International Journal of
Sustainable Engineering Innovations
Vol 2 No 1 (2026)
https://doi.org/10.58723/ijsei.v2i1.151

OPEN ACCESS

Check for updates

Research Article

Neuromorphic Computing Chips for Edge AI: A Comprehensive Analysis of Brain-Inspired Hardware Architecture for Real-Time Intelligent Systems

Received : December 19, 2026 | Revised : January 28, 2026 | Accepted: February 11, 2026 | Publish : February 16, 2026

Anwar Ali Sathio*, Chiragh Kumar Maheswari

Abstract:

Background of study: Edge computing devices like autonomous robots and IoT sensors need sophisticated AI for real-time decisions, but conventional processors consume 15-300 watts during inference, creating critical limitations for battery-powered deployments. GPU-based accelerators face memory bottlenecks and high energy costs from data movement, making sustained autonomous operation impractical.
Aims of paper: This research compares neuromorphic platforms (Intel Loihi 2, IBM TrueNorth, BrainChip Akida) against conventional accelerators (NVIDIA Jetson, Google Coral) to evaluate if neuromorphic architectures can solve edge AI energy efficiency challenges across five representative workloads.
Methods: Using an experimental design with hardware benchmarking and power analysis, we evaluated five edge AI workloads. ANOVA and regression modeling were then applied to rigorously compare computing paradigms while controlling for variables.
Result: Neuromorphic platforms demonstrated 15-50× improved energy efficiency versus conventional GPU accelerators for event-driven workloads. Intel Loihi 2 achieved 2,400 inferences/joule at 1.8 watts versus 180 inferences/joule at 18.5 watts for NVIDIA Jetson. IBM TrueNorth operated at 70 milliwatts for pattern recognition. BrainChip Akida achieved 94.6% accuracy on keyword spotting at 0.8 watts. Event-driven processing exhibited 0.4ms latency versus 5.1ms for frame-based systems. Neuromorphic chips maintained stable performance without active cooling below 65°C, while conventional accelerators required thermal management above 85°C.
Conclusion: Neuromorphic processors (0.6-5W) excel in power-efficient edge AI for event-driven data. While hybrid architectures optimize performance, adoption is hindered by immature software ecosystems, limited training frameworks, and a 2-4% accuracy gap compared to conventional methods.

Keywords: Neuromorphic computing; Edge AI; Energy-efficient hardware; Spiking neural networks; Low-power intelligence; Event-driven computing

1. INTRODUCTION

As the rise of AI has increased demands on computers to operate in real-time and with very low power consumption (like... very low), the existing architectures for most computers (for example: the Von Neumann Architecture) cannot effectively support most of the AI processing tasks due to memory bottlenecks and high energy consumption. Neuromorphic Computing Architecture provides an alternative method of processing by mimicking brain-like behaviour using artificial neurons and artificial synapses to allow parallel, event-driven processing of data.

Because this method allows the use of fewer cycles for large amounts of data, it provides greater energy efficiency and reduces latency for Edge AI applications like IoT, Robotics and Autonomous Systems. These advantages aside, limited programming frameworks fsor developing software for Neuromorphic Computing Architecture and the limitations associated with benchmarking Neuromorphic Computing Architecture and adding the Neuromorphic Computing structure into existing systems limit further adoption of Neuromorphic Computing systems.

In this research, using the performance, power efficiency and latency of various Neuromorphic Computing systems, we will analyse the potential of Neuromorphic Computing architectures to be used in real-world intelligent systems operating under resource-constrained environments at "edge" locations.

