A Comparative Study of SNN Learning Paradigms in Hybrid Graph-Spiking Neural Networks for Biomedical Data Analysis

Dominik Čambál, Monika Herchlová, Juraj Ďuďák
Slovak University of Technology in Bratislava
Faculty of Materials Science and Technology in Trnava
Jána Bottu 25, 917 24 Trnava, Slovakia
{dominik.cambal, monika.herchlova, juraj.dudak}@stuba.com

Abstract. *This study evaluates three spiking neural network (SNN) paradigms within a hybrid Graph-SNN framework for 12-lead ECG arrhythmia classification under severe class imbalance (*$10.3 : 1$*). We benchmark backpropagation through time (BPTT), ANN-to-SNN conversion, and unsupervised STDP on the INCART database across paired partitions. While conversion and BPTT achieve comparable ranking discrimination, conversion yields vastly superior calibration, enabling* 0.933 *minority beat recall and* 97.3% *accuracy. STDP fails to form a usable classifier. Requiring half the training time of BPTT, conversion offers the most effective quality-per-unit-compute trade-off.*

Keywords. graph neural networks, spiking neural networks, ECG classification, SNN learning paradigms

1 Introduction

Electrocardiogram (ECG) signal classification exhibits a dual structure; signals are sequential in time and relational across leads. A standard 12-lead ECG does not represent independent measurements, but 12 geometric projections of a single cardiac dipole in a three-dimensional space, constrained by electrode placement and cardiac physiology (Kligfield et al., 2007).

This structure is particularly evident in the precordial leads (V1-V6), which form a spatial trajectory across the anterior chest wall. As their position shifts relative to cardiac anatomy, so does their perspective on the depolarization wavefront, resulting in systematic morphological variations such as R-wave progression and QRS polarity changes. These patterns do not emerge independently at each lead, but arise from a shared underlying process observed from multiple points of view (Bulíková, 2014). Treating leads as independent channels, therefore, discards structural information that is intrinsic to the signal.

Conventional deep learning approaches only address one dimension of the structure. Convolutional neural networks, whether applied per-lead or over multiple stacked lead inputs, are effective local feature extractors, but without memory the classification depends only on the current receptive field and not on the temporal context accumulated across the beat window (Hannun et al., 2019). Recurrent architectures (e.g. LSTM) address the temporal dimension by maintaining a hidden state across timesteps, but represent leads (inputs) as a flat feature vector, collapsing their inter-lead relational structure.

Graph neural networks (GNNs) applied to a 12-lead ECG, where leads are modelled as nodes and their pairwise statistical dependencies as edges, allow message-passing over the lead topology at each timestep. That way, the GNNs extract inter-lead spatial features at each beat sample, while a downstream stateful temporal module integrates those features across the beat window.

The spatial-relational decomposition processing by a GNN along with a temporal integration by a stateful architecture fits the duality of multi-lead ECGs well. A hybrid GCN-GRU for the detection of atrial fibrillation was demonstrated to outperform an equivalent GRU applied to concatenated lead features by approximately six percentage points of F1, directly attributing the gain to the graph-structured spatial module (Ma and Xia, 2023).

The motivation for instantiating the temporal module as a spiking neural network (SNN) rather than an LSTM brings a two-fold value.

First, SNNs encode information at the rate or timing of discrete binary spike events and enable a sparse, asynchronous computation (Eshraghian et al., 2023) which has been shown to yield great reductions in energy on neuromorphic hardware per inference compared to GPU-based deployment in the Von Neumann sense (Malcolm and Casco-Rodriguez, 2023). The primary clinical context for arrhythmia detection, in continuous monitoring battery-powered wearable devices over days to weeks, directly impacts this efficiency argument (Chu et al., 2022).

Second, the temporal structure of ECG morphology

is naturally aligned with the integrate-and-fire neuron model. The beat features are present as a sequence of transient events in which timing relationships carry diagnostic information with an encoding consistent with the membrane potential dynamics of leaky integrate-and-fire neurons (LIF). Prior work on GNN-SNN hybridisation has addressed the spatial module. Zhu et al. (Zhu et al., 2022) introduced a spiking graph convolutional network that combines GCN-based spatial aggregation with spike-based neural processing, yielding competitive results on benchmark datasets. Wang and Jiang (Wang and Jiang, 2022) embedded spiking dynamics directly into the graph attention mechanism, reporting energy reductions over continuous-valued GATs. Both architectures treat the graph as static and target the spatial component only; neither addresses the problem of integrating a GNN encoder with a spiking temporal integrator over time-varying graph signals, which is the structure required for beat-level ECG classification. Moreover, no existing study compares multiple SNN learning paradigms within a fixed GNN-SNN hybrid framework on a class-imbalanced biomedical task.

