frontiers | Frontiers in Neuroscience

TYPE Methods
PUBLISHED 21 February 2025
DOI 10.3389/fnins.2025.1551656

OPEN ACCESS

EDITED BY
Malu Zhang,
National University of Singapore, Singapore

REVIEWED BY
Elishai Ezra Tsur,
Open University of Israel, Israel
Yufei Guo,
China Aerospace Science and Industry Corporation, China
Shibo Zhou,
Zhejiang Lab, China

*CORRESPONDENCE
Tielin Zhang
✉ zhangtielin@ion.ac.cn
Bo Xu
✉ xubo@ia.ac.cn

RECEIVED 26 December 2024
ACCEPTED 04 February 2025
PUBLISHED 21 February 2025

CITATION
Song Y, Han L, Zhang T and Xu B (2025) Multiscale fusion enhanced spiking neural network for invasive BCI neural signal decoding. *Front. Neurosci.* 19:1551656. doi: 10.3389/fnins.2025.1551656

Multiscale fusion enhanced spiking neural network for invasive BCI neural signal decoding

Yu Song[1,2], Liyuan Han[3], Tielin Zhang[3]* and Bo Xu[1,2]*

[1]Institute of Automation, Chinese Academy of Sciences, Beijing, China, [2]School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China, [3]Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai, China

Brain-computer interfaces (BCIs) are an advanced fusion of neuroscience and artificial intelligence, requiring stable and long-term decoding of neural signals. Spiking Neural Networks (SNNs), with their neuronal dynamics and spike-based signal processing, are inherently well-suited for this task. This paper presents a novel approach utilizing a Multiscale Fusion enhanced Spiking Neural Network (MFSNN). The MFSNN emulates the parallel processing and multiscale feature fusion seen in human visual perception to enable real-time, efficient, and energy-conserving neural signal decoding. Initially, the MFSNN employs temporal convolutional networks and channel attention mechanisms to extract spatiotemporal features from raw data. It then enhances decoding performance by integrating these features through skip connections. Additionally, the MFSNN improves generalizability and robustness in cross-day signal decoding through mini-batch supervised generalization learning. In two benchmark invasive BCI paradigms, including the single-hand grasp-and-touch and center-and-out reach tasks, the MFSNN surpasses traditional artificial neural network methods, such as MLP and GRU, in both accuracy and computational efficiency. Moreover, the MFSNN's multiscale feature fusion framework is well-suited for the implementation on neuromorphic chips, offering an energy-efficient solution for online decoding of invasive BCI signals.

KEYWORDS

BCI decoding, brain-inspired, spiking neural network, feature fusion, energy-efficient computing

1 Introduction

Simulating the human brain remains a key objective in neuroscience and artificial intelligence. While Large Language Models (LLMs), such as GPT (Achiam et al., 2023), aim to replicate the brain's broad functionality on a general-purpose scale, and brain-inspired neural networks (Schmidgall et al., 2024; Zhang et al., 2023; Zhao et al., 2023b) focus on capturing its dynamic complexity, there is also a crucial need to understand and model the brain's inner workings on a microscopic level. Recent advancements in invasive Brain-Computer Interface (BCI) technology allow for the direct recording of spike signals at this microscopic scale. Deep learning models can map these microscopic spike signals to macroscopic behavioral outputs through neural signal decoding. For example, SGLNet (Gong et al., 2023) converts EEG signals into spike trains, utilizing Spiking Neural Networks (SNNs) to extract topological information and spike-based LSTM units to decode temporal dependencies. Similarly, hand gesture decoding has been achieved by

decomposing high-density electromyography signals into motor unit spike trains (Chen et al., 2020), which are classified to estimate each gesture.

In the context of long-term invasive brain-computer interface (BCI) recordings, a significant challenge is the phenomenon of data distribution shifts, which can substantially impair the generalization capabilities of decoding models. These shifts arise from a combination of physiological and technical factors, including electrode drift, inflammatory responses, and neural plasticity (Pun et al., 2024; Kubben et al., 2024). Electrode drift, a prevalent issue in chronic neural implants, refers to the gradual displacement of electrodes within the brain tissue, leading to alterations in the recorded neural activity patterns over time. Inflammatory processes or the formation of fibrotic tissue around the electrodes can degrade signal quality, introducing non-stationarities into the data. Neural plasticity, the brain's inherent capacity to adapt and reorganize its functional architecture, can result in dynamic changes in the neural representations of identical behaviors across different time points (Wen et al., 2023). Collectively, these factors contribute to substantial variability in neural signals, even for the same task performed by the same subject on different days. Consequently, models trained on data from a specific day often exhibit diminished performance when applied to data acquired on subsequent days, rendering cross-day decoding a formidable challenge in BCI research.

Addressing stable cross-day decoding is thus a critical goal in BCI research. For instance, a DRNN (Ran et al., 2019) demonstrated high accuracy and robustness in decoding arm velocity during a macaque monkey's reaching task. Attention-based models, such as the Temporal Attention-aware Timestep Selection (TTS) (Yang et al., 2021), have improved RNN-based neural decoders by selecting key timesteps to enhance accuracy and efficiency. To overcome data limitations, the spatiotemporal Transformer model NDT2 (Ye et al., 2024) leveraged pre-training across sessions, subjects, and tasks, using cross-attention mechanisms and the PerceiverIO architecture to adapt quickly to new sessions with mini-batch labeled data, effectively analyzing diverse neural recordings.

Despite achieving high decoding accuracy, traditional ANN models often suffer from high energy consumption. In contrast, the human brain operates with remarkable energy efficiency. This paper is driven by the need to explore brain-inspired mechanisms for more efficient information processing. As depicted in Figure 1A, the human brain employs parallel processing pathways, specifically the dorsal and ventral streams, to handle visual inputs (Kandel, 2000). These pathways process signals hierarchically in lower-level brain regions, extracting and integrating multiscale features. The functional differentiation between the pathways allows simultaneous extraction of diverse features, which are integrated in higher-level brain regions to form a unified visual perception. This approach contributes to the brain's exceptional signal processing efficiency (Roy et al., 2019). Furthermore, the brain transmits neural signals through discrete action potentials, or "spikes", leading to low energy consumption. SNNs mimic this spike-based communication, where synaptic connections are activated and adjusted only when spikes occur, offering high biological realism. Consequently, SNNs are more energy-efficient in decoding neural signals compared to ANNs (Zhang et al., 2023). In the BCI domain, SNNs play a crucial role in reducing energy consumption, enabling high-throughput invasive BCI systems to achieve high performance while being more compact and extending battery life. Such advancements are essential for the clinical application and commercialization of implantable or portable BCI devices (Makarov et al., 2022).

