

Federated Spiking Neural Networks with Cross-Modal Attention for Energy-Efficient Stroke Detection in Epileptic EEG/ECG Signals

Radha R¹, Raja Sarath Kumar Boddu²

¹Postdoctoral Researcher, Lincoln University; ² Raghu Engineering College, Visakhapatnam, India

Email ID - pdf.radha@lincoln.edu.my; iamrajaboddu@gmail.com

Abstract: People with chronic epilepsy are more likely to have a stroke than those without epilepsy. The need for continuous monitoring for early detection of strokes is clinically essential, but technically difficult because of energy use limitations and privacy concerns with traditional AI models in this group. We propose a model that combines Explainable AI (XAI), federated learning (FL), and spiking neural networks (SNNs). A cross-modal spiking attention mechanism is proposed to integrate the EEG and ECG signals and reject epileptic artifacts. For CHB-MIT data, this method obtains a noise reduction rate of 60 percent and the signal fidelity of 93.1 percent. The federated learning protocol features spike-gradient compression, which lowers the amount of communication bandwidth required by 62% while allowing for collaboration across institutions without compromising privacy. Temporal spike-attribution maps enable clinicians to have meaningful information for making model decisions. The results of Phase 1 indicate an accuracy of 97.2% in spike encoding. Early-stage 2-3 results show an impressive 12.4 mW power usage, or a 23-fold reduction in energy compared to traditional CNNs.

Keywords: Federated Learning, Stroke Detection, Neuromorphic Computing, Explainable AI, EEG/ECG Signal Processing, Cross-modal Attention

Introduction

1.1 Clinical Motivation

Millions of people around the world suffer from epilepsy. Some conditions which make epilepsy difficult to treat are called comorbidities and stroke is one which is particularly serious. Patients with chronic epilepsy have a markedly higher incidence of stroke than the general population. This increased risk places a strong need for ongoing and reliable monitoring to initiate timely interventions.

Healthcare is becoming a stronger industry for intelligent systems to monitor patients and assist in decision making. However, the current AI monitoring technologies face two challenges: wearables use too much power for widespread adoption, and all data is centrally collected, causing privacy concerns.

1.2 The Research Gap

The current strategies have four interrelated weaknesses. Epilepsy monitoring and stroke detection are usually considered as distinct tasks, and this is because of the lack of consideration for the co-occurrence

of these two conditions in real patients. Although accurate, these deep neural networks require a lot of power to be continuously used on a wearable. Centralized training necessitates sharing of private patient information; this is out of the question in many clinical settings. And often high performing models are black boxes that a clinician cannot check.

Such type of gap has been reported in the latest reviews on multimodal EEG analysis and neuromorphic computing in healthcare. The current research tackles all four constraints in one.

1.3 Contributions

By combining the three complementary technologies: energy efficient computation with spiking neural networks, privacy preserving training with federated learning, and clinical interpretability with explainable AI effective outcomes are possible. The architecture aims to detect strokes in epilepsy patients, hence in a noisy neurophysiological environment, caused by epileptic activity.

Heart of this approach is the cross-modal spiking attention mechanism. This dynamically balances the EEG and ECG signals and rejects epileptic artifacts, while maintaining the stroke signals. With federated learning and spike-gradient compression, multi-institutional collaboration is possible without sharing raw data. Temporal spike-attribution maps display the time windows and modalities which activate stroke alerts.

2. Methodology

2.1 Spike Encoding

The IF neuron model is used to transform continuous EEG and ECG signals into sparse binary spike data. The motivation for this step is that it is possible to save energy in event-driven processing. The membrane potential changes over time, as follows:

$$V_{mem}(t) = V_{mem}(t-1) \times \beta + I_{input}(t)$$

When the voltage $V_{mem}(t)$ hits threshold voltage θ , a spike is produced:

$$S(t) = \begin{cases} 1, & V_{mem}(t) \geq \theta \\ 0, & \text{otherwise} \end{cases}$$

Once a spike is fired, the potential becomes reset and the neuron is put into a short refractory period. We assume $\theta=1.0$, $\beta=0.9$ and refractory period is 5 samples, as per the CHB-MIT data.

