropy can be expressed as [24]:

$$h\left(x_f\right)=\frac{1}{2}\log\left(2\pi e\sigma^2\right)$$

where σ^2 is the signal variance, calculated as follows:

$$\sigma^2=\frac{1}{N}\sum_{i=1}^{N}\left(x_f\left(t_i\right)-\bar{x}_f\right)^2$$

where N is the number of sampling points within a single short-time window, and $\bar{x}_f$ is the mean of the signal $x_f(t)$ within that time window.

(2) Power Spectral Density (PSD)

PSD is used to describe the distribution of signal power in the frequency domain; it provides an intuitive representation of the contribution of different frequency components to the total energy of the signal [25]. Its mathematical expression is:

$$PSD_f=\frac{1}{K}\sum_{k=1}^{K}\left|FFT_k\left(x_f\right)\right|^2$$

where $FFT_k\left(x_f\right)$ represents the fast Fourier transform result of the signal x_f within the kth time window, and K is the total number of time windows into which the signal is divided.

Combining these two types of features enables a more comprehensive representation of EEG frequency-domain information, providing richer discriminative criteria for subsequent emotion classification tasks [26].

3.1.2 Spatial Mapping

To preserve the spatial topology and multidimensional features of the EEG signals, the DE and PSD features extracted from the four frequency bands of the 32-channel electrodes based on the 10-20 system [27] were spatially mapped [28] and converted into a 2D feature map with a 9×8 grid. Subsequently, the DE and PSD maps from the four frequency bands were stacked by frequency band to form a 9×8×8 three-dimensional spatio-frequency feature

map, as shown in Fig. 2. This representation integrates spatial, frequency-domain, and feature-type information, providing structured input for deep learning models and supporting the mining of emotion-related features.

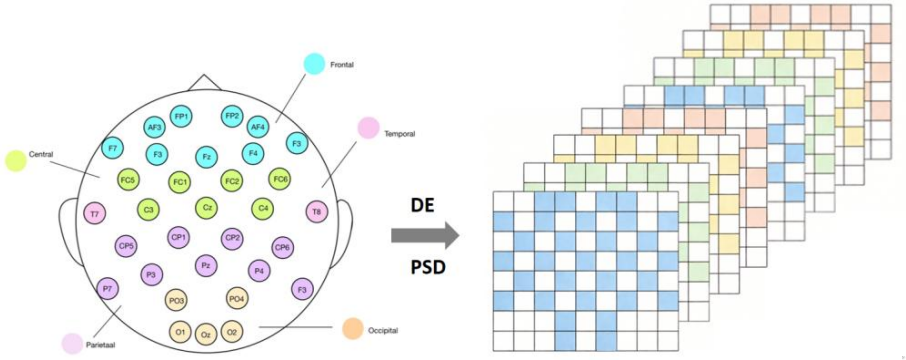


Fig. 2 EEG spatial mapping diagram

3.2 Masked Adversarial Training for Data Augmentation

3.2.1 Random Masking Operation

For 2D EEG feature maps, each value in the matrix is set to a random decimal between 0 and 1, and a threshold r is then randomly selected within a certain range. The value at each position in the matrix is compared with r; if the random number is less than r, that position is marked as masked and set to zero; otherwise, the original value is retained. This process regenerates the mask in every training iteration. This masking mechanism forces the generator to learn universal patterns of the complete data distribution from various partially observed scenarios, rather than memorizing specific data patching patterns, thereby significantly enhancing the model's generalization ability and robustness.

3.2.2 Adversarial Training Network Architecture

This paper proposes a network for enhancing EEG spatial-temporal features based on generative adversarial learning. The network uses the Wasserstein GAN with Gradient Penalty (WGAN-GP) as its basic framework. Through the adversarial interaction between the generator and the discriminator, it learns the true distribution of EEG spatial-temporal features, thereby generating high-quality enhanced feature samples.

