



## OPEN ACCESS

## EDITED BY

Gaetano Di Caterina,  
University of Strathclyde, United Kingdom

## REVIEWED BY

Anguo Zhang,  
Fuzhou University, China  
Zihan Pan,  
Institute for Infocomm Research (A\*STAR),  
Singapore

## \*CORRESPONDENCE

Ying Liu  
✉ hello1668@163.com

RECEIVED 23 June 2025

ACCEPTED 17 September 2025

PUBLISHED 16 October 2025

## CITATION

Yuan L, Wei J and Liu Y (2025) Spiking neural  
networks for EEG signal analysis using  
wavelet transform.  
*Front. Neurosci.* 19:1652274.  
doi: 10.3389/fnins.2025.1652274

## COPYRIGHT

© 2025 Yuan, Wei and Liu. This is an  
open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](#). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic practice.  
No use, distribution or reproduction is  
permitted which does not comply with these  
terms.

# Spiking neural networks for EEG signal analysis using wavelet transform

Li Yuan, Jian Wei and Ying Liu\*

Academy of Military Sciences, Beijing, China

**Introduction:** Brain-computer interfaces (BCIs) leverage EEG signal processing to enable human-machine communication and have broad application potential. However, existing deep learning-based BCI methods face two critical limitations that hinder their practical deployment: reliance on manual EEG feature extraction, which constrains their ability to adaptively capture complex neural patterns, and high energy consumption characteristics that make them unsuitable for resource-constrained portable BCI devices requiring edge deployment.

**Methods:** To address these limitations, this work combines wavelet transform for automatic feature extraction with spiking neural networks for energy-efficient computation. Specifically, we present a novel spiking transformer that integrates a spiking self-attention mechanism with discrete wavelet transform, termed SpikeWavformer. SpikeWavformer enables automatic EEG signal time-frequency decomposition, eliminates manual feature extraction, and provides energy-efficient classification decision-making, thereby enhancing the model's cross-scene generalization while meeting the constraints of portable BCI applications.

**Results:** Experimental results demonstrate the effectiveness and efficiency of SpikeWavformer in emotion recognition and auditory attention decoding tasks.

**Discussion:** These findings indicate that SpikeWavformer can address the key limitations of existing BCI methods and holds promise for practical deployment in portable, resource-constrained scenarios.

## KEYWORDS

spiking neural networks, EEG signal analysis, brain-computer interfaces, discrete wavelet transform, bio-inspired methods

## 1 Introduction

Brain-computer interfaces (BCIs) enable direct communication between the human brain and machines through electroencephalography (EEG) signal processing (Zhang et al., 2020). A typical BCI architecture comprises four functional modules: data acquisition, preprocessing, classification, and a feedback module (Lotte and Guan, 2010). BCI systems have demonstrated extensive real-world applicability in diverse domains including robotic manipulation (Liu et al., 2015), cognitive signal decoding (Cai et al., 2021), and neuropsychiatric interventions for emotional regulation (Zotev et al., 2020; Xing et al., 2019). As a common learning-based BCI method, deep learning methodology has demonstrated superior performance over conventional machine learning approaches across diverse BCI tasks (Ang et al., 2008; Wang et al., 2015), including motor imagery classification (Schirrmeister et al., 2017; Kwon et al., 2019), mental workload monitoring (Jiao et al., 2018), auditory attention decoding (Faghihi et al., 2022; Cai et al., 2024), and emotion recognition (Alarcao and Fonseca, 2017; Li et al., 2018). Nevertheless, previous research has predominantly relied on manually extracted EEG features such as power spectral density (PSD) and differential entropy (DE) (Jiao et al., 2018; Song et al., 2018; Zhong et al., 2020), whose limitations

become increasingly evident. First, these feature extraction paradigms exhibit strong dependence on domain-specific knowledge (Singh and Krishnan, 2023; Subasi, 2019), necessitating task-specific extraction pipelines tailored to distinct experimental protocols, thereby compromising model generalizability across tasks. Second, manually crafted features often fail to capture nonlinear interrelationships in EEG time-frequency characteristics and multiscale dynamic properties (Singh and Krishnan, 2023; Vallabhaneni et al., 2021), potentially leading to critical information loss.

Wavelet Transform (WT) has emerged as a fundamental signal processing tool in EEG analysis (Grobelaar et al., 2022) due to its unique time-frequency analysis capabilities. Unlike conventional Fourier Transform that provides only global frequency-domain information, WT enables multi-scale decomposition through its inherent multi-resolution analysis. This capability permits simultaneous signal characterization at distinct resolution levels—capturing macroscopic patterns (e.g., global trends) at coarse-grained scales while resolving microscopic fluctuations (e.g., localized variations) at fine-grained scales when processing electroencephalographic (EEG) signals. Furthermore, WT achieves adaptive hierarchical representation of non-stationary neural activities by dynamically adjusting the scale and translation parameters of basis functions, thereby effectively characterizing both transient features (e.g., high-frequency oscillations in event-related potentials) and long-range rhythmic patterns (e.g., sustained  $\alpha$ -wave oscillations). Although recent years have witnessed preliminary applications of wavelet transform methodologies in EEG classification tasks. However, their predominant reliance on deep neural networks (DNNs) introduces computationally and resource-intensive demands, conflicting with the low-power objectives of resource-constrained portable BCI devices. Consequently, achieving optimal trade-offs between classification performance, system portability, and energy efficiency remains a critical challenge in practical BCI implementations.

Spiking neural networks (SNNs), recognized as third-generation neural networks, have emerged as a promising alternative in BCI research due to their biologically plausible computation paradigm (Izhikevich, 2003; Maass, 1997; Masquelier et al., 2008). As shown in Figure 1, instead of continuous activations in deep neural networks (DNNs), SNNs employ discrete spike events as neuronal communication media, where spiking neurons activate exclusively upon reaching threshold potentials and remain quiescent otherwise (Gerstner and Kistler, 2002). This event-driven mechanism (Wei et al., 2024) facilitates synaptic computation sparsity while eliminating multiply-accumulate (MAC) operations, thereby achieving superior energy efficiency, which is critical for portable neurotechnological devices. Notably, SNNs have demonstrated remarkable success across multiple computational neuroscience domains in recent years. For instance, the energy-efficient Spike Transformer architectures proposed by Yao et al. (2023, 2024, 2025) and Zhou et al. (2022, 2023) have demonstrated exceptional performance in image classification (Deng et al., 2022; Shi et al., 2024), detection (Luo et al., 2024; Wang et al., 2025), and segmentation (Lei et al., 2025). Similarly, the SNN-based audio processing models developed by Wu et al. (2018); Pan et al. (2020); Wang et al. (2024) have made significant advancements in signal

processing and keyword recognition. These successes establish a solid foundation for the broader adoption and cross-domain application of SNNs.

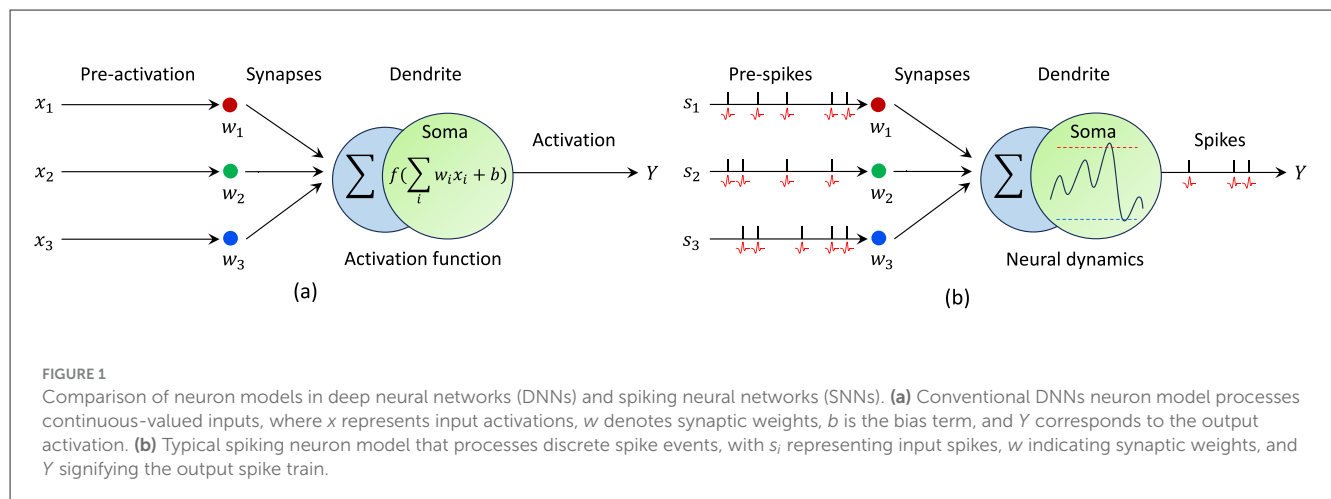
In this paper, we propose a novel BCI signal analysis framework that integrates wavelet transform with an spiking self-attention mechanism. This framework enables dynamic modeling and efficient computation of non-smooth EEG signals by combining brain-inspired spiking neural networks with the global-local feature extraction capabilities of the wavelet domain. Our approach not only overcomes the limitations of traditional manual feature extraction but also demonstrates, for the first time, the synergistic effectiveness of spiking self-attention and wavelet transform in cross-task scenarios through end-to-end training. In experimental evaluations focused on emotion recognition and auditory attention decoding tasks, our method achieves outstanding performance. The main contributions of this work are summarized as follows:

