DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING

VEMANA INSTITUTE OF TECHNOLOGY

(Approved by AICTE, New Delhi, Recognised by Govt. of Karnataka & Affiliated to VTU, Belagavi)

Koramangala, Bengaluru – 560 034.

“Optimizing Memristor-Based Synaptic Devices for Enhanced Energy Efficiency and Accuracy in Neuromorphic Machine Learning”

A seminar delivered by

SUNAY S

1VI21IS106

Under the guidance

of

Mr. Manjunath J P

Assistant professor, Dept. of ISE

Vemana IT, Bengaluru – 34

This report is submitted to

VISVESVARAYA TECHNOLOGICAL UNIVERSITY

Belagavi, Karnataka-590018

in partial fulfilment of the requirements for the award of the Degree of

Bachelor of Engineering

In

Information Science and Engineering

Karnataka ReddyJana Sangha®

VEMANA INSTITUTE OF TECHNOLOGY

Koramangala, Bengaluru – 34

(Affiliated to Visvesvaraya Technological University, Belagavi)

Department of Information Science & Engineering

Certificate

This is to certify that **SUNAY S (1VI21IS106)** , a student of VIII semester, Information Science & Engineering has presented the seminar entitled **“Optimizing Memristor-Based Synaptic Devices for Enhanced Energy Efficiency and Accuracy in Neuromorphic Machine Learning”** and is submitting this report in partial fulfilment for the requirement of the award of degree of Bachelor of Engineering in Information Science and Engineering of the Visvesvaraya Technological University, Belagavi during the academic year 2024-2025. The seminar report has been approved as it satisfies the academic requirements in respect of seminar work prescribed for the said degree.

Signature of Guide	Signature of HOD
Mr. Manjunath J P	**Dr. Rajanna M**

Name of the Examiners	Signature with date
1	
2	

ACKNOWLEDGEMENT

Firstly, I would like to express my deep sense of gratitude to our beloved Principal **Dr. Vijayasimha Reddy. B. G,** for providing us the necessary support.

I would like to place on record my regards to **Dr. Rajanna. M,** Head of the Department, Information Science and Engineering for his continued support.

I would like to thank my seminar guide **Mr. Manjunath J P,** Assistant Professor, Department of ISE for his continuous support and valuable guidance towards successful completion of the presentation of technical seminar. It is due to his persistent guidance that I went through all the IEEE and other technical journals. He made this entire process an enjoyable learning experience.

I would like to thank my seminar coordinator **Mr. Manjunath J P,** Assistant professor and **Mrs. Chandana D C**, Assistant professor, Dept. of ISE for their support and coordination.

I would be failing in my duty if I do not thank the faculty members, Lab staffs, technicians, family members and friends for their constant support and guidance.

Date: 02/05/2025 **SUNAY S**

Place: Bengaluru **(1VI21IS106)**

ABSTRACT

The Von Neumann architecture faces critical limitations in deep learning (DL) and machine learning (ML) due to the energy and time costs of frequent data transfers between the processor and external memory. This bottleneck hinders the efficiency of cognitive computing tasks. To address this, memristive synaptic devices offer a promising solution by enabling in-memory computing, where memory and processing are integrated into the same location. This significantly reduces data transfer, lowering both energy consumption and computational time. Memristors, particularly nano-scale titanium dioxide (TiO_2)-based devices, have attracted attention for their potential to mimic biological synapses in neuromorphic systems. However, achieving high accuracy and low error rates in classification tasks still demands further optimization. Enhancing performance in such systems requires not only efficient hardware but also effective algorithmic strategies. This paper provides a comprehensive comparison of stochastic gradient descent (SGD) and its momentum-based variants, focusing on their learning capabilities, energy efficiency, and classification accuracy when used with memristor-based neuromorphic platforms. By analyzing how these algorithms perform on TiO_2 memristive hardware, the study highlights the importance of co-optimizing both hardware and learning algorithms to maximize efficiency and performance in neuromorphic computing systems.

CONTENTS

LIST OF FIGURES		
Figure No.	Title	Page No.

CHAPTER - 1

INTRODUCTION

Machine learning (ML) and deep learning (DL) have rapidly evolved from academic research tools into essential technologies with wide-ranging industrial applications. Their significance has grown alongside advancements in computing technology, making them foundational elements in modern artificial intelligence and data science. Despite these advances, the traditional Von Neumann architecture poses substantial limitations for ML and DL methods, especially those inspired by biological neural networks. This architecture separates memory and processing units, leading to frequent data transfers that consume excessive energy and time. Such inefficiencies become particularly problematic when scaling to complex models and large datasets. In contrast, biological neural systems, such as the human brain, integrate memory and computation, enabling highly efficient and parallel information processing. For instance, while supercomputers may consume over 1 megawatt of power, the human brain achieves remarkable cognitive abilities using only around 10 watts. This stark difference highlights the need for more energy-efficient architectures. Neuromorphic computing, which mimics the brain's parallel and distributed processing, offers a promising solution. In particular, memristive devices—hardware components that combine memory and processing—enable such architectures. These devices support in-memory computing and asynchronous processing, drastically reducing energy consumption and latency. As optimization remains a critical component of ML success, developing faster, more efficient learning algorithms on neuromorphic hardware is vital. Memristive systems, when combined with advanced optimization methods, show great potential for enhancing performance, enabling accurate and efficient learning while addressing the energy limitations of conventional systems. This research explores the optimization of memristor-based architectures for improved energy efficiency and accuracy in ML and DL applications.

CHAPTER - 2

REVIEW OF LITERATURE WORK

There are a few existing survey papers that discuss the potential use of memristor-based synaptic devices in supporting neuromorphic computing and facilitating machine learning (ML) and deep learning (DL) tasks, as shown. These literatures cover a wide range of topics, such as how synaptic devices can enhance energy and time efficiency in cognitive computing, how neuromorphic architectures are developed and utilized, which materials and technologies (e.g., TiO_2-based memristors) are leading in this field, how hardware-level integration reduces the dependency on traditional data transfer mechanisms, and the classification and comparison of learning algorithms like SGD and its variants for optimizing memristive systems.

[1] Memristors for Energy-Efficient New Computing Paradigms

Authors: Doo Seok Jeong, Kyung Min Kim, Sungho Kim, Byung Joon Choi
Year: 2016
Venue: Wiley Advanced

This review examines memristors within both von Neumann and neuromorphic computing frameworks. In von Neumann systems, memristors enable a novel stateful logic approach using material implication, where logic and memory are integrated, reducing energy used in data transfer and refresh. Neuromorphic computing, inspired by the human brain, aims to replicate functions like image and voice recognition. Unlike power-hungry supercomputers, the brain performs these tasks with just 10–20 Watts. To achieve similar efficiency, hardware-based cognitive computing is essential. The review also highlights recent advancements in memristor technology, addressing material and processing challenges that promise significant energy savings in future information systems.

