Journal of Technological Innovations

Est. 2020

# Emerging Neuromorphic Computing for Edge AI Application: A Systematic Literature Review

***Rajat Suvra Das***

*Senior Director, Business Development L&T*

*Technology Services Email: rajat.tel@gmail.com*

**Abstract:**

Edge AI (Artificial Intelligence) and Neuromorphic computing are two intersected concepts that have significantly impacted in recent years. This is due to the ability of neuromorphic computing to mimic the functionality of the brain for processing information in an energy-efficient and highly parallel way. Likewise, Edge AI refers to deploying AI algorithms or models directly on an edge platform rather than relying on servers or cloud platforms. Thus, to accomplish this, neuromorphic computing and edge AI are combined due to the parallel processing ability of neuromorphic computing, which aligns well with edge AI applications' requirements compared to the traditional von Neumann architecture. Despite its promise for edge AI, very few studies have dealt with neuromorphic computing for edge AI due to limited resources and the unavailability of hardware and software tools, which hinder the progress of this realm of research. Therefore, the current SLR focuses on reviewing studies emphasizing neuromorphic computing for Edge AI, differences in conventional and neuromorphic computing, different chips used for neuromorphic computing, and applications of neuromorphic computing for edge AI. Moreover, the present SLR has scrutinized 25 papers based on inclusion and exclusion criteria. From the obtained 25 papers, different challenges are depicted, and future recommendations for overcoming these shortcomings are provided.

**Keywords**—— Neuromorphic Computing, Edge AI, Applications, Challenges, Survey Method

## 1.Introduction

EC (Edge Computing) is considered the foundation for IoT systems, which fit the idea of smart devices and other factors [1]. The advancement of the IoT (Internet of Things) network and the amount of data transferred in the cloud stress the limits of the DC (Data Center). Unlike CC (Cloud Computing), where the data are processed at DC [2], EC processes the data at the point at which the data are collected. Further, the rise of IoT devices has also resulted in the dire need for data processing on the edge.

Thus, the primary idea of EC is the local processing of the data, which does not entail sending a substantial amount of data into the servers, and all the processing and decision-making mechanisms must be carried out at a low power level; by doing so, the compulsion to have complex DC can be removed. Further, Edge computing has become increasingly pertinent because the power needed for data processing in the DC on the servers has amplified substantially in the past few years [3]. Thus, the main purpose of EC is to make the edge devices perform quicker than before, extra intellectual, and result in less consumption of power [4]. This is primarily due to the ability of neuromorphic architectures to decrease the processing

power and lower energy consumption, which amplifies the life of the battery and ultimately cuts the overall cost of computation [5].

Thus, in recent years, neuromorphic computing technologies have made significant breakthroughs to overcome the power and latency shortfalls of traditional digital computing [6]. This is primarily due to the ability of neuromorphic computing to mimic the efficiencies of neuro-biological architectures like the human brain. urther, it also emphasizes designing software and hardware elements [7]. Thus, Figure.1 shows process of neuromorphic computing.

· To systematically review the concepts associated with neuromorphic and neuromorphic computing in Edge AI.

· To discuss the applications of neuromorphic computing in edge AI applications

· To encounter challenges faced by the prevailing works and deliberate on the future recommendations of the model.

1.1 Paper Organization



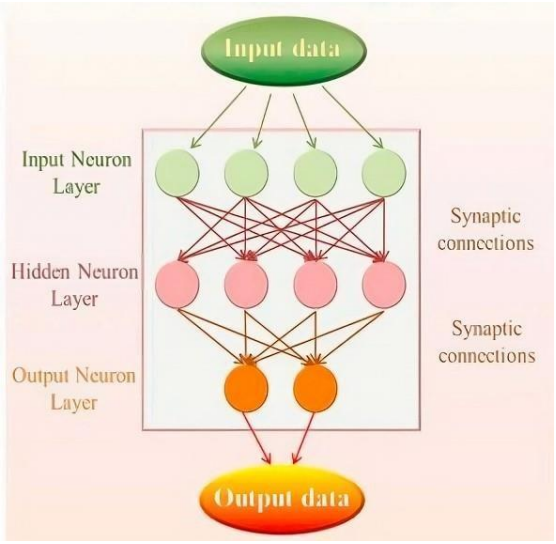


Figure.1. Neuromorphic Computing [8]