- We propose a novel spiking self-attention module integrated with discrete wavelet transform (DWT) for EEG signal processing. This innovative module simultaneously captures global rhythmic patterns and local transient features through multi-scale wavelet decomposition. Leveraging the spatio-temporal dynamics of spiking neurons, it effectively models nonlinear feature dependencies while replacing traditional Transformer's dense attention with efficient sparse pulse sequences.
- We present SpikeWavformer, the first end-to-end spiking neural network framework specifically designed for multi-task BCI analysis. The framework unifies time-frequency decomposition, dynamic feature selection, and classification within a biologically plausible computational paradigm. Its cascade architecture combines reversible wavelet transforms with spiking self-attention layers, enabling adaptive optimization across diverse BCI tasks including emotion recognition and auditory decoding.
- We conduct comprehensive evaluations on multiple public benchmark datasets to validate the effectiveness of SpikeWavformer. Experimental results demonstrate superior performance compared to existing methods, particularly in resource-constrained environments. The framework shows significant practical potential for real-world BCI applications, achieving state-of-the-art results while maintaining low computational overhead.

## 2 Related works

### 2.1 SNNs for EEG signal processing tasks

EEG-based BCIs have demonstrated significant potential across various downstream tasks, with auditory attention decoding (AAD) and emotion recognition representing two prominent application domains. In AAD research, the challenge stems from the cocktail party effect—the neurocognitive ability to selectively focus on target speakers in multi-talker environments (Cherry, 1953), which contrasts with difficulties experienced by hearing-impaired populations (Cai et al., 2024). Neurophysiological signal analyses through ECoG (Mesgarani and Chang, 2012), MEG (Akram



et al., 2016), and EEG (O'sullivan et al., 2015) have enabled AAD implementations, catalyzing developments in neuro-steered hearing aids (Ceolini et al., 2020). For emotion recognition, the field seeks to model higher-order cognitive functions encoded in neurophysiological signals (Tan et al., 2021). While emotional states manifest through various modalities, the susceptibility of physical expressions to masking effects positions non-invasive EEG as a robust solution for emotion decoding (Xu et al., 2024; Li et al., 2019).

SNNs have emerged as a promising computational framework for both applications, leveraging their inherent low-latency processing and energy-efficient characteristics. In AAD research, Faghihi et al. (2022) developed efficient left/right attention pattern decoding, while Cai et al. (2023) proposed BSAnet, integrating biologically plausible mechanisms with attention modeling for temporal dynamics capture. Recent advances include spiking GCNs for spatial feature extraction (Cai et al., 2024), demonstrating promising results in low-density electrode scenarios. In emotion recognition, pioneering SNN applications have shown methodological viability. Tan et al. (2021) implemented NeuroSense achieving 78.97%/67.76% (arousal/valence) accuracy on DEAP, while Alzhrani et al. (2021) attained 94.83% accuracy using bidirectional spiking networks on DREAMER. Recent developments include fractal SNN architectures (Li et al., 2023), SGLNet for spatiotemporal extraction (Gong et al., 2023), and EESCN achieving 94.81% accuracy on DEAP and SEED-IV (Xu et al., 2024). However, previous research has predominantly relied on manually extracted EEG features such as power spectral density (PSD) and differential entropy (DE) (Jiao et al., 2018; Song et al., 2018), and automatic EEG feature extraction in this domain remains largely unexplored.

## 2.2 Spiking self attention mechanism

Traditional SNNs, despite their inherent advantages in energy efficiency and biological plausibility, still exhibit a performance gap compared to their DNN counterparts. Therefore, many recent works have integrated attention mechanisms into SNNs to enhance their performance and capabilities (Yao et al., 2021; Zhu et al.,

2024; Zhou et al., 2024; Lu et al., 2025). Yao et al. (2023) addressed this through Spike-Driven Self-Attention (SDSA), reformulating matrix multiplications as masking operations to ensure purely binary spike signal transmission. Building on this foundation, Yao et al. (2024) introduced the Meta-Spikeformer architecture that extended the SDSA operator. Those advancement inspired subsequent research exploring SNN-specific attention mechanisms. Wang et al. (2023) proposed Spatiotemporal Self-Attention (STSA) for SNNs, maintaining asynchronous transmission while capturing spatiotemporal feature dependencies. More recently, Wang et al. (2025) developed Saccade Spike Self-Attention (SSSA), enabling comprehensive spatiotemporal feature processing for holistic visual scene understanding in SNN paradigms. Overall, these novel spiking self-attention mechanisms have significantly advanced SNN performance. However, there remains a lack of effective spiking self-attention designs specifically tailored for EEG signal processing.

## 3 Preliminary

### 3.1 Leaky integrate-and-fire neuron

SNNs rely on spiking neurons (Maass, 1997) as their basic unit of information transfer, and common spiking neurons include the Hodgkin-Huxley (Abbott and Kepler, 2005), Izhikevich (Izhikevich, 2003), and Leaky Integrate-and-Fire (LIF) (Izhikevich, 2003) model. In this work, we use the LIF model as the spiking neuron in the proposed method. The LIF model is a simple and effective spiking neuron model. When the membrane potential reaches a certain threshold, the neuron emits a spike, followed by a reset of the membrane potential to the resting potential  $V_{reset}$ . The dynamic model of LIF is described as:

$$H[t] = V[t-1] + \frac{1}{\tau}(X[t] - (V[t-1] - V_{reset})), \quad (1)$$

$$S[t] = \Theta(H[t] - V_{th}), \quad (2)$$

$$V[t] = H[t](1 - S[t]) + V_{reset}S[t], \quad (3)$$

where  $\tau$  is the membrane time constant, and  $X[t]$  is the input current at time step  $t$ . When the membrane potential  $H[t]$  exceeds

the firing threshold  $V_{th}$ , the spiking neuron triggers a spike  $S[t]$ .  $\Theta(\cdot)$  is the Heaviside step function which equals 1 for  $v \geq 0$  and 0 otherwise.  $V[t]$  represents the membrane potential after the trigger event which equals  $H[t]$  if no spike is generated, and otherwise equals to  $V_{reset}$ .

### 3.2 Wavelet transform

Wavelet transforms (WTs) are powerful signal-processing tools that enable the localization of signals in both time and frequency domains, which is particularly useful for analyzing non-stationary signals like EEG. The discrete wavelet transform (DWT), in particular, provides an efficient method for multi-resolution analysis by decomposing signals into sub-bands corresponding to different frequency scales. This decomposition enables the extraction of local features at various scales, making it well-suited for EEG signal processing. EEG signals are nonlinear and non-stationary, posing challenges for traditional analysis methods in capturing their time-varying and multiscale nature. Wavelet transforms, and specifically DWT, offer a significant advantage in feature extraction and time-frequency characterization of EEG signals. The DWT decomposes EEG data into frequency bands such as delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (greater than 30 Hz). This decomposition allows us to extract meaningful features from the EEG data that correspond to various cognitive and emotional states.

For our application, we employ the Haar wavelet due to its simplicity and computational efficiency. Haar wavelets are among the earliest and simplest wavelet functions, characterized by a two-tap filter with minimal support, which results in fast computations. Compared to other common wavelets like Daubechies or Morlet, Haar wavelets are computationally less expensive, requiring only additions and binary shifts, which makes them well-suited for real-time, low-power applications such as SNN-based systems. Haar wavelets are particularly efficient in extracting local, low-frequency components (such as delta and theta waves) as well as high-frequency components (like beta and gamma waves), which are essential for distinguishing different cognitive states in EEG analysis. The efficiency and simplicity of Haar wavelets also make them ideal for handling the sparse, event-driven nature of SNNs.

### 3.3 Spiking self attention mechanism

The Transformer architecture, originally devised for natural language processing tasks (Vaswani et al., 2017), has subsequently permeated multiple subfields of artificial intelligence. At its core lies the self-attention mechanism, which facilitates selective information processing by focusing on relevant contextual elements. Spikformer (Zhou et al., 2022) pioneered the integration of self-attention into SNNs through their Spiking Self-Attention (SSA) framework and spikformer architecture. This approach innovatively employs

sparse spiking representations for the query (Q), key (K), and value (V) matrices:

$$Q = \mathcal{SN}(\text{BN}(XW_Q)), \quad (4)$$

$$K = \mathcal{SN}(\text{BN}(XW_K)), \quad (5)$$

$$V = \mathcal{SN}(\text{BN}(XW_V)), \quad (6)$$

here, Q, K, and V form tensors of dimension  $\mathbb{R}^{T \times C \times H \times W}$ , with  $\text{BN}(\cdot)$  representing batch normalization and  $\mathcal{SN}(\cdot)$  denoting the spiking neuron layer that maintains the attention mechanism's spiking nature. The similarity computation between spiking Q and K matrices proceeds via dot-product:

$$\text{Score} = \text{Sim}(Q, K) = QK^T. \quad (7)$$

The attention output is subsequently calculated as a scaled weighted sum of V, transformed through spiking neuron activation, and further processed through linear transformation and batch normalization before final spiking neuron conversion to produce the output Z:

$$\text{Attn} = \mathcal{SN}(s \cdot \text{Score} \cdot V), \quad Z = \mathcal{SN}(\text{BN}(\text{Linear}(\text{Attn}))). \quad (8)$$

## 4 Methods

In this section, we introduce our approach for EEG-based emotion recognition and auditory attention decoding. First, we define the problem formulation in Section 4.1. Then, we describe the overall data processing workflow in Section 4.2. Finally, we present the proposed Spiking Wavelet Transformer (SpikeWavformer) architecture which integrates wavelet transform and self spiking attention mechanisms in Section 4.4.

