

Processing-in-Memory (PIM) Based Defect Prediction of Metal Surfaces Using Spiking Neural Network

Mohammed Siyad B*and R Mohan**

Keywords: Defect prediction, Spiking Neural Network, Processing in Memory, Memristor crossbar array, Neuromorphic Computing.

ABSTRACT

Metal industries have long used the benefits of computer vision to automate various applications in product development, testing, and mistake repair. Most computer vision models for defect prediction use neural networks for accurate and exact defect classification. Reloading weights, saving, and retrieving activations are all impacted by high data processing layers, which results in restricted network bandwidth and latency. To solve this issue, a novel Processing-in-Memory (PIM) based defect prediction using Spiking Neural Network (SNN) has been proposed, in which the weight values in metal surface image processing are given to SNN that carries data only when a specific threshold is reached by LIF neurons, thereby decreasing the processing latency. The emerging non-volatile memory technologies like memristors have been shown to follow biological neurons and synapses which is irreplaceable to the processing of memory concepts in neuromorphic computing. To eliminate the sneak path problem, a novel approach has been implemented which utilizes two memristors and one transistor and reduces the on-chip memory overhead by changing the modes of transistors. Experimental results show an accuracy of 98% and an F-score of 96.5 which outperforms most of the computer vision methods taken for comparison.

Paper Received February, 2023. Revised May, 2023. Accepted June, 2023. Author for Correspondence: Mohammed Siyad B

* Dept. of Computer Science and Engineering,
National Institute of Technology, Tiruchirappalli
email: 406320001@nitt.edu

**Dept. of Computer Science and Engineering,
National Institute of Technology, Tiruchirappalli
email: rmohan@nitt.edu

INTRODUCTION

In traditional Von-Neumann architecture-based general-purpose computing systems, the memory system is detached from the processing unit. Hence a large quantity of data needs to be transferred to and from through the low bandwidth media between the processor and storage or communication units, especially in today's data-centric applications. This "Von-Neumann bottleneck" causes high energy consumption and performance degradation. Processing-in-memory (PIM) architecture, which takes a step away from Von-Neumann systems, is identified as a promising solution to the "memory wall" problem by scaling down the data transfer to a minimum. The idea behind this is that keeping processing elements closer to or within memory architecture minimizes the data travel among them and thus the memory access penalties.

Artificial neural networks (ANNs) are a popular machine learning tool with a wide range of applications in both academia and industries Eurolab4HPC (2020). The increasing size of chip technology and power consumption have limited the growth of large-scale computations, such as real-time image processing, due to the constraints of the traditional Von-Neumann architecture Li, S., et al. (2020). To solve this problem and enable in-memory computing, various technologies have been proposed, and the use of memristors as a fourth fundamental circuit element Chua, L., September (1971) has been established for several processing-in-memory (PIM) solutions, particularly programmable digital PIM systems. Memristors have the potential to address the memory wall problem by minimising the quantity of data transferred between the processor and memory. They can operate as a binary memory element by switching between high (ROFF) and low (RON) resistance values when a voltage is applied to the device. Furthermore, they can produce intermediate resistance values between ROFF and RON, enabling the storage of data in multi-level cells (MLCs) Eliahu, A., et al. (2020).

"Spiking Neural Network" (SNN) is a third-generation artificial neural network that is designed to

imitate the behavior of the human brain. The increased demand for SNN is due to its efficiency in addressing the issues of power consumption and hardware space in neuromorphic computing. SNN utilizes biologically inspired learning mechanisms like Spike Time-Dependent Plasticity (STDP) that provide more hardware support and improve the efficiency of neuromorphic algorithms. Unlike conventional neural networks, SNN only transmits information when the membrane potential of a neuron exceeds a fixed threshold value. When this happens, spikes are generated and sent to neighboring neurons, altering their action potentials. When the action potential exceeds the threshold value, the neuron sends an impulse and enters a “refractory period” before gradually returning to its original stage.

SNN simulations demand enormous endeavors in handling and processing the spatial and temporal information encoded in spike trains. These compute-intensive operations cause performance and efficiency bottlenecks in conventional Von-Neuman architecture. Recent studies have shown the capability of emerging non-volatile memory technologies like memristors to emulate biological neurons and synapses which is irreplaceable to the processing in memory (PIM) concept in neuromorphic computing. Because of its simple structure, compact device footprint, multilayer memristive states, high integration of computing and storage, low transcribe energy and standby power, and other advantages, the memristor is one of the most promising substitutes among in-memory computing ETP4HPC's SRA 4, (2020). Memristors have been studied and are actually real, which has prompted the creation of practical ways to build neuromorphic computing systems that can mimic neuro-biological structures and enable powerful deep neural networks and optimization algorithms Mutlu, O., et al. (2020).

Detecting and measuring faults in metal and metal-coated surfaces is a challenging task in “computer vision-based semantic segmentation” UJITSU LIMITED, (2020). Surface flaw detection is essential in quality control, especially when dealing with precise parts required to build complex machinery. Surface flaws can result from accidental scratches, manufacturing design errors, poor metal coating, surface cracks, and other reasons. Prompt identification and addressing of these flaws are crucial to maintaining a high level of quality in the entire system or machine in which the metal item is used FUJITSU LIMITED, (2020), Kwon, Y., et al., (2021). Surface flaws provide important information about the effectiveness of the metal coating process and the quality of materials used, and prompt detection is necessary to select appropriate solutions and provide feedback on the production process. Many efforts are being made in the industry to detect flaws and preserve product quality Noel, J.-P., et al. (2020), Khaddam-

Aljameh, R., et al. (2021), Khaddam-Aljameh, R., et al., (2021).