Neuromorphic architecture is most typically demonstrated after the neocortex, which is present in the brain. The complex connectivity and encrusted structure of the neocortex are critical due to its capability to process multifaceted information and facilitate human thinking [9]. Similarly, a biological neuron encompasses a cell body (soma) with many dendrites that carry information to other neurons and serve as a connection to other neurons [10]. Thus, the proposed paper focuses on reviewing neuromorphic computing and neuromorphic computing in edge AI applications hence the objective of the paper includes,

The paper is organized as follows. Section II deals with survey methodology, Section III focuses on the rise and evolution of neuromorphic computing, Section IV emphasizes neuromorphic computing for Edge AI, Section V deals with applications of applications, Section VI focuses on challenges and future recommendations to overcome the challenges and conclusion of the work is depicted in Section VII.

## 2. Survey Methodology

Various approaches are employed for fetching the appropriate content. Hence, some techniques utilized for surveying contents are depicted in Figure.2.

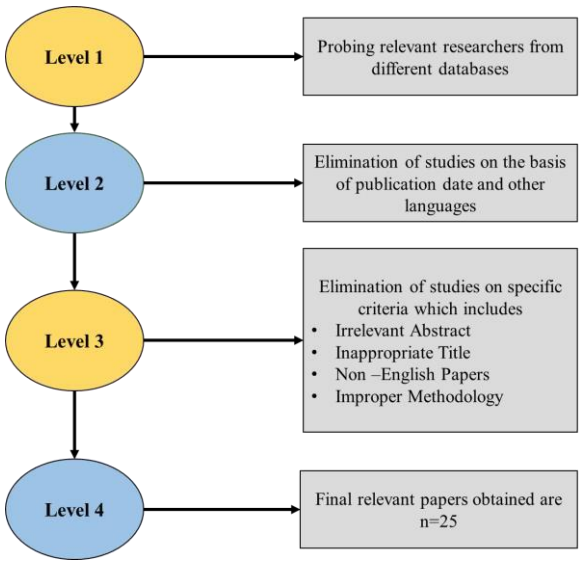

Figure.2. Survey Methodology

**Probing Database**

Different databases are considered for fetching papers, which include IJARET IEEE, Google Scholar, Taylor and Francis, AMC, Springer, Frontier, Elsevier, MDPI, Wiley, Science Direct, and other DBs are used for fetching content.

**Content Fetching**

In order to fetch content in different databases, different keywords are used. Some of the keywords include "Neuromorphic Computing," Edge Computing," "Neuromorphic Computing in Edge AI," "Applications of Neuromorphic Computing in Edge AI Applications," and “Challenges of Neuromorphic Computing in Edge AI."

**Reference Scrutinizing**

The papers fetched for reviewing the studies are from 2019-2024. 25 papers are considered for reviewing the works for neuromorphic computing for Edge AI application.

**Exclusion criteria**

· Papers that are non-English are omitted.

· Papers with irrelevant titles and abstracts are eliminated.

**Inclusion criteria**

· Articles with neuromorphic computing and neuromorphic computing for edge AI are chosen and reviewed.

· Reference papers are taken from the year 2019- 2024.

As per the inclusion criteria, the reference paper taken for the SLR is depicted in Figure.3.