### 4.1 Problem analysis

Given an EEG dataset  $\mathcal{D}_{\text{eeg}}$ , it can be represented as:

$$\mathcal{D}_{\text{eeg}} = \{(\mathbf{x}_i^{\text{eeg}}, y_i)\}_{i=1}^N, \quad (9)$$

where  $\mathbf{x}_i^{\text{eeg}} \in \mathcal{X}_{\text{eeg}}$  denotes the raw EEG input signal for the  $i$ -th sample, and  $y_i \in \mathcal{Y}$  represents its corresponding label (emotion category or auditory attention state). Our objective is to learn a spiking neural network model  $F_\theta$  with parameters  $\theta$  to predict the class label from the EEG input. The model is optimized by minimizing the expected risk based on the cross-entropy loss  $L_{\text{CE}}$ :

$$\arg \min_{\theta} \mathbb{E}_{(\mathbf{x}^{\text{eeg}}, y) \sim \mathcal{D}_{\text{eeg}}} [L_{\text{CE}}(F_\theta(\mathbf{x}^{\text{eeg}}), y)]. \quad (10)$$

In this study, we present a novel spiking transformer model, denoted as  $F_\theta$ , to learn discriminative spatio-temporal representations directly from raw EEG signals for the joint tasks of emotion recognition and auditory attention decoding. To achieve this, we introduce a novel Spiking Wavelet Self-Attention (SWSA) mechanism within a spiking transformer framework. While conventional Spiking Self-Attention (SSA) enables efficient event-driven computation, it is limited in its ability to capture the multi-scale frequency dynamics intrinsic to non-stationary EEG signals.

The proposed SWSA overcomes this limitation by integrating Haar wavelet transforms for joint time–frequency analysis, which offer a minimal filter length and computational simplicity, making them highly efficient for real-time processing. Compared to other wavelet bases, such as Daubechies and Morlet, Haar's shorter filters and multiply-free operations align well with the event-driven, low-power nature of spiking neurons. This integration allows the model to focus on neurophysiologically relevant rhythms (e.g., alpha and beta bands) critical for emotional and attentional processes, while maintaining energy-efficient computation. Finally, a cross-entropy loss function is employed to enable effective gradient-based optimization for learning highly discriminative features across both tasks.

## 4.2 Workflow

The overall workflow of the proposed method is depicted in Figure 2. Raw EEG signals are first preprocessed and segmented into overlapping windows via a sliding window strategy to preserve temporal continuity. To more effectively capture the spatial characteristics of EEG activity,  $\alpha$ -band cortical signals are extracted and projected onto 2D topographic maps, thereby maintaining brain-region dependencies. These maps are subsequently divided into patches and tokenized into fixed-length sequences, which serve as the input to a stack of  $N$  spiking encoder blocks. Finally, the resulting features are fed into an MLP classification head to predict the corresponding emotional or attentional state. In summary, this high-level pipeline constitutes the basis of the proposed model architecture, which is elaborated in the following section.

## 4.3 SpikeWavformer

Building on the workflow described above, we design SpikeWavformer—an end-to-end spiking transformer architecture that combines wavelet-based multiscale analysis with spiking attention to enhance EEG feature representation. As shown in Figure 3, The SpikeWavformer can be written as follows:

$$\begin{aligned} X &= \text{SPS}(X), \\ X'_l &= \text{SWSA}(X_{l-1}) + X_{l-1}, \quad l = 1, \dots, L \\ X_l &= \text{MLP}(X'_l) + X'_l, \quad l = 1, \dots, L \\ Y &= \text{CH}(\text{GAP}(X_L)). \end{aligned} \quad (11)$$

Given the EEG input  $X$ , SpikeWavformer first visualizes the spatial focus position via the topographic distribution of oscillatory cortical activities in the  $\alpha$  band and converts it into a 2D image. Subsequently, the SPS module partitions the input into patches and progressively extracts features, optionally incorporating wavelet transformation to enhance multiscale feature representation. Then,  $L \times$  spiking wavelet encoder blocks with spiking wave attention mechanism are employed to encode the features. Finally, the features obtained from extraction and encoding are compressed into a fixed-dimension vector via global average pooling (GAP) and fed into a fully connected layer classification head (CH) to produce classification results.

## 4.4 Spiking wavelet encoder block

As an essential neurophysiological signal, EEG plays a pivotal role in research areas such as affective computing and auditory attention decoding. Nevertheless, its multi-channel structure, low signal-to-noise ratio (SNR), pronounced temporal non-stationarity, and intricate time–frequency characteristics present substantial challenges for existing analysis techniques. Conventional CNNs are limited in capturing long-range temporal dependencies inherent in EEG data. In contrast, vanilla Transformers possess strong long-range modeling capability but incur prohibitive computational costs when processing long-sequence EEG signals. Furthermore, many existing approaches employ irreversible downsampling during multi-scale feature extraction, resulting in the loss of critical frequency-domain information. This drawback is particularly detrimental to neural decoding tasks that rely on specific frequency bands.

To address these issues, we propose a Spiking Wavelet Self-Attention (SWSA) mechanism for EEG signal processing. It combines the biological plausibility of SNNs with the flexible time–frequency analysis of wavelet transforms, offering an efficient, biologically inspired solution for EEG-based emotion recognition and auditory attention decoding. Specifically, given multi-channel EEG inputs  $X \in \mathbb{R}^{T \times B \times C \times H \times W}$ , where  $T$  denotes time steps,  $B$  batch size,  $C$  EEG channels, and  $H \times W$  spatial-topological 2D arrangement. The frequency-domain features of EEG signals are crucial for neuro-decoding. Different frequency bands correspond to different cognitive states:  $\delta$  with deep sleep,  $\theta$  with memory encoding,  $\alpha$  with relaxation,  $\beta$  with attention and cognitive activities, and  $\gamma$  with perception and higher-order functions. We adopt the Haar wavelet for its minimal filter length and computational simplicity, which enable fast, low-power multiscale decomposition and align well with the event-driven, resource-constrained nature of SNN-based BCI systems. Specifically, the Haar wavelet is used for multiscale decomposition and perform DWT on EEG features at each time step  $t$ :

$$\left[ X_{LL}^{(t)}, X_{LH}^{(t)}, X_{HL}^{(t)}, X_{HH}^{(t)} \right] = \text{DWT} \left( X^{(t)} \right), \quad (12)$$

here,  $X_{LL}^{(t)}$  captures low-frequency components (like  $\delta$ ,  $\theta$ ), while high-frequency sub-bands  $X_{LH}^{(t)}$ ,  $X_{HL}^{(t)}$ ,  $X_{HH}^{(t)}$  retain high-frequency information ( $\beta$ ,  $\gamma$ ). Then, spatial local convolution enhances frequency-band interactions:

