International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 09 | Sept - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Memristor-Based Architectures for AI at the Edge

Nanda Kishor S[1]

[1]*UG Scholar ,Department of ECE,NCERC,Thrissur , Kerala*

Abstract - Memristors, as the fourth fundamental circuit element, have emerged as a transformative technology for enabling efficient in-memory computing and neuromorphic architectures. Unlike conventional CMOS devices, memristors combine storage and processing within a single nanoscale element, allowing massively parallel operations, low leakage, and non-volatile data retention. These properties make them particularly suited for Edge-AI applications, where latency, energy efficiency, and scalability are critical. This paper reviews the role of memristor-based crossbar arrays in implementing synaptic weights for neural networks, highlighting their ability to accelerate matrix–vector multiplications and support adaptive learning mechanisms. Comparative studies against CMOS implementations demonstrate reduced power consumption, higher density, and fault-tolerant performance. Recent research shows applications ranging from healthcare signal analysis to smart sensors and IoT devices, emphasizing memristors as enablers of brain-inspired intelligence at the edge. However, challenges such as variability, limited endurance, and integration with CMOS technology remain significant barriers to commercialization. Future directions include three-dimensional crossbar scaling, hybrid CMOS–memristor systems, and algorithm–hardware co-design to achieve reliable, large-scale deployment. Taken together, memristor-based architectures represent a critical step toward next-generation, low-power, and adaptive edge intelligence.

Key Words: *memristor, in-memory computing, Edge-AI, crossbar architecture, neuromorphic hardware, energy efficiency.*

1.INTRODUCTION

Memristors, which were proposed as the fourth basic circuit device, have garnered a great deal of research interest because they can store and compute data simultaneously in the same nanoscale device. Unlike physically separate units used by conventional CMOS transistors for computation and memory, memristors take advantage of resistance switching to encode and update data states locally. This property is a direct solution to the von Neumann bottleneck, under which repeated data transfer between processing units and memory units in CPUs and GPUs causes latency, high energy, and poor scalability. Due to their non-volatility, small footprint, and ultra-low leakage power consumption, memristors make an appealing option for Edge-AI applications, under which real-time response, energy efficiency, and flexibility are key requirements.

Recent progress in memristor crossbar arrays shows that they can be used as hardware substrates for neural networks. In these architectures, memristors serve as artificial synapses with conductance values storing synaptic weights. The intrinsic parallelism of crossbar architectures allows for matrix–vector multiplications in a single computational step, thus speeding up one of the most computationally expensive operations in deep learning. Additionally, memristors are capable of hosting adaptive learning algorithms like spike-timing dependent plasticity (STDP) and Hebbian learning, which make them well-suited to neuromorphic and spiking neural network (SNN) applications. Experimental prototypes emphasize uses from pattern recognition and biomedical signal analysis to IoT sensing and real-time robotics, illustrating the appropriateness of memristors for localized, low-power intelligence at the edge.

Meanwhile, there are issues that impede the mass deployment of memristor-based systems. Device-level problems including variability, endurance limits, retention failures, and stochastic switching behavior impair performance consistency across arrays. On the system level, incorporating memristors and CMOS technology involves sophisticated fabrication processes, design automation, and new standards for algorithm–hardware co-optimization. In addition, the majority of existing implementations are limited to small-scale demonstrations or toy models with open issues concerning scalability, reliability, and cross-application generalization. Overcoming these constraints will involve hybrid CMOS–memristor platforms, three-dimensional crossbar stacking for improved density, and the co-design of algorithms considering device-level non-idealities.

With these opportunities and challenges in mind, a broad review of memristor-based architectures for Edge-AI is both timely and justified. This seminar paper synthesizes current research for the field, compares relative strengths and limitations to traditional CMOS-based platforms, and emphasizes key applications for healthcare, smart devices, and intelligent robotics. It also determines emerging open issues in device design and architectural design, and points towards potential research directions including hybrid integration, algorithmic adaptation, and runtime validation. Together, memristor-based systems form a promising building block for the future generation of adaptive, energy-efficient, and intelligent edge computing platforms.