All three principal training paradigms; backpropagation through time with surrogate gradients, ANN-to-SNN conversion via threshold-balanced shadow training, and unsupervised spike-timing-dependent plasticity (STDP); differ in computational complexity and cost, efficiency, biological plausibility, i.e. in reliance on supervisory signals, and overall optimisation dynamics (Yi et al., 2023). Their behavior under the pronounced class imbalance typical of arrhythmia datasets, such as the 10:1 Normal-to-Ventricular ratio in the INCART database, is not sufficiently understood.

This work addresses these gaps by implementing a hybrid GNN-SNN architectures in which a graph convolutional encoder processes the 12-lead ECG at each time step, and a spiking temporal module performs sequence integration. The temporal component is trained using three distinct paradigms; firstly, end-to-end BPTT with surrogate gradients (SNNTorch), secondly, ANN-to-SNN conversion via threshold-balanced shadow training (SpikingJelly), and lastly, unsupervised STDP (BindsNET). All variants are evaluated on binary arrhythmia classification (Normal vs Ventricular beats) using the St. Petersburg INCART 12-lead database, consequently, they are compared to a GNN-LSTM baseline, trained under identical conditions.

The following parts of the paper are organised as follows: Section 2 presents the theoretical foundations and methodology, including dataset construction. Section 3 describes the experimental setup and results. Sections 4 and 5 discuss the findings and outline directions for further work.

2 Methodology

In this section, we describe key concepts related to graph neural networks, recurrent and spiking neural networks and the three major learning paradigms of spiking neural networks.

2.1 Graph Neural Networks

Graph neural networks process signals defined on graph structures by iteratively updating node embeddings through neighborhood aggregation. In the 12-lead ECG setting, each lead is represented as a node and inter-lead dependencies as edges, allowing spatial information to be exchanged before temporal integration.

At layer k, the spatial message-passing mechanism computes messages from neighboring nodes $u \in \mathcal{N}(v)$, aggregates them, and updates the hidden representation of node v:

$$m_{u\to v}^{(k)} = \text{Message}^{(k)}(h_u^{(k-1)}, h_v^{(k-1)}, e_{u,v}), \quad (1)$$

$$m_v^{(k)} = \text{Aggregate}^{(k)}(\{m_{u\to v}^{(k)} : u \in \mathcal{N}(v)\}), \quad (2)$$

$$h_v^{(k)} = \text{Update}^{(k)}(h_v^{(k-1)}, m_v^{(k)}). \quad (3)$$

In this work, we use graph convolutional layers because they provide an efficient form of degree-normalized neighborhood aggregation, making them suitable for repeated processing of ECG lead graphs across timesteps.

2.1.1 Recurrent Neural Networks and Backpropagation Through Time

Recurrent neural networks maintain a hidden state across timesteps, allowing predictions to depend on both the current input and previously observed sequence context:

$$\mathbf{h}[t] = \tanh(\mathbf{W}\mathbf{x}[t] + \mathbf{U}\mathbf{h}[t-1]). \quad (4)$$

Gated variants such as LSTMs and GRUs mitigate the vanishing gradient problem by regulating information flow through internal gates. These models are usually trained by backpropagation through time (BPTT), where the recurrent computation is unrolled over the sequence and gradients are propagated through shared temporal parameters (Zhang et al., 2023). The repeated product of temporal Jacobians explains why long-sequence training may still suffer from vanishing or exploding gradients, motivating alternative stateful temporal modules such as spiking neural networks.

2.1.2 Spiking Neural Networks

Spiking neural networks (SNNs) are inspired by biological neurons. Unlike conventional neural networks, their neurons communicate through time-dependent spikes, represented as spike trains of ones and zeroes.

Several spiking neuron models exist, differing in biological plausibility and computational efficiency.

