SpikeEEGformer: A Spike-Driven Transformer for Generalized EEG Recognition

Jiacheng Luo[1], Shuifa Sun[1✉], Qin Tao[1], Wenchao Cui[2], and Ben Wang[1✉]

[1] School of Information Science and Technology, Hangzhou Normal University, Hangzhou, 311121, China
[2] Hubei Key Laboratory of Intelligent Vision Based Monitoring for Hydroelectric Engineering, China Three Gorges University, Yichang, 443002, Hubei, China
{watersun, 20170056}@hznu.edu.cn

Abstract. Spiking Neural Networks (SNNs) enable energy-efficient, interpretable EEG recognition via sparse binary spikes. However, extracting discriminative features through spike-driven computations remains a bottleneck. Existing methods often resort to hybrid ANN-SNN architectures relying on floating-point matrix multiplications, which compromises SNN efficiency and obscures the transparency of spiking dynamics. To address this, we propose SpikeEEGformer, a spike-driven Transformer engineered to strengthen feature extraction from spike sequences. It integrates a Spike EEG Self-Attention for synergistic channel-temporal modeling and a Spatio-Temporal Spike Block to bridge adjacent spiking dependencies. Furthermore, we introduce Latent Token Spike Attention, a multi-timestep spike-friendly mechanism that stabilizes representation learning and improves generalization. Evaluated on five benchmarks, SpikeEEGformer consistently achieves state-of-the-art accuracy with competitive energy efficiency. Critically, analysis of spiking dynamics provides intrinsic interpretability by elucidating the decision-making process via distinctive firing behaviors, while saliency maps effectively identify the critical decision-driving channels.

Keywords: Electroencephalography (EEG) · Spiking Neural Networks (SNNs) · Interpretability.

1 Introduction

Electroencephalography (EEG) recognition serves as a pivotal non-invasive link for translating neural dynamics into diverse applications, including brain-computer interfaces (BCI), cognitive assessment, and clinical monitoring. Despite its significance, conventional Artificial Neural Networks (ANNs) face substantial challenges due to high energy consumption from dense floating-point matrix multiplications (MACs) and an opaque "black-box" decision process [3,23,25]. Spiking Neural Networks (SNNs) offer a neuromorphic solution via event-driven binary spikes [15]. By replacing MACs with low-power additions, SNNs achieve superior energy efficiency. Crucially, SNNs facilitate inherent interpretability: unlike the opaque weight-based logic of ANNs, SNN decisions can be directly elucidated through distinctive spike firing behaviors, providing a transparent view of how the model discriminates between different EEG categories [10].

However, existing SNN-based EEG recognition methods encounter a critical bottleneck in effectively extracting discriminative features from binary spike trains through matrix additive computations. Consequently, many frameworks resort to hybrid ANN-SNN architectures that fall back on floating-point matrix multiplications for signal propagation [2,4,13,19]; this reliance not only deviates from the high-efficiency computational paradigm of SNNs but also obscures the transparency and interpretability of the spike-driven process. Furthermore, while Transformer-based SNNs [21,27,28] have emerged, they often struggle with EEG data owing to their sensitivity to temporal noise, failure to capture simultaneous channel-temporal dependencies, and the destabilizing effect of traditional regularization, such as Dropout, on sparse spiking dynamics.

To address these challenges, we propose SpikeEEGformer, a spike-driven Transformer designed to strengthen feature extraction from spike sequences, thereby elevating EEG recognition performance without sacrificing the intrinsic interpretability of spiking dynamics. Our contributions are:

- To capture synergistic inter-channel correlations and temporal dependencies in EEG signals, we develop the Spike EEG Self-Attention (SESA) mechanism while strictly adhering to spike-driven properties to eliminate floating-point matrix multiplications.
- The Spatio-Temporal Spike Block (STSB) is introduced as a spiking computational unit that integrates spatial feature fusion with dynamic temporal modeling, effectively bridging adjacent spiking states to mitigate information decay.
- We propose the Latent Token Spike Attention (LTSA) as a multi-timestep spike-friendly mechanism that enhances model generalization by reducing data dependency while preserving the inherent temporal sequentiality of spiking sequences.
- Evaluation across five benchmarks shows SpikeEEGformer achieves state-of-the-art accuracy and provides interpretability through spike firing analysis and channel-wise saliency maps, elucidating how EEG categories modulate internal spiking behaviors and identifying critical decision-driving channels while maintaining a competitive energy profile.