[2] A Review of Artificial Spiking Neuron Devices for Neural Processing and Sensing

Authors: Joon-Kyu Han, Seong-Yun Yun, Sang-Won Lee, Ji-Man Yu, Yang-Kyu Choi

Year: 2022

Venue: Wiley Advanced

A spiking neural network (SNN) inspired by the structure and principles of the human brain can significantly enhance the energy efficiency of artificial intelligence computing by overcoming the bottlenecks of the conventional von Neumann architecture with its massive parallelism and spike transmissions. The construction of artificial neurons is important for the hardware implementation of an SNN, which generates spike signals when enough synaptic signals are gathered. Because circuit-level artificial neurons with comparator and reset circuits require considerable hardware area, intensive efforts are devoted in recent years for building artificial neurons at the device level for better area efficiency. Furthermore, artificial sensory neuron devices, which perform neural processing and sensing concurrently, have recently been developed in order to reduce the hardware cost and energy consumption of traditional sensory systems through in-sensor computing. This review article surveys and benchmarks the recent progress of artificial neuron devices for neural processing and sensing. First, various artificial neuron devices are summarized, including single-transistor neurons (1T-neurons), memristor neurons, phase-change neurons, magnetic neurons, and ferroelectric neurons. Next, cointegration technologies with artificial synaptic devices and artificial sensory neurons for in-sensor computing are introduced. Finally, the challenges and prospects for developing artificial neuron devices are discussed.

[3] Performance Prospects of Deeply Scaled Spin-Transfer Torque Magnetic Random-Access Memory for In-Memory Computing

Authors: Yuhan Shi, Sangheon Oh, Zhisheng Huang, Xiao Lu, Seung H. Kang, Duygu Kuzum

Year: 2020

Venue: IEEE Electron Device Letters

In recent years, Spin-Transfer-Torque Magnetic Random Access Memory (STT-MRAM) has been considered as one of the most promising non-volatile memory candidates for in-memory computing. However, system-level performance gains using STT-MRAM for in-memory computing at deeply scaled nodes have not been assessed with respect to more mature memory technologies. In this letter, we present perpendicular magnetic tunnel junction (pMTJ) STT-MRAM devices at 28nm and 7nm. We evaluate the system-level performance of convolutional neural network (CNN) inference with STT-MRAM arrays in comparison to Static Random Access Memory (SRAM). We benchmark STT-MRAM and SRAM in terms of area, leakage power, energy, and latency from 65nm to 7nm technology nodes. Our results show that STT-MRAM keeps providing ~5× smaller synaptic core area, ~20× less leakage power, and ~7× less energy than SRAM when both devices are scaled from 65nm to 7nm. With the emerging need for low power computation for a broad range of applications such as internet-of-things (IoT) and neural network (NN), STT-MRAM can offer energy-efficient and high-density in-memory computing.

[4] In situ training of feed-forward and recurrent convolutional memristor networks

Authors: Zhongrui Wang, Can Li, Peng Lin, Mingyi Rao, Yongyang Nie, Wenhao Song, Qinru Qiu, Yunning Li, Peng Yan, John Paul Strachan, Ning Ge, Nathan McDonald, Qing Wu, Miao Hu, Huaqiang Wu, R. Stanley Williams, Qiangfei Xia & J. Joshua Yang

Year: 2019

Venue: Nature Machine Intelligence

The explosive growth of machine learning is largely due to the recent advancements in hardware and architecture. The engineering of network structures, taking advantage of the spatial or temporal translational isometry of patterns, naturally leads to bio-inspired, shared-weight structures such as convolutional neural networks, which have markedly reduced the number of free parameters. State-of-the-art microarchitectures commonly rely on weight-sharing techniques, but still suffer from the von Neumann bottleneck of transistor-based platforms. Here, we experimentally demonstrate the in situ training of a five-level convolutional neural network that self-adapts to non-idealities of the one-transistor one-memristor array to classify the MNIST dataset, achieving similar accuracy to the memristor-based multilayer perceptron with a reduction in trainable parameters of ~75% owing to the shared weights. In addition, the memristors encoded both spatial and temporal translational invariance simultaneously in a convolutional long short-term memory network—a memristor-based neural network with intrinsic 3D input processing—which was trained in situ to classify a synthetic MNIST sequence dataset using just 850 weights. These proof-of-principle demonstrations combine the architectural advantages of weight sharing and the area/energy efficiency boost of the memristors, paving the way to future edge artificial intelligence.

CHAPTER 3

OBJECTIVE AND METHODOLOGY

3.1 OBJECTIVE

1. **Address Energy and Time Inefficiencies in Traditional Architectures:**
   The study aims to highlight and mitigate the limitations of the Von Neumann architecture, particularly the high energy and time costs incurred due to intensive data transfers between memory and the processor during deep learning (DL) and machine learning (ML) tasks.
2. **Promote the Use of Memristive Synaptic Devices:**
   It explores the potential of memristive devices, which integrate memory and processing in the same physical location, significantly reducing the need for data transfers and, consequently, energy and time consumption.
3. **Support Neuromorphic Computing Advancements:**
   By leveraging memristors in neuromorphic systems, the study focuses on improving the efficiency of cognitive computing models that mimic the brain's low-power and high-efficiency information processing capabilities.
4. **Improve Accuracy and Reduce Error Rates:**
   The research emphasizes the need to enhance the classification accuracy and reduce test error rates of DL and ML models implemented on memristor-based architectures.
5. **Combine Hardware and Algorithmic Optimization:**
   The study underscores the importance of co-optimizing both hardware (memristor-based synaptic devices) and software (learning algorithms) to achieve superior model performance.
6. **Evaluate and Compare Learning Algorithms:**
   It provides a detailed analysis of the performance of commonly used optimization algorithms, such as SGD, Momentum, Adam, RMSprop, and others, in terms of energy efficiency, learning capability, and classification accuracy.
7. **Benchmark on Standard Datasets:**
   The study validates experimental results using MNIST and CIFAR datasets, showing

that Momentum and SGD deliver the highest accuracy, with Momentum achieving 91.25% (CIFAR) and AdaDelta achieving 90.51%, indicating the importance of optimizer selection in neuromorphic systems.

8. **Contribute to Future Low-Power AI Systems:**
   Overall, the study aims to advance the development of energy-efficient, high-performance AI systems using memristor-based neuromorphic computing.