In contrast to typical CMOS technology, memristors may create brain-equivalent learning machines using memristive synapses as non-volatile memory Sebastian, A., et al. (2020). They make it possible to create a cross-point array architecture that is dense, constantly programmable, and somewhat accurate for applications that require large amounts of data Giordano, M., et al. (2021). The non-volatility, low power consumption, low parasitic capacitance, and changeable resistance states of the memristor, as well as its fast speed and versatility, make it ideal for ANN applications UPMEM, (2020). In addition to ANN, memristor-based computing systems have been suggested for a wide range of applications, including dictionary learning, compressive sensing, and sparse coding.

Additionally, it meets the requirement of the community for effective application performance by reducing the amount of pricey data transfers and delivering orders of magnitude productivity and energy savings Radojković, P., et al., (2020). Finally, several industrial prototypes and products demonstrate that the memristor has attained a high technological readiness level Abdelmagid, et al. (2020). Hence, a PIM concept must be incorporated with artificial intelligence techniques. The following are the primary contributions of this paper:

- The SNN-based Memristor cross array reduces the neural network reload weight error by imposing a rigorous latency threshold.
- To prevent sneak path error and bandwidth problems in the neural network, the MTM synapse technique is implemented by utilizing one transistor and two memristors which accurately anticipate metal flaws.
-

The remaining sections are arranged as follows: Survey of the literature is presented in Section 2. The novel solutions and architecture of the proposed model are described in section 3. Section 4 presents the implementation results and their comparison with the contemporary models. The work's concluding observations are included in Section 5.

LITERATURE SURVEY

As of now, surface defect detection by machine vision is one of the growing applications of neural networks in the industry. Surface defect identification is critical in quality control, especially when dealing with precision parts used to construct complicated machinery. We go over a few recent examples of these kind of research projects in this section.

Wang et al Wang, Z., et al. (2020) have proposed “Resistive Switching Materials” (RSMs) for information processing, which are based on various physical principles evolved for memories. These RSMs can enable in-memory computing with low power consumption and occupy minimal space. The four physical processes that cause resistive switching are redox reactions, phase transitions, spin-polarized tunneling, and ferroelectric polarization, which offer RSMs with representation capacity, switching speed, energy efficiency, dependability, and device density. While RSMs have shown greater benefits in metal detection compared to competing technologies, their effectiveness in computing applications should be improved in the future.

A set of tools termed NEUTRAMS has been suggested by Ji et al Ji, Y., et al. (2016, October) for accommodating several neural network (NN) types, such as SNNs and traditional ANNs. The toolset's objective is to decouple NN applications from the execution substrates that are used underneath them. NEUTRAMS has been tested on neuromorphic hardware as well as processor-in-memory architecture for ANNs. The toolset uses NN pruning approaches to improve performance, an improved representation layer, and NN examples to evaluate limitations and sensitivity to various inputs. The computing speed needs to be increased in the future.

Xian Tao et al Tao Xian, et al. (2018) proposed an architecture based on CNN for localizing and classifying defects appearing in metallic surface images. The “cascaded autoencoder” (CASAE) architecture proposes a two-level autoencoder (AE) network which transforms the input defect image into a pixel-wise prediction mask with only damaged pixels and background pixels through semantic segmentation. The defect portions are labeled into corresponding classes via a compact CNN. The major limitation of this method is the need for time- and money-consuming, manually labelled training data for deep networks.

Ling Chen et al Ling Chen, et al. (August 2014) proposed an unsupervised image-learning model that uses a two-layer memristor crossbar arrays combined with CMOS units. The MCA layer stores images while the ICA layer identifies major features, and the similarity factor is used for image recognition. They used controlled pulse and image overlay techniques for noise reduction and a time-slot approach to enhance image processing speed. A two-transistor structure was introduced to reduce the sneak-path error. However, implementing a CMOS unit for industrial use would be a challenge for time efficiency and device performance, and there may be more noise in industrial applications. Therefore, a more precise and estimable memristor model would be advantageous.

Abd et al Abd, et al. (2021) designed an “adaptive spike-to-rank coding” (ASRC), on CMOS memristors that simulate biological synapses with short and long-term plasticities (STP and LTP). The proposed ASRC adjusts synapses' weights to correct discrepancies. Additionally, Cadence design tools and XFAB 0.35 mm CMOS technology are used in the creation of ASRC. However, closed synaptic adaptation circuits without massive predominant digital components is required for better results in future work.

Liao et al Liao, et al. (2021) investigated methods for increasing the energy efficiency and reducing latency time of ANNs in edge computing systems. They have introduced a method of reducing the number of pulses in each stage of the weight update process and tested on a hardware simulator using memristors under various conditions and algorithms. However, memristor-based ANN paves the way for further developing edge computing in IoT systems.

A low-power IoT security module was developed by Rady et al Rady, et al. (2018). Memristors are used in the proposed module to generate AES keys, which mostly depends on the uniqueness of these devices as a result of modifications to the manufacturing process. The outstanding capabilities of the time-based cryptographic algorithms could be used by the proposed hardware security module to meet current technological requirements, such as secure device connection. For more solid security in future research, AES-192 or AES-256 can be employed.