2 Method

2.1 Preliminary

Spiking Self-Attention (SSA) [28] adapts the Transformer mechanism to SNNs by replacing the softmax and floating-point matrix multiplications with spike-driven operations. For an input $\mathbf{X}$, SSA generates binary query ($\mathbf{Q}$), key ($\mathbf{K}$), and value ($\mathbf{V}$) representations and computes the output as:

$$\mathbf{I} = \mathrm{SN}_I\left(\mathrm{BN}\left(\mathbf{X}\mathbf{W}_I\right)\right), \quad \mathbf{I} \in \{\mathbf{Q}, \mathbf{K}, \mathbf{V}\}, \tag{1}$$

$$\mathrm{SSA}\left(\mathbf{Q},\mathbf{K},\mathbf{V}\right)=\mathrm{SN}(\mathbf{Q}\mathbf{K}^{\mathrm{T}}\mathbf{V}\cdot s), \tag{2}$$

where $\mathrm{SN}(\cdot)$ is the spiking neuron function [15], $\mathbf{W}_I$ are weights, and s is a scaling factor. This spike-driven paradigm forms the basis for our SpikeEEGformer.

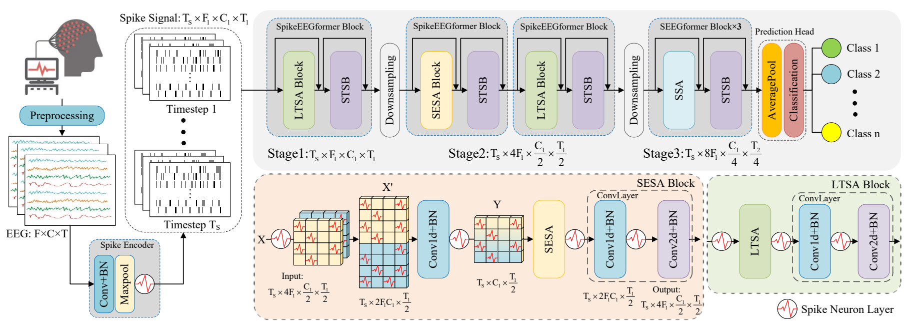

Fig. 1. The overview of SpikeEEGformer.

2.2 Overall Architecture

SpikeEEGformer (Fig. 1) processes input EEG signals $\mathbf{I} \in \mathbb{R}^{F\times C\times T}$, where F, C, and T denote the feature dimension, channel count, and temporal sampling points, respectively. A Spike Encoder first converts the input into a discrete representation in $\{0,1\}^{T_S\times F_1\times C_1\times T_1}$, where T_1 represents the compressed EEG temporal dimension, and T_S denotes the spiking simulation time. A three-stage hierarchical architecture then progressively downsamples spatio-temporal resolutions while increasing feature depth, enabling multi-scale abstraction of neural dynamics. Each stage integrates attention mechanisms with the Spatio-Temporal Spike Block (STSB). Latent Token Spike Attention (LTSA) is adopted in Stage 1, a hybrid of Spike EEG Self-Attention (SESA) and LTSA in Stage 2, while standard SSA [28] is employed in Stage 3 due to the reduced spatio-temporal resolution. The Prediction Head yields final predictions.

2.3 Spike Encoder

To bridge continuous EEG and discrete spikes, a learnable Spike Encoder is employed to mitigate fixed coding loss [16,19]. Raw inputs $\mathbf{I} \in \mathbb{R}^{F\times C\times T}$ are first compressed via conv (kernels $\mathbf{W}$), BN, and max-pool ($\mathcal{P}$) into a $F_1 \times C_1 \times T_1$ representation. A spiking neuron layer ($\mathrm{SN}(\cdot)$) then encodes each floating-point value in this representation into a binary spike sequence over T_S simulation steps, while the EEG temporal dimension T_1 is preserved as a static structural dimension alongside F_1 and C_1, rather than the simulation axis—inspired by