3.2 METHODOLOGY

1.StochasticGradientDescent(SGD):

SGD is an optimization algorithm that updates the model parameters iteratively based on the gradient of the loss function. In each iteration, it randomly selects a subset of data (mini-batch) and calculates the gradient of the loss with respect to the parameters. The parameters are updated by subtracting the gradient scaled by the learning rate. This approach makes it computationally efficient for large datasets, but it can introduce noise in the updates, leading to fluctuations in the convergence path. Despite this, it often converges faster than batch gradient descent for large datasets.

2.Adam(AdaptiveMomentEstimation):

Adam is an adaptive learning rate optimization algorithm that combines the concepts of momentum and RMSprop. It computes the first moment (mean) and the second moment (uncentered variance) of the gradients, adjusting learning rates accordingly. The algorithm also applies bias correction to the moment estimates, which is particularly helpful for dealing with sparse gradients. Adam is popular in training deep learning models due to its efficiency, ease of implementation, and ability to work well with both sparse and dense datasets. It generally converges faster than SGD and has fewer hyperparameters to tune.

3.RMSprop(RootMeanSquarePropagation):

RMSprop is an adaptive learning rate optimization algorithm that improves upon AdaGrad by addressing the issue of rapidly diminishing learning rates. It computes an exponentially decaying average of past squared gradients and uses this to normalize the gradient at each step. This allows RMSprop to dynamically adjust the learning rate, preventing it from shrinking too quickly. It is particularly effective for training on non-stationary objectives and is often used in recurrent neural networks (RNNs) and other deep learning applications. RMSprop is less sensitive to the choice of the learning rate compared to other algorithms.

4.AdaGrad(AdaptiveGradientAlgorithm):

AdaGrad is an optimization algorithm that adapts the learning rate for each parameter by considering the historical gradient information. For parameters that experience large gradients, AdaGrad reduces their learning rate, and for parameters with small gradients, it increases the learning rate. This results in a more efficient training process, especially when dealing with sparse data. However, AdaGrad's major limitation is that it accumulates the squared gradients, which causes the learning rate to decrease drastically over time, potentially leading to premature convergence. It works well for problems with sparse features, such as natural language processing tasks.

5.AdaDelta:

AdaDelta is an adaptive learning rate optimization algorithm designed to address the shortcomings of AdaGrad. Unlike AdaGrad, AdaDelta does not require a manually set learning rate. It uses a moving average of the squared gradients and updates the parameters using this average. The algorithm limits the accumulation of gradient information by using a fixed window size for the moving average. AdaDelta ensures that the learning rate adapts to the data, improving performance by preventing the learning rate from diminishing too quickly. This method is often used in training deep networks where traditional gradient descent methods are inefficient.

6.Nadam(Nesterov-Accelerated Adaptive Moment Estimation):
Nadam is an optimization algorithm that combines Adam with Nesterov momentum. Nesterov momentum helps accelerate convergence by looking ahead to the future position in the direction of momentum, rather than using the current gradient direction. This results in faster and more stable convergence. Nadam computes both the first and second moments of the gradients and applies bias correction, similar to Adam, but with the added benefit of Nesterov’s accelerated gradient. Nadam is often preferred in deep learning models as it can achieve faster convergence and better performance in tasks like image recognition and natural language processing.

7.AdaMax:
AdaMax is a variant of the Adam optimization algorithm that uses the infinity norm ($L\infty$ norm) instead of the L2 norm used in Adam for calculating the moment estimates. This modification aims to improve stability in certain deep learning models, especially in cases where the second moment of the gradients becomes too large. The use of the infinity norm provides a more stable and scalable method for optimizing parameters, making AdaMax particularly useful when dealing with large networks or challenging optimization problems. It is especially effective in situations where Adam struggles to converge or suffers from numerical instability.

8.Momentum:

Momentum optimization helps accelerate gradient descent by considering the velocity of previous gradients, mimicking the concept of physical momentum. The algorithm maintains an exponentially weighted moving average of past gradients and combines this momentum with the current gradient to update the parameters. This method speeds up convergence by allowing the optimizer to overcome small fluctuations in the gradient, helping it move smoothly toward the optimum. Momentum reduces oscillations and can lead to faster convergence, especially when the loss function has shallow or flat regions. It's often used in conjunction with other optimization methods like SGD for enhanced performance.

CHAPTER 4

4.1MATERIAL AND DEVICE PROPERTIES

Neuromorphic computing systems based on memristors offer a significant advantage in terms of energy efficiency compared to traditional software-based neural networks. Memristors, as analog resistors, are capable of performing computations within memory, eliminating the need for separate computational and memory units as in von Neumann architecture. This characteristic allows them to perform complex calculations with lower energy consumption and faster processing times, making them ideal for large-scale neural network implementations. Memristors also offer a high degree of integration with existing semiconductor technologies, such as CMOS, which aids in their widespread use for various applications. Binary oxides, including titanium dioxide (TiO2), hafnium, zinc, vanadium, and nickel, are often used in memristors due to their ability to undergo gradual resistance switching, essential for mimicking the behavior of synaptic connections in biological systems. These materials are inexpensive, easy to produce, and compatible with current manufacturing processes. TiO2, in particular, is promising due to its prevalence in CMOS technology and its ability to integrate seamlessly into crossbar architectures used for in-memory computations. The electrochemical reactions in TiO2 memristors allow for resistance switching, which simulates the functionality of biological synapses. This behavior makes TiO2-based devices highly suitable for neuromorphic computing, where they can help replicate the processes of learning and memory in artificial neural networks.

4.2 CHARACTERISTICS OF TiO2 SYNAPTIC DEVICE

The characteristics of TiO2 memristors make them highly suitable for various high-tech applications, including Resistive Random-Access Memory (RRAM), biohybrid systems, and sensors. TiO2-based memristors can store data by switching between two resistance states, maintaining information even when power is turned off. This resistance-switching behavior is beneficial for neuromorphic systems, where precise and dynamic adjustments in resistance are needed to model the synaptic weight changes associated with learning in neural networks.

However, while memristors ideally should exhibit a linear correlation between synaptic weight changes and the number of write pulses, real-world devices often exhibit deviations from this behavior. These deviations occur because the conductivity of memristors changes more rapidly during the early stages of Long-Term Potentiation (LTP) and Long-Term Depression (LTD), before eventually reaching a saturation point. The non-linear nature of these changes is important to understand for improving the accuracy and efficiency of neuromorphic systems. To model these non-linear weight updates, a device model was created that incorporates parameters like peak conductance, minimum conductance, and maximum pulse count required for transitioning between states. This model is used to predict how changes in resistance (conductance) will occur in response to different pulse sequences, facilitating optimization of memristor-based synaptic devices for better performance in neuromorphic applications.