Inspired by the brain's parallel processing architecture and multiscale feature fusion mechanisms, this paper proposes a method for energy-efficient, invasive cross-day decoding. The main contributions of our work are summarized as follows:

Main contributions:

- We propose a Multiscale Fusion enhanced Spiking Neural Network (MFSNN) that emulates the parallel processing mechanisms of the human visual pathways to achieve efficient signal fusion and feature extraction, thereby significantly enhancing cross-day decoding performance. The MFSNN uses channel attention mechanisms, temporal convolutional networks, and skip connections to capture and integrate the spatial and temporal characteristics of neural signals, demonstrating greater robustness against signal variations caused by electrode drift, inflammation, and neural plasticity.
- Our SNN-based model decodes high-throughput invasive brain signals with reduced energy consumption, offering a practical solution for invasive BCI systems.
- In two invasive BCI paradigms (single-hand grasp-and-touch and center-and-out reach tasks), the MFSNN demonstrates the feasibility and robustness of cross-day decoding through mini-batch supervised generalization learning. This approach enables the model to rapidly adapt to new data distributions with minimal fine-tuning, thereby further enhancing the MFSNN's cross-day generalization capability.

2 Related works

2.1 Brain-inspired computing

The human brain is the only biological system that demonstrates advanced general intelligence with ultra-low power consumption. Insights from the brain have the potential to propel a narrow AI toward a more general one. Brain-inspired computing (BIC) embraces this concept, introducing a new paradigm of computation and learning, inspired by the fundamental structures and information processing mechanisms of the human brain (Liu et al., 2024). It has been found that the parallel processing and multiscale feature fusion are prevalent in brain information processing (Zeki, 2016), and various parallel subsystems extract different aspects of signal features, with feature fusion occurring both within and between subsystems. For example, the human visual system employs complex and organized strategies of parallel processing, hierarchical fusion, and modularity to transform and interpret visual information (Nassi and Callaway, 2009). In the olfactory system, olfactory perception is processed through two parallel pathways within the olfactory bulb (Vaaga and Westbrook, 2016). Similarly, in the pain system, two parallel

pathways, medial and lateral, are responsible for transmitting and processing nociceptive and emotional information, respectively (Wang et al., 2003). Beyond integrating different signal features, the parallel structures in the brain also help reduce operational energy consumption.

With the rapid development of BCI decoding technology, the demand for processing multi-channel and high-sampling-rate signals is increasing. Research has shown that parallel network structures inspired by brain mechanisms have significant advantages over traditional single-network architectures in signal processing and feature extraction. In visual perception (Katsuki and Constantinidis, 2014) and parallel networks in artificial intelligence, the introduction of channel attention mechanisms has become a key approach for handling high-throughput multi-channel signals (Zhao et al., 2023a). This mechanism enhances efficient feature extraction by identifying the relative importance of each parallel channel.

In recent years, channel attention mechanisms have been successfully applied to optimize the performance of Spiking Neural Networks (SNNs). For example, the Multi-dimensional Attention (MA) module proposed by Yao et al. integrates temporal, channel, and spatial attention into SNNs, significantly improving network performance and energy efficiency (Yao et al., 2023). Attention mechanisms enable SNNs to dynamically adjust their responses to different inputs, thereby enhancing the robustness and generalization ability of the network. Cross-Modality Attention (CMA) and other attention mechanisms, have also been successfully applied in SNNs, demonstrating significant advantages in tasks such as image classification and event-based action recognition (Zhou et al., 2024). This paper will further investigate the application advantages of channel attention mechanisms based on SNNs in BCI decoding.

2.2 SNNs in BCI

BCIs are primarily categorized into non-invasive systems, which place electrodes on the scalp but suffer from low signal-to-noise ratios due to transmission losses (Johnson, 2006), and invasive systems, which record directly from the inner brain and have demonstrated remarkable effectiveness in decoding motor signals in animals (Velliste et al., 2008). The integration of CNNs and RNNs has significantly improved the accuracy of decoding invasive signals (Xie et al., 2018; Śliwowski et al., 2022). Emerging methods based on Transformers, such as swin-transformer (Chen et al., 2024) and other transformer alternatives (He, 2021), have shown great potential in managing complex temporal dynamics in large datasets. However, increasing model complexity to handle high-sampling and multi-channel invasive BCIs usually results in substantial computational demands and power consumption, which poses challenges for clinical deployment on chips.

Spiking Neural Networks (SNNs) have garnered significant attention due to their brain-inspired architecture and efficient spatiotemporal processing capabilities. Unlike traditional Artificial Neural Networks (ANNs), SNNs transmit information via binary spike signals, mimicking the behavior of biological neurons. This event-driven characteristic enables SNNs to achieve low power consumption and rapid response times, making them suitable for real-time applications and neuromorphic computing systems (Zhang and Xu, 2021). SNNs have demonstrated broad application prospects in BCIs, effectively processing and decoding neural signals for tasks such as motor control, sensory processing, and cognitive functions (Parameshwara et al., 2021; Guo et al., 2023). For instance, in neurorehabilitation, SNNs can provide real-time feedback to assist in rehabilitation training (Elbasiouny, 2024). In the field of neuroprosthetics, they can control prosthetics with high precision and low latency (Beaubois et al., 2024). In affective computing, SNNs can decode facial expression states to facilitate more intuitive human-machine interactions (Barchid et al., 2023). Additionally, SNNs have shown great potential in the broader field of neuromorphic computing related to BCIs. The Neural Engineering Framework (NEF), a prominent method for neural signal decoding, has been widely applied in neuromorphic computing (Hazan and Ezra Tsur, 2021). The NEF provides a theoretical framework for representing and processing neural signals using spiking neurons, converting high-dimensional neural data into low-dimensional representations to facilitate the design of efficient neuromorphic systems.

This study proposes an SNN-based bio-inspired model for the decoding of invasive BCI signals. This model maintains efficient decoding capabilities while achieving lower power consumption and reduced model complexity, paving the way for the development of efficient and accurate BCI systems.

3 Method

In this section, we provide a detailed description of our proposed method, including the model structure and its operational workflow.

3.1 Overall architecture

Figure 1B presents the overall architecture of our MFSNN algorithm. The input, a recorded spike train with N_c channels from an invasive BCI, is divided into N_s sub-path signals for parallel processing. Each sub-path is then processed by its corresponding sub-encoder for specialized analysis. Each sub-encoder comprises three key components: a linear transformation module for initial signal modification, a channel attention module for enhancing salient features across channels, and a temporal convolution network for extracting temporal dynamics. The outputs from these modules are designated as LT_{out}, CA_{out}, and TCN_{out}, respectively. These outputs are integrated to produce the fusion feature E_i for the ith sub-encoder. The outputs from all sub-encoders, denoted as E_1 through E_{N_s}, are then concatenated to form E_{out}. This consolidated feature vector is passed through a spiking classifier, transforming it into a spiking fusion feature and decoding it to yield the classification result Net_{out}.