The adversarial training network consists of a generator and a discriminator. The generator adopts a U-Net architecture to reconstruct complete feature representations from masked inputs [29]. During the encoding phase, features are progressively extracted through four layers of convolutions to compress the feature space. In the decoding phase, transposed convolutions are used to gradually restore the feature representations. Features from the corresponding layers in the encoding phase are directly concatenated to the decoding layer inputs via skip connections, achieving multi-scale feature fusion and effectively preventing the loss of fine-grained details during reconstruction. This architecture requires the generator to infer global structures from locally visible information, thereby ensuring that the generated EEG features possess semantic consistency and structural rationality.

The generator loss function is:

$$\mathrm{L}_G = -\mathrm{E}_{x\sim P_{\mathrm{Data}}}\left[D\left(G(x_{\mathrm{masked}})\right)\right]$$

where $G(\cdot)$ denotes the generator mapping function, $D(\cdot)$ is the discriminator output, and

x_{masked} is the input sample processed with a random mask.

The discriminator is used to distinguish whether the input spatio-temporal features originate from real data or generated data. The network extracts spatio-temporal joint features through three layers of 2D convolutions and incorporates deep separable convolutions and multi-branch convolutional modules to perform multi-scale feature fusion, thereby reducing computational complexity and enhancing feature representational capacity. Each convolutional layer uses the SELU activation function, and the final discrimination result is output through a fully connected layer.

The optimization objective of the discriminator is based on an estimate of the Wasserstein distance, aiming to accurately distinguish between real and generated samples:

$$\mathrm{L}_D = \mathrm{E}_{x\sim P_{\mathrm{Data}}}\left[D(x)\right] - \mathrm{E}_{x\sim P_{\mathrm{Data}}}\left[D\left(G(x_{\mathrm{masked}})\right)\right] + \lambda_{gp}\cdot GP$$

The gradient penalty term *GP* ensures that the discriminator satisfies the 1-Lipschitz continuity constraint:

$$GP = \mathrm{E}_{\hat{x}\sim p_{\mathrm{interpolation}}}\left[\left(\left\|\nabla_{\hat{x}} D(\hat{x})\right\|_2 - 1\right)^2\right]$$

Interpolated samples $\hat{x}$ are formed by a linear mixture of real and generated samples. λ_{gp}=10.0 is the gradient penalty weight coefficient, which stabilizes the training process by constraining the gradient norm of the discriminator at interpolation points. The discriminator's objective is to minimize$\mathcal{L}_D$, thereby accurately estimating the Wasserstein distance and providing meaningful gradient signals to the generator.

During training, the generator is updated only after the discriminator has been updated multiple times, and an L1 reconstruction constraint is incorporated into the generator's loss to ensure consistency between the generated features and the true features. Through adversarial training, the generator can gradually approximate the distribution of true EEG spatial-temporal features, thereby generating structurally sound augmented samples.

3.3 Neural Architecture Search

3.3.1 Design of the Search Space

This study employs a cell structure based on a directed acyclic graph (DAG), where each cell contains N ordered nodes (N=4). Each edge represents a candidate operation, and the set of candidate operations includes basic operations and attention operations. Basic operation set: 3×3 and 5×5 separable convolutions, dilated convolutions, max pooling, average pooling, skip connections, and zero operations. Attention operation set: Squeeze-Excite (SE), Efficient Channel Attention (ECA), Convolutional Block Attention Module (CBAM), and Double Attention, guiding the network to focus on emotion-related features.

The optimal architecture identified by the network search comprises two types of cells: Normal Cells, which extract and preserve spatial information without altering the feature map dimensions, and Reduction Cells, which reduce resolution through downsampling while simultaneously increasing the receptive field and the number of channels. These two types of cells are stacked according to a predefined fixed pattern, thereby achieving powerful expressive capabilities while ensuring structural simplicity.