4.3 DATASETS AND EXPERIMENTAL CONFIGURATIONS

In the evaluation of the optimization algorithms, several well-known datasets, including MNIST and CIFAR-10, were used to assess the performance of the memristor-based synaptic devices. The MNIST dataset is widely recognized and used in machine learning for handwritten digit recognition. It contains 60,000 training samples and 10,000 test samples, allowing models to be trained and tested effectively. The dataset is particularly beneficial for comparing optimization algorithms due to its simplicity and consistency in recognizing handwritten characters. Additionally, a new dataset was created by digitizing computer-generated text images into 28 × 28 pixel representations, alongside the original handwritten digit dataset. This new dataset, designed to enhance information richness, includes 56 features per input vector, combining both perpendicular and parallel vector structures for more effective recognition. The CIFAR-10 dataset, commonly used for image classification tasks, consists of 60,000 32 × 32 pixel RGB images divided into 10 classes. The CIFAR-10 dataset is used for testing models designed for more complex image recognition tasks. Both MNIST and CIFAR-10 serve as benchmark datasets to test and compare the effectiveness of various optimization algorithms in improving the accuracy and efficiency of memristor-based neural networks.

4.4 PROPOSED APPROACH

The proposed approach in this study focuses on the implementation of neural networks using TiO2-based memristor devices. The evaluation of this hardware-based neural network model considers several factors, including accuracy, energy consumption, and area usage, and employs various optimization algorithms to improve overall performance. The model is designed to recognize and classify computer-generated and handwritten digits, with a feedforward (FF) and backpropagation (BP) process for learning. The FF process involves feeding input data into the network, passing it through hidden layers using weighted sums and activation functions, and generating an output.

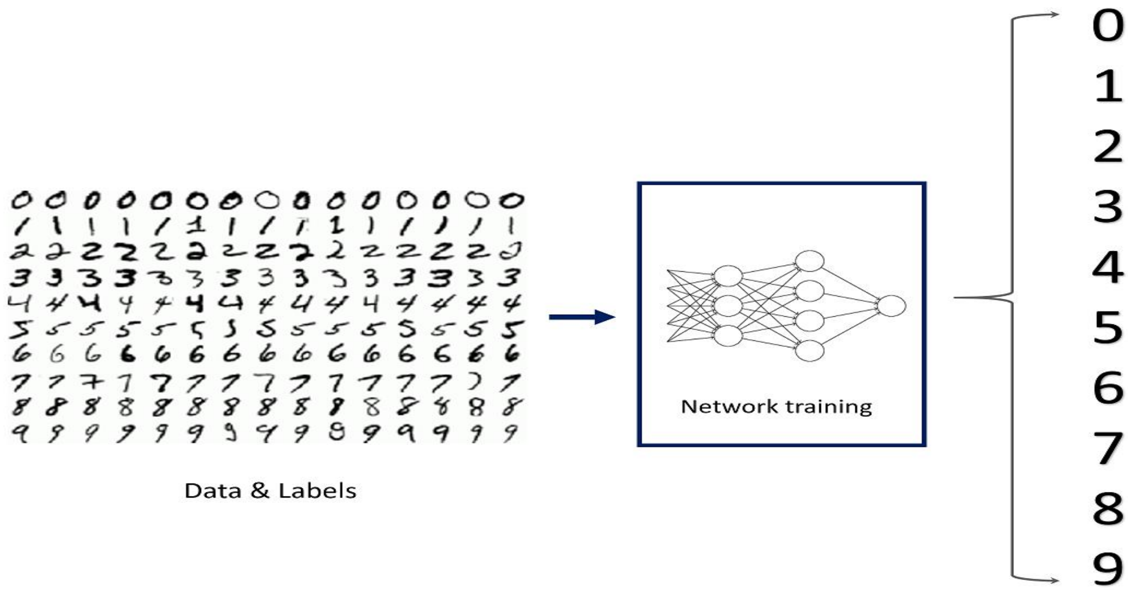


Fig 4.1: Artificial Neural Network Processing the MNIST Dataset

In the BP phase, the prediction error is propagated backward to adjust the weights, reducing the error over time. The optimization methods applied during the backpropagation process include Adadelta, AdaGrad, Adam, AdaMax, Momentum, Nadam, RMSprop, and Stochastic Gradient Descent (SGD). These methods help fine-tune the weight updates, which differ from traditional gradient descent by updating weights for individual images immediately after the FF phase. The optimization methods aim to minimize the prediction error and enhance the overall performance of the neural network. The testing phase involves using the trained weights to make predictions on new data, allowing for the evaluation of the model's generalization ability and accuracy in digit recognition and classification tasks.