In the subsequent sections of this paper, we will provide a comprehensive explanation of the functional roles and contributions of each module within the MFSNN framework, highlighting their synergistic impact on the network's overall

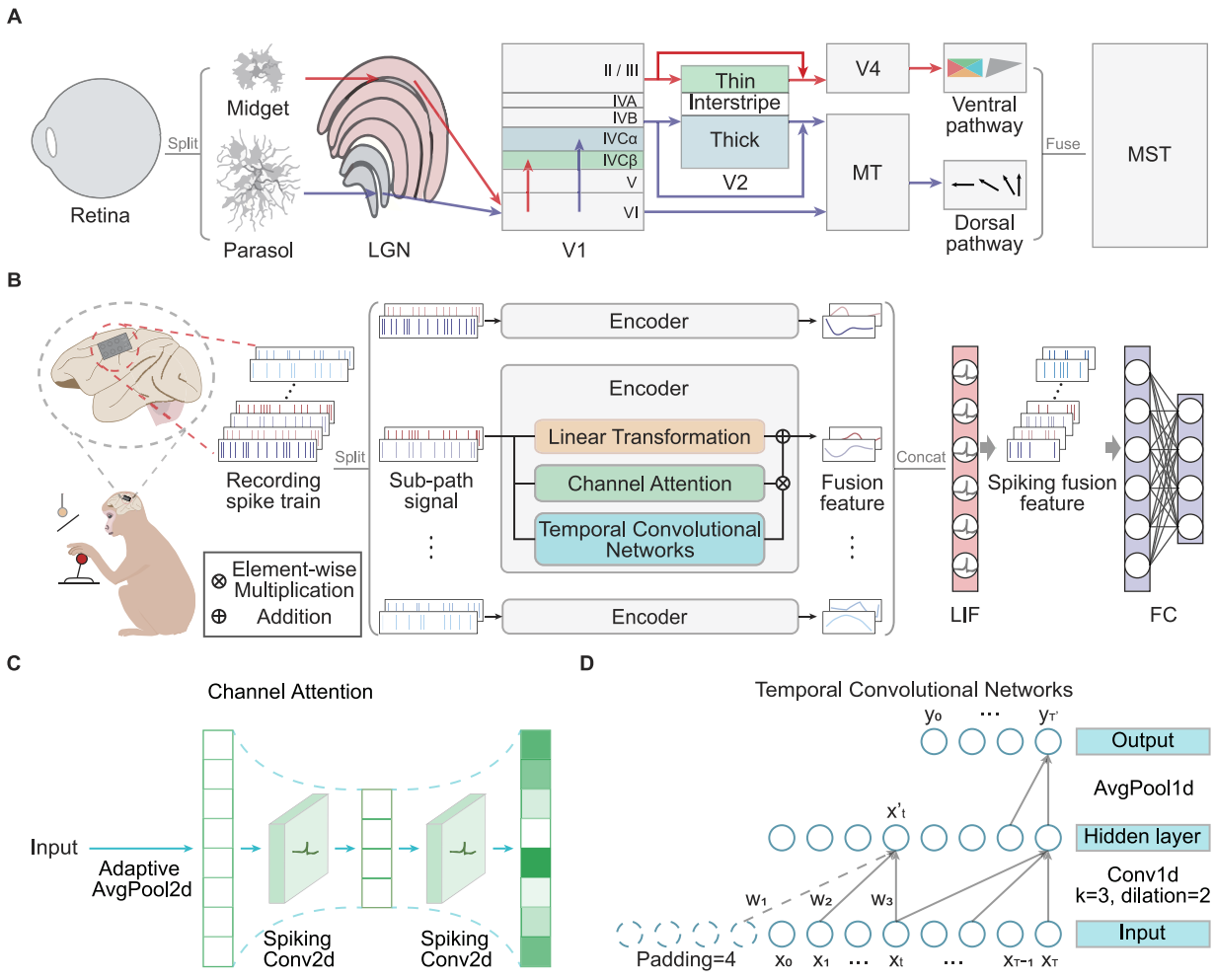

FIGURE 1
(A) The diagram depicts the human brain's visual perception pathway, highlighting its parallel processing mechanisms and multiscale feature fusion. **(B)** The overall MFSNN architecture, where the input consists of multi-channel neural signals from an invasive BCI in monkeys, and the final output represents monkey behavior categories. **(C)** The channel attention module, which includes a bottleneck structure composed of spiking convolutional layers. **(D)** The temporal convolutional network, which incorporates causal and dilated convolutions.

performance in addressing cross-day decoding challenges with high-fidelity neural signal decoding for invasive BCI systems.

3.2 Sub-encoder

3.2.1 Linear transformation (LT)

The given input neural signal is denoted as $Input \in \mathbb{R}^{C\times 1\times T}$, where C represents the channel dimension, jointly determined by the number of raw data channels N_c and the number of sub-encoders N_s, shown as following equations:

$$C = N_c/N_s. \tag{1}$$

At the outset of our processing procedure, a learnable matrix M_l for linear transformation (LT) is implemented. This LT reduces the sequence length from T to T', generating the output $LT_{out} \in \mathbb{R}^{C\times 1\times T'}$.

$$LT_{out} = Input * M_l. \tag{2}$$

This critical step here is to capture the representative features at the raw data level of the sequence, thereby establishing a foundation for the subsequent multiscale feature fusion within the network's architecture.

3.2.2 Channel attention (CA)

In invasive BCI neural signals, each channel captures the activity of different neuronal populations, which contribute variably to the same task. The sub-encoder employs a spiking Channel Attention (CA) module to extract spatial features from the sub-path signal. As depicted in Figure 1C, the CA module first

provides adaptive global average pooling $F_{ap}(\cdot)$ to the $Input \in \mathbb{R}^{C\times1\times T}$ to capture spatial features $f_s \in \mathbb{R}^{C\times1\times1}$ across channels.

$$f_s[C,1,1] = F_{ap}(Input) = \frac{1}{T}\sum_{i=1}^{T} Input[C,1,T]\,. \tag{3}$$

Subsequently, a bottleneck structure, composed of two spiking convolutional layers $SConv2d(\cdot)$ and modified with Leaky Integrate-and-Fire (LIF) neuron-based activation functions $LIF(\cdot)$, is incorporated to compress features first and then expand to derive the channel weight distribution $\tilde{w}_c$.

$$\begin{aligned} SConv2d(\cdot) &= LIF(Conv2d(\cdot)) \\ \tilde{w}_c &= SConv2d(SConv2d(f_s)). \end{aligned} \tag{4}$$

This mechanism allows the model to focus on some specific channels those are more critical for decoding target tasks while effectively suppressing redundancy and noise in the signals. Consequently, this approach enables more precise channel selection, thereby enhancing the accuracy of predictions.