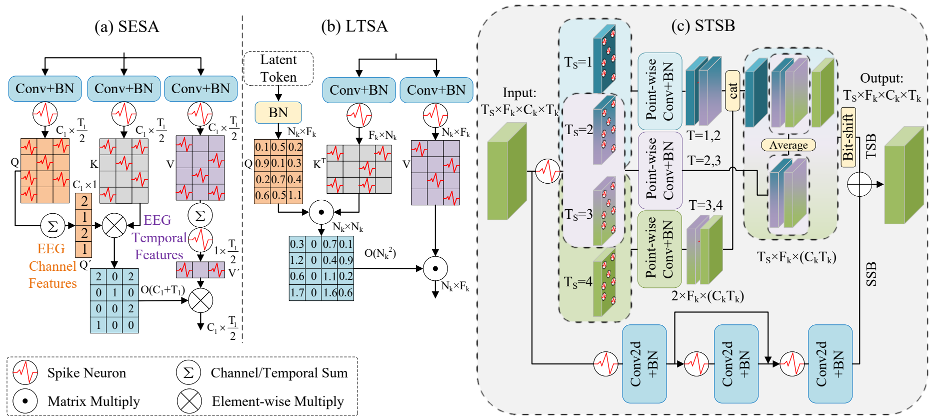

Fig. 2. (a) SESA: Captures channel-temporal features via distinct query-value pairs. (b) LTSA: Decouples query-input dependency via a learnable latent token. (c) STSB: Dual-branch architecture integrating TSB (illustrated with $T_S = 4$) and SSB.

long-sequence SNN-based approaches [8, 14]—to avoid gradient instability and amplitude information loss associated with excessively long spike sequences:

$$\mathbf{S} = \mathrm{SN}\left(\mathcal{P}(\mathrm{BN}(\mathbf{W} * \mathbf{I}))\right), \quad \mathbf{S} \in \{0,1\}^{T_S \times F_1 \times C_1 \times T_1}, \tag{3}$$

This yields a discrete spatio-temporal volume for the spike-driven backbone.

2.4 Spike EEG Self-Attention (SESA)

To effectively model the spatio-temporal coupling in EEG signals under spiking constraints, we propose Spike EEG Self-Attention (SESA).
Feature Remapping. Unlike standard attention that collapses EEG spatio-temporal axes to reduce dimensions, SESA preserves structural integrity for enhanced cross-channel learning (Fig. 1). Input $\mathbf{X} \in \mathbb{R}^{T_S \times 4F_1 \times \frac{C_1}{2} \times \frac{T_1}{2}}$ is discretized and reshaped into $\mathbf{X}' \in \{0,1\}^{T_S \times (2F_1 C_1) \times \frac{T_1}{2}}$ to maintain topological properties via dimensionality reduction. Subsequently, a spike-driven projection $\mathcal{P}_{SD}(\cdot)$ maps $\mathbf{X}'$ into a latent manifold, yielding binary $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ embeddings:

$$\mathbf{X}' = \mathrm{Reshape}(\mathrm{SN}(\mathbf{X})), \quad \mathbf{Q}, \mathbf{K}, \mathbf{V} = \mathcal{P}_{SD}(\mathbf{X}') \in \{0,1\}^{T_S \times C_1 \times \frac{T_1}{2}}. \tag{4}$$

Spatio-Temporal Feature Extraction. Inspired by Q–K Attention [27], the SESA mechanism models channel–temporal dependencies in spiking representations (Fig. 2(a)). Rather than forming full pairwise interactions, SESA decouples channel-wise and temporal-wise modeling by aggregating queries and values along complementary dimensions. Specifically, temporal aggregation of $\mathbf{Q}$ yields a channel descriptor $\mathbf{Q}' \in \mathbb{Z}^{T_S \times C_1 \times 1}$, while channel aggregation of $\mathbf{V}$ produces a temporally modulated representation $\mathbf{V}' \in \{0,1\}^{T_S \times 1 \times \frac{T_1}{2}}$. Leveraging binary $\mathbf{K}$, the Element-wise Multiply ($\otimes$) with integer $\mathbf{Q}'$ is reformulated as multiplier-free conditional accumulations, replacing expensive floating-point matrix products with efficient integer additions:

$$\mathbf{Q}' = \sum_t \mathbf{Q}_{t,c}, \quad \mathbf{V}' = \mathrm{SN}\left(\sum_c \mathbf{V}_{t,c}\right), \quad \mathrm{SESA}(\mathbf{Q}', \mathbf{K}, \mathbf{V}') = \mathrm{SN}((\mathbf{Q}' \otimes \mathbf{K}) \otimes \mathbf{V}')\,. \tag{5}$$

This achieves $O(C_1 + T_1)$ complexity via decoupled channel ($O(C_1)$) and temporal ($O(T_1)$) modulations, bypassing the quadratic bottleneck of full attention matrices. A final ConvLayer restores the feature manifold.