Uddin et al Uddin, et al. (2019) provided a straightforward PUF-based security technique for tiny IoT devices. The goal is to protect the backup data while an embedded processor is in sleep mode or when a battery-less device lacks power. Memristors are being extensively investigated with the development of nanotechnology because of their non-volatility and small environmental impact, among other benefits. Memristors are used as non-volatile backup storage in the proposed security system. As required by this domain, the proposed system is relatively lightweight and offers enough security. More efficient and cutting-edge techniques are to be developed in this domain to reduce costs and boost efficiency.

U. Galan et al Galan, et al. (2018), created a metallic surface flaw detection technique implemented on the NVIDIA Jetson board to achieve the quickest computation time and energy efficiency through the concurrent operation of GPU cores. The connected components of the binary images of the intense and dim portions are processed in the algorithm to find the shadows that originated from the defects.

Weidong Zhao et al Zhao, et al. (2021) propose a reconfigurable network coupled with a multiscale feature fusion approach to address the trouble of microscopic and complicated steel flaws. A deformable convolution which reconstructs the feature extraction network is utilized to enhance the feature extraction capabilities. The multiscale feature graph output is fused to determine the deep semantic aspects of defect features using a feature pyramid network. However, future efforts must focus on improving image quality, detection time, and accuracy.

To sum up, the memristor technology, with its orders of magnitude performance and energy efficiency, can reshape computing, optimization, and AI research. Memristor crossbar applications need to be expanded in order to guarantee more accuracy, cost-effectiveness, security, performance, and power efficiency from the end-user perspective.

DEFECT PREDICTION OF METAL SURFACE WITH PROCESSING IN MEMORY (PIM) IN SPIKE NEURAL NETWORK

Defect detection throughout the production process is essential for assuring the quality of the product. To reduce operational costs and quality-related expenditures, it is critical to detect flaws or defects promptly and take appropriate action. Data processing in neural network layers increases the latency because of the limited bandwidth which weighed down the process with the additional task of continually reloading weights and saving and retrieving activations. To overcome this issue, an **SNN-based Memristor cross-array classifier** is proposed which utilizes a spiking neural network that generates neurons only at a specific threshold value and thus reduces the bandwidth limitation and latency. Moreover, to minimize the reload weight error, the memristor crossbar array is utilized which eliminates the off-chip link thereby increasing the speed of the training process also. The sneak path problem Fatih Gul, (2019) is another crucial challenge in a memristor crossbar array which may cause or prevent the activation or inhibition of a function at an unexpected time. Hence, a novel MTM synapse approach uses two memristors and one transistor in the form of 1M-1T-1M, which eliminates the sneak path and lowers on-chip memory overhead. Figure 1 depicts the proposed model's architecture.

Initially, the metal image is converted as spikes using delta modulation and then given to SNN, which converts real values from image spikes into Spatiotemporal values and forms a weight bias matrix. The vector matrix is multiplied through a linear matrix multiplication by using a memristor crossbar array

synapse. To avoid sneak path issues during the multiplication process in the memristor crossbar array, the proposed crossbar array uses a memristor synapse of 1M-1T-1M and ensures accuracy improvement in defect prediction.

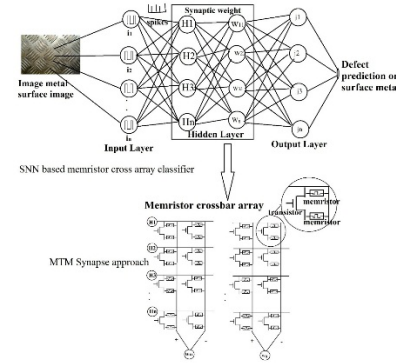


Fig. 1: Architecture of the proposed metal surface defect prediction model

SNN-based Memristor Cross Array classifier

The input image is changed into a spike train with a sequence length via delta modulation, where each pixel or feature is given a discrete value ($X\{i, j\} \in \{0, 1\}$) instead of a continuous value. The difference between each succeeding characteristic is calculated over all time steps and when the difference exceeds the threshold and is positive, a spike is created Jason K. Eshraghian et al., (September, 2021). Then, generated spikes are sent to a spiking neural network based on the “Leaky-Integrate-and-Fire” (LIF) model. The weighted total of inputs is taken into account by the LIF neuron model. The weighted sum of LIF is calculated by the following equation (1) Liu, et al. (2022).

$$y_j = f\left(\sum_{i=1}^M W_{ji} x_i\right) \quad (1)$$

In equation (1),

x_i is i^{th} input

y_j is j^{th} output

W_{ji} represents the weight between the i^{th} input unit and j^{th} output unit

M is several input units

$f(\cdot)$ is the activation function. Here LIF neuron is an activation function

A LIF-based 3-layer fully connected neural network is proposed in which each neuron integrates over many more incoming input spikes. It will emit a voltage spike if the integrated value is sufficient to excite the neuron and the membrane potential of the neuron becomes a threshold value when there is a defect present in the image. An RC circuit is formed as in figure 2 by the capacitive membrane and resistive ion channels. The output spike's size and shape are diminished by the LIF neuron. The output spike's profile details (size, shape, etc.) are processed as a separate event and not stored within it rather the timing

or frequency of the spike thereby, it reduces the on-chip memory overhead. The LIF neuron behavior can be derived as follows:

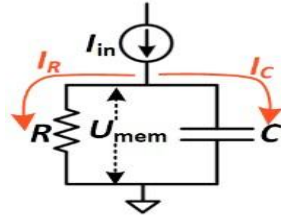


Fig. 2: RC Circuit equivalent of LIF neuron.

$$\text{From figure 2, the input current } I_{in}(t) = I_R + I_C$$

$$\Rightarrow I_{in}(t) = \frac{V_m(t)}{R} + C \frac{dV_m(t)}{dt} \quad (2)$$

$$\Rightarrow R I_{in}(t) = V_m(t) + RC \frac{dV_m(t)}{dt} \quad (3)$$

where $V_m(t)$ is the potential across the membrane, R is the resistance of the memristor and C is the capacitance.