The CHB-MIT Scalp EEG Database contains 22 pediatric patients with intractable seizures recorded at 16-bit resolution and 256 Hz. The data consists of 664 .edf files and 198 annotated seizure events. Some recordings are available with both ECG channels and EEG channels.

2.2 Cross-Modal Spiking Attention

The attention mechanism is used to map EEG and ECG spikes into a common representation:

$$\begin{aligned} P_{eeg}(t) &= W_{eeg} \times S_{eeg}(t) \\ P_{ecg}(t) &= W_{ecg} \times S_{ecg}(t) \end{aligned}$$

The attention score is the result of a context transformation:

$$A(t) = \tanh(W_{context} \times [P_{eeg}(t)P_{ecg}(t)])$$

The attention weights are computed using a spike-based softmax:

$$w_{eeg}(t) = \frac{\exp(A_{eeg}(t))}{\exp(A_{eeg}(t)) + \exp(A_{ecg}(t)) + \epsilon}$$

The weighted contributions are fused together in the output of the spike.

$$S_{fused}(t) = w_{eeg}(t) \times S_{eeg}(t) + w_{ecg}(t) \times S_{ecg}(t)$$

This mechanism focuses on cleaner modalities in the presence of epileptic noise, which is very important for reliable detection of stroke in this patient group.

2.3 Federated Learning with Compression

In federated learning, the raw patient data is not shared and only the models are trained collaboratively. In SNNs, we make two changes to this paradigm. One is essentially that one can only communicate via spike-gradients, which are gradients related to an active neuron. Second, the gradients are clipped in order to keep only top k% of the magnitude values:

$$\text{compressed_grad} = \begin{cases} \text{grad}, & |\text{grad}| > \tau \\ 0, & \text{otherwise} \end{cases}$$

The compression threshold τ can be tuned to the desired compression rate. This technique significantly decreases the bandwidth without losing valuable training signals.

2.4 Energy Benchmarking

The architecture is simulated with a software framework for neuromorphic computing. Energy-per-synaptic-operation (SOP) is:

$$E_{total} = \text{SOP}_{count} \times E_{per_SOP}$$

On the basis of Loihi specifications, $E_{per_SOP} = 1 \text{ pJ}$ sets a theoretical upper bound on the energy consumption.

2.5 Temporal Spike-Attribution Maps

T-SAM visualizes how stroke predictions are different for various time windows and modalities. Attribution scores are:

$$\text{Attribution}(t) = \frac{\partial \text{Output}}{\partial S_{fused}(t)}$$

This method helps clinicians make decisions when using a model, which is a barrier to clinical use of a model.

3. Results

Table I. Phase 1 Validation Results

Metric	Target	Achieved	Status
Spike Encoding Accuracy	>95%	97.2%	Complete
Sparsity	>70%	72.3%	Complete
Noise Reduction	>50%	60.0%	Complete
Signal Fidelity	>90%	93.1%	Complete
Epileptic Artifact Suppression	>70%	91.0%	Complete

Table II. Federated Learning Validation Results

Metric	Target	Achieved	Status
Privacy Budget (ε)	<10	4.2	In Progress
Bandwidth Reduction	>60%	62%	In Progress
Accuracy Gap (vs Centralized)	<3%	1.8%	In Progress
Non-IID Robustness	>95%	96.8%	In Progress

Table III. Power Consumption Comparison

Architecture	Power (mW)	Battery Life (Days)	Energy Gain
CNN	285.0	2.8	Baseline
LSTM	245.0	3.2	1.16×
GRU	210.0	3.8	1.35×
Vanilla SNN	52.0	15.4	5.48×
SNN + Attention	38.5	20.8	7.40×
Proposed	12.4	64.5	23.0×

4. Discussion

4.1 Research Questions

Energy efficiency: The proposed architecture has 23× more energy efficiency compared to CNN-based models (estimated battery life of 64.5 days). This implies that it is appropriate for long-term patient monitoring.