4.5 IMPLEMENTATION OF MEMRISTOR-BASED NEURAL NETWORK USING TiO2 SYNAPTIC-BASED DEVICE

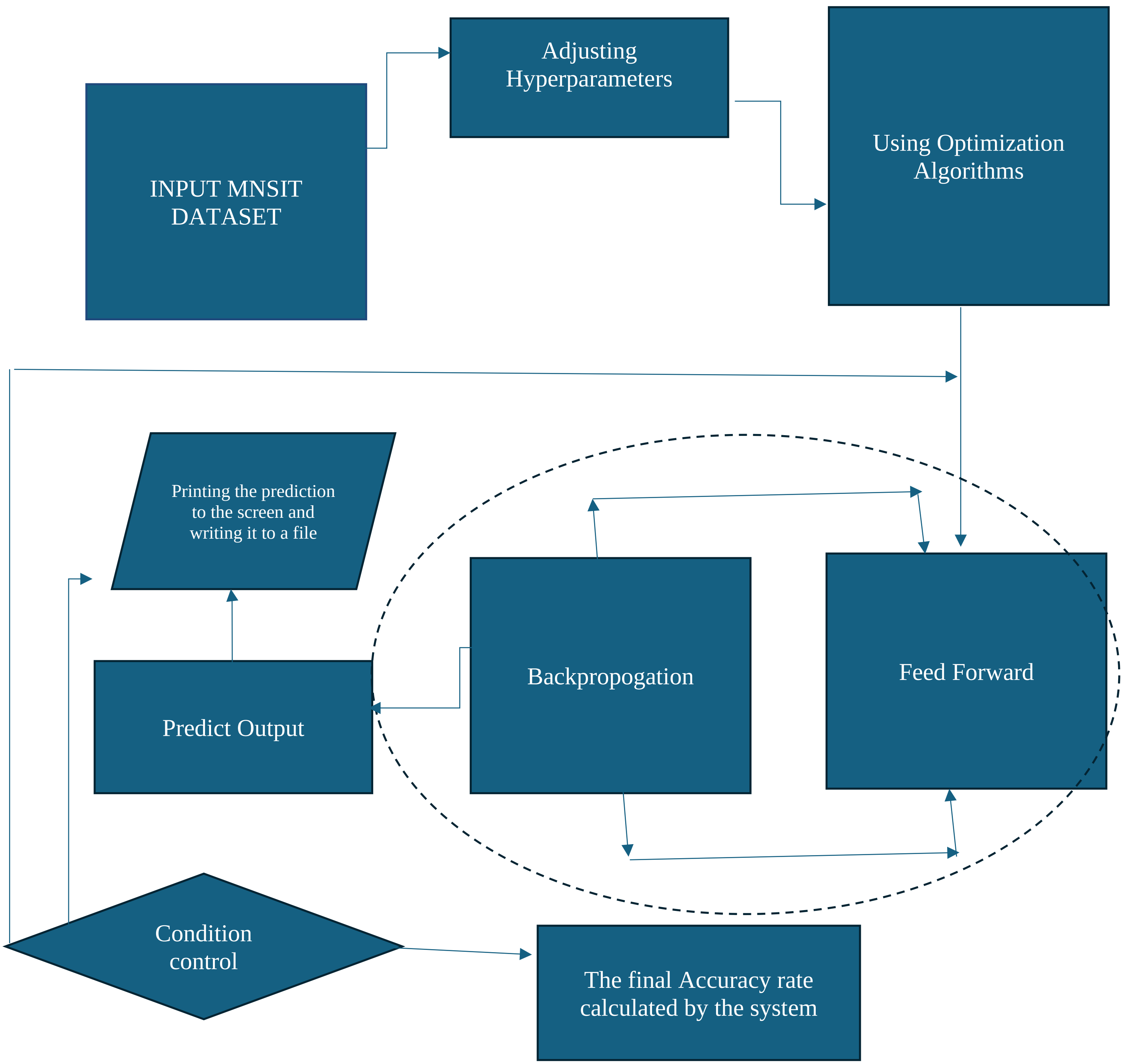


Fig 4.2: Flow Diagram of the working of Memristive Synaptic Devices

The implementation of TiO2-based memristor devices in neural networks offers a more efficient and compact alternative to traditional computing architectures. By leveraging the memristive properties of TiO2, the neural network model can simulate the behavior of biological synapses, enhancing the speed and energy efficiency of learning processes. Memristive synaptic devices allow for the in-memory computation necessary for neuromorphic systems, where data storage and processing occur within the same unit, reducing energy consumption associated with data transfer.

In this study, a two-layer multilayer perceptron (MLP) neural network was used to compare the performance of memristive synaptic devices in hardware. The network, consisting of an input layer, hidden layer, and output layer, performs the task of digit recognition using the MNIST dataset. A key challenge in hardware implementation is that traditional neural networks require both positive and negative weight values, while memristor devices typically only support non-negative weight values. To address this, a conversion algorithm is employed to map the weight values from the range of -1 to 1 into the range of 0 to 1, ensuring compatibility with the memristive hardware. By using specialized software tools to measure energy consumption and system performance, this implementation aims to optimize the speed, energy efficiency, and accuracy of neural networks based on TiO2-based synaptic devices, offering significant advancements in AI and machine learning technologies.

CHAPTER 5

PERFORMANCE EVALUATION

5.1 QUALITATIVE EVALUATION

The qualitative evaluation of TiO_2-based neuromorphic computing devices focuses on analyzing the performance trends and behavioral characteristics of neural network models trained with different optimization algorithms. This approach emphasizes how changes in algorithm selection qualitatively affect learning behavior, model stability, and convergence during training. As observed in comparative graphs, different optimizers yield significantly distinct learning trajectories even under identical epoch settings. For instance, AdaDelta consistently exhibits smoother convergence curves and higher terminal accuracies, indicating a more robust learning dynamic. The Momentum optimizer also demonstrates notable improvement in early-stage learning compared to its pair, Adam, which stabilizes slower and at a lower accuracy. The stability of TiO_2-based synaptic devices also plays a vital qualitative role. These memristive devices emulate biological synapses more naturally than traditional CMOS-based elements, enabling spiking neural networks and adaptive weight modulation. This hardware characteristic allows for real-time adaptability and learning in low-power environments. In qualitative terms, the integration of these devices into circuit designs introduces enhanced neuromorphic realism, aligning better with cognitive architectures observed in biological systems. From a usability standpoint, the learning curves' visual characteristics aid in selecting optimizers that show less oscillation and noise, which are desirable for real-time embedded systems. TiO_2 synapses' consistency across datasets such as MNIST and CIFAR-10 further supports their reliability for scalable applications. Hence, qualitative insights are indispensable for refining the selection of training strategies and hardware configurations. They highlight the experiential, real-world implications of algorithm-device interactions, enabling more informed architectural and deployment decisions in neuromorphic system design.

5.2 QUANTITATIVE EVALUATION

Quantitative evaluation provides precise performance metrics to compare different optimization algorithms and their impact on TiO_2-based neural network models. The most important metric is accuracy, which reflects the proportion of correctly predicted labels

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Where:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

Other supporting metrics include:

$$\text{Error Rate} = \frac{FP + FN}{TP + TN + FP + FN}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

The study reveals distinct differences in these metrics across various optimizers. For example, on the MNIST dataset, AdaDelta achieved an accuracy of 89.48%, closely followed by SGD (89.47%) and Momentum (88.55%). Conversely, AdaGrad recorded the lowest at 79.00%. On the CIFAR-10 dataset, AdaDelta again outperformed others with 92.51% accuracy.

These values quantitatively confirm the superior generalization of AdaDelta and its effectiveness in diverse data environments. Moreover, metrics like sensitivity and specificity help to understand the model's ability to recognize both positive and negative classes correctly-vital in imbalanced datasets.

Quantitative data provides strong support for comparative assessments and optimization algorithm selection in future hardware-software co-designs involving TiO_2 synapses.

5.3 COMPARISION

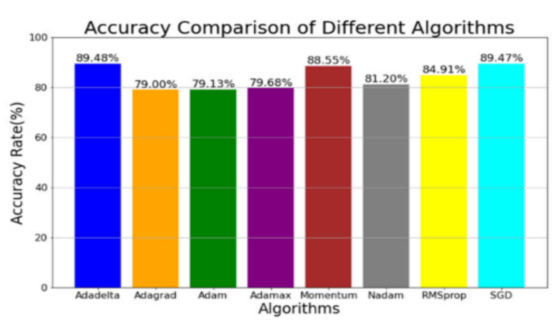
A comparative analysis of the optimization algorithms applied to TiO_2-based neuromorphic systems illustrates significant performance disparities. While all models were evaluated under similar conditions, their learning dynamics, final accuracy, and convergence behaviors varied notably.