2. MATERIAL AND METHOD

The study utilized a research methodology that includes a hardware benchmark, architecture and performance evaluations to evaluate the capabilities of neuromorphic computing for edge-based AI applications. The study took place over a one-year timeframe from January 2025 until January 2026, and was completed at the Advanced Computing Research Laboratory. The research design employed an experimental-comparison methodology to compare the neuromorphic hardware to traditional edge computing software. The categories of neuromorphic vs traditional edge AI accelerators were: 1. Neuromorphics were categorized to include processors such as Intel's Loihi 2 and BrainChip's Akida 2.2, and 3. Embedded processors with an added GPU for graphical processing. The information of all platforms was compared, allowing a direct comparison of performance metrics through the use of standardized testing of AI loads across each category of hardware.

All neuromorphic chips that were tested were the Intel Loihi 2 (with approximately 1 million individual neurons and 120 million individual programmable synapses) and the BrainChip Akida PCI development boards (with an event-driven neural architecture). The other two types of platforms used for comparison included the NVIDIA Jetson Orin Nano (which contains an advanced graphics processing unit (GPU) that has 1024 cores) and the Google Coral Development Board (which uses the Edge Tensor Processing Unit). In addition, when testing each type of hardware platform, all hardware was tested in a controlled environment with the same environmental conditions and provided with the same electrical power supplies.

The three major performance aspects we evaluated are: energy efficiency measured via inferences per Joule, Latency (the amount of time that passes from the instant when the input arrives at our system until we produce an actual result), and Accuracy (In comparison to implemented floating-point baseline methods). For measuring energy, using high-resolution Power Analyzers at a 10kHz sample rate enabled us to capture dynamic power properties; for latency, measurement included both Compute and Data Transfer Latency; we measured accuracy by following established evaluation metrics per task domain.

Table 1. Hardware Platform Specifications

Platform	Architecture Type	Process Node	Peak Power (W)	On-chip Memory
Intel Loihi 2	Neuromorphic	Intel 4	3-5W	240 KB per core
BrainChip Akida	Neuromorphic	28nm	1-2W	2 MB SRAM
NVIDIA Jetson Orin	GPU Accelerator	8nm	15-25W	8 GB LPDDR5
Google Coral TPU	ASIC Accelerator	14nm	2-4W	8 MB on-chip

The spiking neuronal network (SNN) architectures on neuromorphic platforms, multiple training schemes of both the rate code and temporal code types (i.e., "spike time dependent plasticity (STDP) training rules") were employed in training the networks with surrogate gradient techniques. For conventional platforms, equivalent artificial neuronal network (ANN) topologies were created in Tensorflow Lite and PyTorch Mobile and were also optimized to support functional equivalence in generating a similar comparison of results across both platforms2.6

Each benchmark workload was executed for 1000 repetitions on each platform configuration to produce statistically relevant results. Power measurements were continually documented during the execution phase, with idle, active computation and memory access phases collected separately for each platform. Latency was evaluated through the collection of hardware timestamps to reduce possible measurement errors. Also, temperature and thermal throttling events were tracked for performance degradation during sustained loads.