A commonly used model is the Leaky Integrate-and-Fire (LIF) neuron. Its internal state, called the membrane potential $U[t]$, accumulates input and decays over time. When it crosses a threshold θ, a spike is emitted and the membrane potential is reset:

$$U[t] = \beta U[t-1] + WX[t] - S_{out}[t-1]\theta. \quad (5)$$

Here, β is the decay rate, W are synaptic weights, $X[t]$ are inputs, and $S_{out}[t-1]$ denotes whether a spike occurred in the previous time step. If the decay is removed ($\beta = 1$), the model becomes the simpler Integrate-and-Fire (IF) model.

SNNs are motivated mainly by sparsity and energy efficiency. Information may be encoded in the rate or timing of spikes, allowing asynchronous and parallel computation. This makes SNNs particularly suitable for neuromorphic hardware, where memory and computation are integrated locally and energy consumption can be significantly reduced (Malcolm and Casco-Rodriguez, 2023).

However, SNNs still face important challenges, especially in training. Standard backpropagation is difficult to apply because spike generation is non-differentiable, leading to the so-called dead neuron problem. As a result, SNNs often underperform conventional neural networks on large benchmark datasets, although high accuracy has been reported on simpler datasets such as MNIST and CIFAR-10 (Eshraghian et al., 2023; Yi et al., 2023).

2.2 SNN Learning Paradigms

Training conventional neural networks relies on backpropagation and gradient descent, which require continuous and differentiable activation functions. In SNNs, however, the activation function is usually the Heaviside step function. Since it is not differentiable in the classical sense, gradients become either zero or undefined, which prevents direct application of standard backpropagation and causes the dead neuron problem (Eshraghian et al., 2023).

In practice, several approaches are used to train SNNs, including surrogate gradients, ANN-to-SNN conversion, and more biologically plausible local learning rules. These methods are often heuristic rather than fully rigorous.

2.2.1 Surrogate Gradient

The surrogate gradient method replaces the derivative of the non-differentiable spike function during training with the derivative of a smooth approximation, such as a sigmoid, arctangent, or triangular function (Eshraghian et al., 2023). The forward pass still uses spikes, but the backward pass uses the surrogate derivative.

This allows BPTT to be applied to SNNs by treating the membrane potential similarly to the hidden state in RNNs. For the LIF model, the temporal derivative includes the reset mechanism:

$$\frac{\partial U[k+1]}{\partial U[k]} = \beta - \theta \frac{\partial S[k]}{\partial U[k]}, \quad (6)$$

where the non-differentiable term is replaced by a surrogate gradient during training.

Surrogate gradients are currently considered the de facto standard for training SNNs because they can achieve sufficient accuracy. Their disadvantages include biological implausibility, increased memory and computational cost, and the need to tune surrogate functions, thresholds, and time constants carefully (Ayasi et al., 2025; Zenke and Vogels, 2023).

2.2.2 Shadow Training

Shadow training is based on training a conventional artificial neural network and then converting it into an SNN. Neuron activations are transformed into spikes using rate or temporal coding. The main advantage is that efficient ANN training methods can be used, which often improves accuracy on complex datasets. However, the conversion itself is approximate and may reduce the energy-efficiency benefits of SNNs, while also lacking biological plausibility (Eshraghian et al., 2023).

2.2.3 Spike-timing-dependent Plasticity

Spike-timing-dependent plasticity (STDP) is an unsupervised learning rule inspired by biological synaptic plasticity and the Hebbian principle. It updates synaptic weights according to the relative timing of presynaptic and postsynaptic spikes. If the presynaptic spike occurs shortly before the postsynaptic spike, the synapse is strengthened; if it occurs after, the synapse is weakened.

This mechanism is often described by

$$\Delta W = \begin{cases} A_+ e^{\Delta t/\tau_+}, & t_{post} > t_{pre}, \\ -A_- e^{-\Delta t/\tau_-}, & t_{post} < t_{pre}. \end{cases} \quad (7)$$

Here, ΔW is the weight update, A_+ and A_- determine the magnitudes of potentiation and depression, and τ_+ and τ_- are time constants. STDP is biologically plausible and local, but on its own it is usually insufficient for competitive performance and is often combined with supervisory signals or architectural modifications (Eshraghian et al., 2023; Hebb, 1949).

2.3 Dataset Description and Class Imbalance

We utilized the St. Petersburg INCART 12-lead Arrhythmia Database (Tihonenko et al., 2007), which

comprises 75 annotated 30-minute recordings extracted from 32 Holter records at a sampling rate of 257 Hz. To construct our dataset, we combined data from the first five recordings. Because the first two recordings contain exclusively normal (`N`) and ventricular ectopic (`V`) beats, all other beat classes were omitted to formulate a binary classification problem.