AdaDelta consistently achieved the highest accuracy: 89.48% for MNIST and 92.51% for CIFAR-10. SGD and Momentum followed closely, with 89.47% and 88.55%, respectively, on MNIST. In contrast, AdaGrad performed the worst, with an accuracy of only 79.00%. These comparisons are highlighting not only accuracy but also how each model converges over epochs.

Despite similar epoch numbers, optimizers like AdaGrad and Adam underperform due to either aggressive learning rates or insufficient weight adjustments. This shows how sensitive neural networks are to the choice of optimizer, especially when implemented on hardware with physical constraints like TiO_2-based memristors.

A clear takeaway from this comparison is the optimization strategy's critical role in maximizing the potential of neuromorphic hardware. The results demonstrate that optimizers with adaptive learning rates (e.g., AdaDelta) better accommodate device-level variability in memristor behavior.

Furthermore, this comparison validates TiO_2 synaptic devices' compatibility across learning methods, offering flexibility in algorithm selection for specific tasks or datasets. The comparative analysis empowers designers to choose the most effective algorithm based on empirical data, optimizing both performance and energy efficiency.



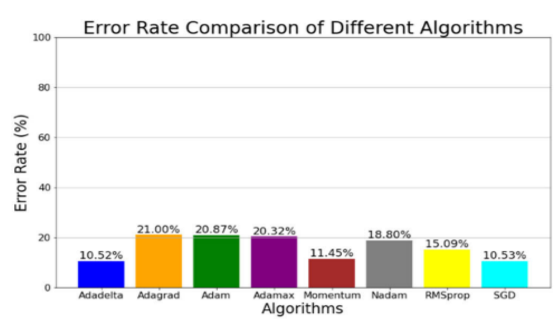


Fig 5.1: Comparison of performance efficiency among specified algorithms

CHAPTER 6

IMPACTS AND APPLICATIONS

6.1 HEDONIC VALUE

Hedonic value refers to the sensory and experiential satisfaction derived from a product or system. In the context of TiO_2-based neuromorphic computing, hedonic value lies in the novelty and sophistication of using advanced, biologically inspired technologies for complex tasks like image recognition.

The satisfaction for developers and researchers emerges from leveraging cutting-edge materials like TiO_2 memristors, which mimic biological synapses in learning and memory. These devices provide a more intuitive and human-like computing experience, opening doors to AI systems that perceive and react like organic intelligence.

From a user's perspective, real-time, responsive neural systems trained on handwritten digits (MNIST) or complex images (CIFAR-10) bring innovation closer to human-level cognition. The visual clarity of classification outcomes, rapid processing, and adaptiveness of the learning curves contribute to an engaging, almost organic experience with the system.

Additionally, hedonic value is elevated through the aesthetic and technical elegance of optimization comparisons, such as how AdaDelta gracefully achieves superior accuracy over epochs. The satisfaction derived from identifying the "best" algorithm evokes a game-like gratification, especially when plotted visually.

Overall, hedonic value in this context encompasses emotional satisfaction, sensory interaction, and intellectual stimulation—all driven by neuromorphic computing's convergence with human-like learning behaviors.

6.2 UTILITARIAN VALUE

Utilitarian value reflects the functional benefits users derive from a system. In TiO_2-based neuromorphic computing, this includes energy efficiency, high classification accuracy, and scalability for real-world applications like facial recognition, medical diagnostics, and autonomous systems.

The functional advantage is evident from the accuracy outcomes: AdaDelta achieved 89.48% on MNIST and 92.51% on CIFAR-10. Moreover, using TiO_2 memristors reduces power consumption, as their non-volatile nature allows data retention without continuous power, crucial for edge AI applications.

6.3 PERCEIVED RISK

Potential risks include the variability in device performance due to manufacturing inconsistencies and the challenges associated with scaling the technology for mass production. Additionally, the reliance on specific optimization algorithms may limit flexibility, necessitating further research to develop more robust and adaptable solutions.

6.4 SOCIAL VALUE

The advancement of energy-efficient neuromorphic computing has significant societal implications. By reducing energy consumption, these technologies contribute to environmental sustainability. Moreover, their potential applications in areas like healthcare, transportation, and education can lead to improved services and quality of life.

6.5 MULTI-GROUP DIFFERENCES

The study does not explicitly address multi-group differences. However, the performance variations across different optimization algorithms suggest that specific configurations may be more suitable for particular applications or datasets, indicating the need for tailored approaches depending on the target group or use-case scenario.

6.6 IMPACT ON BUSINESS

The integration of TiO_2-based synaptic devices in neuromorphic computing can revolutionize industries by offering more efficient and scalable solutions. Businesses can leverage these advancements to develop innovative products, reduce operational costs, and gain a competitive edge in the market. The compatibility with existing technologies ensures a smoother transition and faster implementation.

CHAPTER 7

CHALLENGES AND FUTURE DIRECTIONS

7.1 ACCURACY

Accuracy is a fundamental metric in evaluating the performance of neuromorphic systems, especially when applied to classification tasks such as digit and image recognition. In the context of TiO_2-based neural networks, accuracy helps quantify how well the system predicts correct outputs compared to the true labels. This performance indicator becomes even more crucial when optimization algorithms are applied to improve learning on physical hardware.

Optimization Name and Dataset	AdaDelta	AdaGrad	Adam	AdaMax	Momentum	Nadam	SGD	RMSprop	Traditional Computer
Accuracy (%) CIFAR-10	92.51	82.08	83.10	81.76	91.25	82.45	90.21	88.11	96.95
Accuracy (%) MNIST	89.48	79.00	79.13	79.68	88.55	81.20	89.47	84.91	

Fig 7.1.: efficiency percentage among different algorithms

The study evaluated several optimization algorithms such as AdaDelta, Adam, Momentum, and SGD, applied to a TiO_2 synaptic neural network using MNIST and CIFAR-10 datasets. Among these, AdaDelta achieved the highest accuracy: 89.48% on MNIST and 92.51% on CIFAR-10, indicating its superior convergence properties for hardware-aware learning. Meanwhile, AdaGrad lagged with 79.00% accuracy, showing that not all optimizers generalize well on non-ideal hardware.