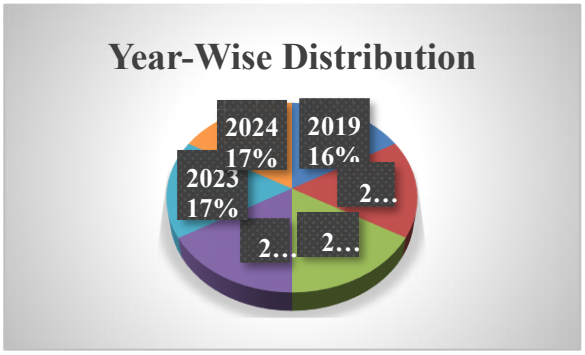


Figure.3. Year-Wise Distribution

From Figure.3, it can be identified that 2019 and 2020 years have covered considerably minimum papers for neuromorphic computing, whereas papers above 2020, such as 2021, 2022, 2023, and 2024, have covered a lot in terms of neuromorphic computing and neuromorphic computing in Edge AI applications. Similarly, the present SLR also classifies the works according to journals cited in the paper, which is depicted in Figure.4. Figure.4 shows that different journals have covered the concepts of neuromorphic computing, including IJARET, ACM, IEEE, MDPI, Springer, and many more. Likewise, another journal in Figure.4 includes journals like Neuromorphic Computing and Engineering and many more.

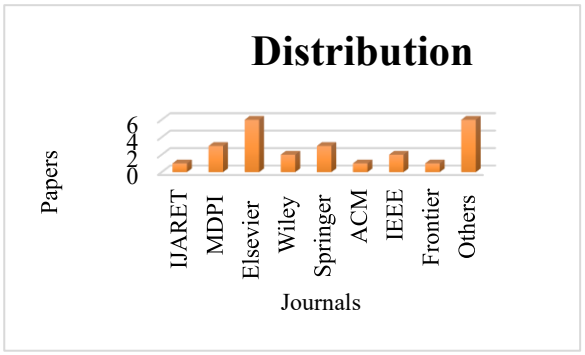


Figure.4. Journal-wise distribution

Year-wise and journal-wise distribution helps to understand the advancements of neuromorphic computing in recent times. Further, reviewing the papers systematically can

offer clear insights required for the study and aid in providing a better understanding of the concept.

## 3. Upsurge of Neuromorphic computing

Neuromorphic computing has developed in modern years as a harmonizing architecture to Von Neumann systems. The term neuromorphic computing was devised by Carever Mead in the year 1990 [11]. At the time, Mead referred to VSLI with components that mimicked biological neural systems as neuromorphic systems. Lately, the term has come to incorporate applications based on non-von Neumann architecture. These neuromorphic architectures are prominent for requiring less power, collocating memory, and processing [12].

Further, Neuromorphic architecture has received better consideration due to the impending end of Moore's law and also due to the low bandwidth between memory and CPU. Figure.5 shows the difference between conventional von Neumann architecture and neuromorphic computing.

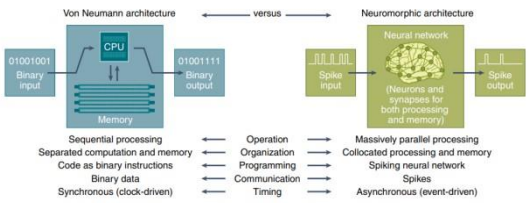


Figure.5. Von-Neumann vs Neuromorphic architecture [13]

Figure.5 shows the differences between neuromorphic architecture and von Neumann architecture, in which the classic von Neumann model is based on resistors, transistors, inductors, and other communication connections [14], though the conventional model advantages like the range of operation, resistors, transistors and many other, there are few limitations which need to be taken into consideration such as connectivity, fault tolerance, energy consumption, functionality and many more. Hence, the characteristics of neuromorphic computing are depicted as follows,

**Brain-inspired architecture** – Neuromorphic computing systems are primarily designed to mimic the structure and functionality of the human brain.

**Highly parallel**- up to 1 million neurons can be found on neuromorphic semiconductors. Each neuron carries out a simultaneous task. Tolerance of Faults – The fault tolerance of neuromorphic computers is exceedingly high.

