

Digital Implementation of a Spiking Convolutional Neural Network for Tumor Detection

Ayoub Adineh-vand¹, Gholamreza Karimi¹ and Mozafar Khazaei²

¹Razi University, Faculty of Engineering, Electrical Engineering Department, Kermanshah, Iran

²Kermanshah University of Medical Sciences, Reproduction Research Center, Kermanshah, Iran

Abstract: The structural variation of the brain tissue creates challenges for detection of tumors in MRI images. In this paper, an architecture for spiking convolutional neural networks (SCNNs) is implemented in an embedded system and their potential is evaluated in terms of hardware utilization and power consumption in complex applications such as tumor detection. Accordingly, the structure of the proposed SCNN is implemented on a field-programmable gate array (FPGA) using fixed point arithmetic. To evaluate the speed, accuracy and flexibility of the proposed SCNN, Izhikevich neuron model is used with the spike-timing-dependent plasticity (STDP) learning rule. The suggested neural network is explored for digital implementation possibility and costs. Results of the hardware synthesis and digital implementation are presented on an FPGA.

Keywords: Brain tissue; MRI images; Spiking Neural Network; Digital Implementation; STDP

Digitalna implementacija sunkovnih nevronskih mrež za detekcijo tumorjev

Izvleček: Strukturne razlike v možganskem tkivu predstavljajo iziv pri detekciji tumorjev v MRI slikah. Članek opisuje arhitekturo implementacije sunkovnih konvolucijskih nevronskih mrež v vgradnih sistemih. Njihov potencial je ocenjen na osnovi strojne uporabnosti in porabe pri detekciji tumorjev. Struktura je implementirana v FPGA okolju. Ocena hitrosti, natančnosti in fleksibilnosti je opravljena z Izhikevichevim nevronskim modelom.

Ključne besede: možgansko tkivo; MRI slike; sunkovne nevronske mreže; digitalizacija; STDP

*Corresponding Author's e-mail: ghkarimi@razi.ac.ir

1 Introduction

Most clinical reports struggle with the problem of large volume of data about patient records in the form of medical imaging such as MRI images and CT scans. Analyzing and efficient processing of these huge data opens up a research avenue motivating researchers to explore possible solutions and help physicians to have a better diagnosis particularly in case of emergencies when no expert is available. Convolutional neural networks (CNNs) as a type of deep neural network (DNN) present a premium performance in machine learning fields including pattern recognition, speech and image processing, and natural language processing.

Simulation and implementation of brain-like networks are vital for perceiving the way the brain processes information. The third generation of neural network models, called spiking neural networks (SNNs), improved the level of biological realism in neural simulations. SNNs have provided many opportunities for opening up a totally new field in artificial intelligence research. Currently, spiking neural networks (SNNs) have gained popularity because of their biological plausibility. Practically, when a neuron model is selected for large SNN, there is always an exchange between the biological plausibility and computational efficiency [6].

Alan Lloyd Hodgkin and Andrew Huxley suggested the first scientific model of spike neurons named Hodgkin-Huxley (HH) model in 1952 [7]. This model presented the procedure of spike generation with a set of four differential equations by describing how action potentials take place and reproduce. Considering accuracy and computational complexity, various biological models such as Izhikevich model [8], [9], Integrate and Fire model [8], FitzHugh-Nagumo (FHN) model [10], [11], and Hindmarsh-Rose model [12] are available. Effective tools for analysis of primary procedures in the brain are provided by spiking models and solutions are suggested for a wide range of special problems in engineering, including fast signal processing and pattern and speech recognition [13]. The procedure of data processing in the brain can be simulated outside the brain through analog or digital circuits if the effective model and detailed condition of neuron connections are selected. Through targeting various platforms, hardware realization of biological neuron models has been examined. VLSI systems are notable options for the neural systems' direct implementations. Rapid prototyping of neural algorithms to realize theories of computational neuroscience, network architecture, and learning system is made by a VLSI implementation as it enjoys high performance and remarkably improved technology [14]. Digitally implemented neurobiological networks possess shorter development times and are more flexible while they consume more silicon area and power compared with their analog counterparts.

Nowadays, breakthroughs in circuits and systems such as application specific integrated circuits (ASIC), graphical processing units (GPU), and custom hardware accelerators have been proposed as methods for implementing CNNs in practical applications [1-2]. A high accuracy digital implementation makes it possible to develop networks with high dynamic range and stability. Recently, in order to realize neural system models, reconfigurable digital platforms have been utilized [15]-[22]. Critical challenges of the digital implementation include Through FPGA it is possible to achieve lower power consumption [3, 4, 5].