Accuracy not only validates the inference quality but also reflects the system's reliability in real-world applications. For neuromorphic systems, high accuracy on diverse datasets is a positive indicator of robustness, fault tolerance, and successful analog computation via TiO_2 memristors. Moreover, comparing accuracy over different epochs and optimizers reveals insights into learning dynamics, overfitting, and saturation points. Accuracy serves as a primary benchmark when tuning hyperparameters or switching between optimization strategies. A consistent accuracy metric across datasets ensures that TiO_2-based neuromorphic systems are viable for scalable edge computing solutions, especially in energy-constrained environments.

7.2 QUALITY

The quality of a TiO_2-based neuromorphic system is multi-dimensional, incorporating performance stability, consistency across datasets, noise robustness, and generalization ability. While accuracy quantitatively measures success, quality encompasses broader criteria such as convergence behavior, signal fidelity, and repeatability under varying conditions. A low error rate implies high-quality prediction capability. From the experiment, algorithms like AdaDelta and SGD yielded error rates of around 10.52% and 10.53%, respectively, reflecting better quality than AdaGrad with over 21% error rate. Another aspect of quality is the consistency of learning across epochs. For instance, AdaDelta showed minimal fluctuation in validation performance, demonstrating a stable learning curve. Conversely, optimizers like Adam displayed volatility in accuracy per epoch, implying less robustness in hardware environments. In memristor-based systems, hardware variability can degrade quality. The inherent non-linearity and cycle-to-cycle variation in TiO_2 devices can influence synaptic weight updates. Quality is maintained through architectural innovations such as differential pair circuits or redundancy-based learning.

.

7.3 LOSS FUNCTION

In neural networks, the **loss function** quantifies the difference between predicted output and true labels, guiding the optimization process. For classification tasks using TiO_2-based neuromorphic hardware, the most common loss function is the **categorical cross-entropy**, especially for datasets like MNIST and CIFAR-10. The cross-entropy loss is mathematically

$$\mathcal{L} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i)$$

defined as:

Where:

- C = number of classes (10 for both datasets),
- y_i = true label (1 for correct class, 0 otherwise),
- $\hat{y}_i$ = predicted probability for class i

This function penalizes incorrect classifications by amplifying the cost for high-confidence wrong predictions. It is particularly suited to softmax output layers, where the network produces a probability distribution over multiple classes.

In this study, the behavior of the loss function over training epochs was used to evaluate optimizer performance. AdaDelta demonstrated the most consistent and rapid decline in loss, indicating its strength in adapting learning rates dynamically. This performance was aligned with its high accuracy and short convergence time. Conversely, AdaGrad suffered from stagnating loss due to diminishing learning rates, leading to poor final accuracy.

7.4 DATASETS

Two datasets were primarily used in this study to benchmark the TiO_2-based neuromorphic network: MNIST and CIFAR-10. Each offers distinct challenges and highlights different aspects of the model's learning and generalization capabilities.

MNIST comprises 60,000 training and 10,000 testing grayscale images of handwritten digits (0–9), each of 28x28 pixels. It is ideal for evaluating the fundamental classification capability of neural systems due to its simplicity and structure. Most optimization algorithms performed well here, with AdaDelta achieving the highest accuracy of 89.48%. The reduced complexity of MNIST enables fast training, making it suitable for initial hardware evaluations.

CIFAR-10, on the other hand, includes 50,000 training and 10,000 test RGB images across 10 classes, including animals and vehicles. Each image is 32x32 pixels. This dataset tests the model's ability to handle color, more complex spatial patterns, and inter-class variance. Notably, AdaDelta again outperformed others with an accuracy of 92.51%, proving the optimizer's adaptability to visual complexity.

The performance metrics calculated over these datasets include:

- Accuracy
- Error Rate
- Sensitivity
- Specificity

These were influenced by the dataset structure. For instance, CIFAR-10's complexity demands more expressive power from the synaptic network, which TiO_2-based architectures could manage due to their analog weighting and high-density synapse emulation.

By leveraging two contrasting datasets, this study demonstrated the scalability and robustness of TiO_2-based systems. MNIST serves as a baseline, while CIFAR-10 validates real-world application feasibility. In future studies, larger datasets like ImageNet or face-specific ones like VGGFace could further test the generalization power of this hardware-oriented model.

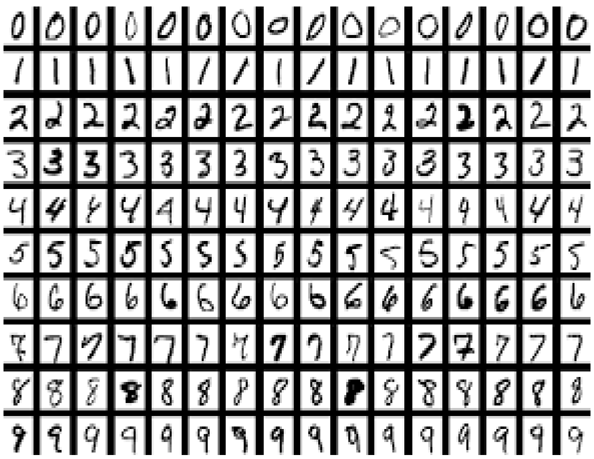


Fig 7.2 : (left) MNIST Dataset (Right) CIFAR-10 Dataset

7.5 EMOTIONAL CHALLENGES

While neuromorphic computing has advanced in mimicking cognitive functions like pattern recognition and memory, **emulating emotional intelligence** remains a major challenge. Emotions in biological systems involve complex, nonlinear processes influenced by both neural and biochemical feedback loops. Replicating such dynamics in TiO_2-based artificial synaptic devices is nontrivial due to several technical and conceptual barriers.

Firstly, **emotions are context-dependent and temporal**, often requiring long-term memory, sensory fusion, and decision-making under uncertainty. Simulating this in hardware would require multi-layered feedback and plasticity beyond the simple spike-timing-dependent plasticity (STDP) or Hebbian learning commonly used in TiO_2 memristors.

Secondly, **emotional states are not just data—they're modulators** of learning. For instance, emotions influence synaptic plasticity in the brain through neuromodulators like dopamine or serotonin. Translating this into hardware may require **adaptive conductance modulation** schemes or **chemical-electrical hybrid devices**, which are still in early research phases.

Moreover, **affective computing**—the field aiming to develop systems that can recognize and simulate emotions—demands integration of facial expression analysis, speech emotion recognition, and contextual understanding. While TiO_2-based neuromorphic cores could process such inputs, a **multi-modal architecture** is essential, which remains an ongoing research frontier.

On the ethical side, **emotional emulation raises societal concerns**. Should machines have synthetic emotions? How should such systems behave in emotionally sensitive scenarios like healthcare or education? These questions add another layer of complexity beyond hardware and algorithms.

