Probabilistic Spiking Neural Networks on Spintronic Accelerators vs. Monte Carlo Dropout in LLMs

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: How does the inference latency of probabilistic spiking neural networks on specialized spintronic accelerators compare to software-based Monte Carlo dropout in large language models under equivalent. 9 claims were extracted from source literature; 5 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 7.0/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: SpinAPS: A High-Performance Spintronic Accelerator for Probabilistic Spiking Neural Networks. Research question: How does the inference latency of probabilistic spiking neural networks on specialized spintronic accelerators compare to software-based Monte Carlo dropout in large language models under equivalent accuracy constraints?.

2 Methodology

Systematic literature search across multiple databases yielded 15 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.0/10.

3 Results

15 papers retrieved. 9 claims extracted; 5 independently verified. Quality review score: 7.0/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
The research was supported in part by the National Science Foundation grant #1710009.	×	0.01
The research was supported in part by the CAMPUSENSE project grant from CISCO Systems Inc.	×	0.03
Resources of the High-Performance Computing facility at NJIT were used in this work.	×	0.06
SpinAPS implements a principled direct learning rule for first-to-spike decoding without the need for conversion from pr	✓	0.30
SpinAPS achieves comparable performance with an equivalent ANN on handwritten digit and human activity recognition bench	✓	0.27
Software emulation shows SpinAPS achieves a 4$\times$ performance improvement in terms of GSOPS/W/mm2 compared to an equivalent	✓	0.26
SpinAPS achieves 75% of the test performance in as few as 4 algorithmic time steps on the handwritten digit benchmark.	✓	0.25
The performance of probabilistic SNNs is evaluated on two benchmarks: handwritten digit recognition and human activity r	✓	0.18
The first-to-spike rule is evaluated for the first time in this work.	×	0.05

References

- http://arxiv.org/abs/2008.02189v1
- http://arxiv.org/abs/2604.22179v1
- http://arxiv.org/abs/2004.14942v1