2.5 Spatio-Temporal Spike Block (STSB)

To mitigate information decay across discrete time steps in SNNs, we propose a dedicated Temporal Spike Block (TSB). As illustrated in Fig. 2, this TSB operates in parallel with a Spatial Spike Block (SSB) to form the dual-branch STSB, synergistically decoupling temporal and spatial feature extraction.

Temporal Dynamics Modeling. The TSB captures dependencies between adjacent spiking time steps T_S by processing overlapping segments. For an input $\mathbf{X}$ with n spiking steps, the sequence is split and reshaped into adjacent pairs $\{\mathbf{B}_{i,i+1} \in \{0,1\}^{2\times F_k \times (C_k T_k)}\}_{i=1}^{n-1}$. Temporal correlations are enhanced via spike-driven kernels W_{pw} applied along the T_S axis:

$$\{\mathbf{B}_{i,i+1}\}_{i=1}^{n-1} = \mathrm{Reshape}(\mathrm{Split}_{T_S}(\mathrm{SN}(\mathbf{X}))), \quad \mathbf{B}'_{i,i+1} = \mathrm{BN}(\mathbf{W}_{pw} * \mathbf{B}_{i,i+1}) \tag{6}$$

To ensure global temporal consistency, segments are merged and averaged at overlapping steps (e.g., step 2 in pairs $[1,2]$ and $[2,3]$), effectively smoothing inter-block discontinuities. This averaging is implemented via exponent-bit adjustment to maintain a multiplier-free flow.

Spatial Feature Extraction. The SSB branch targets spatial patterns using a residual hierarchy of spike-driven 1×1 and 3×3 kernels. This branch preserves electrode-level topological features while the TSB provides temporal context.

Structural Integration. The final STSB output fuses both branches, where the TSB contribution is scaled by 0.5 via a hardware-friendly right bit-shift operator ($\gg 1$) to ensure a computation flow free of floating-point matrix multiplications:

$$\mathrm{STSB}(\mathbf{X}) = (\mathrm{TSB}(\mathbf{X}) \gg 1) + \mathrm{SSB}(\mathbf{X}). \tag{7}$$

2.6 Latent Token Spike Attention (LTSA)

To stabilize sparse spiking dynamics where traditional Dropout fails in SNNs [12], we propose Latent Token Spike Attention (LTSA). Inspired by [9], LTSA decouples attention from input noise by deriving queries $\mathbf{Q}$ from a learnable latent token $\mathbf{L} \in \mathbb{R}^{N_k \times F_k}$ rather than input $\mathbf{X}$, significantly enhancing generalization. As shown in Fig. 2(b), $\mathbf{Q} = \mathrm{BN}(\mathbf{L})$ is broadcast across T_S to align with binary $\mathbf{K}$:

$$\mathrm{LTSA} = \mathrm{SN}(\mathbf{Q}\mathbf{K}^\top \mathbf{V}), \quad \mathbf{K}, \mathbf{V} \in \{0,1\}^{T_S \times N_k \times F_k}. \tag{8}$$

Binary $\mathbf{K}$ ensures $\mathbf{Q}\mathbf{K}^\top$ is free of floating-point matrix multiplications [7].

Table 1. Performance on HUSM dataset (Mean/Std). Params (M), Energy (mJ).

Type	Model	Param	Energy	Acc(%)	Sen(%)	F1(%)
ANN	Deprnet [20]	55.13	32.73	88.99/6.55	88.87/6.75	88.90/6.64
	C-Attn [25]	7.56	36.85	90.01/9.25	89.86/9.39	89.89/9.39
	Double ViT [17]	175.73	117.85	88.87/8.81	88.59/9.24	85.46/9.32
	TSception [6]	0.87	10.63	89.48/8.22	89.42/8.27	89.38/8.32
	MixNet [1]	6.71	55.46	86.24/6.37	85.94/6.71	85.88/6.82
	MSTAN [24]	59.84	4.80	85.62/7.95	85.27/8.37	85.25/8.40
SNN	Spikformer [28]	65.60	34.29	89.80/6.77	90.00/6.70	90.00/6.77
	QKFormer [27]	131.36	40.01	89.66/7.13	89.66/7.06	89.62/7.14
	SpikingResformer [21]	59.35	50.05	91.12/7.73	91.01/7.92	91.02/7.89
	SpikeEEGformer-S	15.90	15.11	93.12/7.11	93.19/6.98	93.11/7.12
	SpikeEEGformer-L	61.18	40.26	**93.89**/6.62	**93.96**/6.46	**93.88**/6.62