3.3.2 DARTS Search Method

To efficiently search for the optimal architecture, this study employs the Differentiable

ARchiTecture Search (DARTS) method. By using a continuous relaxation strategy, this method transforms the discrete architecture selection problem into a differentiable optimization problem, enabling the simultaneous optimization of architecture parameters and network weights via gradient descent, thereby significantly reducing computational costs [30].

During the search phase, the output of each edge is composed of a weighted combination of all candidate operations, where the weights of each operation are derived from the corresponding architecture parameters via a normalization function. This ensures that the search space is continuous and differentiable, and allows the architecture's performance to be evaluated using the validation set loss function. During the search process, the network weights are updated under fixed architectural parameters, and then the architectural parameters are updated under fixed weights. As the search progresses, the weights of the candidate operations gradually converge. Ultimately, by selecting the operation combination with the highest weight for each edge, the continuous architecture is transformed into a discrete network structure.

3.3.3 Personalized Search Strategy

To minimize the impact of variations in EEG data distributions across participants on the architecture search, this study performed the architecture search independently for each participant. During the search process, an independent neural network model was initialized for each participant, and the network weights and architecture parameters were updated using only their individual training and validation data. By evaluating the performance of different operational combinations on the individual's data, the architecture parameters were gradually optimized, ultimately yielding a personalized network architecture capable of effectively capturing the relationship between the participant's EEG features and emotional states.

4. Experiments and Results

4.1 Experimental Setup

All experiments were conducted on a Windows 10 computer equipped with an NVIDIA 3060 GPU and a CPU. The code was developed using Python 3.9 and the PyTorch deep learning framework.

Experiments were performed on the DEAP [31] public EEG emotion dataset. After preprocessing the data [32], the dataset was split into training and testing sets in an 8:2 ratio, and five-fold cross-validation was used for model selection and parameter tuning. In masked generative adversarial training, both the generator and discriminator used the Adam optimizer with learning rates of 1×10^{-5} and 5×10^{-5}, respectively, trained for 300 epochs with a batch size of 64. During the neural architecture search phase, network weights were optimized using stochastic gradient descent with an initial learning rate of 0.1 and momentum of 0.9, while architecture parameters were optimized using the Adam optimizer with a learning rate of 6×10^{-4}. During the architecture evaluation phase, the learning rate was adjusted to 2.5×10^{-3}. In this phase, the batch size was fixed at 256, and architecture search and evaluation were conducted for 100 and 200 training epochs, respectively.

4.2 Experimental Results

4.2.1 Quality Assessment of Enhanced EEG Feature Maps

By randomly selecting several generated feature maps and comparing them with the actual feature maps, as shown in Fig. 3, it can be observed that the generated feature maps are highly similar to the actual feature maps in terms of spatial patterns and channel characteristics, successfully preserving the activation patterns across different frequency bands and the spatial distribution across different brain regions. In particular, at unmasked locations, the generated features are almost identical to the original features. At masked locations, the generator successfully infers reasonable feature values from information in the surrounding unmasked areas, resulting in smooth reconstructions that align with the spatial continuity characteristics of EEG signals.