The resulting dataset exhibits severe class imbalance, containing 10,967 `N` beats and 1,066 `V` beats (an imbalance ratio of approximately 10.3:1).

2.4 Preprocessing and Data Partitioning

During preprocessing, each lead within every beat window was independently normalized using z-score normalization:

$$z = \frac{x - \mu}{\sigma} \tag{8}$$

where μ and σ denote the mean and standard deviation of the specific lead window. Independent lead scaling eliminates amplitude variance across leads and subjects, forcing models to learn structural waveform morphology. Inter-lead graph edges were established based on spatial correlation, retaining connections with a correlation coefficient exceeding 0.5.

The preprocessed beat dataset was partitioned into training, validation, and test sets using an 70:10:20 ratio. Data partitioning was performed on a beat-wise basis rather than a patient-wise basis. This design decision was made to establish a controlled intra-dataset comparison of SNN learning paradigms under identical sample dynamics rather than a clinical generalization study. We discuss the implications of beat-wise splitting regarding patient generalization in Section 6.

2.5 Training Setup, Optimization, and Dynamic Thresholding

All models were trained using an original pipeline designed to produce the initial benchmarks and an updated pipeline designed to handle class imbalance, prevent overfitting, and optimize decision boundaries. Presented in this study are only the results of the updated pipeline, although the results from both were comparable.

To prevent gradient bias toward the majority class during training, mini-batches were drawn using PyTorch's `WeightedRandomSampler`. This forces an approximately balanced 50:50 distribution of `N` and `V` beats in every training batch without altering the underlying loss function formulation. Validation and test sets maintained their original 10.3:1 natural class distribution to accurately evaluate real-world performance.

Models were optimized using the Adam optimizer with an initial learning rate of $\eta = 10^{-3}$. A dynamic learning rate scheduler reduced η by half (factor of 0.5) whenever validation performance plateaued for 3 consecutive epochs.

To prevent overfitting, early stopping was implemented based on the Precision-Recall Area Under the Curve (PR-AUC) evaluated on the un-resampled validation set. PR-AUC was chosen over standard accuracy or ROC-AUC due to its sensitivity under severe class imbalance. Training was halted if validation PR-AUC failed to improve for 8 consecutive epochs, and the model weights corresponding to the highest validation PR-AUC were restored for testing.

Gradient-based architectures (Base GNN-LSTM, GNN-SNN BPTT, and ANN-to-SNN conversion) were allocated a maximum budget of 30 epochs. In contrast, the unsupervised STDP model was capped at 10 epochs. This difference was dictated by computational complexity and learning dynamics:

1. **Computational Cost:** Due to BindsNet's unbatched, sample-level execution requirement (batch size $B = 1$), each epoch required significant processing time ($\sim$ 42 minutes per epoch, totaling 7 hours for 10 epochs).

2. **STDP Instability:** Empirical tracking revealed that the STDP model achieved its highest validation metrics during the first epoch. Subsequent epochs caused monotonic performance degradation, confirming that standard unsupervised Hebbian weight updates without supervised lateral inhibition or threshold stabilization cannot converge over multi-epoch temporal training.

Because models trained under 50:50 oversampling output predicted probabilities calibrated to a balanced prior, applying the default classification threshold (0.5) to un-resampled evaluation sets causes severe overprediction of the minority class (resulting in high false positives).

To restore optimal decision boundaries, threshold calibration was conducted on the validation set prior to test evaluation:

- Probability decision thresholds were swept from 0.05 to 0.95 in increments of 0.005.

- The optimal threshold T^* was selected by maximizing the Matthews Correlation Coefficient (MCC) on validation data. MCC was selected because it symmetrically evaluates all four quadrants of the confusion matrix and remains robust under class imbalance.

- The tuned threshold T^* was subsequently locked and applied to the test set for final reporting.

3 Experimental Setup and Results

This section describes our base architecture, based on which the three GNN-SNN architectures were developed, and the results from our experiments.