Kirchhoff's law is employed to multiply the memristor crossbar array by the output weight bias matrix. The establishment of the memristor model enables the proposed defect prediction model to be developed more quickly and to reflect realistic behavior. The architecture of the proposed memristor cross array for the fully connected layer is shown in figure 3.

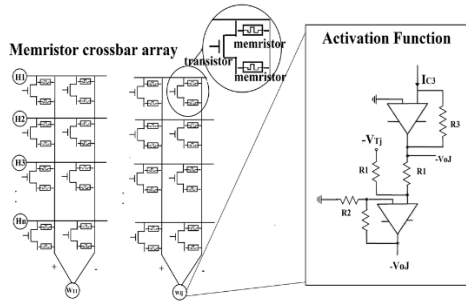


Fig. 3: Architecture of proposed memristor cross array

The computational complexity of the weighted summing operation is also reduced to $O(1)$ by using Kirchhoff's law Liu, et al. (2022).

At the classification layer of the spiking neural network, the proposed model utilized LIF neurons as the activation function. Figure 4 depicts the activation function in the crossbar configuration.

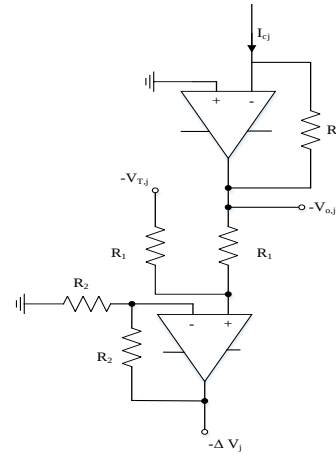


Fig. 4: Activation function in the crossbar

The proposed crossbar array Sigmoid function with LIF neuron is used as an activation function. The sigmoid function is mathematically expressed as in equation (4) Wei, et al. (2020),

$$f(x) = \frac{1}{1+e^{-x}} \quad (4)$$

Each input x has a corresponding weight, which is changed to model the plasticity of synapses. The LIF neuron unit transforms the integrated signal into the weighted sum of all the inputs in order to produce its output using an activation function $f(x)$. The activation function determines, following integration, how the input and output are related. The function takes neuron values and produces an output value range of 0 to 1. LIF neuron is activated for positive input and produces 1 as output; negative input produces 0 as output. Further, to predict defects in the metal surface, SNN is trained with a surrogate gradient descent algorithm. Training SNNs is difficult because of the discrete character of the number of spikes in a given interval. The derivatives of these discrete values are nearly 0 everywhere, hence the surrogate gradient (SG) approaches are necessary. The continuous-time structure of SNNs leads to incredibly sparse network activity because even a single spike's emission duration includes information. These spike timings are smooth, steady values that change in response to neuronal input. As a result, the continuous derivatives between the network's inputs and outputs were made possible by using spike timings along with SG. The following equation (5) describes non-leaky neurons in the network post-training Neftci, et al. (2019).

$$\frac{dU_i}{dt} = I_i = \sum_j W_{ij} \sum_r \theta(t - t_j^r) \exp(-(t - t_j^r)) \quad (5)$$

In equation (5),

$\theta(\cdot)$ is the Heaviside step function

t_j^r is a time of the r^{th} spike from neuron

SG descent training algorithm increased the neural network training process speed with the memristor crossbar array. However, the memristor's presence in the crossbars reduces the accuracy of prediction because of the sneak path. Hence, to avoid sneak paths, the cross MTM synapse approach is

utilized in the crossbar, which is explained in the upcoming section.

MTM synapse approach

One of the major challenges of the crossbar array structure other than the memristor-intrinsic problems like asymmetry and nonlinearity of conductance modulation is the sneak path problem. “Sneak paths” are unwanted paths alongside the selected path since the memristors in the crossbar are bidirectional [32]. The sneak path error in the crossbar array and corresponding circuit is shown in figure 5.

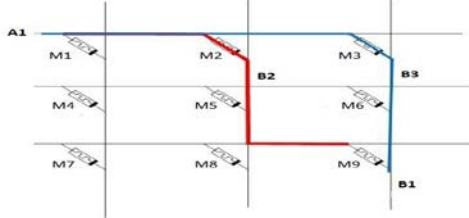


Fig. 5: Sneak path error in conventional crossbar array circuit

The desired flow of the current is through path A1-B3-B1, shown in blue color in figure 5. But the current flow through an undesired path A1-B2-B1 which is a sneak path, colored red in figure 5. To solve these, a two-memristor one-transistor (1M-1T-1M) synaptic device is proposed, where the series transistor restricts the sneak path access by intentionally manipulating the switching ON-OFF states. When the series transistor in the particular row changed to the ON state, the memristor switches the state due to the voltage drop across it. In the OFF state of the transistor, zero current flows through the cell, and no voltage drop across the memristor. The switching ON-OFF process of the transistor enables accurate resistance programming and reading in the crossbar array which leads to the precise classification of images.