3.2.3 Temporal convolution network(TCN)

Invasive BCI signals are temporal sequences, making the features along the time dimension particularly important. The sub-encoder utilizes a Temporal Convolution Network (TCN) to effectively capture long-term dependencies along with the time dimension through causal convolution and dilated convolution. As shown in Figure 1D, taking a single channel from the $Input \in \mathbb{R}^{C\times1\times T}$ for example, the temporal sequence $X = (x_1, x_2, ..., x_T)$ is subjected to a time convolution with a kernel size of $k = 3$, a dilation rate of $d = 2$, and padding defined as $(k-1)\times d + 1 = 4$. The output dimension of the hidden layer remains unchanged. For any moment x'_t in the hidden layer output $X' = (x'_1, x'_2, ..., x'_T)$ and its corresponding convolution kernel $F_t = (w_1, w_2, w_3)$, the following equation holds:

$$x'_t = b + \sum_{i=1}^{3} w_i \times x_{t-(3-i)\times d}$$

Then, an average pooling operation with a window size of p, denoted as $F_{ap}(\cdot)$, is applied to X' to obtain the module output $Y = (y_0, ..., y_{T'})$ for the temporal sequence of a single channel.

$$\begin{aligned} Y(y_0, ..., y_{T'}) &= F_{ap}(X'(x'_1, x'_2, ..., x'_T)) \\ y_t &= \frac{1}{p}\sum_{i=t-p+1}^{t} x'_t,\ \ p = \frac{T}{T'}. \end{aligned} \tag{5}$$

By performing the aforementioned temporal convolution simultaneously across the C channels of the sub-path signal $Input \in \mathbb{R}^{C\times1\times T}$, the module output is obtained as $TCN_{out} \in \mathbb{R}^{C\times1\times T'}$.

Compared to traditional models for processing temporal signals such as RNN, GRU, and Transformer, the proposed TCN offer superior parallel processing capabilities. All convolutional operations can be computed simultaneously, thereby enhancing computational efficiency and processing speed.

3.2.4 Feature fusion

We integrate the raw data-level features $LT_{out} \in \mathbb{R}^{C\times1\times T'}$ generated by the LT module, the spatial features $\tilde{w}_c \in \mathbb{R}^{C\times1\times1}$ produced by the CA module, and the temporal features $TCN_{out} \in \mathbb{R}^{C\times1\times T'}$ derived from the TCN module to obtain the integrated features of ith sub-encoder $E_i \in \mathbb{R}^{C\times1\times T'}$.

$$E_i = LT_{out} + \tilde{w}_c * TCN_{out}. \tag{6}$$

Subsequently, we concatenate the outputs of all N_s sub-encoders to form the output of the entire signal encoder. The resulting output is represented as $E_{out} \in \mathbb{R}^{N_c\times1\times T'}$.

$$E_{out} = Concat(E_i) \qquad (i = 1, 2, 3, ..., N_s). \tag{7}$$

3.3 Spiking classifier

The spiking classifier is composed of a LIF neuron layer and a fully connected layer, which takes the fused feature E_{out} as the input current, leading to fluctuations in membrane potential and the generation of spikes. Utilizing a spiking layer to extract sparse spike features, the classifier then decodes to achieve the classification result Net_{out}. Moreover, the spiking classifier can generalize with fine-tuning on small samples after pre-training. Its advantages include reduced computational costs and enhanced model adaptability, making it particularly suitable for addressing cross-day decoding issues and for future deployment on neuromorphic chips.

4 Experiments

4.1 Experimental setting

4.1.1 Data details

Dataset 1: The experimental paradigm is depicted in Figure 2A. The macaque monkeys are used as subjects, with each trial divided into four phases. Baseline–the task initiation is marked by the monkey pulling the joystick for 1 s until the appearance of a cue. Preparation–following the cue, the target object appears between 0.1 and 0.5 seconds, after which the monkey releases the joystick. Reach–the monkey releases the joystick and reaches for the position of the ball within approximately 0.5–1 s. Task action–the task concludes upon completion of the touch or grasp action, which lasts over 1 second. Each trial lasts 2–4 s in total. The tasks are categorized into four types: right-hand touch, right-hand grasp, left-hand touch, and left-hand grasp, with each trial conducted independently. Concurrently, 128-channel neural signals from the monkey's motor cortex (M1) are recorded during the trial at a sampling rate of 30 kHz. The experimental data were collected over eight separate days, from 01/26/2022 to 03/09/2022, with an average of approximately 300 trials per day.

Dataset 2: The experimental paradigm is depicted in Figure 3A. The dataset from the research conducted by Churchland et al. (2012), pertains to the collection of neural signals extracted from two rhesus monkeys, identified as J and N, during their

performance of the center-and-out task to eight directions. The signals were recorded using a pair of 96-channel electrode arrays implanted in the M1, with an average of 2,000 trials per day.

4.1.2 Implementation details

All training and testing were conducted on two NVIDIA GeForce RTX 4060 graphics cards. Within the MFSNN network, the spiking neuron model employs the LIF neuron model, with a time window set to 20 ms. During the training process, we train the MFSNN with the Adam optimizer and a batch size of 32. The learning rate was dynamically adjusted using the cosine annealing learning rate schedule, starting from 0.01 and ranging down to 0.0001. In comparative experiments, we adopted a similar training strategy for MLP and GRU.

4.2 Neural signal decoding

4.2.1 Single-hand grasp-and-touch task

In the study of neural signal decoding for single-hand grasp-and-touch task, we focus our analysis on decoding signals from the "Task action" phase. Acknowledging the variability of neural signal characteristics over time, we employ two testing methods: single-day and cross-day decoding. The single-day decoding uses training and testing sets from the same day with an 8:2 ratio; the cross-day decoding uses sets from different days. The data includes eight days spanning from January 26 to March 9, 2022. The single-day decoding experiments were conducted daily, amounting to a total of eight sets.