3.1 Base GNN-LSTM Architecture

As a baseline, we designed and implemented an end-to-end hybrid GNN-LSTM architecture using PyTorch and PyTorch Geometric (PyG). The model operates directly on raw graph-structured ECG windows without manual feature extraction and consists of three components:

- **Spatial GNN:** Composed of two graph convolution layers with ReLU activation functions to extract spatial inter-lead relationships at each timestep.
- **Temporal GNN:** Iterates sequentially through each timestep t in the ECG window, passing temporal slices through the Spatial GNN to form a sequence of 768-dimensional node-aggregated embeddings.
- **Hybrid Classifier Head:** Flattens the Temporal GNN output sequence and feeds it into a two-layer LSTM (hidden size 256), followed by a linear classification head for binary output.

After hyperparameter tuning, this baseline achieved a training accuracy of 99.88% and a testing accuracy of 99.75%.

3.2 GNN-SNN Architectures

To evaluate different spiking neural network paradigms, we redesigned the temporal integration head of the base GNN-LSTM into three distinct GNN-SNN architectures, maintaining end-to-end training across all variants. Each architecture adapts the presentation of spatial features to accommodate the specific learning mechanisms of its underlying SNN framework. All models preserved equivalent parameter capacities in their temporal heads for fair comparison.

3.2.1 BPTT with Surrogate Gradient

For gradient-based recurrent spiking, we implemented a GNN-SNN model using the snnTorch library. The spatial GNN feature extractor remains unchanged, while the temporal LSTM is replaced by a two-layer spiking recurrent network with Leaky Integrate-and-Fire (LIF) neurons (hidden size 256).

Information is presented sequentially across time steps. The temporal loop carries the continuous LIF membrane potentials forward across timesteps t, mirroring the hidden state persistence of an LSTM while matching its parameter count. To optimize the network via BPTT, fast sigmoid is utilized as a surrogate activation function during the backward pass. Rather than averaging discrete spike counts over short observation windows – which introduces high variance – classification is performed on the final LIF layer's continuous membrane potential. This final membrane voltage provides a smooth, differentiable continuous signal ideal for standard cross entropy loss optimization.

3.2.2 ANN-to-SNN Conversion

To leverage direct rate-coded conversion, a shadow ANN was implemented in SpikingJelly and converted to an SNN via threshold balancing. Because standard graph convolution operations are not natively supported in SpikingJelly's conversion pipeline, the Spatial GNN remains an ANN feature extractor, while the temporal head is converted to SNN format.

The temporal LSTM was replaced with a 1D CNN comprising three convolutional blocks ($768 \to 512 \to 256$ channels). The kernel sizes narrow progressively ($7 \to 5 \to 3$) to capture broad temporal context before fine-grained refinement. At the resampled sampling frequency ($f_s = 250$ Hz), a kernel size of 7 spans $\sim$ 27 ms, explicitly matching the duration of high-frequency ECG components such as the QRS complex. Three blocks provide sufficient depth without gradient saturation over short input windows ($T = f_s/2 \approx 125$ samples).

Information is presented as a 2D spatial-temporal feature map. Following convolution, a Global Average Pooling layer collapses the temporal axis into a 256-dimensional vector before passing to a linear classifier. Global average pooling naturally translates to mean firing rates in spiking space, making the representation optimal for the `ann2snn` conversion framework. The shadow network is trained using ReLU activations – ensuring mathematical equivalence to Integrate-and-Fire (IF) neuron firing rates (Pietrzak et al., 2023) – and is converted by calibrating maximum pre-activations across a calibration set (Rueckauer et al., 2017).

3.2.3 STDP

To evaluate local biological learning, we constructed a GNN-SNN model utilizing spike-timing-dependent plasticity (STDP) via the BindsNet library. The LSTM block is replaced with an unsupervised spiking layer consisting of 768 InputNodes connected to 256 LIFNodes via PostPre STDP connections, followed by a supervised `Linear(256, 2)` classification head.

Because BindsNet operates at the sample level, batch sizes are reduced to 1. At each timestep t, the continuous GNN feature embedding is converted into a probabilistic Poisson spike train and fed sequentially into the network.

To prevent activity state leakage between distinct ECG samples, a fresh BindsNet network instance is initialized per sample. However, to preserve unsupervised learning across the dataset without LSTM hidden states, we introduce a persistent weights buffer (768×256). This weight matrix is updated after each beat sample via an Exponential Moving Average, simulating slow Hebbian consolidation. STDP weight updates occur automatically and unsupervised during the forward pass. Supervised learning is isolated strictly to the final linear classifier, which maps the output LIF layer's final membrane voltages to class logits via

Adam optimization.