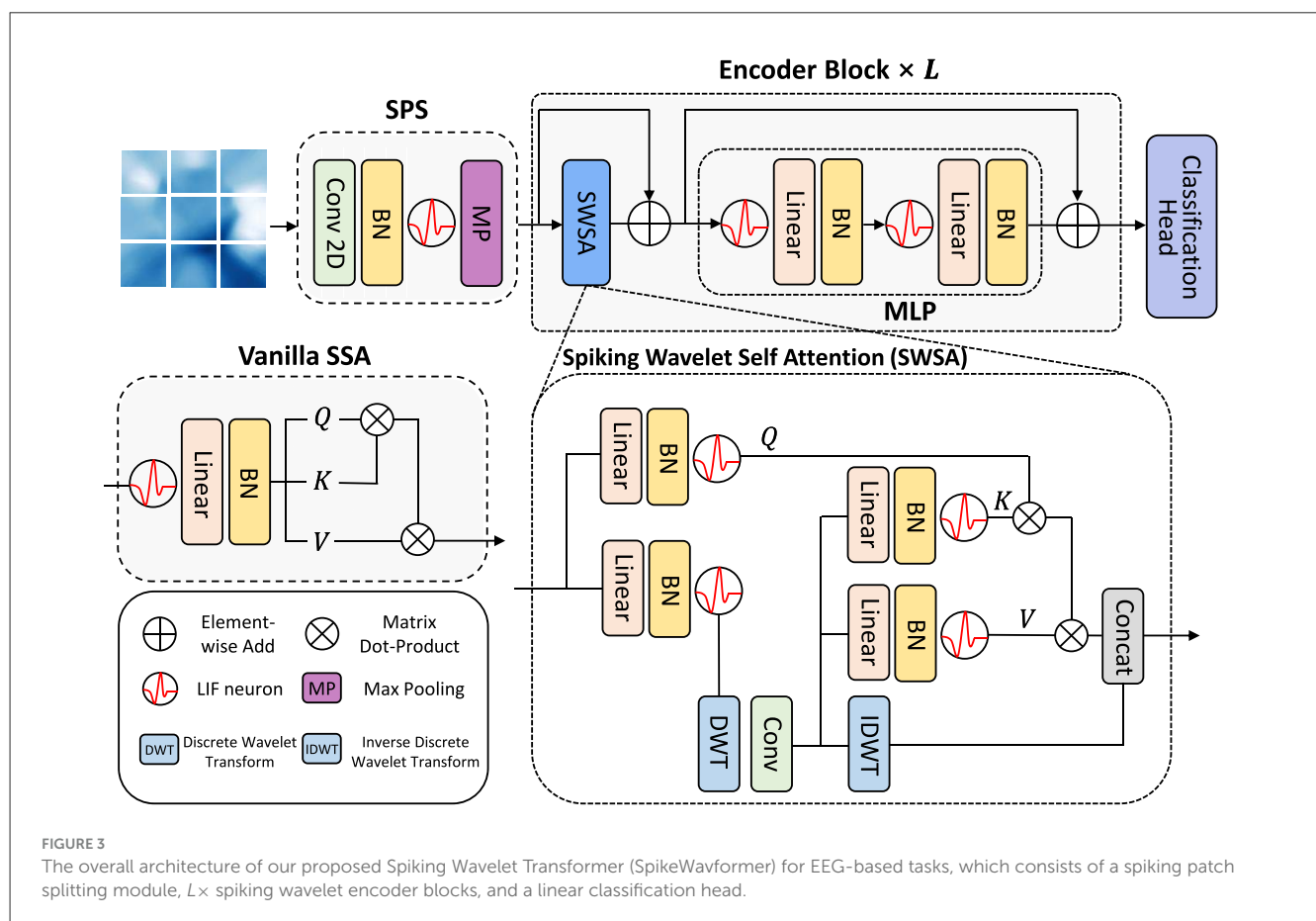
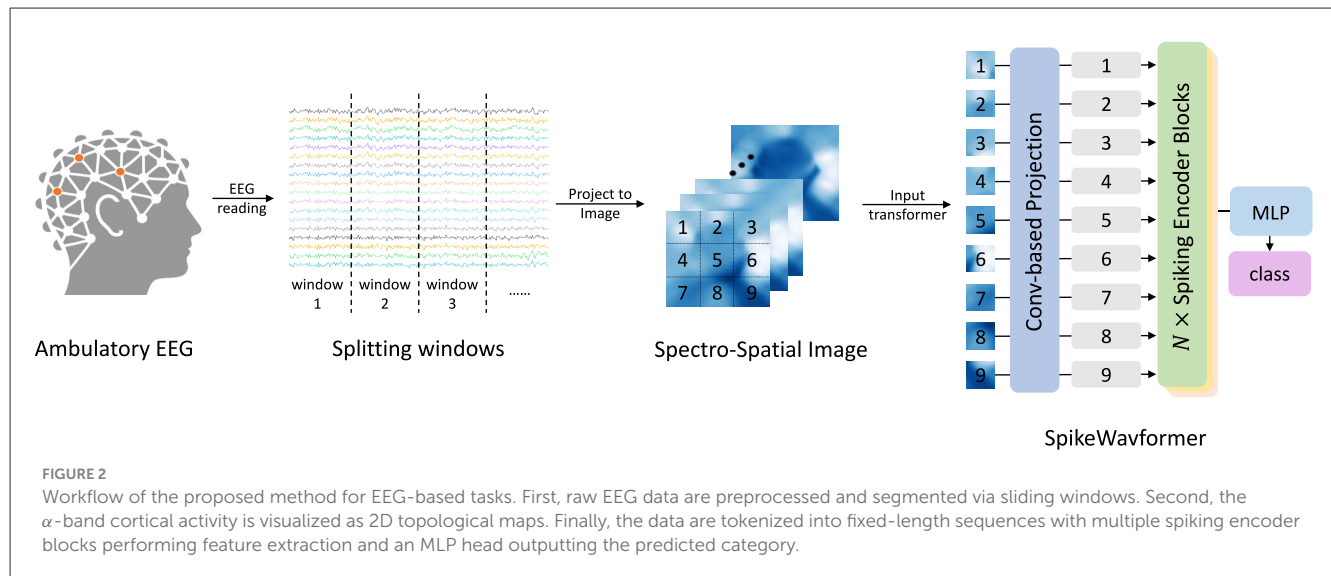
$$X_{filt}^{(t)} = \text{LIF} \left( \text{BN} \left( \text{Conv} \left( \text{Concat} \left( \left[ X_{LL}^{(t)}, X_{LH}^{(t)}, X_{HL}^{(t)} \right] \right) \right) \right) \right), \quad (13)$$

here, BN is batch normalization, LIF a spiking neuron layer. IDWT reconstructs spatial-domain features:

$$X_{recon}^{(t)} = \text{IDWT} \left( X_{filt}^{(t)} \right). \quad (14)$$

Our encoder, inspired by vanilla encoder (Vaswani et al., 2017), first calculates block-input spikes for self-attention. Three matrices  $W_q \in \mathbb{R}^{d \times d_q}$ ,  $W_k \in \mathbb{R}^{d \times d_k}$ ,  $W_v \in \mathbb{R}^{d \times d_v}$  map tokens to vectors. Spiking neurons convert vectors to spiking sequences  $Q, K, V$ :

$$Q = \text{LIF}(\text{BN}(XW_q)), K = \text{LIF}(\text{BN}(XW_k)), V = \text{LIF}(\text{BN}(XW_v)). \quad (15)$$



Next, we compute  $Q$ - $K$  similarity. Following Zhou et al. (2022), a scaling factor  $s$  controls matrix-multiplication magnitude without affecting attention properties:

$$X_{\text{attn}} = \text{LIF}(QK^T V * s), \quad (16)$$

$$X'_{\text{attn}} = \text{LIF}(\text{BN}(\text{Linear}(X_{\text{attn}}))). \quad (17)$$

To integrate wavelet and attention features effectively, we use channel-wise concatenation:

$$X_{\text{combined}} = \text{Concat}(X'_{\text{attn}}, X_{\text{recon}}^{(t)}), \quad (18)$$

$$\text{SWSA}(X) = \text{LIF}(\text{BN}(X_{\text{combined}})). \quad (19)$$

By integrating wavelet decomposition with spiking mechanisms, SpikeWavformer enables efficient processing of

long-sequence EEG data while facilitating the analysis of cross-frequency neural dynamics, thereby providing richer feature representations for complex neuro decoding tasks. Specifically, we analyze the advantages of integrating wavelet transform into SNNs from the perspectives of convergence and convergence speed. First, We define the EEG signal space as  $X = \{x \in \mathbb{R}^{T \times C \times H \times W}\}$ , where  $T$  represents time steps,  $C$  denotes channels, and  $H \times W$  represents spatial dimensions. The discrete wavelet transform operator is defined as:

$$W: X \rightarrow Y, \quad (20)$$

where  $Y = \{(X_{LL}, X_{LH}, X_{HL}, X_{HH})\}$  represents the wavelet coefficient space.

The SWSA mechanism can be formalized as a composite operator:

$$\text{SWSA}(X) = F(\text{Attn}(W(X))) \oplus W^{-1}(W(X)), \quad (21)$$

where  $W$  is the DWT operator,  $\text{Attn}$  is the spiking attention operator,  $F$  is the fusion operator,  $\oplus$  denotes concatenation and  $W^{-1}$  is the inverse DWT (IDWT).

**Theorem 1 (Lipschitz Continuity of SWSA):** The SWSA mechanism satisfies Lipschitz continuity (Hager, 1979; Gouk et al., 2021; Goldstein, 1977) with constant  $L_{\text{SWSA}}$ , ensuring stable convergence during training.

**Proof:** First, we establish the Lipschitz properties of individual components: **Haar Wavelet Transform Lipschitz Property:** For the Haar wavelet transform  $W$ , we have:  $\|W(x_1) - W(x_2)\|_2 \leq L_W \|x_1 - x_2\|_2$ . Since Haar wavelets are orthonormal,  $L_W = 1$ . **Spiking Attention Lipschitz Property:** For the spiking attention mechanism with LIF neuron, let  $\phi(\mu) = \Theta(\mu - V_{th})$  be the spike generation function. The membrane potential dynamics:  $V[t] = \tau V[t-1] + X[t] - v_{reset} S[t-1]$ . For bounded inputs, the LIF neuron satisfies:  $|\phi(\mu_1) - \phi(\mu_2)| \leq \frac{1}{V_{th}} |\mu_1 - \mu_2|$ . Therefore,  $L_A = \frac{1}{V_{th}}$ . **Combined Operator:** The SWSA operator combines these components:  $\|\text{SWSA}(x_1) - \text{SWSA}(x_2)\|_2 \leq L_{\text{SWSA}} \|x_1 - x_2\|_2$ , where  $L_{\text{SWSA}} = L_W \cdot L_A \cdot L_F = \frac{L_F}{V_{th}}$  with  $L_F$  being the Lipschitz constant of the fusion operation.