The cross-day decoding experiment is further divided into two parts: one part trained with data from January 26 and tested with data from January 30 to February 9. The left data were trained with data from March 3 and tested with data from March 6 to 9. Each part consisted of three experiments, making a total of six cross-day decoding tests. During the cross-day decoding process, the model utilized a mini-batch of data for fine-tuning to rapidly adapt to new sessions.

We compare the performance of three algorithms(MLP, GRU, and MFSNN) under the two testing paradigms. As shown in Figure 2B, the average accuracy of the three algorithms under single-day decoding is comparable and notably high(>95%), attributed to the stability of neural signal feature distribution within the same day. That is also why the accuracy rates of single-day decoding are higher than those of cross-day decoding. In cross-day decoding, despite the same fine-tuning apply to MLP and GRU as MFSNN, MFSNN still show a higher accuracy rate(>80%), demonstrating its superior decoding capability in the face of cross-day changes in neural signal characteristics.

Furthermore, we conduct gradient testing on the ratio of fine-tuning data for MFSNN in cross-day decoding to analyze its impact on performance. As shown in Figure 2C, when the fine-tuning data ratio reaches 7.8%, the model performance could be stabilized and good enough(>80%). This slightly higher fine-tuning ratio may be due to the minimal differences among the four types of actions in the task, resulting in more similar neural signals, thus requiring more fine-tuning data to achieve good generalization.

4.2.2 Center-and-out task

In the center-and-out task experiments conducted on monkeys J and N, we test both single-day and cross-day decoding. The single-day decoding task is similar to the single-hand grasp-and-touch task. For cross-day decoding, the model is trained on the first day's data and test on the remaining days. For example, monkey J had data from four days. The network will be trained on the first day and tested on the second, third, and fourth days.

The results, shown in Figure 2B, indicate that all three models achieved very high accuracy rates (≥95%) in single-day decoding, with no significant differences among them. In cross-day decoding, under the same fine-tuning conditions, average accuracy rate of MFSNN is significantly higher than that of MLP and GRU, and the results are more concentrated, indicating stronger generalization and higher robustness.

We also conduct a gradient test of the fine-tuning data ratio for MFSNN on both monkeys to assess its impact on performance. As shown in Figure 3C, with a fine-tuning data ratio of only 3.2%, stable and effective generalization is achieved in both monkeys.

In summary, comparing the results of the two experiments, we find that due to the short time span of single-day data, the distribution of neural signal features do not change significantly, resulting in high and similar accuracy rates for all models. However, in cross-day decoding, the significant changes in the distribution of neural signal features due to the longer time span led to a decrease in the average accuracy rate of all models. Nevertheless, MFSNN, with its excellent multi-level feature fusion mechanism, can stably capture the distribution of neural signals on different days, and with a small sample (<8%) of fine-tuning, it can achieve efficient and robust decoding performance. In practical BCI systems, taking the first task as an example, there are on average 300 trials per day, with each trial lasting 2-4s. With a fine-tuning data ratio of 8%, the total duration amounts to only 48-96s. This indicates that the model can rapidly adapt to the neural signal data distribution of a new session within an extremely short time, thereby achieving long-term stable neural signal decoding and demonstrating highly efficient adaptability.

4.3 Energy consumption

We estimate the theoretical energy consumption of MFSNN and Multiscale Fusion enhanced Artificial Neural Network(MFANN) by the following two equations (Mark, 2014; Yao et al., 2024):

$$SOP_s(l) = Rate \times T \times FLOP_s(l)$$

$$E_{MFSNN} = E_{AC} \times \sum_{i=1}^{16}(SOP^i_{LT} + SOP^i_{CA} + SOP^i_{TCN}). \quad (8)$$

$SOPs(l)$ means synaptic operations (the number of spike-based accumulate(AC) operations) of layer l, Rate is the average firing rate of input spike train to layer l, T is the time window of LIF neurons, and $FLOPs(l)$ refers to the floating point operations (the number of multiply-and-accumulate (MAC) operations) of layer l. We assume

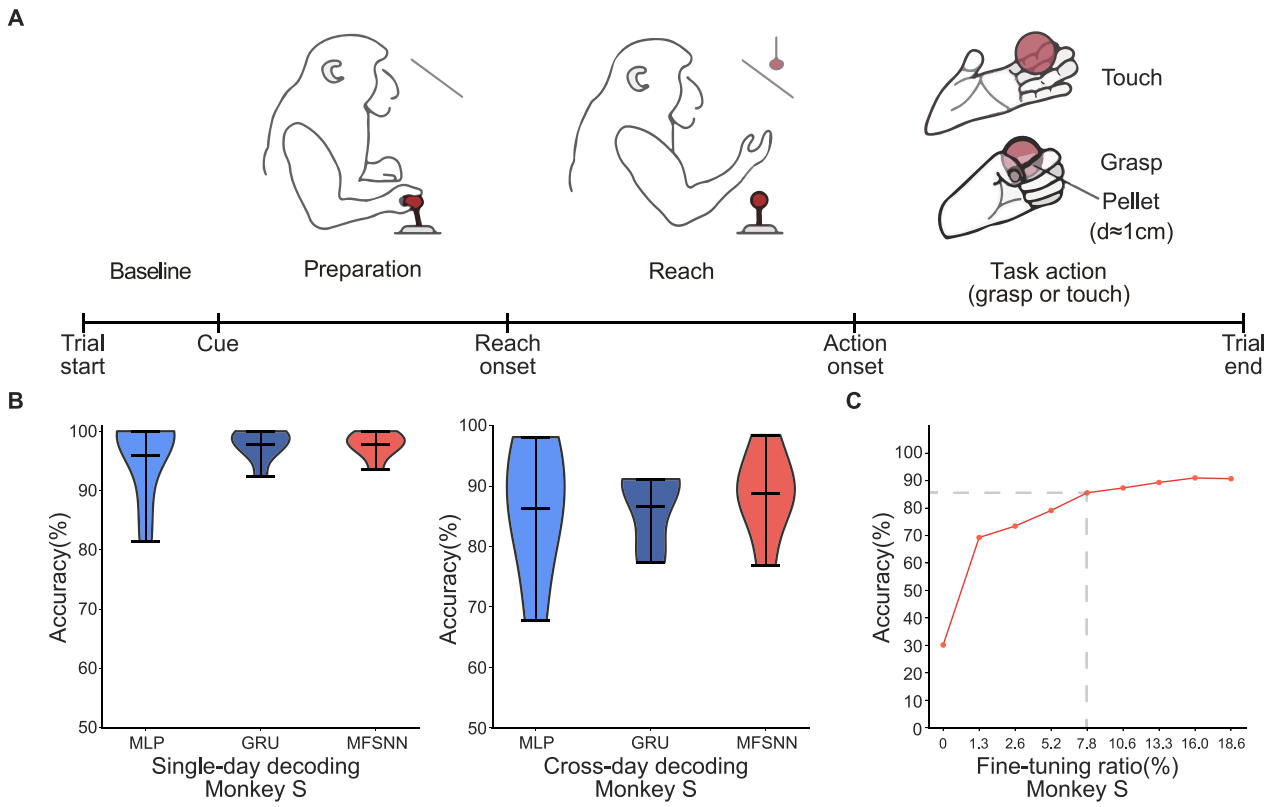

FIGURE 2
(A) The single-hand grasp-and-touch experimental paradigm. **(B)** MLP, GRU, and MFSNN single-day and cross-day decoding experiments on neural signals of monkey S. **(C)** The accuracy under different fine-tuning ratios.

that the MAC and AC operations are implemented on the 45nm hardware (Mark, 2014), with $E_{MAC} = 4.6pJ$ and $E_{AC} = 0.9pJ$.