4 Results

Each paradigm was evaluated across three paired experimental runs utilizing distinct random data partitions and weight initializations governed by split seeds (42, 101, and 2023). All paradigms share identical data partitions to enable paired contrasts, thereby isolating paradigm behavior from dataset partition difficulty. Table 1 summarizes test-set performance, calibration, and training duration across all paradigms.

4.1 Ranking Discrimination vs. Calibration Dynamics

Under threshold-independent ranking metrics (PR-AUC and ROC-AUC), the two gradient-based paradigms perform comparably. Although shadow training achieves a higher marginal mean PR-AUC (0.779 vs. 0.697), the paired per-seed PR-AUC differences are $+0.008$, $+0.245$, and -0.007. The entire marginal gap is driven by a single partition (seed 101); excluding this outlier, shadow training averages 0.701 PR-AUC against 0.697 for BPTT with surrogate gradient. Given an observed within-paradigm spread of ± 0.028 for BPTT, performance differences below $\sim$ 0.06 PR-AUC fall within variance noise. Consequently, ranking quality between conversion and surrogate gradient optimization is statistically indistinguishable in this setup.

However, a marked performance divergence emerges in thresholded decision boundaries and output calibration:

- **Minority Class Detection:** Shadow training achieves a `V`-class recall of 0.933 ± 0.038 compared to 0.610 ± 0.027 for BPTT. The run ranges do not overlap.
- **Calibration Error:** Expected Calibration Error (ECE) for shadow training (0.007 ± 0.004) is an order of magnitude lower than BPTT (0.073 ± 0.016).

Because both paradigms order beats similarly well, this divergence indicates that their output probabilities occupy fundamentally different scales. Validation-tuned decision thresholds (T^*) transfer effectively to test sets under shadow training but underperform under BPTT. For BPTT, the selected operating point ($T^* = 0.825$) forfeited approximately one-third of available minority beat recall.

4.2 Partition Stability and Supervised Pre-Training

The paired experimental design reveals a distinct difference in stability across data partitions. Across the three seeds, BPTT exhibits consistent average precision (0.674, 0.688, 0.728; $\sigma = 0.028$). Conversely, shadow training displays high dispersion (0.682, 0.933, 0.721; $\sigma = 0.135$) – a variance nearly five times larger on identical splits.

This instability originates during the supervised source ANN pre-training phase rather than during conversion. The source ANNs achieved validation average precision scores of 0.575, 0.974, and 0.861 across partitions. Thus, while shadow training provides superior mean calibration and reduced computational overhead, individual runs exhibit lower predictability compared to end-to-end BPTT.

4.3 Unsupervised STDP Failure Modes

The unsupervised STDP paradigm failed to establish a functional decision boundary, yielding a PR-AUC of 0.133 ± 0.017 (near the no-skill baseline of 0.0885) and ROC-AUC of 0.603 ± 0.011. Because local Hebbian updates operate without class label supervision under severe imbalance ($10.3 : 1$), synaptic weight updates are dominated by majority-class waveform dynamics. Multi-epoch training exacerbated this failure: validation performance peaked during Epoch 1 and deteriorated monotonically thereafter, confirming that unsupervised local plasticity without lateral inhibition or supervised error feedback cannot converge on imbalanced sequence classification tasks.

5 Sensitivity Analysis and Readout Ablation

To determine whether the superior calibration of ANN-to-SNN conversion stems from its architecture or its output readout mechanism, we conducted an ablation study. BPTT outputs classification logits directly from unbounded, unnormalized LIF membrane potentials. Conversely, the converted SNN classifies from rate-normalized spike counts accumulated over time.

The conversion pipeline provides an internal control: the exact same trained parameters and extracted GNN features were evaluated under two conditions: (1) as the source ANN with a continuous logit readout, and (2) as the converted SNN with a spike-rate readout. Table 2 details the paired results.