**Corollary 1:** Under the assumption that  $L_{\text{SWSA}} < 1$ , the SWSA operator is a contraction mapping, guaranteeing convergence to a unique fixed point.

**Theorem 2 (Accelerated convergence):** The SWSA mechanism achieves faster convergence compared to vanilla spiking self attention.

**Proof:** Consider the optimization landscape with loss function  $L(\theta)$ . The gradient update for SWSA parameters follows:  $\theta_{t+1} = \theta_t - \alpha \nabla_{\theta} L(\theta_t)$ . The wavelet decomposition provides a natural regularization through frequency localization:

$$L_{\text{SWSA}}(\theta) = L_{\text{data}}(\theta) + \lambda \sum_j \|W_j\|_1. \quad (22)$$

This  $L_1$  regularization on wavelet coefficients promotes sparsity. The convergence rate is bounded by:

$$L(\theta_t) - L^* \leq \frac{1}{2\alpha T} \|\theta_0 - \theta^*\|^2 + \frac{\alpha L}{2} \sigma^2, \quad (23)$$

where the wavelet regularization reduces the effective variance  $\sigma^2$ , leading to faster convergence.

## 5 Experiment

This section presents comprehensive experiments to evaluate the effectiveness and efficiency of the proposed SpikeWayformer model. First, we detail the experimental setup, including datasets, preprocessing, and implementation specifics. Second, comparative studies are conducted on the DEAP and KUL datasets, demonstrating superior performance over existing methods in both emotion recognition and auditory attention decoding tasks. Additionally, we provide an analysis of the model's energy efficiency, highlighting its advantages in low-power computing environments.

### 5.1 Experimental setup

#### 5.1.1 Datasets

**DEAP.** The DEAP dataset (Koelstra et al., 2011), widely used in emotion recognition research, examines emotional responses to multimedia stimuli by employing peripheral physiological data and EEG signals. It includes 32-channel EEG recordings and various physiological signals, such as skin temperature, blood volume pulse (BVP), respiratory rate, galvanic skin response (GSR), electrooculogram (EOG), and video clips of facial expressions. The facial expressions of the first 22 participants were also recorded. Each participant completed 40 trials, with each trial lasting 1 min and a 3-second baseline recorded before the start of each trial. After each trial, participants filled out a questionnaire to self-report their emotional state in terms of arousal, valence, dominance, and liking, with each dimension rated on a 9-point scale. EEG data were collected using a 32-channel device at a sampling rate of 512 Hz.

**KUL.** The KUL dataset (Das et al., 2019) comprises EEG data collected using the BioSemi ActivateTwo device. The experimental environment was electromagnetically shielded and soundproofed to minimize potential noise interference. Data were collected from 16 subjects with normal hearing, who were instructed to focus on a specific speaker amidst two speakers. The speakers narrated four Dutch stories. Each subject participated in 8 trials, each lasting 6 min. Auditory stimuli, filtered through HRTF, were presented to the subjects in two forms: from the left or right side, in a randomized manner.

#### 5.1.2 Implementation details

The EEG data from each channel was first re-referenced to the average response of all electrodes. Given that the analyzed EEG signals were collected at different sampling rates, they were all band-pass filtered between 1 and 32 Hz using a 6th-order Chebyshev Type II filter and down sampled to a 128 Hz sampling rate. The frequency range was chosen based on previous nonlinear AAD studies. Finally, the EEG data channels were normalized to ensure a mean of zero and unit variance for each trial. The study on the KUL dataset analyzed seven decision window sizes: 0.1, 0.2, 0.5, 1, 2, 5, and 10 seconds. Experiments were conducted using two NVIDIA RTX 4090 GPUs. The model was optimized using the Adam optimizer with an initial learning rate of  $1 \times 10^{-4}$  and trained for 200 epochs. For the SNN model parameters, LIF neurons were

set with an initial membrane potential of 0, a spiking threshold of 0.5, and a simulation time step of 4. To facilitate effective backpropagation, a sigmoid function with parameter  $\alpha = 4$  was used as the surrogate gradient function, expressed as  $\text{sigmoid}(x) = 1/(1 + \exp(-\alpha x))$ . The remaining setup of spiking transformer architecture follows spikformer (Zhou et al., 2022).

### 5.2 Comparative study

We conduct experiments on the DEAP and KUL datasets using proposed SpikeWavformer and compare the results with existing methods for emotion recognition and auditory attention decoding. As shown in Tables 1, 2, our method achieves state-of-the-art performance on all datasets. On the DEAP dataset for emotion recognition, the SpikeWavformer method reaches an Arousal accuracy of 76.51% (std: 5.48%) and a Valence accuracy of 77.10% (std: 5.68%). Existing methods like EEGNet (Lawhern et al., 2018) achieve 58.29% (std: 8.60%) for Arousal and 54.56% (std: 8.14%) for Valence. SCN (Schirrmester et al., 2017) attains 61.19% (std: 10.28%) for Arousal and 59.42% (std: 8.30%) for Valence. DCN (Schirrmester et al., 2017) gets 61.03% (std: 8.58%) for Arousal and 59.92% (std: 7.82%) for Valence. Tsception (Ding et al., 2022) achieves 61.57% (std: 11.04%) for Arousal and 59.14% (std: 7.60%) for Valence.

We further compared the performance of the SpikeWavformer for different detection window sizes, ranging from 0.1 to 10 seconds, with the results presented in Table 2. On the KUL dataset, the SpikeWavformer achieved an average decoding accuracy of 96.5% across all subjects for a 1-second decision window, 97.1% for a 2-second decision window, 97.3% for a 5-second decision window, and 98.6% for a 10-second decision window. Generally, larger decision windows yielded better results, corroborating findings from previous studies (De Tallez et al., 2020; Ciccarelli

et al., 2019; Vandecappelle et al., 2021). Notably, our proposed method is capable of decoding auditory spatial attention with a very short decision window of less than 1 second. For decision windows of 0.5 seconds and 0.2 seconds, the SpikeWavformer attained high accuracy rates of 94.2% and 86.7%, respectively. Although the accuracy for the 0.1-second decision window was lower than that of the 1-second decision window, SpikeWavformer maintained a high accuracy rate of 80.5%. In all comparisons with related work (De Cheveigné et al., 2018; Cai et al., 2021; Su et al., 2022), the SpikeWavformer demonstrated competitive performance.

### 5.3 Energy consumption comparison

In this section, we validate the energy efficiency of our proposed model over its ANN counterpart. Based on the energy calculation standard in neuromorphic computing (Sengupta et al., 2019), we use the method proposed by Wang et al. (2024) to compute the energy consumption ratio between our model and the equivalent ANN model:

$$\text{Energy}_{\text{rate}} = \frac{\text{AC}}{\text{MAC}} * \text{SpikingRate} * \text{TimeSteps}.$$

(24)

In the equation,  $\frac{\text{AC}}{\text{MAC}}$  denotes the energy consumption ratio of an accumulate (AC) operation in SNNs to a multiplication (MAC) in ANNs. Extensive studies confirm the theoretical value of  $\frac{\text{AC}}{\text{MAC}}$  is  $\frac{1}{7}$  (Horowitz, 2014). Here, SpikingRate is the average spiking rate, and TimeSteps the simulation time window. In our model, SpikingRate is 12.3%, and TimeSteps is set to 4. Based on Equation 24, our model achieves over  $7\times$  energy efficiency compared to its ANN counterpart.

TABLE 1 Comparison of different methods on DEAP dataset.

| Dataset | Method                          | Arousal       |              | Valence       |              |
|---------|---------------------------------|---------------|--------------|---------------|--------------|
|         |                                 | Acc.          | Std          | Acc.          | Std          |
| DEAP    | EEGNet (Lawhern et al., 2018)   | 58.29%        | 8.60%        | 54.56%        | 8.14%        |
|         | SCN (Schirrmester et al., 2017) | 61.19%        | 10.28%       | 59.42%        | 8.30%        |
|         | DCN (Schirrmester et al., 2017) | 61.03%        | 8.58%        | 59.92%        | 7.82%        |
|         | Tsception (Ding et al., 2022)   | 61.57%        | 11.04%       | 59.14%        | 7.60%        |
|         | <b>SpikeWavformer</b>           | <b>76.51%</b> | <b>5.48%</b> | <b>77.10%</b> | <b>5.68%</b> |

The bold text refers to the method proposed in this paper.