Under the cross-day decoding test paradigm of dataset 2 with monkey J, the computational energy consumption on 45nm hardware for a single spike train is simulated for MFANN and MFSNN. The results, as shown in Figure 4, indicate that the energy consumption of MFSNN is reduced by 90.9% compared to that of the similarly structured MFANN.

4.4 Ablation study

In the ablation study conduct on the cross-day decoding experiment of monkey J from dataset 2, we scrutinize the roles of the CA, TCN, and LT modules within the MFSNN. Analysis of Figure 5 reveals that all three modules significantly contribute to the model's performance. Particularly, the contributions of LT and TCN are substantial. Given the temporal nature of neural signals, the feature extraction along the temporal dimension by TCN is of paramount importance. LT, on the other hand, focuses on extracting deep features from the raw data, capturing nuances that may be overlooked by temporal and spatial features. These two modules serve as two critical tiers in the multi-level feature fusion strategy of the MFSNN, essential for the model's overall performance.

Upon examining Figure 5, it is observed that the performance distribution of MFSNN without the CA module is represented by a violin plot that is wider at the bottom and narrower at the top, whereas the inclusion of the CA module reverses this, presenting a plot that is narrower at the bottom and wider at the top. This indicates that although the CA module has a limited effect on enhancing the average accuracy, it effectively elevates a number of cross-day decoding outcomes from below to over 80%, a change that holds significant practical implications.

In summary, the results of the ablation study demonstrate that the CA, TCN, and LT modules, which target different aspects of feature extraction, collectively underpin the efficacy of the MFSNN.

5 Conclusion

Stable and long-term decoding of neural signals is crucial for BCIs, which represents an advanced fusion of neuroscience and artificial intelligence. In this paper, we propose the Multiscale Fusion enhanced Spiking Neural Network (MFSNN) framework, which emulates the parallel processing and multiscale feature fusion mechanisms of human visual perception, enabling real-time and energy-efficient neural signal decoding. Our experiments demonstrate that the MFSNN achieves robust cross-day decoding performance in two invasive BCI paradigms involving macaque monkeys performing simple motor tasks (grasp-and-touch and

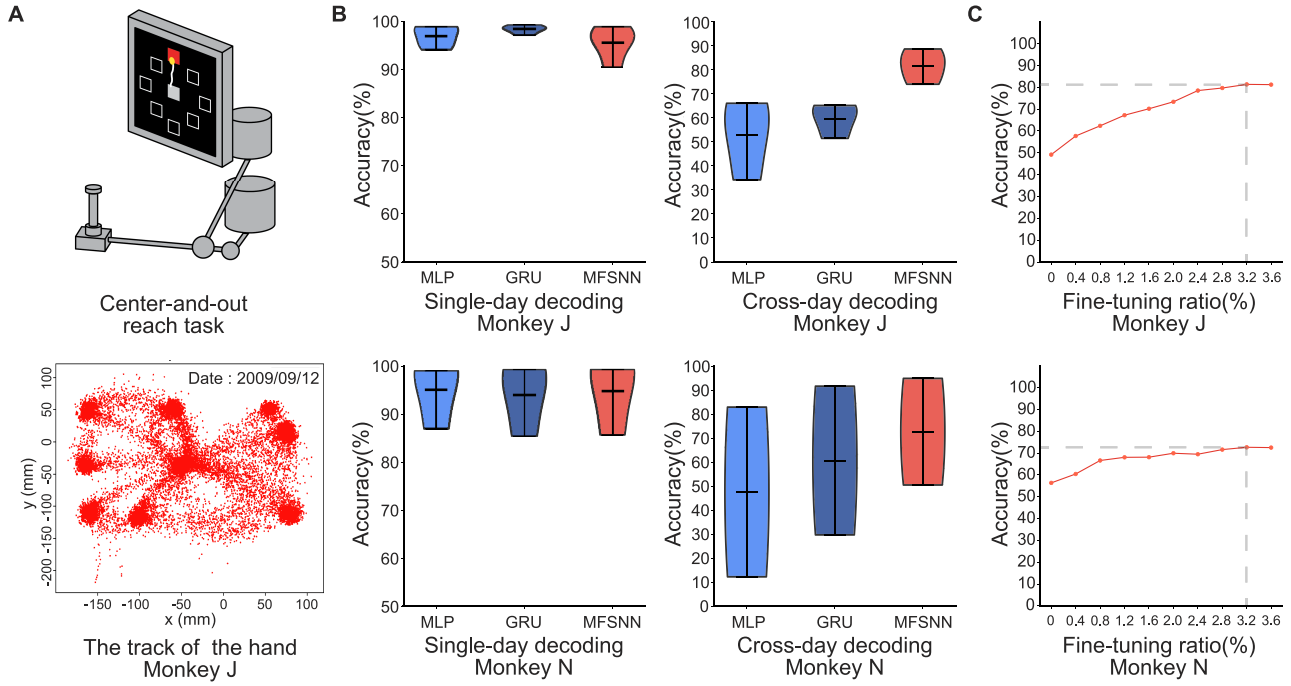

FIGURE 3
(A) Center-and-out experimental paradigm. **(B)** MLP, GRU, and MFSNN single-day and cross-day decoding experiments on neural signals of monkey J and N. **(C)** The accuracy under different fine-tuning ratios.

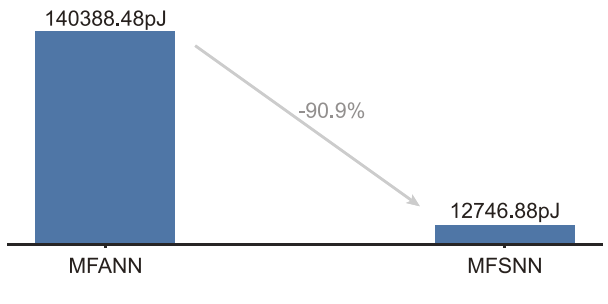

FIGURE 4
Energy consumption for processing a single data entry.