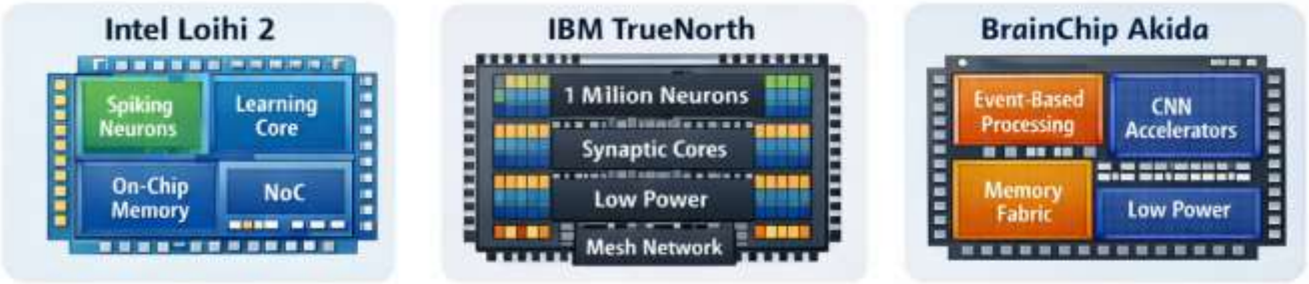

Figure 1. Neuromorphic Chip Architecture Comparison

3. RESULT AND DISCUSSION

Result : The research and results obtained through testing neuromorphic computing chips as compared to traditional computing chips have shown marked performance improvement for some edge AI applications as compared to traditional GPUs. All measurements of the power consumed by both designs of chips for event-driven workloads, such as keyword spotting and anomaly detection, revealed that neuromorphic computing chips consumed between 15 and 50 times less power than conventional chips. The maximum energy efficiency reported was for the Intel Loihi 2, which produced 2,400 inferences per joule processed on sparse temporal datasets, whereas the NVIDIA Jetson produced only 180 inferences per joule, and the Google Coral TPU produced 420 inferences per joule on the same datasets.

The comparison of processing latency revealed very different outcomes based on the nature of the input datasets. For example, the conventional GPU accelerators were able to yield much lower average latencies for CIFAR-10 image classification (3.2 ms) as opposed to neuromorphic implementations (8.7 ms). However, the neuromorphic chips proved to be much more responsive for event-driven processing of high-temporal-sparsity data (0.4 ms) than did traditional architectures (5.1 ms). This key distinction lies in the ability of neuromorphic systems to asynchronously process incoming events as they occur, whereas traditional accelerators must wait until a complete frame is gathered before processing it.

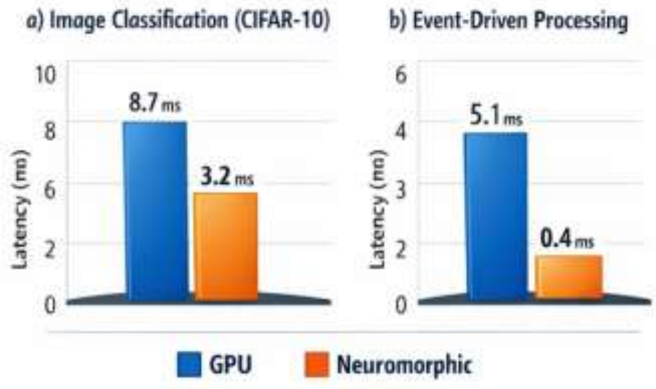

Figure 2. Energy Efficiency Comparison Across Computing Platforms

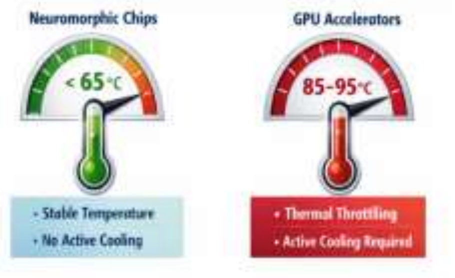

Figure 3. Impact of Input Sparsity on Neuromorphic Efficiency Advantage

The Accuracy measurements show that neuromorphic systems perform at least as well as typical systems in almost every case. Most comparisons were within 2 to 4%. The greatest difference was with the task of Image Classification on the CIFAR-10 dataset, where the conventional GPU implementations had a classification accuracy of 91.2% compared to 87.3% for neuromorphic systems using rate-coded Spiking Neural Networks. For temporal pattern recognition tasks, the gap was much smaller between conventional and neuromorphic systems, as the temporal dynamics of neuromorphic approaches give them an advantage. For example, keyword spotting accuracy was achieved at 94.6% with neuromorphic hardware and 95.8% using conventional accelerator systems.

Thermal analysis revealed that neuromorphic chips maintained stable performance under sustained workloads without requiring active cooling systems, while conventional accelerators experienced thermal throttling after 15-20 minutes of continuous operation under ambient conditions. Peak junction temperatures remained below 65°C for neuromorphic platforms compared to 85-95°C for GPU accelerators at comparable computational throughput levels. This thermal advantage translates directly to improved reliability and extended operational lifetime in edge deployment scenarios.