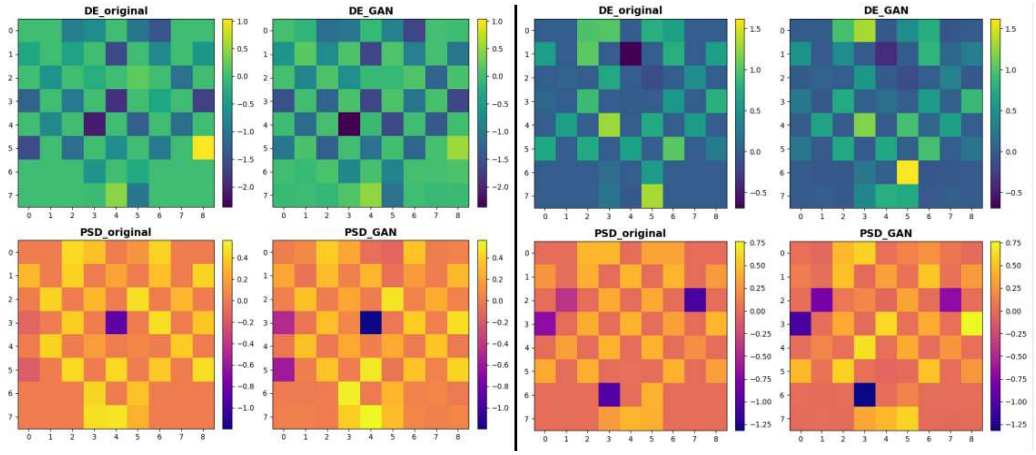
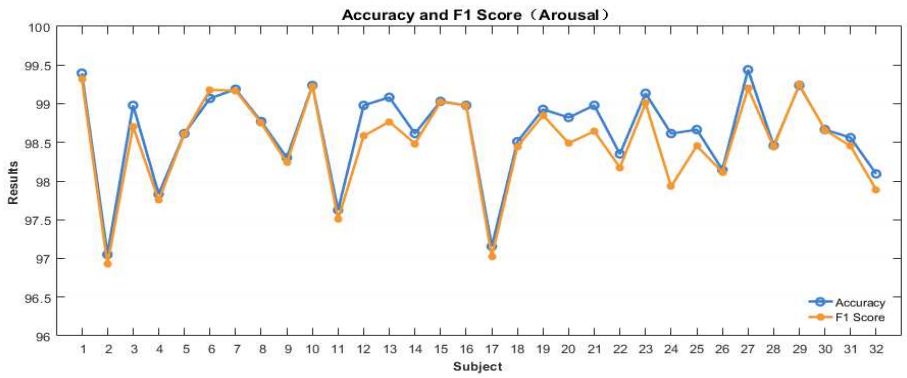


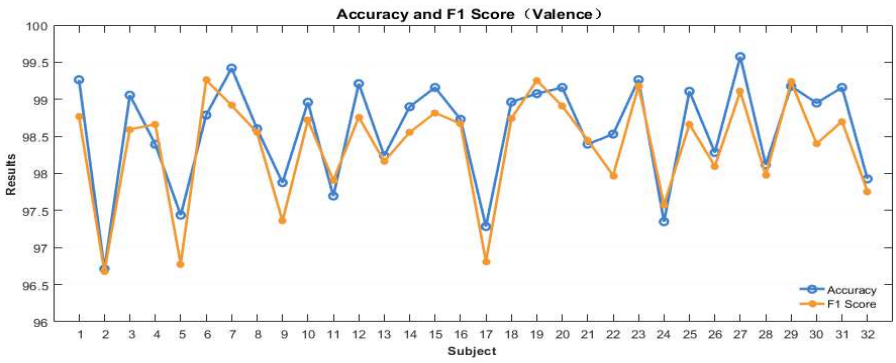
Fig. 3 Feature maps before and after data augmentation

4.2.2 Overall Performance

To comprehensively evaluate the model's overall performance in the EEG emotion recognition task, this paper adopts accuracy and F1-score as the primary evaluation metrics. Accuracy measures the proportion of correct predictions across all test samples and provides an intuitive reflection of the model's overall classification capability; while the F1-score comprehensively considers both precision and recall, allowing for a more objective evaluation of the model's discriminative performance across different categories when sample distributions are imbalanced or class boundaries are ambiguous. The accuracy results are shown in Fig. 4, and the confusion matrix for the arousal dimension is shown in Fig. 5.