Table 2. Performance on the MODMA and PRED+CT datasets (Mean/Std).

Model	MODMA			PRED+CT		
	Acc(%)	Sen(%)	F1(%)	Acc(%)	Sen(%)	F1(%)
EEGNet [11]	72.22/3.45	70.60/3.56	69.80/3.91	74.54/2.90	63.46/4.76	61.98/3.93
Deprnet [20]	59.98/6.55	58.44/3.60	56.50/4.95	72.96/2.47	60.17/4.56	57.56/3.64
C-LSTM [18]	68.12/5.78	65.62/5.53	63.90/6.63	74.18/3.29	54.50/2.88	53.50/2.72
C-Attn [25]	71.10/3.56	67.43/6.38	65.36/6.62	76.43/3.91	59.36/6.34	58.58/6.68
TSception [6]	72.58/5.69	71.53/5.12	71.07/5.08	73.83/3.64	56.83/2.57	55.88/3.96
MixNet [1]	70.61/6.46	69.02/7.27	66.76/9.63	76.15/2.57	60.43/6.43	59.63/5.99
MSTAN [24]	59.15/3.08	57.36/2.22	56.54/2.72	72.87/4.77	53.37/1.71	51.88/0.89
SpikeEEG former-L	**80.64**/5.60	**80.11**/5.49	**79.93**/5.71	**79.32**/3.17	**68.69**/4.54	**65.91**/3.77

3 Experiment

3.1 Datasets

SpikeEEGformer is benchmarked on five datasets across three clinical paradigms: (i) Psychiatric Diagnosis (HUSM, $n = 64$; MODMA, $n = 53$; PRED+CT, $n = 118$), (ii) Emotion Recognition (DEAP, $n = 32$; binary valence), and (iii) Motor Imagery (MI) (PhysioNet, $n = 105$; 4-class: L/R/both fists, feet). Performance is validated via subject-independent k-fold cross-validation ($k = 7$ for diagnosis; $k = 5$ for emotion) to assess generalization. For MI, a subject-dependent 7-fold strategy is employed to accommodate highly individualized neural signatures.

3.2 Experiment Setting

Two model variants are evaluated: SpikeEEGformer-L ($F_1 = 128$) and Spike-EEGformer-S ($F_1 = 64$), both with a fixed simulation resolution of $T_S = 4$.

Table 3. DEAP and PhysioNet MI results (Mean/Std). "–" denotes not applicable.

Model	DEAP			PhysioNet MI		
	Acc(%)	Sen(%)	F1(%)	Acc(%)	Sen(%)	F1(%)
EEGNet [11]	62.75/4.50	54.05/1.53	52.16/1.18	62.15/1.42	61.86/1.68	62.08/1.66
C-LSTM [18]	65.71/4.98	53.22/1.57	52.46/1.69	52.25/7.20	52.25/7.20	52.42/7.18
C-Attn [25]	65.12/4.88	56.43/3.33	55.10/2.48	59.08/1.18	59.08/1.18	59.26/1.34
TSception [6]	67.80/4.22	57.55/2.72	56.04/2.11	54.69/1.14	54.69/1.14	54.68/1.02
MixNet [1]	63.62/5.32	53.49/1.89	51.96/0.92	61.47/1.45	61.47/1.45	61.53/1.47
BISNN [22]	62.91/3.86	54.00/2.44	52.87/2.43	–	–	–
MI SNN [26]	–	–	–	50.00/1.12	50.00/1.12	49.88/1.10
SpikeEEG former-L	**71.40**/4.05	**57.79**/2.65	**57.26**/2.11	**65.13**/0.84	**65.13**/0.85	**65.14**/0.84

Spiking dynamics are governed by the LIF model, using the ATan surrogate gradient to enable backpropagation. The framework is trained for 120 epochs via AdamW (LR $= 5 \times 10^{-4}$, batch size 16) to minimize the MSE loss.