TABLE 2 Performance comparison across different decision windows.

| Dataset | Model                                    | Decision window (second) |             |             |             |             |             |             |
|---------|------------------------------------------|--------------------------|-------------|-------------|-------------|-------------|-------------|-------------|
|         |                                          | 0.1                      | 0.2         | 0.5         | 1           | 2           | 5           | 10          |
| KUL     | Linear (CCA) (De Cheveigné et al., 2018) | 50.9                     | 53.6        | 55.7        | 60.2        | 63.5        | 69.4        | 75.9        |
|         | Non-linear (CNN) (Cai et al., 2021)      | 74.3                     | 78.2        | 80.6        | 84.1        | 85.7        | 86.9        | 87.9        |
|         | STAnet (Su et al., 2022)                 | 80.8                     | 84.3        | 87.2        | 90.1        | 91.4        | 92.6        | 93.9        |
|         | <b>SpikeWavformer</b>                    | <b>80.5</b>              | <b>86.7</b> | <b>94.2</b> | <b>96.5</b> | <b>97.1</b> | <b>97.3</b> | <b>98.6</b> |

The bold text refers to the method proposed in this paper.

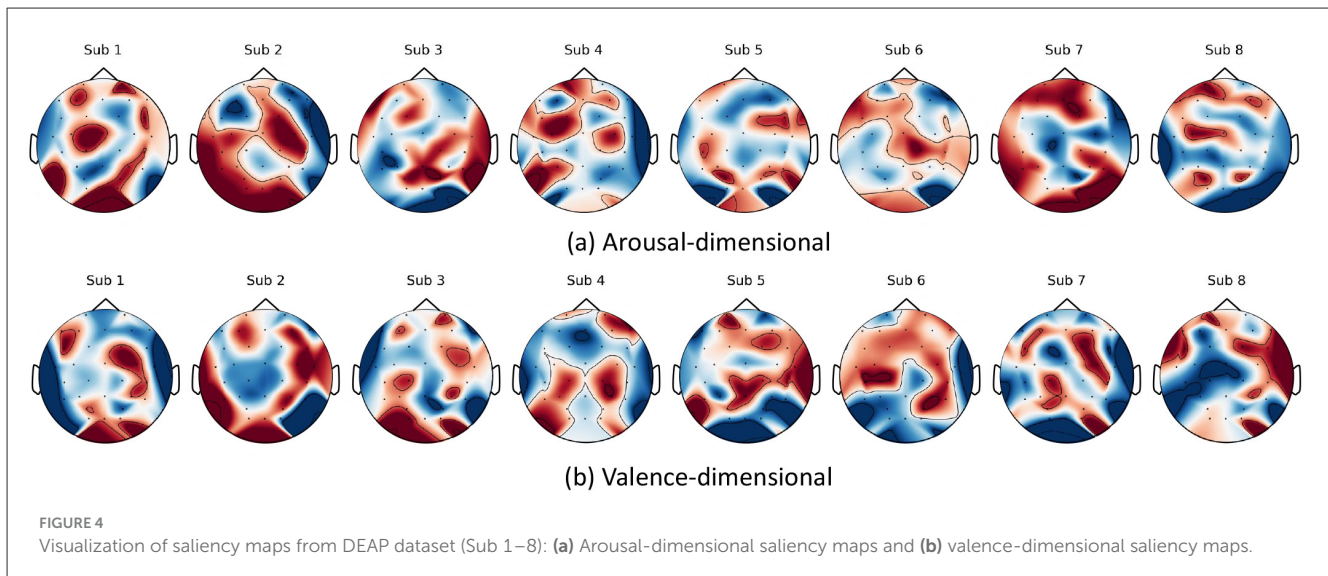


FIGURE 4  
Visualization of saliency maps from DEAP dataset (Sub 1–8): (a) Arousal-dimensional saliency maps and (b) valence-dimensional saliency maps.

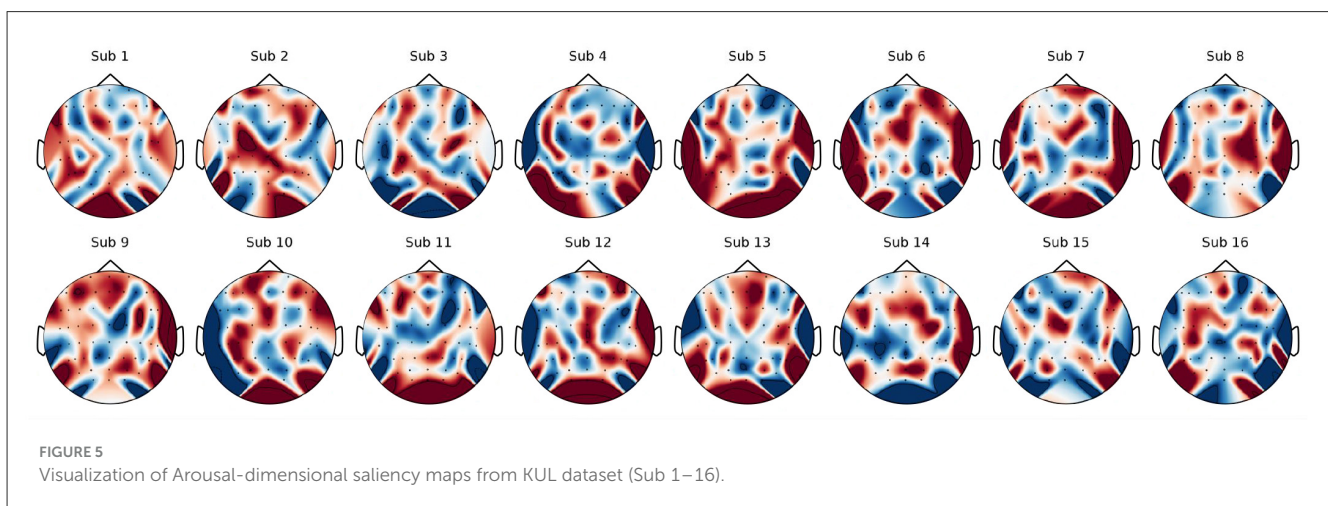


FIGURE 5  
Visualization of Arousal-dimensional saliency maps from KUL dataset (Sub 1–16).

## 5.4 Interpretability

In this section, saliency maps (Simonyan et al., 2013) are employed to visualize the areas of the data that contain the most information and contribute to classification performance. The saliency map is one of the most widely used tools for illustrating which regions of the input data hold classification-relevant information. To enhance the visualization of the saliency maps, the original maps were averaged along the time dimension to capture the topology of the EEG channels. Additionally, the normalized saliency maps were averaged across different samples for each subject to produce generalized average saliency maps. The average saliency maps for the DEAP dataset and the KUL dataset are presented in Figures 4, 5, respectively.

**DEAP.** For arousal, as illustrated in Figure 4a, the temporal and frontal regions of the brain contain a wealth of information. This indicates that these regions are more involved in processing emotions, aligning with findings from previous studies (Gao et al., 2021; Huang et al., 2012; Mickley Steinmetz and Kensinger, 2009). Emotional arousal is predominantly represented in the temporal and frontal lobes. The asymmetry between the frontal and temporal lobes is closely associated with emotion recognition within the

arousal dimension. In terms of valence, Figure 4b shows that the parietal and temporal lobes are also rich in information. This observation is consistent with earlier research (Huang et al., 2012), suggesting that the network effectively learns from these relevant regions.

**KUL.** It is expected that the areas of neural activity contributing to speech processing will exhibit greater significance. As illustrated in Figure 5, the average saliency map of the KUL dataset reveals that the frontal and temporal regions contain more substantial information. These findings align with previous research indicating that activation is prominently observed in the frontal and temporal cortices (Ciccarelli et al., 2019; Geirnaert et al., 2020; Vandecappelle et al., 2021).

## 6 Conclusion

This paper presents SpikeWavformer, an end-to-end deep learning SNN model that integrates the wavelet transform with spiking transformer architecture. The model combines the global-local feature extraction capability of the wavelet transform with the low-power, event-driven computation of spiking neurons,

enabling dynamic modeling and efficient processing of EEG signals. This integration supports effective time–frequency decomposition, automatic feature extraction, and classification, thereby improving generalization across diverse scenarios. Experiments on two publicly available datasets demonstrate that SpikeWavformer consistently outperforms established methods. The experimental results validate its effectiveness in both emotion recognition and auditory attention decoding tasks, highlighting its potential for deployment in resource-constrained brain–computer interface applications. Future deployment of SpikeWavformer on neuromorphic hardware platforms presents both promising opportunities and technical challenges. The energy-efficient characteristics of the approach make it particularly well-suited for implementation on neuromorphic chips, potentially enabling low-power BCI applications in portable devices. However, contemporary neuromorphic architectures are primarily optimized for convolution-based SNNs, necessitating further hardware–software co-design efforts to fully realize the benefits of Transformer-based spiking architectures. Overall, this study advances the development of energy-efficient, high-performance brain–computer interfaces suitable for resource-constrained practical deployment.

## Data availability statement

The datasets used in this study are publicly available. The dataset DEAP for this study can be found at <https://www.eecs.qmul.ac.uk/mmv/datasets/deap/>. The dataset KUL for this study can be found at <https://zenodo.org/records/4004271>.

## Author contributions

LY: Writing – review & editing, Writing – original draft, Software, Methodology. JW: Writing – original draft, Formal analysis. YL: Writing – review & editing, Supervision, Investigation.