In the proposed model, two memristors (M_1, M_2) are identical in every way and are connected to the emitter and collector terminal of the transistor which acts as a switch to control a weight modification and a synapse output. Conduction of the n and p channels of the transistor is controlled by a positive or negative value greater than the transistor's threshold which has a positive value or takes a negative value and is in a high-impedance state (no signal). According to the transistor properties, when receiving a signal either positive or negative, the n or p channel will shut down (allow no more current to pass) and the control threshold will be sent to input to change the weight of the synapse. In contrast, the channels of the transistor are conductive when there is no input signal (high impedance), and the synapse output is based on a synapse weight and input. Because its value is much lower than the transistor threshold, the input signal has

no impact on the transistor's n and p channels. The architecture of the proposed one memristor-one transistor-one memristor (1M-1T-1M) synapse in the crossbar array has been shown in figure 6.

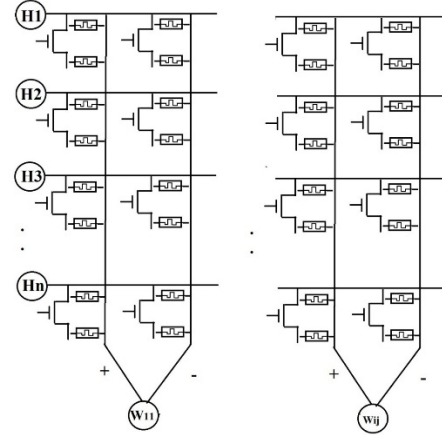


Fig. 6: Proposed one memristor one transistor one memristor (1M-1T-1M) Architecture

The total memristance of the 1M-1T-1M synapse at the initial state is given by equation (6)

$$M_{initial} = M_{1,initial} + M_{2,initial} \quad (6)$$

where

$$M_{1,initial} = r_{on}i_{10} + r_{off}(1 - i_{10}) \quad (7)$$

$$M_{2,initial} = r_{off}i_{20} + r_{on}(1 - i_{20}) \quad (8)$$

In equations (7) and (8), $M_{1,initial}$ and $M_{2,initial}$ are the memristance values at the initial state, r_{on} is the maximum resistance, r_{off} is the minimum resistance, and i_{10}, i_{20} are the state variables at the initial state. The memristors at any state have been given in equations (9) and (10) as follows:

$$M_1(\Delta T) = (r_{on} - r_{off})(i_{10} + \Delta i) + r_{off} \quad (9)$$

Similarly, the memristor M_2 is connected in parallel hence it performs the reverse operation of the memristor M_1

$$M_2(\Delta T) = (r_{off} - r_{on})(i_{20} - \Delta i) + r_{on} \quad (10)$$

Equation (11) provides the total memristance, $M(\Delta T)$, of the two memristors.

$$M(\Delta T) = M_{1,initial} + M_{2,initial} + k(r_{on} - r_{off}) \times \int \Delta T (f(+x)(i_{10} + \Delta i) - f(-x)(i_{20} - \Delta i)) \times x(\Delta T) dT \quad (11)$$

where $f(+x), f(-x)$ denote the activation function's value when the memristor is conducting forwardly or backward, respectively. When the term $(f(+x)(i_{10} + \Delta i) - f(-x)(i_{20} - \Delta i))$ is 0, it can be seen that the total memristance of the synapse will equal the sum of the two initial memristances and this term lies between 0 to 1. As a result, when a constant-voltage source is applied to the synapse, both the total resistance of the synapse and its current remain constant. Hence it is noted that when the input voltage is positive, the output voltages of the synapses, which represent the weight values, will rise linearly. The weights of the synapses in a circuit of synapses won't

change in the absence of input voltage. In contrast, the synaptic weights will linearly drop when a negative input voltage is provided to the synaptic circuit.

Furthermore, this crossbar array circuit utilizes memristors monolithically integrated into three dimensions on the transistor to obtain unit cells with the same size as 1T1M devices. Additionally, there is no longer a requirement for extra read-before-write operations because linear conductance modulation has been accomplished utilising the same programming pulse scheme as incremental amplitude or width modulation. Based on these computations, defects in the metal surface are predicted with the SNN-based Memristor cross-array classifier model and MTM synapse approach. There the classification of defects is done precisely without any errors. The following section covered the findings.

RESULTS AND DISCUSSION

This section provides a thorough explanation of the implementation outcomes and a performance assessment of the proposed model.

Experimental Setup

This proposed model has been implemented in the working platform of MATLAB. Initially, dataset RGB 3D images are converted to a grayscale image. From the converted grayscale image neural network is trained with the proposed model where neurons are generated for the trained model. Then the input metal image is given to the trained model to predict the defect. Initially, the network assigns neurons for the input image based on the dimensions of the image. Then each neuron of the network predicts the defect and forms a confusion matrix. Initially, the values of TP, TN, FP, and FN are zero, if the actual value and the predicted values are 1 then it increases the TP value by one. If the predicted value is 1 and the actual value is 0 then it increases the FP value by one. If the actual value and predicted value are 0 then the FN value rises. It raises the TN value by one if the predicted value is 0 and the actual value is 1. From the final values in the confusion matrix the proposed model's accuracy, recall, f1-score, and precision are calculated.

Dataset Description

In this work, the detection technique can efficiently identify minor target defects on the metal surfaces, which may be a reference for automatic metal defect identification. For this prediction, we used insulator Dataset- "Chinese power line insulator dataset" (CPLID) António Raimundo, (February 11, 2020) which contains UAV-captured normal insulator images and synthetic flawed insulator images. A training set and a test set were created from the CPLID dataset for the proposed model, with 80% of the dataset going to the training set and 20% to the test set.