**Combined memory and processing unit** – Instead of having distinct sections, neuromorphic computer chips, inspired by the human brain, process, and store data jointly on each neuron. Inherent scalability- Due to the extra neuromorphic chips, which emphasize amplifying the number of neurons and synapses, neuromorphic computers are typically intended to be intrinsically scalable.

**Event-driven computation** [15]– Distinct neurons and synapses calculate a response to spikes from other neurons, which means only a tiny portion of neurons essentially process spikes and use energy, and the computer remains idle the rest of the time. This results in tremendous usage of power.

Neuromorphic computing generally uses ANN (Artificial Neural Network) for performing computational tasks. Among the different types of ANN, SNN (Spiking Neural Network) is used, based on artificial neurons that communicate via electric signals called spikes. Hence, SNNs mimic NL (Natural Language) processes by animatedly plotting synapses between artificial neurons in response to stimuli by interpreting the data within its signals and time.

## 4. Neuromorphic Computing for Edge AI

As the demand for IoT increases, existing architectures face different challenges like overload of servers, cost of high bandwidth, and many more. Hence, the recommended study [16] has used the neuromorphic computing hardware method for Albased edge computing. Similarly, the suggested study [17] has used the NeuRRAM model for nextgeneration edge AI hardware. The performance of SNNs on neuromorphic hardware in edge computing setup has been focused on in the suggested study. Thus, the EdgeMap framework [18] has been used. Further, it also predominantly focused on minimizing congestion and spike latency, which is extremely important for edge applications. Figure.6 depicts the illustration of EdgeMap.

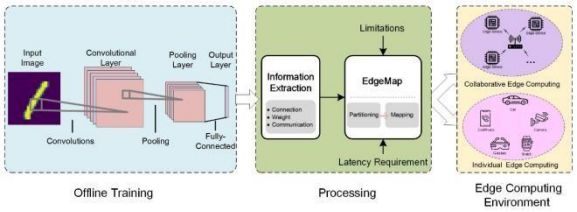


Figure.6. Overview of EdgeMap [18]

Figure.6 shows the overview of EdgeMap. Three stages of the model have taken place, which include the offline training stage, processing stage, and edge computing stage. Correspondingly, the neuromorphic technology-based NeuroEdge has been introduced in the existing work [19] for Edge AI. NeuroEdge and NM500 chips have been used in the study, in which the effectiveness of NM500 using a face recognition system has been employed with the aim to prove that the chip has the potential in hardware implementation of AI applications. NeuroEdge was based on an openboard Raspberry Pi4 and two neuromorphic chips known as NM500. Although the NM500 chip has delivered a better performance for face recognition systems, Figure.6 illustrates the application and prospects of the NM500 chip beyond face recognition.

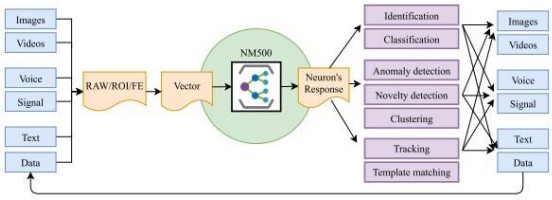


Figure.7. Applications of NM500 Chip [19]

Figure.7 shows that neuron response can be applied for tracking, novelty detection, clustering, detection, classification, and many more. Similarly, study [20] has used NeuMap for Edge AI. NeuMap initially obtained the communication pattern of SNN by estimation rather than simulation process. In order to accomplish better performance, a spike firing rate is applied to SNN with feed-forward topology like SCNN (Spiking –Convolutional Neural Network). Experimental outcomes have demonstrated that NeuMap minimizes the average energy consumption by 84%, resulting in lower spike latency.

**Merits of employing Neuromorphic computing in Edge AI**

· *Lower power consumption* – Edge based devices like embedded systems are often considered as limited power resources. SNN can significantly reduce the consumption of power when compared with the conventional computing models which makes it well suited for edge AI applications.