(a)



(b)

Fig. 4 Model performance results

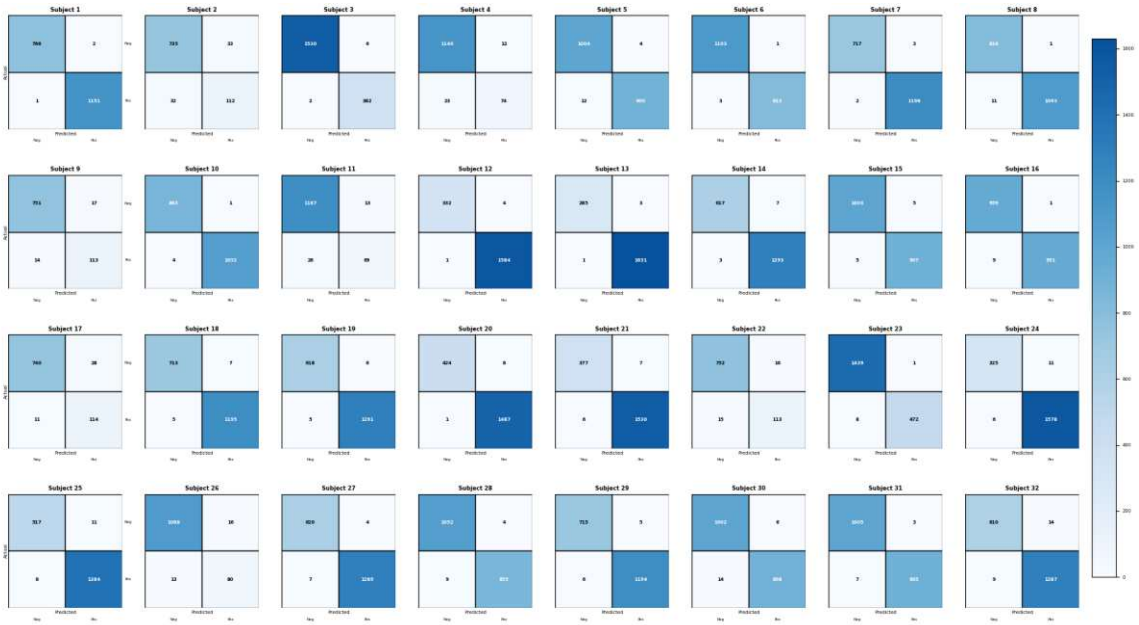


Fig. 5 Confusion matrix for the arousal dimension

The experimental results demonstrate that the method proposed in this paper achieves excellent performance on both the arousal (Fig. 4.a) and valence (Fig. 4.b) emotional dimensions. In the arousal recognition task, the model achieved an average classification accuracy of 98.64% with a corresponding F1 score of 98.51%; in the valence recognition task, the average accuracy and F1 score reached 98.59% and 98.37%, respectively. These results indicate that the proposed method can effectively learn discriminative spatial-frequency features in EEG signals and maintain stable recognition performance across different emotional dimensions. Overall, performance on the arousal dimension is slightly higher than on the valence dimension, which is consistent with existing research concluding that neural activity related to arousal is more pronounced and highly distinguishable.

Furthermore, the relationship between accuracy and F1 score reveals that the numerical differences between the two metrics across both emotional dimensions are kept within a narrow range. This indicates that the model exhibits good predictive balance across categories. This result validates that the proposed adversarial training and spatiotemporal feature modeling method can enhance overall recognition accuracy while maintaining good generalization ability and classification stability.

Although accuracy varies across different subjects, the method proposed in this paper outperforms the baseline methods for each subject, demonstrating strong cross-subject generalization capability.

4.2.3 Method Comparison

Based on comparative experimental data, this paper conducts a comprehensive evaluation and analysis of the overall performance of 10 EEG emotion recognition methods, focusing on the performance differences among models based on different technical approaches, as shown in Table 1.