The ablation separates ranking quality from calibration:

1. **Ranking Quality:** Rate coding slightly reduces ROC-AUC across all partitions, reflecting discrete quantization loss during conversion. PR-AUC paired differences alternate in sign ($+0.098$, -0.033, -0.150), indicating no systematic ranking advantage for rate readouts.
2. **Calibration Improvement:** Calibration improves consistently across every partition. ECE

Table 1. Test-Set Performance and Efficiency Across Three Paired Runs (Split Seeds 42 / 101 / 2023)

Paradigm	Parameters	PR-AUC	ROC-AUC	V-Recall	V-F1 Score	Accuracy	ECE	Train Time (h)
Shadow Training (Conversion)	1,907,650	**0.779 ± 0.135**	**0.966 ± 0.016**	**0.933 ± 0.038**	**0.860 ± 0.087**	**0.973 ± 0.015**	**0.007 ± 0.004**	**2.52 ± 0.37**
BPTT w/ Surrogate Gradient	267,460	0.697 ± 0.028	0.903 ± 0.023	0.610 ± 0.027	0.657 ± 0.015	0.940 ± 0.000	0.073 ± 0.016	5.03 ± 0.04
STDP (Unsupervised)	4,802 + 196,608*	0.133 ± 0.017	0.603 ± 0.011	–	–	–	–	7.91 ± 0.20

* Indicates 4,802 trainable parameters (GNN + classification head) and 196,608 persistent unsupervised STDP weight buffer parameters.

Table 2. Readout Ablation: Source ANN (Logit Readout) vs. Converted SNN (Rate Readout) across Three Paired Partitions using Identical Weights and Features.

Metric	ANN (Logit)	SNN (Rate)
PR-AUC	0.807 ± 0.199	0.779 ± 0.135
ROC-AUC	0.982 ± 0.011	0.966 ± 0.016
Sensitivity	0.923 ± 0.042	0.933 ± 0.036
ECE	0.032 ± 0.013	**0.007 ± 0.004**

drops by factors of 6.5, 4.7, and 3.9 under rate readout, with non-overlapping ranges (0.018–0.043 for logit vs. 0.0038–0.0112 for rate).

This demonstrates that accumulating discrete spikes over time naturally normalizes output variance, yielding well-calibrated probabilities that transfer robustly across decision thresholds.

6 Threats to Validity

We identify several threats to validity:

- **Sample Size and Significance:** Evaluation across three paired seeds provides descriptive spread and paired ordering but lacks statistical power for formal significance testing.
- **BPTT Optimization Constraints:** The BPTT model reached the 30-epoch ceiling without triggering early stopping. Reported BPTT metrics represent a lower bound on performance; extended training budgets could narrow the observed calibration gap, though at increased computational cost.
- **Beat-Wise Data Partitioning:** Splitting data at the beat level rather than the patient level allows subject-specific morphological features to appear in both training and test sets. While valid for comparative benchmarking of SNN learning paradigms under identical feature distributions, absolute metrics should not be interpreted as clinical cross-subject generalization performance.

7 Code Availability

To promote scientific transparency and reproducibility, all source code is publicly available at: https://github.com/xcambald/snn-gnn-paradigms.

8 Discussion

The experimental results demonstrate that temporal learning rules dominate overall performance and operational reliability in hybrid GNN-SNN architectures. While graph convolutions effectively capture spatial inter-lead relationships, temporal integration under severe class imbalance remains the primary bottleneck.

ANN-to-SNN conversion offers an advantageous efficiency-performance trade-off. Conventional optimization avoids surrogate gradient approximations, while rate-coded spike accumulation provides intrinsic regularizing effects that yield well-calibrated output probabilities. However, initialization sensitivity during source ANN training requires multiple restarts to ensure stability.

For surrogate-gradient BPTT, optimization instability and poor probability scaling highlight the need for specialized loss formulations (e.g., focal loss or class-weighted cross-entropy adapted for membrane dynamics) and dynamic neuron parameters (such as learnable firing thresholds or adaptive decay rates). Unsupervised STDP in its standard formulation is unsuited for imbalanced temporal classification, necessitating supervised feedback mechanisms such as reward-modulated STDP.

9 Conclusion

This study presented a controlled comparative analysis of three SNN learning paradigms integrated with spatial GNN encoders for ECG arrhythmia classification. Under realistic class imbalance (10.3 : 1), the ANN-to-SNN conversion paradigm achieves superior classification performance, minority beat recall, calibration, and training efficiency. End-to-end BPTT achieves competitive discrimination ranking but suffers from output miscalibration. Unsupervised STDP fails to converge to a usable decision boundary. Future work should focus on neuromorphic graph operations, hardware deployment metrics, and advanced gradient formulations tailored to event-driven processing.