2. LITERATURE REVIEW

Wang et al. [1] constructed a near-threshold memristive computing-in-memory (mCIM) engine to counter the high energy dissipation and latency of AI applications. Their on-chip device consists of 256 input channels and reports 10.49 TOPS and 88.51 TOPS/W performance, with hybrid control coupled with analog–digital methods to counter mismatch and variability. The research illustrates howmeticulously designed architectures can convert memristor devices into highly efficient computing substrates for Edge-AI.

Jiang et al. [2] gave an in-depth overview of memristor-enabled machine learning accelerators, comparing current prototype chips with traditional CMOS-based solutions. Their discussion highlights accuracy, precision, and power efficiency trade-offs, citing the notable advantages that memristors bring in terms of density and scalability. The review points to significant shortcomings in standardization, benchmarking, and model robustness that need to be addressed before wide-scale uptake.

Ding et al. [3] investigated the unorthodox function of memristor noise by demonstrating how device fluctuations can be translated into computational innovations in stochastic computing, Bayesian inference, and random number generation, where noise is a desirable property instead of an imperfection. It emphasizes the promise of accepting device-level flaws as a pathway to new Edge-AI computational paradigms.

Yousuf, Adam, et al. [4] suggested a fault-tolerant scheme for memristive neural networks based on layer ensemble averaging. The approach enhances inference resilience against device noise and defects by improving the methodology, which has been validated by simulation and hardware prototype. The method shows that redundancy in architecture can significantly enhance the reliability of edge AI systems based on memristors.

Taskov et al. [5] tested simulated memristor reservoir architectures to evaluate their performance under non-ideal circumstances. Their paper demonstrates how reservoir computing can restore system performance even under device variability and noise. This indicates the appropriateness of memristor reservoirs for reliable edge intelligence where imperfect devices are unavoidable.

Qahtani et al. [6] illustrated a neuromorphic edge-AI system for surgery decision support based on spiking neural networks (SNNs) and symbolic inference. Deployed on Intel's Loihi 2 chip, the system delivered decision latencies under 50 ms with much lower power consumption compared to GPU-based solutions. Their research shows how memristor-motivated neuromorphic hardware is able to deliver robust real-time aid in life-critical healthcare scenarios.

Jebali et al. [7] implemented a memristor-based binarized neural network (BNN) supplied by an integrated solar cell to investigate energy autonomy for edge devices. Their device comprises 32,768 memristors and is insensitive to varying illumination, obviating the need for elaborate analog calibration. The research offers a template for sustainable low-power AI systems that are self-sufficient at the edge.

Turck et al. [8] suggested the Logarithmic Memristor-Based Bayesian Machine, which is an architecture based on memristors for Bayesian inference applications. The system enhances classification accuracy and energy efficiency in tasks like gesture recognition and sleep stage monitoring. Their findings demonstrate how logarithmic computation with memristors can be used to improve probabilistic reasoning in edge devices.

Li, Fu, and colleagues [9] constructed a memristor-based circuit realization of MobileNetV3 for image classification. The circuit design obtained more than 90% accuracy on CIFAR-10 and significantly lowered the computational time and energy with respect to digital-only baselines. Their study demonstrates the promise of memristor-augmented neural designs for lean vision applications at the edge.

Aguirre et al. [10] surveyed hardware implementations of memristive artificial neural networks, highlighting design options, peripheral circuits, and performance compromises. Their discussion highlighted that memristor crossbars allow dense integration but variability and non-idealities are still significant challenges. The survey concludes that efficient peripheral design and judicious mapping strategies are crucial for scalable architectures.

Jiang et al. [11] presented up-to-date views of memristor-based machine learning accelerators, as well as opportunities and limitations of recent prototypes. They addressed precision, power efficiency, and accuracy trade-offs and compared different design methods across platforms. The review indicates that memristor accelerators can address the increasing need for embedded intelligence.

James [12] wrote a monograph on robust analog in-memory computing with memristor crossbars. The paper discusses methods to counter device variability, such as echo-state networks, redundancy, and three-dimensional crossbar stacking. It highlights that analog in-memory computing, if made robust, would be able to significantly improve edge AI deployments.