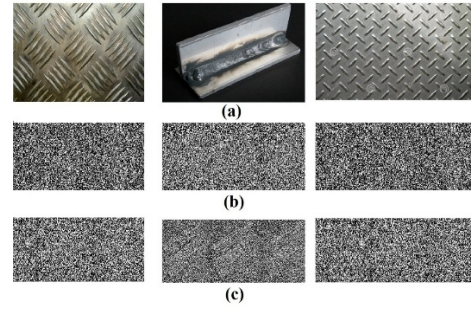


Fig. 7: (a) Input images (b) Output images with noise (c) Output images without noise

The above figure 7a & 7b shows some sample input images and the corresponding output images. The noise-canceled output images are shown in figure 7c.

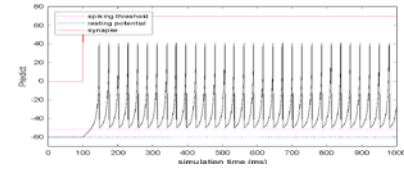


Fig. 8: Proposed method of defect prediction

Figure 8 shows the prediction of defects for a particular simulation time by setting a constant threshold value of -50 and a resting potential value of -60. When simulation time increases from 0 ms, the model prediction has risen above the threshold value. The simulation time from 100ms to 1000ms 40 image prediction has been done. The proposed method achieved this high prediction value by setting a constant threshold value.

Performance Analysis of the proposed model

The performance of the proposed model is analyzed by calculating the following measures.

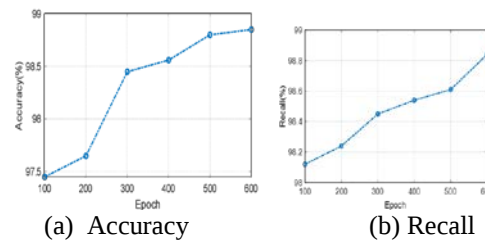
$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (12)$$

where, TP, TN, FP and FN respectively stands for "True Positive", "True Negative", "False Positive", and "False Negative" Values.

$$Recall = \frac{TP}{TP+FN} \quad (13)$$

$$F\text{-score} = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (14)$$

$$\text{where Precision} = \frac{TP}{TP+FP}$$



(a) Accuracy

(b) Recall

Model	Recall(%)	Precision (%)
SSD Lv,et al. (2020)	57	72
Faster-RCNN Wang, et al. (2021)	40	71
YOLO-V2 Lv,et al. (2020)	43	50
YOLO-V3 Lv,et al. (2020)	43	45
EDNN Lv,et al. (2020)	85	72
Proposed	95	85

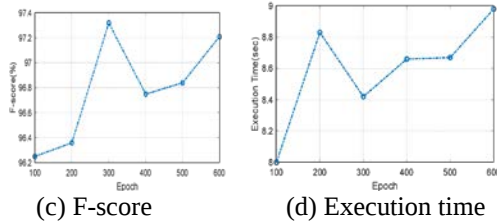


Fig. 9: Performance measures of the proposed Model

Figures 9a–d show how the suggested model performs in terms of accuracy, recall, f-score, and execution time across various epochs. An epoch refers to the number of times the learning algorithm processes the full training dataset, and it demonstrates how the model parameters are adjusted for each sample in the training dataset. In the proposed method, the recall value is improved by transmitting data with a threshold value to avoid the propagation cycle in each iteration. The proposed method also uses a more flexible weight range and has an increased f-score by avoiding the use of one transistor with two memristors. The MTM synapse approach is utilized to reduce sneak path error, resulting in faster execution. The following table 1 displays the overall performance metrics across 600 epochs.

Table 1: Overall performance measures of the proposed model

Epochs	Accuracy (%)	Recall (%)	F-score(%)	Execution Time(sec)
100	97.5	98.13	96.25	8
200	97.7	98.22	96.36	8.83
300	98.48	98.42	97.32	8.42
400	98.6	98.52	96.7	8.66
500	98.8	98.61	96.84	8.67
600	98.82	98.82	97.21	8.98

4.4 Comparative analysis of the proposed model

The efficiency of the proposed model in identifying diverse metal surface flaws was demonstrated by comparing its performance to widely used defect classification and detection models. The comparison with the contrastive models is listed in tables 2 and 3.

Table 2: Accuracy and Time comparison

Model	Accuracy(%)	Execution Time(sec)
Improved ResNet50	97.6	7
Generative Adversarial Networks	97.5	210

Faster R-CNN	97.2	220
Classification and object detection	98.3	60
Proposed	98.82	6

Table 3: Recall and Precision comparison

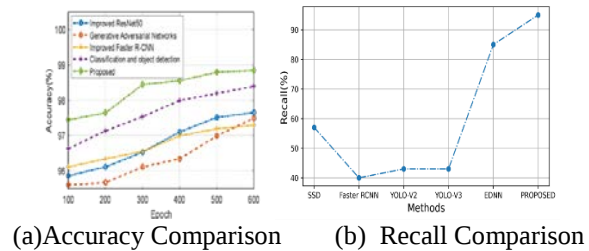


Fig. 10: Comparison of Accuracy and Recall

The accuracy comparison of the proposed model with existing models is presented in Figure 10(a). The proposed model outperforms all other models in terms of accuracy, with a maximum accuracy of 98.82%, while the highest accuracy achieved by improved ResNet50, Generative Adversarial Networks, faster R-CNN, and classification and object detection is 97.6%, 97.5%, 97.2%, and 98.3% respectively. Figure 10(b) compares the proposed model's recall to those of existing models, and the model demonstrates its superior recall ability, with a recall percentage of 95%, compared to popular models listed in table 3.