3.3 Experimental Results and Analysis

Tables 1–3 demonstrate that both SpikeEEGformer-L and -S achieve highly competitive performance ($p < 0.01$, paired two-tailed t-test). For fair comparison, all baselines were re-trained from scratch on identical data splits and preprocessing. While the L-variant provides peak accuracy, the S-variant serves as a high-efficiency alternative. Specifically, on the HUSM dataset (Table 1), SpikeEEGformer-S reduces energy consumption by 59.0% and 69.8% compared to the premier attention-based ANN [25] and SNN [21], respectively, while yielding modest yet consistent average metric improvements of 3.22% and 2.09%. Furthermore, when compared to the highly lightweight TSception [6], SpikeEEGformer-S incurs a slight energy overhead of 5 mJ. However, it achieves a 3.75% accuracy gain, presenting an acceptable trade-off when precise diagnosis is prioritized over absolute minimal power. Following established literature, theoretical energy is quantified as 4.6 pJ $\times$ MAC for ANNs [27], where MAC denotes floating-point multiply-accumulate operations, and 0.9 pJ $\times$ SOP for SNN components, with SOP representing event-driven, addition-only synaptic operations triggered by spikes. This scalability underscores a favorable performance-to-power trade-off tailored for resource-constrained clinical deployment.

3.4 Ablation Study

Table 4 underscores the contribution of SpikeEEGformer's core components. Substituting SESA with standard SSA [28] incurs the most significant accuracy drop (3.36% and 3.06% on HUSM and PhysioNet MI), validating its efficacy in modeling EEG-specific spatio-temporal dynamics. Furthermore, removing the TSB branch or replacing LTSA with standard SSA paired with Dropout leads

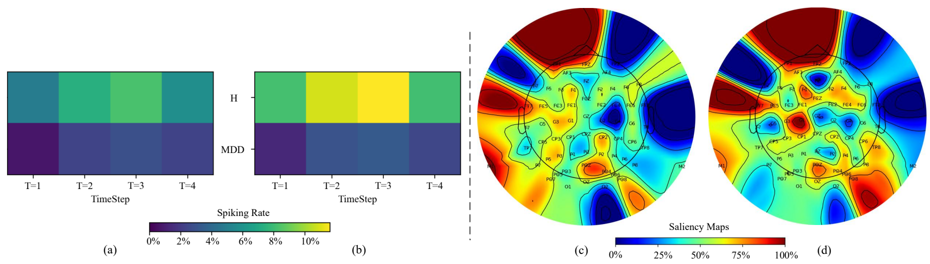

Fig. 3. Interpretability analysis via spiking dynamics and channel saliency. (a–b) Comparison of average firing rates in the final spiking attention and STSB for HC and MDD groups (HUSM). (c–d) Channel saliency maps for HC and MDD subjects (PRED+CT), where darker intensities indicate higher feature relevance for classification.

to consistent performance degradation (1.28%–2.47%), confirming their respective roles in mitigating intrinsic temporal information loss and enhancing cross-paradigm generalization.

3.5 Spike Firing Analysis

To analyze class-dependent spiking behavior on the HUSM dataset, average firing rates over $T_S = 4$ simulation steps were measured from the final attention layers and the final STSB module (Fig. 3(a,b)). For healthy control (HC) samples, the final attention layers and STSB exhibited average firing rates of 6.56% and 9.69%, respectively, whereas for MDD samples, the corresponding rates decreased to 2.06% and 2.52%. These results indicate that SpikeEEGformer forms separable spike-domain representations characterized by distinct activation sparsity across categories.

3.6 EEG Channel Analysis

Saliency maps, computed via surrogate gradient-based backpropagation, are used to analyze channel-wise contributions to classification on the PRED+CT dataset. Fig. 3(c,d) present class-specific saliency maps for healthy control (HC)

Table 4. Ablation study on HUSM and PhysioNet MI (Mean/Std).