## References

- Abbott, L. F., and Kepler, T. B. (2005). “Model neurons: from hodgkin-huxley to hopfield,” in *Statistical Mechanics of Neural Networks: Proceedings of the Xlth Sitges Conference Sitges, Barcelona, Spain, 3–7 June 1990* (Springer), 5–18. doi: 10.1007/3540532676\_37
- Akram, S., Presacco, A., Simon, J. Z., Shamma, S. A., and Babadi, B. (2016). Robust decoding of selective auditory attention from meg in a competing-speaker environment via state-space modeling. *Neuroimage* 124, 906–917. doi: 10.1016/j.neuroimage.2015.09.048
- Alarcao, S. M., and Fonseca, M. J. (2017). Emotions recognition using EEG signals: a survey. *IEEE Trans. Affect. Comput.* 10, 374–393. doi: 10.1109/TAFFC.2017.2714671
- Alzhrani, W., Doborjeh, M., Doborjeh, Z., and Kasabov, N. (2021). “Emotion recognition and understanding using EEG data in a brain-inspired spiking neural network architecture,” in *2021 International Joint Conference on Neural Networks (IJCNN)* (IEEE), 1–9. doi: 10.1109/IJCNN52387.2021.9533368
- Ang, K. K., Chin, Z. Y., Zhang, H., and Guan, C. (2008). “Filter bank common spatial pattern (fbcs) in brain-computer interface,” in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)* (IEEE), 2390–2397. doi: 10.1109/IJCNN.2008.4634130
- Cai, S., Li, P., and Li, H. (2023). A bio-inspired spiking attentional neural network for attentional selection in the listening brain. *IEEE Trans. Neural Netw. Learn. Syst.* 35, 17387–17397. doi: 10.1109/TNNLS.2023.3303308
- Cai, S., Su, E., Xie, L., and Li, H. (2021). EEG-based auditory attention detection via frequency and channel neural attention. *IEEE Trans. Hum.-Mach. Syst.* 52, 256–266. doi: 10.1109/THMS.2021.3125283
- Cai, S., Zhang, R., Zhang, M., Wu, J., and Li, H. (2024). EEG-based auditory attention detection with spiking graph convolutional network. *IEEE Trans. Cogn. Dev. Syst.* 16, 1698–1706. doi: 10.1109/TCDS.2024.3376433
- Ceolini, E., Hjortkjær, J., Wong, D. D., O’Sullivan, J., Raghavan, V. S., Herrero, J., et al. (2020). Brain-informed speech separation (biss) for enhancement of target speaker in multitalker speech perception. *Neuroimage* 223:117282. doi: 10.1016/j.neuroimage.2020.117282
- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 25, 975–979. doi: 10.1121/1.1907229
- Ciccarelli, G., Nolan, M., Perricone, J., Calamia, P. T., Haro, S., O’Sullivan, J., et al. (2019). Comparison of two-talker attention decoding from EEG with nonlinear neural networks and linear methods. *Sci. Rep.* 9:11538. doi: 10.1038/s41598-019-47795-0
- Das, N., Francart, T., and Bertrand, A. (2019). *Auditory Attention Detection Dataset Kuleuven*. London: Zenodo.
- De Cheveigné, A., Wong, D. D., Di Liberto, G. M., Hjortkjær, J., Slaney, M., and Lalor, E. (2018). Decoding the auditory brain with canonical component analysis. *Neuroimage* 172, 206–216. doi: 10.1016/j.neuroimage.2018.01.033

## Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- De Taillez, T., Kollmeier, B., and Meyer, B. T. (2020). Machine learning for decoding listeners' attention from electroencephalography evoked by continuous speech. *Eur. J. Neurosci.* 51, 1234–1241. doi: 10.1111/ejn.13790
- Deng, S., Li, Y., Zhang, S., and Gu, S. (2022). Temporal efficient training of spiking neural network via gradient re-weighting. *arXiv preprint arXiv:2202.11946*.
- Ding, Y., Robinson, N., Zhang, S., Zeng, Q., and Guan, C. (2022). Tsception: capturing temporal dynamics and spatial asymmetry from EEG for emotion recognition. *IEEE Trans. Affect. Comput.* 14, 2238–2250. doi: 10.1109/TAFFC.2022.3169001
- Faghihi, F., Cai, S., and Moustafa, A. A. (2022). A neuroscience-inspired spiking neural network for EEG-based auditory spatial attention detection. *Neural Netw.* 152, 555–565. doi: 10.1016/j.neunet.2022.05.003
- Gao, Y., Cao, Z., Liu, J., and Zhang, J. (2021). A novel dynamic brain network in arousal for brain states and emotion analysis. *Mathem. Biosci. Eng.* 18, 7440–7463. doi: 10.3934/mbe.2021368
- Geirnaert, S., Francart, T., and Bertrand, A. (2020). Fast EEG-based decoding of the directional focus of auditory attention using common spatial patterns. *IEEE Trans. Biomed. Eng.* 68, 1557–1568. doi: 10.1109/TBME.2020.3033446
- Gerstner, W., and Kistler, W. M. (2002). *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511815706
- Goldstein, A. A. (1977). Optimization of lipschitz continuous functions. *Mathem. Program.* 13, 14–22.
- Gong, P., Wang, P., Zhou, Y., and Zhang, D. (2023). A spiking neural network with adaptive graph convolution and LSTM for EEG-based brain-computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* 31, 1440–1450. doi: 10.1109/TNSRE.2023.3246989
- Gouk, H., Frank, E., Pfahringer, B., and Cree, M. J. (2021). Regularisation of neural networks by enforcing lipschitz continuity. *Mach. Learn.* 110, 393–416. doi: 10.1007/s10994-020-05929-w
- Grobbelaar, M., Phadikar, S., Ghaderpour, E., Struck, A. F., Sinha, N., Ghosh, R., et al. (2022). A survey on denoising techniques of electroencephalogram signals using wavelet transform. *Signals* 3, 577–586. doi: 10.3390/signals3030035
- Hager, W. W. (1979). Lipschitz continuity for constrained processes. *SIAM J. Control Optim.* 17, 321–338.
- Horowitz, M. (2014). “1.1 computing's energy problem (and what we can do about it),” in *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC) (IEEE)*, 10–14. doi: 10.1109/ISSCC.2014.6757323
- Huang, D., Guan, C., Ang, K. K., Zhang, H., and Pan, Y. (2012). “Asymmetric spatial pattern for EEG-based emotion detection,” in *The 2012 International Joint Conference on Neural Networks (IJCNN) (IEEE)*, 1–7. doi: 10.1109/IJCNN.2012.6252390
- Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE Trans. Neural Netw.* 14, 1569–1572. doi: 10.1109/TNN.2003.820440
- Jiao, Z., Gao, X., Wang, Y., Li, J., and Xu, H. (2018). Deep convolutional neural networks for mental load classification based on EEG data. *Pattern Recognit.* 76, 582–595. doi: 10.1016/j.patcog.2017.12.002
- Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., et al. (2011). Deap: a database for emotion analysis; using physiological signals. *IEEE Trans. Affect. Comput.* 3, 18–31. doi: 10.1109/TAFFC.2011.15
- Kwon, O.-Y., Lee, M.-H., Guan, C., and Lee, S.-W. (2019). Subject-independent brain-computer interfaces based on deep convolutional neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* 31, 3839–3852. doi: 10.1109/TNNLS.2019.2946869
- Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., and Lance, B. J. (2018). EEGnet: a compact convolutional neural network for EEG-based brain-computer interfaces. *J. Neural Eng.* 15:056013. doi: 10.1088/1741-2552/aace8c
- Lei, Z., Yao, M., Hu, J., Luo, X., Lu, Y., Xu, B., et al. (2025). “Spike2former: efficient spiking transformer for high-performance image segmentation,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, 1364–1372. doi: 10.1609/aaai.v39i2.32126
- Li, J., Zhang, Z., and He, H. (2018). Hierarchical convolutional neural networks for EEG-based emotion recognition. *Cognit. Comput.* 10, 368–380. doi: 10.1007/s12559-017-9533-x
- Li, P., Liu, H., Si, Y., Li, C., Li, F., Zhu, X., et al. (2019). EEG based emotion recognition by combining functional connectivity network and local activations. *IEEE Trans. Biomed. Eng.* 66, 2869–2881. doi: 10.1109/TBME.2019.2897651
- Li, W., Fang, C., Zhu, Z., Chen, C., and Song, A. (2023). Fractal spiking neural network scheme for EEG-based emotion recognition. *IEEE J. Translat. Eng. Health Med.* 12, 106–118. doi: 10.1109/JTEHM.2023.3320132
- Liu, R., Wang, Y.-X., and Zhang, L. (2015). An FDES-based shared control method for asynchronous brain-actuated robot. *IEEE Trans. Cybern.* 46, 1452–1462. doi: 10.1109/TCYB.2015.2469278
- Lotte, F., and Guan, C. (2010). Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms. *IEEE Trans. Biomed. Eng.* 58, 355–362. doi: 10.1109/TBME.2010.2082539
- Lu, C., Du, H., Wei, W., Sun, Q., Wang, Y., Zeng, D., et al. (2025). Estsformer: efficient spatio-temporal spiking transformer. *Neural Netw.* 191:107786. doi: 10.1016/j.neunet.2025.107786
- Luo, X., Yao, M., Chou, Y., Xu, B., and Li, G. (2024). “Integer-valued training and spike-driven inference spiking neural network for high-performance and energy-efficient object detection,” in *European Conference on Computer Vision* (Springer), 253–272. doi: 10.1007/978-3-031-73411-3\_15
- Maass, W. (1997). Networks of spiking neurons: the third generation of neural network models. *Neural Netw.* 10, 1659–1671. doi: 10.1016/S0893-6080(97)00011-7
- Masquelier, T., Guyonnet, R., and Thorpe, S. J. (2008). Spike timing dependent plasticity finds the start of repeating patterns in continuous spike trains. *PLoS ONE* 3:e1377. doi: 10.1371/journal.pone.0001377
- Mesgarani, N., and Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* 485, 233–236. doi: 10.1038/nature11020
- Mickley Steinmetz, K. R., and Kensinger, E. A. (2009). The effects of valence and arousal on the neural activity leading to subsequent memory. *Psychophysiology* 46, 1190–1199. doi: 10.1111/j.1469-8986.2009.00868.x
- O'sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., et al. (2015). Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cerebral cortex* 25, 1697–1706. doi: 10.1093/cercor/bht355
- Pan, Z., Chua, Y., Wu, J., Zhang, M., Li, H., and Ambikairajah, E. (2020). An efficient and perceptually motivated auditory neural encoding and decoding algorithm for spiking neural networks. *Front. Neurosci.* 13:1420. doi: 10.3389/fnins.2019.01420
- Schirrmester, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggenberger, K., Tangermann, M., et al. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization. *Hum. Brain Mapp.* 38, 5391–5420. doi: 10.1002/hbm.23730
- Sengupta, A., Ye, Y., Wang, R., Liu, C., and Roy, K. (2019). Going deeper in spiking neural networks: VGG and residual architectures. *Front. Neurosci.* 13:95. doi: 10.3389/fnins.2019.00095
- Shi, X., Hao, Z., and Yu, Z. (2024). “Spikingresformer: bridging resnet and vision transformer in spiking neural networks,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5610–5619. doi: 10.1109/CVPR52733.2024.00536
- Simonyan, K., Vedaldi, A., and Zisserman, A. (2013). Deep inside convolutional networks: visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*.
- Singh, A. K., and Krishnan, S. (2023). Trends in EEG signal feature extraction applications. *Front. Artif. Intell.* 5:1072801. doi: 10.3389/frai.2022.1072801
- Song, T., Zheng, W., Song, P., and Cui, Z. (2018). EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Trans. Affect. Comput.* 11, 532–541. doi: 10.1109/TAFFC.2018.2817622
- Su, E., Cai, S., Xie, L., Li, H., and Schultz, T. (2022). Stanet: a spatiotemporal attention network for decoding auditory spatial attention from EEG. *IEEE Trans. Biomed. Eng.* 69, 2233–2242. doi: 10.1109/TBME.2022.3140246
- Subasi, A. (2019). *Practical Guide for Biomedical Signals Analysis Using Machine Learning Techniques: A MATLAB Based Approach*. New York: Academic Press. doi: 10.1016/B978-0-12-817444-9.00002-7
- Tan, C., Šarlija, M., and Kasabov, N. (2021). Neurosense: short-term emotion recognition and understanding based on spiking neural network modelling of spatio-temporal EEG patterns. *Neurocomputing* 434, 137–148. doi: 10.1016/j.neucom.2020.12.098
- Vallabhaneni, R. B., Sharma, P., Kumar, V., Kulshreshtha, V., Reddy, K. J., Kumar, S. S., et al. (2021). Deep learning algorithms in EEG signal decoding application: a review. *IEEE Access* 9, 125778–125786. doi: 10.1109/ACCESS.2021.3105917
- Vandecappelle, S., Deckers, L., Das, N., Ansari, A. H., Bertrand, A., and Francart, T. (2021). EEG-based detection of the locus of auditory attention with convolutional neural networks. *Elife* 10:e56481. doi: 10.7554/eLife.56481
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). “Attention is all you need,” in *Advances in Neural Information Processing Systems*, 30.
- Wang, S., Zhang, D., Shi, K., Wang, Y., Wei, W., Wu, J., et al. (2024). Global-local convolution with spiking neural networks for energy-efficient keyword spotting. *arXiv preprint arXiv:2406.13179*.
- Wang, S., Zhang, M., Zhang, D., Belatreche, A., Xiao, Y., Liang, Y., et al. (2025). Spiking vision transformer with saccadic attention. *arXiv preprint arXiv:2502.12677*.
- Wang, Y., Shi, K., Lu, C., Liu, Y., Zhang, M., and Qu, H. (2023). “Spatial-temporal self-attention for asynchronous spiking neural networks,” in *IJCAI*, 3085–3093. doi: 10.24963/ijcai.2023/344