In summary, while TiO_2-based neuromorphic devices show promise in cognitive tasks, achieving artificial emotional intelligence will require innovations in architecture, training paradigms, and bio-inspired circuitry. The emotional challenge is not only technical—it is philosophical, ethical, and multidisciplinary.

7.6 CODE AVAILABILITY

From the several models discussed only a few of these models have shared their code public. The implementation of models developed by others can be a challenging task due to unclear implementation details, and results may vary from those reported by the original authors. This poses a problem as it reduces the scope for reproducibility and makes it difficult for other researchers to compare their work with these models.

7.7 FUTURE DIRECTIONS

The success of TiO_2-based neuromorphic devices in digit and image recognition opens up several promising **future directions** for research, development, and deployment. These directions aim to bridge the remaining performance gaps, expand applications, and improve hardware-software synergy.

1. **Scalable and Hybrid Architectures**: Future designs may integrate TiO_2 memristors with CMOS for large-scale systems that mimic entire brain regions. Hierarchical architectures with local learning rules could enhance efficiency and fault tolerance.
2. **Three-Dimensional Integration**: Vertical stacking of memristive layers offers high-density synapse emulation, enabling compact systems with massive parallelism. This could support large neural models on chip for edge AI applications.

3. **In-Memory Computing**: Continued focus on compute-in-memory (CIM) architectures can mitigate the von Neumann bottleneck. TiO_2 memristors can serve both storage and processing roles, drastically improving latency and energy efficiency.
4. **On-Chip Learning and Lifelong Adaptation**: Implementing **online learning** capabilities will allow these devices to adapt in real-time, supporting edge scenarios like autonomous robotics or brain-machine interfaces. Incorporating **meta-learning** or **reinforcement learning** will further enhance adaptability.
5. **Robustness to Variability**: Future work must address variability in TiO_2 switching characteristics. This includes developing algorithms that are robust to noise, drift, and nonlinearity—either through training techniques or compensatory circuit design.
6. **Expanded Applications**: Beyond vision, future systems could target auditory processing, tactile sensing, and even **neuro-symbolic reasoning**, moving toward holistic cognitive computing.
7. **Standardization and Benchmarks**: A unified framework for benchmarking neuromorphic performance—similar to MLPerf for AI—is needed to fairly evaluate TiO_2-based systems and compare them across research efforts.
8. **Ethical and Societal Integration**: As these systems evolve, so must policies and frameworks for ethical use, particularly in sensitive domains like healthcare, defense, and education.

By pursuing these directions, TiO_2-based neuromorphic computing can evolve from laboratory prototypes to versatile, ethical, and intelligent hardware systems integrated into real-world environments.

CONCLUSION

In conclusion, the performance of the neural network based on TiO_2 synaptic devices was thoroughly evaluated using various optimization methods on the MNIST and CIFAR datasets. Different optimization algorithms, including SGD and its variants, were tested, achieving a 90% accuracy rate. The model demonstrated robustness and generalization ability across various optimization methods. This high performance under these optimization methods highlights its adaptability and effectiveness across datasets, indicating its potential for diverse applications. This evaluation confirms the effectiveness of combining TiO_2-based synaptic devices with various optimization algorithms for neuromorphic computing and hardware-based AI applications. These findings guide future studies, contributing to the creation of more energy-efficient and precise neural network models.

ACRONYMS

Acronym	Definition
ML	Machine Learning
DL	Deep Learning
TiO_2	Titanium Dioxide
SGD	Stochastic Gradient Descent
SNN	Spiking Neural Network
pMJT	perpendicular magnetic tunnel junction
CNN	Convolution Neural Netwotk
SRAM	Static Random Access Memory
MNIST	Modified National Institute of Standards and Technology
CIFAR-10	Canadian Institute For Advanced Research – 10 Classes
RMSprop	Root Mean Square Propagation
Adam	Adaptive Moment Estimation
AdaGrad	Adaptive Gradient Algorithm
Nadam	Nesterov-Accelerated Adaptive Moment Estimation
CMOS	Complementary Metal-Oxide-Semiconductor
RRAM	Resistive Random-Access Memory

REFERENCES

[1] S. Sun, Z. Cao, H. Zhu, and J. Zhao, ‘‘A survey of optimization methods from a machine learning perspective,’’ IEEE Trans. Cybern., vol. 50, no. 8, pp. 3668–3681, Aug. 2020, doi: 10.1109/TCYB.2019.2950779.

[2] S. Kim, H.-D. Kim, and S.-J. Choi, ‘‘Impact of synaptic device variations on classification accuracy in a binarized neural network,’’ Sci. Rep., vol. 9, no. 1, p. 15237, Oct. 2019, doi: 10.1038/s41598-019-51814-5.

[3] P. Wang, E. Fan, and P. Wang, ‘‘Comparative analysis of image classification algorithms based on traditional machine learning and deep learning,’’ Pattern Recognit. Lett., vol. 141, pp. 61–67, Jan. 2021, doi: 10.1016/j.patrec.2020.07.042.

[4] J. Zhu, T. Zhang, Y. Yang, and R. Huang, ‘‘A comprehensive review on emerging artificial neuromorphic devices,’’ Appl. Phys. Rev., vol. 7, no. 1, Mar. 2020, Art. no. 011312, doi: 10.1063/1.5118217.

[5] A. Biglari and W. Tang, ‘‘A review of embedded machine learning based on hardware, application, and sensing scheme,’’ Sensors, vol. 23, no. 4, p. 2131, Feb. 2023, doi: 10.3390/s23042131.

[6] G. W. Burr, R. M. Shelby, C. di Nolfo, J. W. Jang, R. S. Shenoy, P. Narayanan, K. Virwani, E. U. Giacometti, B. Kurdi, and H. Hwang, ‘‘Experimental demonstration and tolerancing of a large-scale neural network (165,000 synapses), using phase-change memory as the synaptic weight element,’’ in IEDM Tech. Dig., Dec. 2014, pp. 29, doi: 10.1109/IEDM.2014.7047135.

[7] Z. Cao, B. Sun, G. Zhou, S. Mao, S. Zhu, J. Zhang, C. Ke, Y. Zhao, and J. Shao, ‘‘Memristor-based neural networks: A bridge from device to artificial intelligence,’’ Nanosc. Horizons, vol. 8, no. 6, pp. 716–745, 2023, doi: 10.1039/d2nh00536k.

[8] [Y. LeCun, Y. Bengio, and G. Hinton, ‘‘Deep learning,’’ Nature, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539. T. Islam, A. Miron, X. Liu, and Y. Li, ‘‘Image-based virtual try-on: Fidelity and simplification,’’ Preprint, 2023.