Model	**HUSM**			**PhysioNet MI**		
	Acc(%)	**Sen(%)**	**F1(%)**	**Acc(%)**	**Sen(%)**	**F1(%)**
Full Model	**93.89**/6.62	**93.96**/6.46	**93.88**/6.62	**65.13**/0.84	**65.13**/0.85	**65.14**/0.84
SESA→SSA	90.53/6.93	90.41/7.06	90.45/7.03	62.07/1.29	62.07/1.29	61.96/1.35
w/o TSB	92.61/7.35	92.62/7.22	92.55/7.32	63.15/0.70	63.15/0.70	63.22/0.76
LTSA→SSA	92.22/7.20	92.18/7.28	92.18/7.26	62.66/1.45	62.65/1.46	62.49/1.69

and MDD samples, respectively, where darker regions indicate higher relevance. For SpikeEEGformer, frontal and central channels (e.g., FZ, FPZ, FP1/2, CZ, C3/4, FT7/8, F7/8) show consistently higher saliency, which aligns with clinical findings that these regions are particularly implicated in depression-related neural dysfunction [5].

4 Conclusion

This work presents SpikeEEGformer, a directly trained spiking Transformer for EEG classification. The SESA module models channel-temporal interactions, while the STSB preserves local spiking dependencies to reduce information loss in spike-based representations. The LTSA further enhances robustness and generalization across tasks. SpikeEEGformer achieves competitive performance and energy efficiency on multiple benchmarks, with improved interpretability via structured spike activation patterns. Future work will explore further reductions in energy consumption.

Disclosure of Interests The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Autthasan, P., Chaisaen, R., Phan, H., De Vos, M., Wilaiprasitporn, T.: MixNet: Joining force of classical and modern approaches toward the comprehensive pipeline in motor imagery eeg classification. IEEE Internet of Things Journal **11**(17), 28539–28554 (2024)
2. Banovoth, R.S., K V, K.: Roman domination-based spiking neural network for optimized eeg signal classification of four class motor imagery. Computers in Biology and Medicine **194**, 110397 (2025). https://doi.org/https://doi.org/10.1016/j.compbiomed.2025.110397, https://www.sciencedirect.com/science/article/pii/S0010482525007486
3. Cao, L., Zhao, W., Sun, B.: Emotion recognition using multi-scale eeg features through graph convolutional attention network. Neural Networks **184**, 107060 (2025)
4. Chen, X., Kong, Y., Chang, H., Gao, Y., Liu, Z., Coatrieux, J.L., Shu, H.: MGSN: Depression EEG lightweight detection based on multiscale DGCN and SNN for multichannel topology. Biomedical Signal Processing and Control **92**, 106051 (2024)
5. Cook, I.A., Hunter, A.M., Korb, A.S., Leuchter, A.F.: Do prefrontal midline electrodes provide unique neurophysiologic information in major depressive disorder? Journal of Psychiatric Research **53**, 69–75 (2014). https://doi.org/https://doi.org/10.1016/j.jpsychires.2014.01.018, https://www.sciencedirect.com/science/article/pii/S0022395614000272
6. Ding, Y., Robinson, N., Zhang, S., Zeng, Q., Guan, C.: Tsception: Capturing temporal dynamics and spatial asymmetry from eeg for emotion recognition. IEEE Transactions on Affective Computing **14**(3), 2238–2250 (2022)