- Wang, Y.-K., Jung, T.-P., and Lin, C.-T. (2015). EEG-based attention tracking during distracted driving. *IEEE Trans. Neural Syst. Rehabil. Eng.* 23, 1085–1094. doi: 10.1109/TNSRE.2015.2415520
- Wei, W., Zhang, M., Zhang, J., Belatreche, A., Wu, J., Xu, Z., et al. (2024). Event-driven learning for spiking neural networks. *arXiv preprint arXiv:2403.00270*.
- Wu, J., Chua, Y., Zhang, M., Li, H., and Tan, K. C. (2018). A spiking neural network framework for robust sound classification. *Front. Neurosci.* 12:836. doi: 10.3389/fnins.2018.00836
- Xing, M., Lee, H., Morrissey, Z., Chung, M. K., Phan, K. L., Klumpp, H., et al. (2019). Altered dynamic electroencephalography connectome phase-space features of emotion regulation in social anxiety. *Neuroimage* 186, 338–349. doi: 10.1016/j.neuroimage.2018.10.073
- Xu, F., Pan, D., Zheng, H., Ouyang, Y., Jia, Z., and Zeng, H. (2024). EESCN: a novel spiking neural network method for EEG-based emotion recognition. *Comput. Methods Programs Biomed.* 243:107927. doi: 10.1016/j.cmpb.2023.107927
- Yao, M., Gao, H., Zhao, G., Wang, D., Lin, Y., Yang, Z., et al. (2021). “Temporal-wise attention spiking neural networks for event streams classification,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10221–10230. doi: 10.1109/ICCV48922.2021.01006
- Yao, M., Hu, J., Hu, T., Xu, Y., Zhou, Z., Tian, Y., et al. (2024). Spike-driven transformer v2: meta spiking neural network architecture inspiring the design of next-generation neuromorphic chips. *arXiv preprint arXiv:2404.03663*.
- Yao, M., Hu, J., Zhou, Z., Yuan, L., Tian, Y., Xu, B., et al. (2023). “Spike-driven transformer. *Advances in Neural Information Processing Systems*, 64043–64058.
- Yao, M., Qiu, X., Hu, T., Hu, J., Chou, Y., Tian, K., et al. (2025). Scaling spike-driven transformer with efficient spike firing approximation training. *IEEE Trans. Pattern Anal. Mach. Intell.* 47, 2973–2990. doi: 10.1109/TPAMI.2025.3530246
- Zhang, X., Liu, J., Shen, J., Li, S., Hou, K., Hu, B., et al. (2020). Emotion recognition from multimodal physiological signals using a regularized deep fusion of kernel machine. *IEEE Trans. Cybern.* 51, 4386–4399. doi: 10.1109/TCYB.2020.2987575
- Zhong, P., Wang, D., and Miao, C. (2020). EEG-based emotion recognition using regularized graph neural networks. *IEEE Trans. Affect. Comput.* 13, 1290–1301. doi: 10.1109/TAFFC.2020.2994159
- Zhou, C., Yu, L., Zhou, Z., Ma, Z., Zhang, H., Zhou, H., et al. (2023). Spikingformer: spike-driven residual learning for transformer-based spiking neural network. *arXiv preprint arXiv:2304.11954*.
- Zhou, C., Zhang, H., Zhou, Z., Yu, L., Huang, L., Fan, X., et al. (2024). Qkformer: Hierarchical spiking transformer using qk attention. *arXiv preprint arXiv:2403.16552*.
- Zhou, Z., Zhu, Y., He, C., Wang, Y., Yan, S., Tian, Y., et al. (2022). Spikformer: When spiking neural network meets transformer. *arXiv preprint arXiv:2209.15425*.
- Zhu, R.-J., Zhang, M., Zhao, Q., Deng, H., Duan, Y., and Deng, L.-J. (2024). TCJA-SNN: Temporal-channel joint attention for spiking neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* 36, 5112–5125. doi: 10.1109/TNNLS.2024.3377717
- Zotev, V., Mayeli, A., Misaki, M., and Bodurka, J. (2020). Emotion self-regulation training in major depressive disorder using simultaneous real-time fMRI and EEG neurofeedback. *NeuroImage* 27:102331. doi: 10.1016/j.nicl.2020.102331