· *Robustness and fault tolerance* – Neuromorphic computing aids in improving fault tolerance and robustness of edge AI applications. This resilience is vital in an edge environment where dependability is essential.

· *Sensor fusion and perception* – Edge AI application typically involves processing data from various sensors like cameras and other IoT sensors.

· *Reduced latency* – By performing the computations locally at the edge, SNN can considerably reduce latency, as SNN aids in reducing the need for data transmission to a remote server.

· *Reduced Bandwidth and storage requirement* –Data processing and data analysis happen at the edge, reducing the need for transmission of data and storage on a control server.

· *Proficient resource utilization* – Localized processing and smaller NN on edge devices usually require less energy and computational resources when compared to the central sensors.

Due to these massive advantages, neuromorphic computing is implemented in Edge AI applications for various purposes described in the subsequent section.

## 5. Applications of Neuromorphic Computing in Edge Applications

Due to the low energy consumption, neuromorphic computing is considered an appropriate approach, which can be used on edge AI applications [21]. Thus, the facilitation of efficient on-device intelligence aids neuromorphic chips to unravel more value from IoT networks.

**Autonomous Vehicles**

Another great potential application for neuromorphic computing at the edge is AV (Autonomous Vehicles) [22] such as UAV (Unmanned Aerial Vehicles) or drones. If the consumption of power needed for autonomous vehicles is minimized using neuromorphic computing techniques, then it will have a great impact on the usage of batteries for upcoming autonomous vehicles.

**Aerospace**

Further, in aerospace, sensor fusion is identified as a crucial component for gathering and incorporating data from various sensors for obtaining a thorough understanding of situation.

**Autonomous Driving**

Likewise, aerospace industry, autonomous driving is characterized as one of the aspects of automobile industry in which neuromorphic computing can be used for enhancing the capabilities of the autonomous vehicles by enabling real-time operations. Further, brain inspired framework of the neuromorphic computing can improve the safety and reliability of autonomous driving system.

In recent times, smart cities are used for enhancing the quality of human life for the residents and optimize resource management. Thus, neuromorphic computing can facilitate real time analytics by processing data at the edge closer to the sources. This ability is identified to be extremely crucial for smart cities as it permits for immediate insights and responses to events like traffic congestion or other such emergencies and many more.

**Fraudulent Activities**

Edge fault detection and AI security play a huge role in order to detect anomalies and prevent fraudulent activities. This can be achieved by primarily using neuromorphic chips, which aid in analyzing financial transactions and network traffic. Moreover, it is also useful for securing sensitive data and transactions at the edge without uploading everything to the cloud for analysis. Different chips used are illustrated as follows,

· Akida [23] is considered the first commercial neuromorphic processor. Akida is advertised as a power-efficient event-based processor for edge computing, which does not require any external CPU. Akida is capable of processing at 1000 watts. Around 80 Neural processing units are used in 1 Akida chip in a mesh network, which enables 1,200,000 neurons and 10,000,000,000 synapses. Akida chip was primarily built at TSMC nm. Further, a free chip emulator is also provided using the Akida ecosystem.

· NM500 chip [24] enables EC, an extension of IoT, to spread throughout the industry. Thus, a small-scale vehicle has been assembled with the help of Arduino peripherals and NM500 to discuss real-time performance and hardware implementation. This NM500 completed the learning process and surely identified the road signals in all

possible cases, even with minimal no. of. neurons. Table 1 depicts the common applications and descriptions of neuromorphic computing in Edge AI.

Table-1. Application and Description

| Application | Description |
|---|---|
| Robotics | Typically, neuromorphic controllers are trained to handle complex and dynamic environments, which enables the robots to learn from previous experiences and adapt behavior in real-time. |
| Real-time Anomaly detection and sensor fusion | Conventional techniques mostly slip in real time due to the implementation of weak approaches. This often leads to missing substantial prompts in the cascade of information.. |

| | |
|---|---|
| Image recognition and video recognition at the edge | Neuromorphic computing excels in the realm of image recognition and video recognition, especially for low-power object recognition and image classification. |

Table 1 deals with different applications of neuromorphic computing for edge AI, where different applications like image and video recognition, realtime anomaly, and robotics are discussed.