Eslami et al. [13] examined on-chip learning in memristor neural networks under noise and variation. Their accelerator is capable of training as well as inference and demonstrates near 97% accuracy even with moderate device non-idealities. This effort depicts the viability of adaptive,

resilient to noise learning systems for edge AI real-time applications.

Alam et al. [14] presented a memristor-based unsupervised learning system for anomaly detection on edge devices. Their algorithm achieved good detection accuracy in resource-limited settings, indicating the potential for memristors in online, self-tuning learning applications. This indicates practical potential in IoT and smart sensing.

Song et al. [15] have designed a memristor-based stochastic edge detection system that is lightweight and error-resilient. Their system demonstrated huge energy efficiency and robust bit-error tolerance and is well suited for visual edge detection in low-power embedded applications. This work demonstrates how memristor-based stochastic logic can produce elegant yet effective solutions for edge AI.

3. COMPARISONS OF VARIOUS WORK

Literature examined indicates a discernible transition from simple memristor device physics studies to sophisticated architectures for Edge-AI implementation. Wang et al. (2025) demonstrated a near-threshold memristive computing-in-memory engine with capabilities for high throughput as well as energy efficiency, while Jiang et al. (2025) synthesized broad advances in memristor-facilitated machine learning and determined primary trade-offs in accuracy and power efficiency against CMOS. Ding et al. (2025) further developed this view by reinterpreting noise in memristors as a computational advantage, and proposing its use in Bayesian and stochastic environments.

Simultaneously, Yousuf et al. (2025) and Taskov et al. (2025) highlighted reliability and robustness by applying ensemble averaging and reservoir computing to manage defects and variability in hardware. At the application level, Qahtani et al. (2025) had insisted on neuromorphic edge AI for surgical decision-making on Loihi 2 with millisecond latency and huge energy efficiency, while Jebali et al. (2024) and Turck et al. (2024) illustrated edge-oriented deployments in the shape of solar-powered binarized neural networks and logarithmic Bayesian inference machines with real-world applicability to autonomous and probabilistic edge intelligence. Li et al. (2024) subsequently applied system-level integration to a MobileNetV3 model with an embedded memristor, which provided high accuracy in image recognition at reduced energy and latency. Supporting surveys by Aguirre et al. (2024) and Jiang et al. (2025) emphasized the importance of peripheral circuits and system-level design choices, while James (2025) presented guidelines for secure analog in-memory computing in accordance with crossbars. Concurrently with this, Eslami et al. (2025) addressed on-chip learning in noisy environments, Alam et al. (2025) applied memristor-based structures for unsupervised anomaly detection using edge devices, and Song et al. (2025) presented error-resilient visual edge detection via stochastic computing. Collectively, these studies highlight memristors as enablers of edge dense, energy-efficient, adaptive computing, and point to unresolved challenges in device variability, endurance, hybrid integration, and large-scale system standardization that must be addressed in order to succeed in their transition from laboratory prototypes to commercial products.

Table 1: Table of comparison

REF NO.	YEAR	AUTHORS (main)	SHORT TITLE / TOPIC	WHAT THE PAPER SUGGESTS / PROPOSES
1	2025	L. Wang et al.	Near-threshold memristive computing-in-memory engine	Fabricated mCIM engine with 256 inputs; achieves 10.49 TOPS and 88.51 TOPS/W; mitigates mismatch and variability with hybrid control.
2	2025	M. Jiang et al.	Review: Memristor-accelerated ML accelerators	Surveys ML accelerators using memristors; compares with CMOS; highlights precision–power trade-offs.
3	2025	C. Ding et al.	Transforming memristor noise into computational innovations	Explores how noise can be exploited for randomness, Bayesian inference, and reservoir computing.