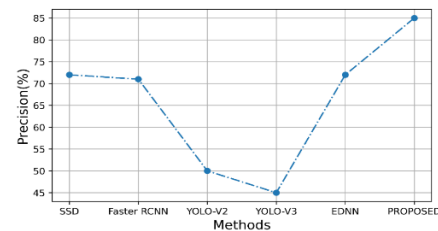


Fig. 11: Comparison of Precision

In Figure 11, a comparison of the precision of the proposed model is presented. The model's precision is 85% when compared with popular models given in table 3. This high precision value highlights the proposed model's potential in accurately identifying defects.

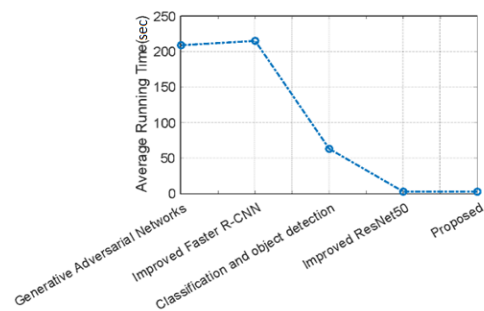


Fig.12: Comparison of the Running Time

In Figure 12, the running time comparison between the proposed model and other existing methods listed in table 2 is displayed. The proposed model has the shortest running time among all techniques as demonstrated in the figure. This makes the proposed model more effective and efficient in detecting defects on metal surfaces.

Overall, the proposed SNN-based memristor crossbar array uses an activation function to process metal images and identify defects. The neural network is trained with features extracted from the metal image. When the network encounters a defective area, the LIF activation function triggers a spiking neuron in the spiking neural network, allowing it to accurately classify the defective area of the input metal image.

CONCLUSION

In summary, we have demonstrated an efficient metal surface defect recognition model using SNN via a memristive crossbar array aiming to reduce latency, resolve bandwidth concerns, and get rid of sneak path issues. The SNN-based memristor cross-array classifier utilizes a spiking neural network that generates neurons only at a specific threshold value and thus overcomes the bandwidth limitation and latency. Emerging non-volatile memristors, which are considered invaluable to the processing in memory (PIM) technology in neuromorphic computing, are used to emulate biological neurons and synapses.

To eliminate the “sneak path” problem in memristor crossbar arrays, a novel MTM synapse approach has been introduced that uses two memristors and one transistor in the form of 1M-1T-1M and thereby promising accuracy. The results of experiments prove that the SNN-based model proposed performs significantly better than the state-of-the-art models currently available. The success of the proposed model highlights the practicality and usefulness of the memristor-based PIM concept in real-world scenarios. Moreover, the advantages provided by the neuromorphic chips make SNNs a reasonable choice for different viable applications.

REFERENCES

- Abd, Hamam, and Andreas König, "Adaptive Spiking Sensor System Based on CMOS Memristors Emulating Long and Short-Term Plasticity of Biological Synapses for Industry 4.0 Applications," *tm-Technisches Messen*, Vol. 88, No. s1, pp. s114-s119, (2021).
- Abdelmagid, Yasmin, K., et al. "Investigation of DW Spintronic Memristor performance in 2T1M Neuromorphic Synapse." *2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES). IEEE*, (2020).
- António Raimundo, "Insulator Data Set - Chinese Power Line Insulator Dataset (CPLID)", *IEEE Dataport*, doi: <https://dx.doi.org/10.21227/qtxb-2s61>, (February 11, 2020),
- Chua, L., "Memristor-The missing circuit element," in *IEEE Transactions on Circuit Theory*, Vol. 18, No. 5, pp. 507-519, (September 1971), doi: 10.1109/TCT.1971.1083337
- Eliahu, A., Ben-Hur, R., Ronen, R and Kvatinisky, S., "abstractPIM: Bridging the Gap Between Processing-In-Memory Technology and Instruction Set Architecture," *2020 IFIP/IEEE 28th International Conference on Very Large-Scale Integration (VLSI-SOC)*, pp. 28-33, (2020).
- ETP4HPC's SRA 4, "Strategic Research Agenda for High-performance Computing in Europe," *White Paper*, (2020).
- Eurolab4HPC Long-Term Vision on High-Performance Computing (2nd Edition), (2020).
- Fatih Gul, "Addressing the sneak-path problem in crossbar RRAM devices using memristor-based one Schottky diode-one resistor array", *Results in Physics*, Vol. 12, pp. 1091-1096, (2019).
- FUJITSU LIMITED, "FUJITSU Supercomputer PRIMEHPC FX1000," *White Paper*, (2020).
- Galan, Ulises, et al., "Surface defect identification and measurement for metal castings by vision system," *Manufacturing Letters*, Vol. 15, pp. 5-8, (2018).
- Giordano, M., Prabhu, K., Koul, K., Radway, R. M., Gural, A., Doshi, R., Khan, Z. F., Kustin, J. W., Liu, T., Lopes, G. B., Turbiner, V., Khwa, W.-S., Chih, Y.-D., Chang, M.-F., Lallement, G., Murmann, B., Mitra, S and Raina, P., "CHIMERA: A 0.92 TOPS, 2.2 TOPS/W Edge AI Accelerator with 2 MByte On-Chip Foundry Resistive RAM for Efficient Training and Inference," *Symposium on VLSI Circuits (VLSI)*, (2021).
- Jason K. Eshraghian et al., "Training Spiking Neural Networks Using Lessons from Deep Learning", *arXiv preprint arXiv:2109.12894*, (September 2021).
- Ji, Y., Zhang, Y., Li, S., Chi, P., Jiang, C., Qu, P., Xie, Y. and Chen, W., "NEUTRAMS: Neural network transformation and co-design under neuromorphic hardware constraints," *In 2016 49th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO), IEEE*, pp. 1-13, (2016, October).
- Khaddam-Aljameh, R., et al., "HERMES Core – A 14nm CMOS and PCM-based In-Memory Compute Core using an array of 300ps/LSB Linearized CCO-based ADCs and local