7. Guo, Y., Liu, X., Chen, Y., Peng, W., Zhang, Y., Ma, Z.: Spiking Transformer: Introducing accurate addition-only spiking self-attention for Transformer. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 24398–24408 (2025)
8. Hu, B., Lv, C., Li, M., Liu, Y., Zheng, X., Zhang, F., Zhang, F., et al.: Spikestag: Spatial-temporal forecasting via gnn-snn collaboration. arXiv preprint arXiv:2508.02069 (2025)
9. Jaegle, A., Gimeno, F., Brock, A., Vinyals, O., Zisserman, A., Carreira, J.: Perceiver: General perception with iterative attention. In: International conference on machine learning. pp. 4651–4664. PMLR (2021)
10. Jia, Z., Ji, J., Zhou, X., Zhou, Y.: Hybrid spiking neural network for sleep electroencephalogram signals. Science China Information Sciences **65**(4), 140403 (2022)
11. Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., Lance, B.J.: EEGNet: a compact convolutional neural network for eeg-based brain–computer interfaces. Journal of neural engineering **15**(5), 056013 (2018)
12. Lee, C., Sarwar, S.S., Panda, P., Srinivasan, G., Roy, K.: Enabling spike-based backpropagation for training deep neural network architectures. Frontiers in neuroscience **14**, 497482 (2020)
13. Li, W., Zhu, Z., Shao, S., Lu, Y., Song, A.: Spiking spatiotemporal neural architecture search for eeg-based emotion recognition. IEEE Transactions on Instrumentation and Measurement **74**, 1–14 (2025). https://doi.org/10.1109/TIM.2024.3472838
14. Lv, C., Wang, Y., Han, D., Zheng, X., Huang, X., Li, D.: Efficient and effective time-series forecasting with spiking neural networks. In: Proceedings of the 41st International Conference on Machine Learning. ICML'24, JMLR.org (2024)
15. Maass, W.: Networks of spiking neurons: The third generation of neural network models. Neural networks **10**(9), 1659–1671 (1997)
16. Petro, B., Kasabov, N., Kiss, R.M.: Selection and optimization of temporal spike encoding methods for spiking neural networks. IEEE Transactions on Neural Networks and Learning Systems **31**(2), 358–370 (2020). https://doi.org/10.1109/TNNLS.2019.2906158
17. Qayyum, A., Razzak, I., Tanveer, M., Mazher, M., Alhaqbani, B.: High-density electroencephalography and speech signal based deep framework for clinical depression diagnosis. IEEE/ACM transactions on computational biology and bioinformatics **20**(4), 2587–2597 (2023)
18. Saeedi, A., Saeedi, M., Maghsoudi, A., Shalbaf, A.: Major depressive disorder diagnosis based on effective connectivity in EEG signals: A convolutional neural network and long short-term memory approach. Cognitive neurodynamics **15**(2), 239–252 (2021)
19. Sam, A., Boostani, R., Hashempour, S., Taghavi, M., Sanei, S.: Depression identification using EEG signals via a hybrid of LSTM and spiking neural networks. IEEE Transactions on Neural Systems and Rehabilitation Engineering **31**, 4725–4737 (2023)
20. Seal, A., Bajpai, R., Agnihotri, J., Yazidi, A., Herrera-Viedma, E., Krejcar, O.: Deprnet: A deep convolution neural network framework for detecting depression using EEG. IEEE Transactions on Instrumentation and Measurement **70**, 1–13 (2021)
21. Shi, X., Hao, Z., Yu, Z.: Spikingresformer: Bridging resnet and vision transformer in spiking neural networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5610–5619 (2024)

22. Sun, H., Mao, S., Cai, W., Cui, Y., Chen, D., Yao, D., Guo, D.: Bisnn: bio-information-fused spiking neural networks for enhanced eeg-based emotion recognition. Cognitive Neurodynamics **19**(1), 1–12 (2025)
23. Wang, Y., Zhao, S., Jiang, H., Li, S., Luo, B., Li, T., Pan, G.: Diffmdd: A diffusion-based deep learning framework for mdd diagnosis using eeg. IEEE Transactions on Neural Systems and Rehabilitation Engineering **32**, 728–738 (2024)
24. Xi, Y., Chen, Y., Meng, T., Lan, Z., Zhang, L.: Depression detection based on the temporal-spatial-frequency feature fusion of eeg. Biomedical Signal Processing and Control **100**, 106930 (2025)
25. Xia, M., Zhang, Y., Wu, Y., Wang, X.: An end-to-end deep learning model for EEG-based major depressive disorder classification. IEEE Access **11**, 41337–41347 (2023)
26. Zhang, H., Wang, H., An, J., Zheng, S., Wu, D.: A lightweight spiking neural network for eeg-based motor imagery classification. Neural Networks p. 107741 (2025)
27. Zhou, C., Zhang, H., Zhou, Z., Yu, L., Huang, L., Fan, X., Yuan, L., Ma, Z., Zhou, H., Tian, Y.: QKFormer: Hierarchical spiking Transformer using Q-K attention. In: The Thirty-eighth Annual Conference on Neural Information Processing Systems (2024), https://openreview.net/forum?id=AVd7DpiooC
28. Zhou, Z., Zhu, Y., He, C., Wang, Y., YAN, S., Tian, Y., Yuan, L.: Spikformer: When spiking neural network meets Transformer. In: The Eleventh International Conference on Learning Representations (2023)