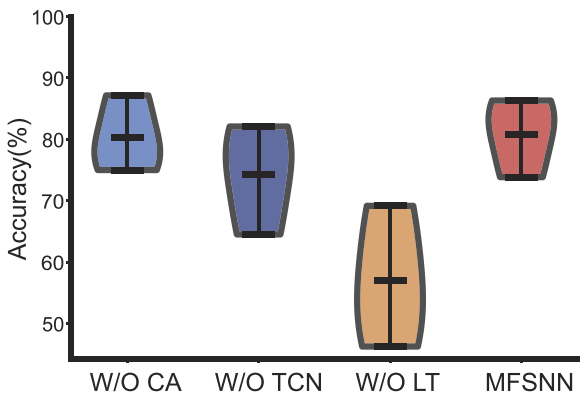

FIGURE 5
The ablation experiments of MFSNN.

center-and-out tasks). However, while the results are promising, the model's transferability to human neural data or more complex real-world scenarios remains untested. Future work should validate the MFSNN's efficacy in human subjects and extend its application to diverse behavioral contexts, such as neuroprosthetics or affective computing, to assess its broader applicability.

To maintain high cross-day decoding accuracy, the MFSNN requires fine-tuning with 8% of the new session's data (equivalent to 48-96 seconds of neural recordings). While this represents a minimal calibration effort compared to full retraining, it may still pose practical challenges for long-term implantable systems where frequent recalibration is infeasible. Reducing reliance on fine-tuning through self-supervised adaptation or leveraging invariant neural representations could enhance the model's autonomy. Additionally, the energy-efficient architecture of the MFSNN, which reduces computational costs by 90.9% compared to ANN-based counterparts, positions it as a viable candidate for deployment on neuromorphic hardware, further supporting long-term usability.

Finally, we propose a future direction inspired by the neural manifold hypothesis (Gallego et al., 2020; Zhao et al., 2024), which posits that high-dimensional neural signals are inherently embedded within low-dimensional latent spaces. The neural manifold, represented in these low-dimensional spaces, encodes the stable behavioral features underlying naturalistic actions. Integrating manifold learning into the MFSNN framework could

involve two stages: 1. dimensionality reduction of raw spike trains to extract invariant neural manifolds, and 2. decoding macroscopic behaviors from these manifolds. This approach may enhance the model's ability to generalize across sessions and subjects by isolating task-relevant neural dynamics from non-stationary noise. Combining multiscale fusion with manifold-based decoding could unlock new avenues for designing stable, long-term BCI systems.

Data availability statement

The datasets presented in this article are not readily available because they intersect with unpublished work from collaborating teams, and public deposition risks compromising ongoing studies. Requests to access the datasets should be directed to songyu2022@ia.ac.cn.

Ethics statement

Ethical approval was not required for the study involving animals in accordance with the local legislation and institutional requirements because This study focuses on the research of signal decoding algorithms and models, without conducting any animal experiments.

Author contributions

YS: Writing – original draft, Writing – review & editing. LH: Writing – original draft, Writing – review & editing. TZ: Writing – review & editing. BX: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was financially supported by Center for Excellence in Brain Science and Intelligence Technology under Project No. LG-GG-202402-06-07.

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.

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.

References

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., et al. (2023). Gpt-4 technical report. *arXiv* [preprint] arXiv:2303.08774. doi: 10.48550/arXiv.2303.08774

Barchid, S., Allaert, B., Aissaoui, A., Mennesson, J., and Djeraba, C. C. (2023). "Spiking-fer: spiking neural network for facial expression recognition with event cameras," in *Proceedings of the 20th International Conference on Content-based Multimedia Indexing* (New York, NY: Association for Computing Machinery), 1–7.

Beaubois, R., Cheslet, J., Duenki, T., De Venuto, G., Carè, M., Khoyratee, F., et al. (2024). Bióemus: a new tool for neurological disorders studies through real-time emulation and hybridization using biomimetic spiking neural network. *Nat. Commun.* 15:5142. doi: 10.1038/s41467-024-48905-x

Chen, C., Yu, Y., Ma, S., Sheng, X., Lin, C., Farina, D., et al. (2020). Hand gesture recognition based on motor unit spike trains decoded from high-density electromyography. *Biomed. Signal Process. Control* 55:101637. doi: 10.1016/j.bspc.2019.101637

Chen, J., Chen, X., Wang, R., Le, C., Khalilian-Gourtani, A., Jensen, E., et al. (2024). Subject-agnostic transformer-based neural speech decoding from surface and depth electrode signals. *bioRxiv*. doi: 10.1101/2024.03.11.584533

Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Foster, J. D., Nuyujukian, P., Ryu, S. I., et al. (2012). Neural population dynamics during reaching. *Nature* 487, 51–56. doi: 10.1038/nature11129

Elbasiouny, S. M. (2024). The neurophysiology of sensorimotor prosthetic control. *BMC Biomed. Eng.* 6:9. doi: 10.1186/s42490-024-00084-y

Gallego, J. A., Perich, M. G., Chowdhury, R. H., Solla, S. A., and Miller, L. E. (2020). Long-term stability of cortical population dynamics underlying consistent behavior. *Nat. Neurosci.* 23, 260–270. doi: 10.1038/s41593-019-0555-4

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. Rehabilitat. Eng.* 31, 1440–1450. doi: 10.1109/TNSRE.2023.3246989

Guo, Y., Huang, X., and Ma, Z. (2023). Direct learning-based deep spiking neural networks: a review. *Front. Neurosci.* 17:1209795. doi: 10.3389/fnins.2023.1209795

Hazan, A., and Ezra Tsur, E. (2021). Neuromorphic analog implementation of neural engineering framework-inspired spiking neuron for high-dimensional representation. *Front. Neurosci.* 15:627221. doi: 10.3389/fnins.2021.627221

He, H. (2021). *Transformer-Based Methods for Neural Decoding.*

Johnson, D. H. (2006). Signal-to-noise ratio. *Scholarpedia* 1:2088. doi: 10.4249/scholarpedia.2088

Kandel, E. (2000). *Principles of Neural Science*. McGraw-Hill.