- digital processing," in *Proc. Symposium on VLSI Circuits*, (2021).
- Khaddam-Aljameh, R., Francese, P.-A., Benini, L and Eleftheriou, E., "An SRAM-Based Multibit In-Memory MatrixVector Multiplier with a Precision that Scales Linearly in Area, Time, and Power," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 29, (2021).
- Kwon, Y., et al., "A 20nm 6GB Function-In-Memory DRAM, based on HBM2 with a 1.2TFLOPS Programmable Computing Unit Using Bank-Level Parallelism, for Machine Learning Applications," in *Proceedings of IEEE International Solid-State Circuits Conference (ISSCC)*, (2021).
- Li, S., Yang, Z., Reddy, D., Srivastava, A and Jacob, B., "DRAMsim3: A Cycle-Accurate, Thermal-Capable DRAM Simulator," *IEEE Computer Architecture Letters*, Vol. 19, (2020).
- Liao, Zhiheng, Jingyan Fu, and Jinhui Wang, "Ameliorate Performance of Memristor Based ANNs in Edge Computing," *IEEE Transactions on Computers*, (2021).
- Ling Chen, Chuandong Li, Tingwen Huang, Yiran Chen, and Xin Wang, "Memristor crossbar-based unsupervised image learning", *Neural Comput. Appl.*, Vol. 25, No. 2, pp. 393–400, (August 2014).
- Liu, Xiaoyang, and Zhigang Zeng, "Memristor crossbar architectures for implementing deep neural networks," *Complex Intell. Syst.*, Vol. 8, pp. 787–802, (2022).
- Lv, Xiaoming, et al., "Deep metallic surface defect detection: The new benchmark and detection network," *Sensors*, Vol. 20, No. 6, pp. 1562, (2020).
- Mutlu, O., Ghose, S., Gómez-Luna, J and Ausavarungnirun, R., "A Modern Primer on Processing in Memory," in *arXiv*, (2020).
- Neftci, Emre O., Hesham Mostafa, and Friedemann Zenke, "Surrogate gradient learning in spiking neural networks: Bringing the power of gradient-based optimization to spiking neural networks," *IEEE Signal Processing Magazine*, Vol. 36, No. 6, pp. 51-63, (2019).
- Noel, J.-P., Pezzin, M., Gauchi, R., Christmann, J.-F., Kooli, M., Charles, H.-P., Ciampolini, L., Diallo, M., Lepin, F., Blampey, B., Vivet, P., Mitra, S and Giraud, B., "A 35.6 TOPS/W/mm² 3-Stage Pipelined Computational SRAM with Adjustable Form Factor for Highly Data-Centric Applications," *IEEE Solid-State Circuits Letters*, Vol. 3, (2020).
- Radojković, P., et al., "Towards Resilient EU HPC Systems: A Blueprint," *European HPC resilience initiative*, (2020).
- Rady, Hanan, et al., "Memristor-Based AES Key Generation for Low Power IoT Hardware Security Modules." *2019 IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS)*. IEEE, (2019).
- Rasika Joshi & John M Acken, "Sneak Path Characterization in Memristor Crossbar Circuits", *International Journal of Electronics*, Vol. 108, No. 8, pp. 1255-1272, (2021).
- Sebastian, A., Gallo, M. L., Khaddam-Aljameh, R and Eleftheriou, E., "Memory devices and applications for in-memory computing," *Nature Nanotechnology*, No. 7, pp. 529–544, (2020).
- Shulz, D and Feldman, D., "Spike timing-dependent plasticity", *Neural Circuit Development and Function in the Brain*, pp. 155-181, (2013).
- Tao Xian, Dapeng Zhang, Wenzhi Ma, Xilong Liu, and De Xu, "Automatic Metallic Surface Defect Detection and Recognition with Convolutional Neural Networks," *Applied Sciences*, Vol. 8, No. 9, pp. 1575, (2018).
- Uddin, Mesbah, et al., "Memristor crossbar PUF based lightweight hardware security for IoT," *2019 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, (2019).
- UJITSU LIMITED, "FUJITSU Supercomputer PRIMEHPC Specifications," *White Paper*, (2020).
- UPMEM, "UPMEM PIM Security Benefits - Architecture and Features Overview," *White Paper*, (2020).
- Wang, Shuai, et al., "Automatic detection and classification of steel surface defect using deep convolutional neural networks." *Metals*, Vol. 11, No. 3, pp. 388, (2021).
- Wang, Z., Wu, H., Burr, G.W., Hwang, C.S., Wang, K.L., Xia, Q. and Yang, J.J., "Resistive switching materials for information processing," *Nature Reviews Materials*, Vol. 5, No. 3, pp.173-195, 2020.
- Wei, Linyu, et al., "P-SFA: Probability based sigmoid function approximation for low-complexity hardware implementation," *Microprocessors and Microsystems*, Vol. 76, pp. 103105, (2020).
- Zhao, Weidong, et al., "A new steel defect detection algorithm based on deep learning." *Computational Intelligence and Neuroscience*, (2021).