Table 1 Accuracy of arousal and valence for the proposed method and other methods on the DEAP dataset

Methods	Arousal	Valence
SVM	84.84	84.54
CNN	87.03	87.14
DGCNN	88.93	89.59
WGANDA	66.85	67.99
CWGAN-T(BN)	93.52	92.75
GANSER	93.52	94.21
NAS	97.94	97.82
SSTNAS	63.52	63.05
AutoEER	94.65	94.77
Ours	98.64	98.59

Experimental results indicate significant performance differences among EEG-based emotion recognition methods using different technical approaches. Traditional machine learning methods, represented by Support Vector Machines (SVM), achieve an identification accuracy of approximately 84%. Although these models feature simple structures and low computational overhead, their overall performance is inferior to deep learning methods due to limitations in their ability to capture features manually. Deep learning models demonstrate stronger feature learning capabilities. Among them, Convolutional Neural Networks (CNNs) and Dynamic Graph Convolutional Neural Networks (DGCNNs) achieved recognition accuracies of approximately 87% and 89%, respectively, indicating their advantage in capturing the spatiotemporal features of EEG signals. Generative adversarial network (GAN)-based methods exhibit significant performance variations: WGANDA and SSTNAS achieve accuracy rates of approximately 63% – 67%, whereas structurally optimized models such as CWGAN-T(BN) and GANSER can reach over 93%, highlighting the critical impact of network architecture design on generative model performance. Methods based on neural architecture search further enhance recognition performance, achieving accuracy rates of 94% – 97%. In contrast, the method proposed in this paper achieves an accuracy of approximately 88% on both the arousal and valence dimensions, significantly outperforming existing methods and demonstrating stronger emotional feature extraction capabilities and classification performance.

4.2.4 Ablation Studies

To gain a deeper understanding of the mechanisms and contributions of each module in the method described in this paper, a series of ablation experiments were conducted to evaluate the effects of feature extraction, masked data augmentation, and neural architecture

search, respectively. The results are shown in Fig. 6.

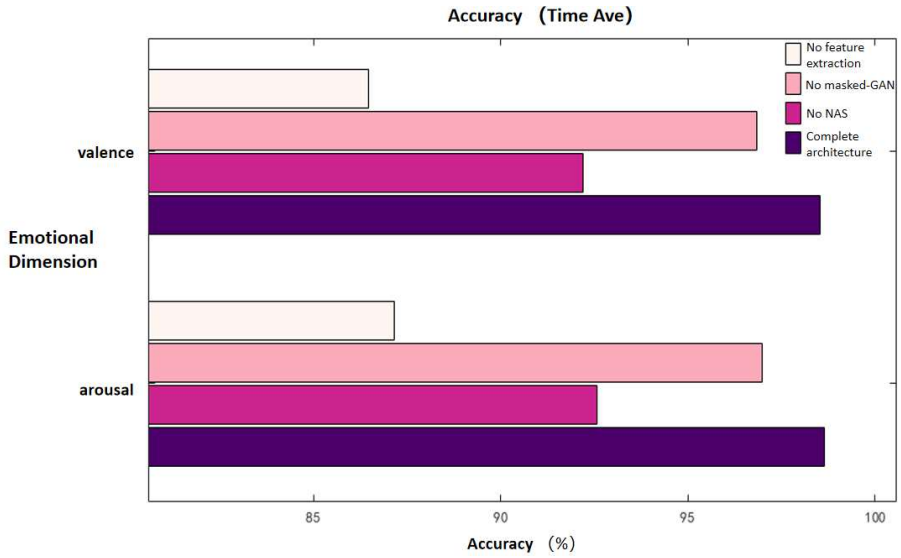


Fig. 6 Results of ablation experiments

Under conditions without feature extraction, the classifier was trained using only raw EEG features, achieving accuracy rates of 87.16% and 86.46% for arousal and valence, respectively. These results are significantly lower than those of the full architecture, indicating that relying solely on raw signals makes it difficult to fully capture subtle changes in frequency-domain features, thereby highlighting the necessity of feature enhancement. Under conditions without masked adversarial training, the model did not generate augmented samples and relied solely on the original data, achieving accuracy rates of 96.98% and 96.84%. While these results were better than those without feature extraction, they remained lower than those of the full architecture, proving that data augmentation plays a critical role in expanding the training sample size and enriching feature information. Under conditions without NAS, although the full feature generation and masking enhancement were employed, the accuracy dropped to 92.57% and 92.18% due to the fixed network structure, indicating that NAS can automatically optimize network topology, improve feature utilization efficiency, and adapt to individual differences.