REF NO.	YEAR	AUTHORS (main)	SHORT TITLE / TOPIC	WHAT THE PAPER SUGGESTS / PROPOSES
4	2025	O. Yousuf, G. Adam, et al.	Layer ensemble averaging for fault tolerance in MNNs	Proposes ensemble averaging to tolerate device faults; improves inference robustness in prototypes.
5	2025	T. Taskov et al.	Simulated memristor reservoir architecture	Uses reservoir computing to tolerate non-idealities; improves performance despite device variations.
6	2025	A. H. Qahtani et al.	Neuromorphic Edge-AI for surgical decision support	Demonstrates SNN + symbolic reasoning on Loihi 2; <50 ms latency and significant energy savings vs GPU.
7	2024	F. Jebali et al.	Memristor BNN + solar powering	Implements 32,768-memristor BNN powered by solar cell; robust under illumination; low-power design.
8	2024	C. Turck et al.	Logarithmic memristor-based Bayesian Machine	Proposes Bayesian inference machine; shows improved accuracy and efficiency in gesture/sleep analysis.
9	2024	J. Li, Y. Fu, et al.	Memristor-based MobileNetV3 for image classification	Designs MobileNetV3 circuit with memristors; >90% CIFAR-10 accuracy; reduced time and energy costs.
10	2024	F. Aguirre et al.	Hardware implementation of memristive neural networks	Reviews design alternatives and trade-offs for memristive ANNs; analyzes peripheral circuits/tools.
11	2025	M. Jiang et al.	Current opinions on memristor ML accelerators	Highlights prototype accelerators; surveys power vs accuracy trade-offs in ML workloads.
12	2025	A. James	Reliable analog in-memory computing with crossbars	Monograph on analog memristor crossbars; addresses variability; proposes echo-state and 3D arrays.
13	2025	M. R. Eslami et al.	On-chip learning under variations	Memristor accelerator with on-chip training; maintains ~97% accuracy under moderate device noise.
14	2025	M. S. Alam et al.	Unsupervised learning & anomaly detection on edge devices	Proposes memristor-based unsupervised learning system for anomaly detection at the edge.
15	2025	L. Song et al.	Error-tolerant edge detection with memristors	Introduces memristor-enabled stochastic logic; robust error tolerance and major energy savings.

4. DISCUSSIONS

Literature considered here indicates that memristor work has rapidly evolved from theoretical device research to realistic architectures intended for Edge-AI deployment. Wang et al. (2025) and Jiang et al. (2025) set the foundation for memristors as extremely efficient in-memory computing devices and proved that computing-in-memory engines can offer high throughput and energy efficiency, while solving the von Neumann bottleneck. Simultaneously, Ding et al. (2025) rewrote device noise as a computational feature and proposed its incorporation into probabilistic and stochastic models. The above studies unanimously highlight that memristors are no longer restricted to debates in device physics, but are being folded into frameworks that facilitate real-time intelligence on edge platforms.

One of the significant research agendas includes the usage of memristor crossbars for adaptive learning and acceleration in neural networks. Studies like Jebali et al. (2024), Turck et al. (2024), and Li et al. (2024) presented real-world applications of binarized neural networks, Bayesian machines, and low-weight convolutional networks implemented via memristor modules. These studies highlight the benefits of compact design, low latency, and reduced power consumption, making memristors suitable for vision recognition, biomedical signal analysis, and IoT sensing tasks. However, challenges such as variability, endurance limitations, and integration with CMOS remain obstacles to large-scale deployment. Studies by Yousuf et al. (2025) and Taskov et al. (2025) demonstrated how methods such as ensemble averaging and reservoir computing can partially address reliability issues, which implies system-level innovations are necessary in order to fully leverage memristor hardware.

Apart from performance improvements, newer research also highlights memristors' versatility for practical edge intelligence. Qahtani et al. (2025) confirmed neuromorphic systems in essential healthcare operations, demonstrating millisecond-order decision latency with significant energy efficiency, while Alam et al. (2025) took the paradigm to unsupervised anomaly detection for IoT devices. Song et al. (2025) even expanded applications by exhibiting error-resistant visual edge detection using stochastic logic. Cumulatively, these research works confirm that memristors are able to provide adaptive, energy-efficient, and domain-specific intelligence at the edge. However, open issues in device reliability, hybrid CMOS–memristor integration, and standardization underscore the necessity for further research. Collectively, the surveyed papers indicate memristors as a foundation technology for next-generation edge AI systems, providing scalability and efficiency while necessitating innovations in materials, architectures, and co-designed algorithms.