## 6. Challenges and Future Recommendations

Despite reasonable advantages and exciting applications created by using Neuromorphic computing, there are a few shortcomings that can be overcome in the future,

**Constrained Resources**

When compared to server or cloud-based systems, edge devices often have limited computational resources, power, and memory. Thus, neuromorphic computing algorithms need to be tailored to overcome the challenging aspects.

**Hardware Restrictions**

Developing scalable and competent hardware for neuromorphic computing can be considered a stimulating task. Thus, designing specialized circuits and chips that can mimic the complicated NN of the human brain while uploading computational power and energy efficacy of the model is demonstrated as a

substantial obstacle.

**Lack of clearly established benchmarks**

Without standard metrics and benchmarks, evaluating which hardware system is appropriate for a given application is tremendously challenging. Further, assessing a model becomes extremely perplexing without predefined metrics [13]. **Energy efficacy optimization**

Optimizing the power consumption of neuromorphic hardware and algorithms is considered an important aspect of resource-constrained edge devices. Moreover, another major challenge of hardware implementation comprises of reconfigurability and flexibility of the network. Thus,

to become industrially relevant, neuromorphic chips based on resistive devices should take network reconfigurability into consideration [25].

**Algorithm complexity and software tools**

Implementing and optimizing algorithms for neuromorphic computing can be daunting. Thus, specialized training should be provided for NN implemented in neuromorphic systems. Further, the development of effective and scalable algorithms for SNN remains vigorous research. Hence, appropriate algorithms will be focused on resource-intensive tasks.

These challenges can be overcome by enhancing the existing works in future as future work. Thus, some of the future recommendations include,

Ø In the future, amalgaming strengths of conventional algorithms and neuromorphic algorithms can leverage the efficacy of SNN for particular tasks while utilizing the computational power of CPU/GPU for complex processing.

Ø Unrelenting developments in hardware design and fabrication methods play a significant role in neuromorphic computing for developing effective and specialized neuromorphic chips and architecture.

Ø Recognizing and focusing on specific domains or industries is considered to be another future direction.

## 7. Conclusion

Neuromorphic computing holds enormous potential for transfiguring Edge AI applications due to various merits like the capability to handle complicated data, less power consumption, immense execution speed, energy efficiency, and many more when compared to the existing methods as it functions like the human

brain. Due to these exciting aspects of neuromorphic computing, the present SLR focused on reviewing studies that employed neuromorphic computing and neuromorphic computing for Edge AI. Therefore, different aspects, like using neuromorphic chips and controllers, are reviewed. Further, applications of neuromorphic computing for Edge AI were also discussed. However, limited studies have dealt with neuromorphic computing for Edge AI as it is an impending research work, and the field of neuromorphic computing for Edge AI is still relatively new. Thus, the present SLR recommends future directions to overcome the existing challenges and drawbacks, which helps researchers work on neuromorphic computing efficiently.

## References


[1] L. Liu, H. Zhu, T. Wang, and M. J. E. Tang, "A Fast and Efficient Task Offloading Approach in EdgeCloud Collaboration Environment," vol. 13, no. 2, p. 313, 2024.

[2] R. S. Rajan, V. J. I. J. o. A. R. i. E. Vasudevan, and Technology, "Enhanced priority based load balance scheduling of parallel work load in cloud computing," vol. 11, no. 4, 2020.

[3] Y. Himeur, A. Sayed, A. Alsalemi, F. Bensaali, and A. J. I. o. T. Amira, "Edge AI for Internet of Energy: Challenges and perspectives," p. 101035, 2023.

[4] S. S. Bhandari, R. Devullapalli, A. Swapnil, R. Karri, and C. S. Gopi, "EDGE COMPUTING," 2023.