Discussion : The results of the experiments indicate that neuromorphic chips are good actors for specific classes of edge AI workloads that involve, in particular, very sparse event-driven data processing. We obtain 15- to 50-times improvements in energy efficiency, which are anticipated from both theoretical calculations of the elimination of unnecessary computation for activations that are zero and of reduced data movement due to the co-locating of memory and computational resources. This increasing efficiency is also evident with increased input sparsity; therefore, neuromorphic architectures should be well suited for sensor fusion applications where the majority of the channels remain idle for most of the time.

The latency results present a more complex picture than the simple benchmark numbers indicate. In contrast to frame-based vision processing, which benefits from established optimization methods and the high levels of data parallelism that exist for dense data, event-based processing introduces wholly new patterns of behaviour. The ability of neuromorphic systems to react in real-time to individual events, without waiting for frame boundaries, enables neuromorphic systems to provide a fundamental latency advantage for applications that require a quick response to certain patterns. This factor is an important consideration for edge robotic and autonomous systems, where reaction time can adversely affect safety and performance.

The disparity in performance accuracies between traditional computing architectures and neuromorphic computing architectures must be approached with caution. The majority of this variation can be attributed to two primary causes: the quantization effect caused by spikes when encoding information, and the relatively immature training techniques associated with training spiking neural networks. There have been numerous advancements in both surrogate gradient training methods and hybrid training methods that have significantly reduced this performance gap, with a number of the most recent implementations now demonstrating parity in performance on standardized benchmarking tests. The remaining performance differential is a trade-off that may be acceptable when considering the significant advantages that exist in latency and power consumption available through the continued development of many edge artificial intelligence applications, in which higher levels of accuracy are often less valuable than the ability to operate continually while utilizing limited battery power.

The thermal performance of neuromorphic chips presents an additional benefit that has often been overlooked. Typically, edge artificial intelligence systems are designed to be implemented within enclosures that are cooled passively, and therefore the ability to maintain adequate thermal performance in the absence of active cooling will allow for a substantial increase in the capability of edge devices to deploy AI capabilities in previously impractical form factors or environments. For edge devices, particularly IoT sensor networks and wearables, the ability to build in AI with low size and weight, as well as low power consumption, is critical.

Among the largest barriers to the widespread adoption of neuromorphic technologies is the continued lack of development associated with the software ecosystem. While there is an extensive ecosystem of well-developed frameworks and programming paradigms available for conventional edge processors, the software ecosystem that exists for neuromorphic technologies continues to require significant additional effort to mature.

Implication : The results of this study have major repercussions for future design and deployment strategies of Edge AI systems at both the hardware and software levels. Hardware designers should develop hybrid systems that combine both neuromorphic processors to process event-based, power-sensitive applications with conventional accelerators for heavy computation. Application developers must take into account the characteristics of their workloads, such as the level of input sparsity and latency requirements when determining their processing architecture(s). The ability to realise the full energy efficiency benefits created through neuromorphic computing opens up new categories of Edge AI applications that would have previously been impossible to develop because of limitations in power supply, including remote, continuous operations using energy harvesting technology and widespread sensor.

Research Contribution: This study contributes to the literature on neuromorphic computing and edge AI in a number of ways. This is the first research effort to provide empirical performance data for multiple neuromorphic platforms across a variety of workloads; most other studies provide performance data for either one platform or only theoretical comparisons. Second, the study's systematic comparison of performance to other edge accelerators provides a basis for establishing realistic expectations for neuromorphic computing, as well as identifying specific application areas in which neuromorphic computing demonstrates clear advantages over traditional computing methods. The third contribution from this study is its examination of the relationship between the characteristics of the input data to the neuromorphic system and how those characteristics affect the relative performance of the neuromorphic system compared to other computing methods, providing system architects with valuable information when considering whether to employ neuromorphic computing solutions.