5. Conclusion

This paper addresses three core challenges in EEG-based emotion recognition—data scarcity, network architecture design, and individual differences—by proposing a comprehensive solution that combines data augmentation via masked adversarial training with neural architecture search. Extensive experiments on the public DEAP dataset validate the effectiveness of our method.

Of course, this study also has some limitations. First, although personalized architecture search can significantly improve performance, it incurs high computational costs, as each subject requires independent search and training, which limits its feasibility for large-scale applications. Second, the current method primarily focuses on intra-subject emotion recognition tasks, and there is still room for improvement in its cross-subject generalization capabilities. Although the personalized strategy achieves excellent performance within a single subject, personalized search cannot be performed in cross-subject scenarios due to the

lack of labeled data for the target subjects. Third, the current method employs discrete emotion classification models and does not account for the continuity and dynamics of emotions. Fourth, the current method is primarily based on single-modality EEG signals; future research could explore multimodal fusion by integrating other physiological signals to build a more robust multimodal emotion recognition system.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J. Chen, Y. Cui, C. Wei, K. Polat, F. Alenezi, Advances in EEG-based emotion recognition: challenges, methodologies, and future directions, Applied Soft Computing 180 (2025) 113478.

[2] S. K. Mudgal, S. K. Sharma, J. Chaturvedi, A. Sharma, Brain computer interface advancement in neurosciences: applications and issues, Interdisciplinary Neurosurgery 20 (2020) 100694.

[3] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, S. D. Pollak, Emotional expressions reconsidered: challenges to inferring emotion from human facial movements, Psychological Science in the Public Interest 20 (2019) 1–68.

[4] X.-L. Quan, Z.-G. Zeng, J.-H. Jiang, Y.-Q. Zhang, B.-L. Lv, D.-R. Wu, Physiological signals based affective computing: a systematic review, Acta Automatica Sinica 47 (2024) 1769–1784.

[5] X. Li, et al., EEG based emotion recognition: a tutorial and review, ACM Computing Surveys 55 (2022) 1–57.

[6] C. Mühl, B. Allison, A. Nijholt, G. Chanel, A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges, Brain-Computer Interfaces 1 (2014) 66–84.

[7] S. Makeig, K. Gramann, T.-P. Jung, T. J. Sejnowski, H. Poizner, Linking brain, mind and behavior, International Journal of Psychophysiology 73 (2009) 95–100.

[8] G. Zhao, et al., Emotion analysis for personality inference from EEG signals, IEEE Transactions on Affective Computing 9 (2018) 362–371.

[9] A. Craik, Y. He, J. L. Contreras-Vidal, Deep learning for electroencephalogram (EEG) classification tasks: a review, Journal of Neural Engineering 16 (2019) 031001.

[10] J. Pan, W. Fang, Z. Zhang, B. Chen, Z. Zhang, Multimodal emotion recognition based on facial expressions, speech, and EEG, IEEE Open Journal of Engineering in Medicine and Biology 5 (2023) 396–403.

[11] C. Li, Z. Zhang, R. Song, J. Cheng, Y. Liu, X. Chen, EEG-based emotion recognition via neural architecture search, IEEE Transactions on Affective Computing 14 (2023) 957–968.

[12] R. Li, L. Wang, P. N. Suganthan, O. Sourina, Sample-based data augmentation based on electroencephalogram intrinsic characteristics, IEEE Journal of Biomedical and Health Informatics 26 (2022) 4996–5003.