5. CONCLUSION

This review has canvassed the evolution of memristor-based architectures from their beginnings as new nanoscale devices into their new role as viable enablers of Edge-AI systems. One overriding theme among the studies that were reviewed is that memristors couple storage and computation in a single component, facilitating in-memory computing, extreme parallelism, and ultra-low power operation. Researchers have shown memristor crossbar arrays are capable of speeding up neural network functions like matrix–vector multiplication, accommodating adaptive learning rules, and running in real time within edge platforms where latency and efficiency are paramount. Demonstrated applications included image recognition, biomedical signal processing, anomaly detection, and healthcare decision support, all of which indicate the versatility of memristors within edge intelligence. In addition, hybrid methods that bridge between device-level innovation and architectural solutions, including ensemble averaging, stochastic logic, and Bayesian inference, increase adaptability and resilience in the presence of hardware variability.

While these improvements have been made, the literature also points to seminal challenges yet to be solved before wide-scale deployment. Device variability, low endurance, and stochastic switching are still significant impediments to reliability, and integration with CMOS technology means increasing fabrication complexity and scalability issues. Most present-day implementations are still limited to small-scale prototypes or reduced models, which emphasize the discrepancy between laboratory showcases and potentially marketable products. However, the overall evidence suggests that memristors are gradually moving in the direction of application in Edge-AI. Progress in material engineering, 3D crossbar stacking, and hardware co-design with algorithms will be fundamental in bringing about their full potential. Finally, memristor-based systems provide a future-looking approach towards the design of the next generation of adaptive, efficient, and intelligent edge devices.

REFERENCES

1. Wang L., et al., 2025. A near-threshold memristive computing-in-memory engine for AI acceleration. *Nature Communications*, 16: 61025.

2. Jiang M., et al., 2025. Memristor-accelerated machine learning accelerators: Current status and challenges. *Microelectronics Journal*, Elsevier.

3. Ding C., et al., 2025. Transforming memristor noises into computational innovations. *Communications Materials*, Nature Portfolio.

4. Yousuf O., Adam G., et al., 2025. Layer ensemble averaging for fault tolerance in memristive neural networks. *IEEE Transactions on Neural Networks and Learning Systems*

5. Taskov T., et al., 2025. Simulated memristor reservoir architectures for robust neural computing. *Neurocomputing*, Elsevier.

6. Qahtani A. H., et al., 2025. Neuromorphic edge AI for surgical decision support. *Artificial Intelligence in Medicine*, Wolters Kluwer.

7. Jebali F., Majumdar A., Turck C., et al., 2024. Powering AI at the edge using memristor-based binarized neural networks and solar cells. *Nature Communications*, 15: 4476.

8. Turck C., Harabi K. E., et al., 2024. The Logarithmic Memristor-Based Bayesian Machine. *arXiv preprint*, arXiv:2406.03492.

9. Li J., Fu Y., et al., 2024. Memristor-based MobileNetV3 circuit design for image classification. *arXiv preprint*, arXiv:2402.10512.

10. Aguirre F., et al., 2024. Hardware implementation of memristive artificial neural networks: architectures and trade-offs. *Nature Communications*, 15: 45670.

11. Jiang M., et al., 2025. Current perspectives on memristor machine learning accelerators. *Microelectronics Journal*, Elsevier.

12. James A., 2025. Reliable Analog In-Memory Computing with Crossbars. *Foundations and Trends in Integrated Circuits and Systems*, Now Publishers.

13. Eslami M. R., et al., 2025. On-chip learning with memristor-based neural networks under variations. *arXiv preprint*, arXiv:2408.14680.

14. Alam M. S., et al., 2025. Unsupervised online learning and anomaly detection on edge devices using memristor-based neuromorphic systems. *Information*, MDPI, 16(3): 222.

15. Song L., et al., 2025. Lightweight error-tolerant edge detection using memristor-enabled stochastic computing. *Nature Communications*, 16: 59872.