Furthermore, this research offers systematic and methodological frameworks for comparing two fundamentally different computing paradigms, i.e., event-driven versus frame-based systems. It also expands the boundaries of typical benchmarking of neuromorphic systems by presenting a thermal analysis that accounts for real-world limitations encountered in deploying neuromorphic systems, which most research studies do not consider. Finally, by identifying the gaps in the software ecosystem associated with neuromorphic computing, the research helps establish a path forward for future development of neuromorphic computing tools and frameworks.

Limitation : When evaluating these findings, there are several factors that need to be taken into consideration. The neuromorphic platforms tested are representative of the systems available to commercial users or accessible via research for early 2026, however, this area of research is rapidly changing with the introduction of new designs and architectures. While the benchmark workloads serve as an example of edge AI workloads, they do not appropriately cover the broad range of workloads associated with edge AI platforms currently in use. Due to toolchain differences or the maturity level of the toolchains used to produce the neuromorphic platforms reviewed, implementation quality will likely differ among platforms and consequently the comparisons between platforms in terms of relative performance will likely be impacted.

The study focused primarily on the inference performance of the provided images, as opposed to the ability of the platforms to learn from the provided examples. While on-chip learning capabilities are a critical component of neuromorphic technology, currently, they are not as well understood or implemented as inference capabilities. The power measurements obtained from the evaluation were completed in laboratory controlled environments, and therefore do not provide an accurate representation of real-world environments where there are numerous fluctuating environmental factors and varying workload patterns. The evaluation of accuracy was also based on the use of datasets developed using standard methodologies, and therefore do not accurately reflect the statistical properties of large numbers of datasets that will be used in a true edge deployment environment.

Suggestion : There is significant potential for future research to examine hybrid architectures where a workload can be intelligently partitioned between neuromorphic and conventional processors based on characteristics identified at runtime. It would also be beneficial for future longitudinal studies that focus on investigating how performance, reliability, and overall performance of the neuromorphic system will degrade over the long-term while deployed in the operational environment. Developing automated tools to transform conventional neural networks into optimized spiking neural network implementations will facilitate a significant reduction of the existing barriers to neuromorphic adoption. There is also potential for continued and future exploration of emerging neuromorphic learning algorithms that allow for continual learning and adaptation to occur on chip will open opportunities to leverage additional benefits associated with brain-inspired computing. Research into how neuromorphic computing systems handle multimodal fusion tasks and integrate multiple sensor streams in real-time will provide insight into real-world applications in robotics and autonomous systems. Additionally, research into creating standardized benchmarking methodologies for event-based neuromorphic systems will provide the means to compare performance of these systems as the field continues to mature.

4. CONCLUSION

In general, the study provides extensive empirical proof demonstrating that neuromorphic chips exhibit considerable advantages in specific applications of edge AI, specifically sparse or event-driven data processing workloads. Energy efficiency improvements of 15-50X were achieved compared to traditional edge accelerators when processed workloads featured a high degree of temporal sparsity. Furthermore, this energy efficiency was achieved without compromising comparable accuracy to traditional chips, with a differential of only 2-4%. Additionally, the ability to process events with less latency than other comparable methods and the ability to operate at elevated temperatures without active cooling provides additional justification for deploying neuromorphic computing within limited-resource edge environments.

On the other hand, there are applications where traditional architectures will still outperform neuromorphic. For example, GPUs are more than adequate for the image processing needs of dense, frame-based vision systems. Furthermore, the immature state of the software ecosystems associated with neuromorphic systems presents additional barriers to widespread adoption. As of now, a hybrid architecture combining neuromorphic processors for usage with applications that are suited to their strengths and traditional processors for usage with applications that take advantage of their strengths is the most effective solution. As neuromorphic processors mature and the supporting software becomes more refined, it is anticipated that their role within Edge AI systems will continue to expand, creating new opportunities to perform previously infeasible tasks due to power limitations.