Acknowledgments

This work was supported by the Slovak University of Technology in Bratislava (STU) under the internal grant scheme “Mladý výskumník” (Project No. 1364) awarded to Dominik Čambál.

References

Ayasi, B., Carmona, C. J., Saleh, M., & García-Vico, A. M. (2025). A Practical Tutorial on Spiking Neural Networks: Comprehensive Review, Models, Experiments, Software Tools, and Implementation Guidelines. *Eng*, *6*(11), 304. https://doi.org/10.3390/eng6110304

Bulíková, T. (2014). *EKG pre záchranárov nekardiológov* [ECG for Non-Cardiologist EMS Workers]. Grada Slovakia; Grada.

Chu, H., Yan, Y., Gan, L., Jia, H., Qian, L., Huan, Y., Zheng, L., & Zou, Z. (2022). A Neuromorphic Processing System With Spike-Driven SNN Processor for Wearable ECG Classification. *IEEE Transactions on Biomedical Circuits and Systems*, *16*(4), 511–523. https://doi.org/10.1109/tbcas.2022.3189364

Eshraghian, J. K., Ward, M., Neftci, E. O., Wang, X., Lenz, G., Dwivedi, G., Bennamoun, M., Jeong, D. S., & Lu, W. D. (2023). Training Spiking Neural Networks Using Lessons From Deep Learning. *Proceedings of the IEEE*, *111*(9), 1016–1054. https://doi.org/10.1109/jproc.2023.3308088

Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., & Ng, A. Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine*, *25*(1), 65–69. https://doi.org/10.1038/s41591-018-0268-3

Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. John Wiley & Sons.

Kligfield, P., Gettes, L. S., Bailey, J. J., Childers, R., Deal, B. J., Hancock, E. W., van Herpen, G., Kors, J. A., Macfarlane, P., Mirvis, D. M., Pahlm, O., Rautaharju, P., & Wagner, G. S. (2007). Recommendations for the Standardization and Interpretation of the Electrocardiogram. *Journal of the American College of Cardiology*, *49*(10), 1109–1127. https://doi.org/10.1016/j.jacc.2007.01.024

Ma, H., & Xia, L. (2023). Atrial Fibrillation Detection Algorithm Based on Graph Convolution Network. *IEEE Access*, *11*, 67191–67200. https://doi.org/10.1109/ACCESS.2023.3291352

Malcolm, K., & Casco-Rodriguez, J. (2023). A Comprehensive Review of Spiking Neural Networks: Interpretation, Optimization, Efficiency, and Best Practices. https://doi.org/10.48550/ARXIV.2303.10780

Pietrzak, P., Szczęsny, S., Huderek, D., & Przyborowski, Ł. (2023). Overview of Spiking Neural Network Learning Approaches and Their Computational Complexities. *Sensors*, *23*(6), 3037. https://doi.org/10.3390/s23063037

Rueckauer, B., Lungu, I.-A., Hu, Y., Pfeiffer, M., & Liu, S.-C. (2017). Conversion of Continuous-Valued Deep Networks to Efficient Event-Driven Networks for Image Classification. *Frontiers in Neuroscience*, *11*. https://doi.org/10.3389/fnins.2017.00682

Tihonenko, V., Khaustov, A., Ivanov, S., & Rivin, A. (2007). St.-Petersburg Institute of Cardiological Technics 12-lead Arrhythmia Database. https://doi.org/10.13026/C2V88N

Wang, B., & Jiang, B. (2022). Spiking gats: Learning graph attentions via spiking neural network. https://arxiv.org/abs/2209.13539

Yi, Z., Lian, J., Liu, Q., Zhu, H., Liang, D., & Liu, J. (2023). Learning rules in spiking neural networks: A survey. *Neurocomputing*, *531*, 163–179. https://doi.org/10.1016/j.neucom.2023.02.026

Zenke, F., & Vogels, T. P. (2023). The Remarkable Robustness of Surrogate Gradient Learning for Spiking Neural Networks. *Neural Computation*, *35*(4), 899–925. https://doi.org/10.1162/neco_a_01566

Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2023, December). *Dive into Deep Learning*. Cambridge University Press.

Zhu, Z., Peng, J., Li, J., Chen, L., Yu, Q., & Luo, S. (2022). Spiking graph convolutional networks. https://arxiv.org/abs/2205.02767