Katsuki, F., and Constantinidis, C. (2014). Bottom-up and top-down attention: different processes and overlapping neural systems. *Neuroscientist* 20, 509–521. doi: 10.1177/1073858413514136

Kubben, P. (2024). Invasive brain-computer interfaces: a critical assessment of current developments and future prospects. *JMIR Neurotechnol.* 3:e60151. doi: 10.2196/60151

Liu, F., Zheng, H., Ma, S., Zhang, W., Liu, X., Chua, Y., et al. (2024). Advancing brain-inspired computing with hybrid neural networks. *Nat. Sci. Rev.* 11:nwae066. doi: 10.1093/nsr/nwae066

Makarov, V. A., Lobov, S. A., Shchanikov, S., Mikhaylov, A., and Kazantsev, V. B. (2022). Toward reflective spiking neural networks exploiting memristive devices. *Front. Comput. Neurosci.* 16:859874. doi: 10.3389/fncom.2022.859874

Mark, H. (2014). "Computing's energy problem (and what we can do about it)," in *Proceedings of the IEEE International Solid-State Circuits Conference* (San Francisco, CA: IEEE), 9–13.

Nassi, J. J., and Callaway, E. M. (2009). Parallel processing strategies of the primate visual system. *Nat. Rev. Neurosci.* 10, 360–372. doi: 10.1038/nrn2619

Parameshwara, C. M., Li, S., Fermller, C., Sanket, N. J., Evanusa, M. S., and Aloimonos, Y. (2021). "Spikems: deep spiking neural network for motion segmentation," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Prague: IEEE), 3414–3420.

Pun, T. K., Khoshnevis, M., Hosman, T., Wilson, G. H., Kapitonava, A., Kamdar, F., et al. (2024). Measuring instability in chronic human intracortical neural recordings towards stable, long-term brain-computer interfaces. *Commun. Biol.* 7:1363. doi: 10.1038/s42003-024-06784-4

Ran, X., Zhang, S., Sun, G., Zhu, J., Chen, W., and Pan, G. (2019). "Decoding velocity from spikes using a new architecture of recurrent neural network," in *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)* (San Francisco, CA: IEEE), 231–234.

Roy, K., Jaiswal, A., and Panda, P. (2019). Towards spike-based machine intelligence with neuromorphic computing. *Nature* 575, 607–617. doi: 10.1038/s41586-019-1677-2

Schmidgall, S., Ziaei, R., Achterberg, J., Kirsch, L., Hajiseyedrazi, S., and Eshraghian, J. (2024). Brain-inspired learning in artificial neural networks: a review. *APL Mach. Learn.* 2:2. doi: 10.1063/5.0186054

Śliwowski, M., Martin, M., Souloumiac, A., Blanchart, P., and Aksenova, T. (2022). Decoding ecog signal into 3d hand translation using deep learning. *J. Neural Eng.* 19:026023. doi: 10.1088/1741-2552/ac5d69

Vaaga, C. E., and Westbrook, G. L. (2016). Parallel processing of afferent olfactory sensory information. *J. Physiol.* 594, 6715–6732. doi: 10.1113/JP272755

Velliste, M., Perel, S., Spalding, M. C., Whitford, A. S., and Schwartz, A. B. (2008). Cortical control of a prosthetic arm for self-feeding. *Nature* 453, 1098–1101. doi: 10.1038/nature06996

Wang, J.-Y., Luo, F., and Han, J. (2003). The medial and lateral pain system - parallel processing of nociceptive information. *Chin. J. Neurosci.* 19, 416–419.

Wen, S., Yin, A., Furlanello, T., Perich, M. G., Miller, L. E., and Itti, L. (2023). Rapid adaptation of brain-computer interfaces to new neuronal ensembles or participants via generative modelling. *Nat. Biomed. Eng.* 7, 546–558. doi: 10.1038/s41551-021-00811-z

Xie, Z., Schwartz, O., and Prasad, A. (2018). Decoding of finger trajectory from ecog using deep learning. *J. Neural Eng.* 15:036009. doi: 10.1088/1741-2552/aa9dbe

Yang, S.-H., Huang, J.-W., Huang, C.-J., Chiu, P.-H., Lai, H.-Y., and Chen, Y.-Y. (2021). Selection of essential neural activity timesteps for intracortical brain-computer interface based on recurrent neural network. *Sensors* 21:6372. doi: 10.3390/s21196372

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. doi: 10.48550/arXiv.2404.03663

Yao, M., Zhao, G., Zhang, H., Hu, Y., Deng, L., Tian, Y., et al. (2023). Attention spiking neural networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 9393–9410. doi: 10.1109/TPAMI.2023.3241201

Ye, J., Collinger, J., Wehbe, L., and Gaunt, R. (2024). Neural data transformer 2: multi-context pretraining for neural spiking activity. *Adv. Neural Inf. Process. Syst.* 36:558113. doi: 10.1101/2023.09.18.558113

Zeki, S. (2016). Multiple asynchronous stimulus-and task-dependent hierarchies (stdh) within the visual brain's parallel processing systems. *Eur. J. Neurosci.* 44, 2515–2527. doi: 10.1111/ejn.13270

Zhang, T., Cheng, X., Jia, S., Li, C. T., Poo, M.-,m., and Xu, B. (2023). A brain-inspired algorithm that mitigates catastrophic forgetting of artificial and spiking neural networks with low computational cost. *Sci. Adv.* 9:eadi2947. doi: 10.1126/sciadv.adi2947

Zhang, T.-L., and Xu, B. (2021). Research advances and perspectives on spiking neural networks. *Chin. J. Comp.* 44, 1767–1785. doi: 10.11897/SP.J.1016.2021.01767

Zhao, X., Sun, Y., Zhang, T., and Xu, B. (2023a). Local convolution enhanced global fourier neural operator for multiscale dynamic spaces prediction. *arXiv* [preprint] arXiv:2311.12902. doi: 10.48550/arXiv.2311.12902

Zhao, X., Sun, Y., Zhang, T., and Xu, B. (2024). Enhanced spatiotemporal prediction using physical-guided and frequency-enhanced recurrent neural networks. *arXiv* [preprint] arXiv:2405.14504. doi: 10.48550/arXiv.2405.14504

Zhao, X., Zhang, D., Liyuan, H., Zhang, T., and Xu, B. (2023b). Ode-based recurrent model-free reinforcement learning for pomdps. *Adv. Neural Inf. Process. Syst.* 36, 65801–65817. doi: 10.5555/3666122.3668994

Zhou, S., Yang, B., Yuan, M., Jiang, R., Yan, R., Pan, G., et al. (2024). Enhancing snn-based spatio-temporal learning: A benchmark dataset and cross-modality attention model. *Neural Netw.* 180:106677. doi: 10.1016/j.neunet.2024.106677