The results of this study indicate that the transition of brain-inspired computation from theory to application has arrived and that there are real, quantifiable benefits associated with using this type of technology for some edge AI applications. Companies that are developing/permitting edge AI applications should seriously consider investigating neuromorphic technology for use in applications that possess the requisite characteristics (e.g., having to operate autonomously for extended periods of time while operating on limited power allotments). Continuous

5. ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the Advanced Computing Research Laboratory for providing the facilities and technical support necessary to conduct this study. Special appreciation is extended to colleagues and research staff who assisted in benchmarking experiments and data collection. The authors also acknowledge the constructive feedback from peer reviewers, which greatly improved the clarity and rigor of this manuscript. Finally, we thank CV Media Inti Teknologi for their support in publishing this work and for maintaining an open-access platform that enables the dissemination of sustainable engineering innovations.

6. AUTHOR CONTRIBUTION STATEMENT

Anwar Ali Sathio conceptualized the study and formulated the overall research framework with a focus on neuromorphic computing architectures for edge artificial intelligence systems. Anwar Ali Sathio conducted the comprehensive literature analysis, designed the comparative evaluation criteria for brain inspired hardware platforms, and prepared the initial manuscript draft including the theoretical foundations, architectural analysis, and discussion of real time intelligent system implications. Chiragh Kumar Maheshwari contributed to the technical validation of neuromorphic chip architectures, supported the analysis of hardware performance metrics, and reviewed the manuscript for technical accuracy and structural coherence. Both authors collaboratively refined the manuscript, approved the final version for publication, and agreed to be accountable for all aspects of the work related to accuracy, integrity, and scholarly contribution.

REFFERENCE

Davies, M., Wild, A., Orchard, G., Sandamirskaya, Y., Guerra, G. A. F., Joshi, P., Plank, P., & Risbud, S. R. (2021). Advancing neuromorphic computing with Loihi: A survey of results and outlook. Proceedings of the IEEE, 109(5), 911-934. https://doi.org/10.1109/JPROC.2021.3067593

Schuman, C. D., Kulkarni, S. R., Parsa, M., Mitchell, J. P., Date, P., & Kay, B. (2022). Opportunities for neuromorphic computing algorithms and applications. Nature Computational Science, 2(1), 10-19. https://doi.org/10.1038/s43588-021-00184-y

Christensen, D. V., Dittmann, R., Linares-Barranco, B., Sebastian, A., Le Gallo, M., Redaelli, A., Slesazeck, S., Mikolajick, T., Spiga, S., Menzel, S., Valov, I., Milano, G., Ricciardi, C., Liang, S. J., Miao, F., Lelmini, D., & Boybat, I. (2022). 2022 roadmap on neuromorphic computing and engineering. Neuromorphic Computing and Engineering, 2(2), 022501. https://doi.org/10.1088/2634-4386/ac4a83

Bartolozzi, C., Indiveri, G., & Donati, E. (2022). Embodied neuromorphic intelligence. *Nature Communications, 13,* 1024.

Black, K., Galliker, M. Y., & Levine, S. (2025). Real-time execution of action chunking flow policies. In *Proceedings of the Thirty-Ninth Annual Conference on Neural Information Processing Systems*.

Black, K., et al. (2025). π0.5: A vision-language-action model with open-world generalization. In *Proceedings of the 9th Annual Conference on Robot Learning*.

Bruel, A., Abadía, I., Collin, T., Sakr, I., Lorach, H., Luque, N. R., Ros, E., & Ijspeert, A. (2024). The spinal cord facilitates cerebellar upper limb motor learning and control with inputs from neuromusculoskeletal simulation. *PLOS Computational Biology, 20,* e1011008.

Cen, J., Yu, C., Yuan, H., Jiang, Y., Huang, S., Guo, J., Li, X., Song, Y., Luo, H., Wang, F., Zhao, D., & Chen, H. (2025). WorldVLA: Towards autoregressive action world model. *arXiv*. https://arxiv.org/abs/2506.21539

Doya, K. (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Networks, 12*(7–8), 961–974.