Privacy preservation: The federated learning protocol reduces bandwidth by 62%, and keeps privacy budget $\epsilon = 4.2$. This allows for hands-on multi-institutional training without the risks of exposing patient information.

Clinical interpretability: T-SAM offers visual attribution maps which enable the clinician to find out which segments of the signals led to an alert. This is one way to address the black-box concern that has constrained the clinical use of AI systems.

4.2 Significance

Cross-modal spiking attention achieves a 60% noise reduction, which is clinically relevant. Its anti-epileptic function eliminates epileptic artifacts, keeping the number of false alarms low but maintaining the indicators of a stroke. This is an improvement over traditional fusion method.

4.3 Limitations

The number of subjects in the CHB-MIT dataset is small (22 subjects), which reduces generalizability. Not all recordings contain electrocardiogram (ECG) signals, limiting cross-modal training. Commercial neuromorphic chips need to be further validated for clinical use. An evaluation by practicing neurologists is required for the T-SAM module.

5. Conclusion

This paper presented a federated energy-efficient explainable spiking neural network (SNN) for cross-modal stroke detection in epilepsy patients. According to the Phase 1 results, 97.2% of the data is encoded correctly, and the noise reduction is 60%. Initial Phase 2-3 results indicate a 62% bandwidth reduction, and 12.4 mW power consumption, which is 23× more efficient than CNN-based approaches.

The article addresses key challenges of energy efficiency, privacy protection and clinical interpretability in continuous patient monitoring. This work offers a benchmark for developing next-generation intelligent healthcare systems by showing how feasible stroke detection is in a low power, privacy-preserving manner.

References

- [1] Zhao, S., Li, W., Wang, X., & Chen, Y. (2025). A systematic review of machine learning methods for multimodal EEG data in clinical application. *arXiv*. <https://doi.org/10.48550/arXiv.2501.08585>
- [2] Ma, N., Wang, W., et al. (2026). SCAF-Net: A spiking cross-modal attention fusion network for multimodal emotion recognition. *Biomedical Signal Processing and Control*, 118.
- [3] Intelligent Computing for Healthcare: A Survey. (2025). *IEEE Xplore*.
- [4] Yu, D., Du, X., et al. (2024). Exploiting Label Skewness for Spiking Neural Networks in Federated Learning. *arXiv*, accepted at IJCAI 2025.
- [5] Spiking Neural Networks for the Analysis of Physiological Signals. (2026). *Directory of Open Access Journals*.
- [6] Pan, Y., Zou, S.-H., Yang, J.-W., et al. (2026). Surgical Data Science in Time-Critical Contexts: A Roadmap Toward Brain-Inspired Computing. *Journal of Computer Science and Technology*, 41(1): 428-446. <https://doi.org/10.1007/s11390-025-5908-8>
- [7] Rohi, G., & Sille, R. (2025). Neuromorphic horizons: Exploring the intersection of generative AI and neuromorphic computing in healthcare. *Taylor & Francis*.
- [8] Emerging memory devices for neuromorphic computing in the Internet of Medical Things. (2025). *Cell*.
- [9] Akram, M. S., Varma, B. S., Javed, A., Harkin, J., & Finlay, D. (2025). Toward TinyDPFL systems for real-time cardiac healthcare: Trends, challenges, and system-level perspectives. *Journal of Systems Architecture*, 168.

- [10] Jenab, K. (2026). Neuromorphic computing for healthcare engineering: A bibliometric analysis. *Healthcare Engineering*, 2(4): 199-216.
- [11] Towards neuromorphic neurotechnologies: integrating brain-inspired computing with brain-computer interfaces. (2026). *npj Biomedical Innovations*, 3, 51. <https://doi.org/10.1038/s44385-026-00074-w>