[13] R. Wardoyo, et al., Oversampling approach using radius-smote for imbalance electroencephalography datasets, Emerging Science Journal 6 (2022) 382–398.

[14] C. Liao, S. Zhao, X. Wang, J. Zhang, Y. Liao, X. Wu, EEG data augmentation method based on the gaussian mixture model, Mathematics 13 (2025) 729.

[15] Y. Luo, S.-Y. Zhang, W.-L. Zheng, B.-L. Lu, WGAN domain adaptation for EEG-based emotion recognition, in: Neural Information Processing, Springer, Cham, 2018: pp. 275–286.
[16] Z. Zhang, Y. Liu, S. Zhong, Ganser: a self-supervised data augmentation framework for EEG-based emotion recognition, IEEE Transactions on Affective Computing 14 (2023) 2048–2063.
[17] J. Tang, et al., SDC-Net: a domain adaptation framework with semantic-dynamic consistency for cross-subject EEG emotion recognition, arXiv:2507.17524, 2025.
[18] C. Li, Z. Zhang, X. Zhang, G. Huang, Y. Liu, X. Chen, EEG-based emotion recognition via transformer neural architecture search, IEEE Transactions on Industrial Informatics 19 (2023) 6016–6025.
[19] W. Li, Z. Zhu, S. Shao, Y. Lu, A. Song, Spiking spatiotemporal neural architecture search for EEG-based emotion recognition, IEEE Transactions on Instrumentation and Measurement 74 (2025) 1–14.
[20] Y. Wu, et al., AutoEER: automatic EEG-based emotion recognition with neural architecture search, Journal of Neural Engineering 20 (2023).
[21] A.-C. Conneau, S. Essid, Assessment of new spectral features for EEG-based emotion recognition, in: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2014: pp. 4698–4702.
[22] J. Wang, M. Wang, Review of the emotional feature extraction and classification using EEG signals, Cognitive Robotics 1 (2021) 29–40.
[23] R.-N. Duan, J.-Y. Zhu, B.-L. Lu, Differential entropy feature for EEG-based emotion classification, in: 2013 6th International IEEE/EMBS Conference on Neural Engineering, IEEE, 2013: pp. 81–84.
[24] S. Hwang, K. Hong, G. Son, H. Byun, Learning CNN features from DE features for EEG-based emotion recognition, Pattern Analysis and Applications 23 (2020) 1323–1335.
[25] R. Jenke, A. Peer, M. Buss, Feature extraction and selection for emotion recognition from EEG, IEEE Transactions on Affective Computing 5 (2014) 327–339.
[26] X. Wang, Y. Ren, Z. Luo, W. He, J. Hong, Y. Huang, Deep learning-based EEG emotion recognition: current trends and future perspectives, Frontiers in Psychology 14 (2023) 1126994.
[27] J. N. Acharya, A. Hani, J. Cheek, P. Thirumala, T. N. Tsuchida, American clinical neurophysiology society guideline 2: guidelines for standard electrode position nomenclature, Journal of Clinical Neurophysiology 33 (2016) 308–311.
[28] M. Li, M. Qiu, W. Kong, L. Zhu, Y. Ding, Fusion graph representation of EEG for emotion recognition, Sensors 23 (2023) 1404.
[29] O. Ronneberger, P. Fischerr, and T. Brox, U-Net: convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer Assisted Intervention, Springer, Cham, 2015: pp. 234–241.
[30] H. Liu, K. Simonyan, Y. Yang, Darts: differentiable architecture search, arXiv:180 6.09055, 2018.
[31] S. Koelstra et al., DEAP: a database for emotion analysis using physiological signals, IEEE Transactions on Affective Computing 3 (2012) 18–31.
[32] J. A. Urigüen, B. Garcia-Zapirain, EEG artifact removal-state-of-the-art and guidelines, Journal of Neural Engineering 12 (2015) 031001.