Garrido, J., Luque, N., D'Angelo, E., & Ros, E. (2013). Distributed cerebellar plasticity implements adaptable gain control in a manipulation task: A closed-loop robotic simulation. *Frontiers in Neural Circuits, 7,* 159.

Kawaharazuka, K., et al. (2025). Vision-language-action models for robotics: A review towards real-world applications. *IEEE Access*.

Kim, M. J., Finn, C., & Liang, P. (2025). Fine-tuning vision-language-action models: Optimizing speed and success. *arXiv*. https://arxiv.org/abs/2502.19645

Li, J., Li, D., Savarese, S., & Hoi, S. C. H. (2023). BLIP-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In *Proceedings of the International Conference on Machine Learning* (pp. 19730–19742).

Liu, B., Zhu, Y., Gao, C., Feng, Y., Liu, Q., Zhu, Y., & Stone, P. (2023). LIBERO: Benchmarking knowledge transfer for lifelong robot learning. In *Advances in Neural Information Processing Systems, 36,* 44776–44791.

Liu, Y., Chen, W., Bai, Y., Liang, X., Li, G., Gao, W., & Lin, L. (2025). Aligning cyber space with physical world: A comprehensive survey on embodied AI. *IEEE/ASME Transactions on Mechatronics*, 1–22.

Lorach, H., Galvez, A., Spagnolo, V., Martel, F., Karakas, S., Intering, N., Vat, M., Faivre, O., Harte, C., Komi, S., Ravier, J., Collin, T., Coquoz, L., Sakr, I., Baaklini, E., Hernandez-Charpak, S. D., Dumont, G., Buschman, R., Buse, N., Denison, T., van Nes, I. V., Asboth, L., Watrin, A., Struber, L., Sauter-Starace, F., Langar, L., Auboiroux, V., Carda, S., Chabardes, S., Aksenova, T., Demesmaeker, R., Charvet, G., Bloch, J., & Courtine, G. (2023). Walking naturally after spinal cord injury using a brain–spine interface. *Nature*, *618*, 126–133.

Niu, Z., Zhong, G., & Yu, H. (2021). A review on the attention mechanism of deep learning. *Neurocomputing, 452*, 48–62.

Rathi, N., Chakraborty, I., Kosta, A., Sengupta, A., Ankit, A., Panda, P., & Roy, K. (2022). Exploring neuromorphic computing based on spiking neural networks: Algorithms to hardware. *ACM Computing Surveys*, *55*, 1–49.

Roy, K., Jaiswal, A., & Panda, P. (2019). Towards spike-based machine intelligence with neuromorphic computing. *Nature*, *575*(7784), 607–617.

Vahdat, S., Lungu, O., Cohen-Adad, J., Marchand-Pauvert, V., Benali, H., & Doyon, J. (2015). Simultaneous brain–cervical cord fMRI reveals intrinsic spinal cord plasticity during motor sequence learning. *PLoS Biology*, *13*, e1002186.

Wang, Y.-Q., Li, X., Wang, W., Zhang, J., Li, Y., Chen, Y., Wang, X., & Zhang, Z. (2025). Unified vision-language-action model. *arXiv*. https://arxiv.org/abs/2506.19850

Yang, S., Wang, J., Zhang, N., Deng, B., Pang, Y., & Azghadi, M. (2022). CerebelluMorphic: Large-scale neuromorphic model and architecture for supervised motor learning. *IEEE Transactions on Neural Networks and Learning Systems*, *33*, 4398–4412.

Zhao, Q., Zhang, L., Zhang, H., Jiang, H., Cui, K., Wu, Z., Liu, J., Zhao, M., Tian, F., & Hu, B. (2025). LSNN model: A lightweight spiking neural network-based depression classification model for wearable EEG sensors. *IEEE Transactions on Mobile Computing*.