[5] B. Gökgöz, F. Gül, T. J. C. Aydın, C. Practice, and Experience, "An overview memristor based hardware accelerators for deep neural

[6] W. Wang, H. Zhou, W. Li, and E. Goi, "Neuromorphic computing," in Neuromorphic Photonic Devices and Applications: Elsevier, 2024, pp. 27-45.

[7] Y. Tuchman et al., "Organic neuromorphic devices: Past, present, and future challenges," vol. 45, no. 8, pp. 619-630, 2020.

[8] B. Sun et al., "Synaptic devices based neuromorphic computing applications in artificial intelligence," vol. 18, p. 100393, 2021.

[9] J. Q. Yang et al., "Neuromorphic engineering: from biological to spike-based hardware nervous systems," vol. 32, no. 52, p. 2003610, 2020.

[10] L. J. S. Luo, "Architectures of neuronal circuits," vol. 373, no. 6559, p. eabg7285, 2021.

[11] W. Shen and Q. Zhang, "Large-scale neuromorphic systems enabled by integrated photonics," in Neuromorphic Photonic Devices and Applications: Elsevier, 2024, pp. 191-220.

[12] E. Goi and M. Gu, "Perspective on photonic neuromorphic computing," in Neuromorphic Photonic Devices and Applications: Elsevier, 2024, pp. 353375.

[13] C. D. Schuman, S. R. Kulkarni, M. Parsa, J. P. Mitchell, and B. J. N. C. S. Kay, "Opportunities for neuromorphic computing algorithms and applications," vol. 2, no. 1, pp. 10-19, 2022.

[14] N. Zins, Y. Zhang, C. Yu, and H. An, "Neuromorphic computing: A path to artificial intelligence through emulating human brains," in Frontiers of Quality Electronic Design (QED) AI, IoT and Hardware Security: Springer, 2023, pp. 259-296.

[15] N. Rathi et al., "Exploring neuromorphic computing based on spiking neural networks: Algorithms to hardware," vol. 55, no. 12, pp. 1-49, 2023.

[16] B.-S. Lin et al., "Fall detection system with artificial intelligence-based edge computing," vol. 10, pp. 4328-4339, 2022.

[17] W. Wan et al., "Edge AI without compromise: efficient, versatile and accurate neurocomputing in resistive random-access memory," 2021.

[18] J. Xue et al., "EdgeMap: An Optimized Mapping Toolchain for Spiking Neural Network in Edge Computing," vol. 23, no. 14, p. 6548, 2023.

[19] C. I. Nwakanma, J.-W. Kim, J.-M. Lee, and D.-S. J. I. E. Kim, "Edge AI prospect using the NeuroEdge computing system: Introducing a novel neuromorphic technology," vol. 7, no. 2, pp. 152-157, 2021.

[20] C. Xiao, J. Chen, and L. J. S. Wang, "Optimal Mapping of Spiking Neural Network to Neuromorphic Hardware for Edge-AI," vol. 22, no. 19, p. 7248, 2022.

[21] Z. Chang, S. Liu, X. Xiong, Z. Cai, and G. J. I. I. o. T. J. Tu, "A survey of recent advances in edgecomputing-powered artificial intelligence of things," vol. 8, no. 18, pp. 13849-13875, 2021.

[22] C. Schuman et al., "Evolutionary vs imitation learning for neuromorphic control at the edge," vol. 2, no. 1, p. 014002, 2022.

[23] D. Ivanov, A. Chezhegov, M. Kiselev, A. Grunin, and D. J. F. i. N. Larionov, "Neuromorphic artificial intelligence systems," vol. 16, p. 1513, 2022.

[24] J. Kim, "New neuromorphic AI NM500 and its ADAS application," in AETA 2018-Recent Advances in Electrical Engineering and Related Sciences: Theory and Application, 2020, pp. 3-12: Springer.

[25] D. Ielmini and S. J. N. Ambrogio, "Emerging neuromorphic devices," vol. 31, no. 9, p. 092001, 2019.