In this paper, a spiking convolutional network has been proposed based on Izhikevich neuron. By using STDP as a learning rule, the network was trained to achieve the higher accuracy in tumor detection. Furthermore, an architecture is presented for Izhikevich neuron. Accordingly, synthesis and physical implementation have been done on the FPGA board.

2 The proposed neural network

Considering the biological plausibility and power efficiency of neuromorphic platforms, developing deep

SNNs for these platforms is inevitable. These types of neural networks are not precise and are not considered as deep learning methods. On the other side, SNNs are efficient networks for simulating the brain performances to solve complicated problems in the field of intelligent objects. The proposed architecture has been presented as the simplest deep structure which is fully connected and consists of input, hidden, and output layer. Fig. 1 shows the overall structure of the deep SNN with its layers. The structure of the suggested deep spiking neural network offers a neuronal population with hidden layers which is capable of being employed in the medical images. The input layer learns to perform pre-processing on the input. Information is then sent to a series of hidden layers. These layers can vary in number. As the information disseminates through hidden layers, more complex features are extracted and learned. The output layer performs classifications and detects the tissues of the input images, usually by Soft-max. The proposed SNN network contains Izhikevich neurons. In this structure, the data flows in a completely one-way flow from the input to the output units. Data processing can be performed at several layers of neurons, but there is no feedback in this structure. For bridging biologically plausible learning algorithms and traditional learning methods in neural networks, deep spiking neural networks can be an ideal choice. An important restricting parameter is lack of training algorithms that have specific uses in the capability of spiking neuron models. Most methods use rate-based approximations of conventional DNNs. Deep SNNs might still be suitable because approximating the results could be achieved more efficiently and faster than traditional systems, especially if the SNN is implemented on a neuromorphic hardware platform. Furthermore, designing and analyzing the training algorithms for SNNs and their employment are more difficult because they are discontinuous and asynchronous methods for computing [23]. In the last decades, a new learning approach has been emerged in cellular learning according to which temporal order has been focused instead of frequency. This novel learning rule has been known as spike-timing dependent plasticity (STDP). STDP process presents the activity-dependent development of neural systems by considering long-term potentiation and long-term depression. Also, it has obtained great popularity because of having the mixture of computational power and biological plausibility [24]. To train network weights, an STDP algorithm has been applied along with a gradient algorithm that is a kind of reinforcement training. This network is used to achieve a better classification in detecting benign and malignant tumors. The output results of the SNN are used for categorizing two types of images which are recognized at frequencies 11 Hz and 80 Hz for images with and without tumors, respectively.

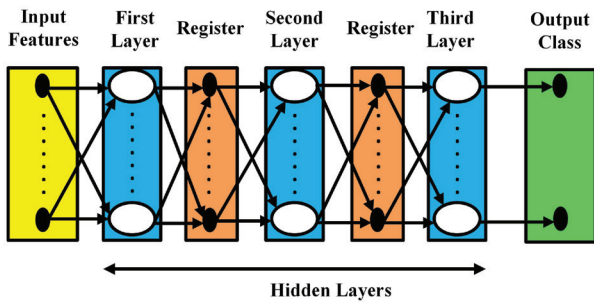


Figure 1: The suggested deep spiking neural network

By using Izhikevich neurons as biologically plausible units, the equations of the neuron are as follows [8], [9]:

$$\frac{dv}{dt} = (0.04 \times v^2) + (5 \times v) + 140 - u + I \quad (1)$$

$$\frac{du}{dt} = a \times ((b \times v) - u) \quad (2)$$

$$\text{if } v \geq v_{th} : \quad \begin{cases} v \leftarrow c \\ u \leftarrow u + d \end{cases} \quad (3)$$

Also STDP algorithm is presented by [25]:

$$\left\{ \begin{array}{ll} W(x) = A_+ \exp\left(-\frac{x}{\tau_+}\right) & \text{for } x > 0 \\ W(x) = -A_- \exp\left(-\frac{x}{\tau_-}\right) & \text{for } x < 0 \end{array} \right\} \quad (4)$$

where, A_+ and A_- are the domains of weight changes, τ_+ and τ_- are 10 ms, and W is the synaptic weight. To evaluate the speed, accuracy, and flexibility of the proposed spiking neural network, it is implemented on FPGA.

3 Hardware design

Countless analog and digital brain-inspired electronic systems have been put forward as solutions for brisk simulations of spiking neural networks. While these architectures are proper for realizing the computational features of large scale models of the nervous system, the challenge of constructing physical devices that are able to operate intelligently in the real world and display cognitive competence is still kept open. Designing and efficient implementation of these structures in hardware provide us with the benefit of presenting a processing system based on the structure of brains. Analog circuits require precision in terms of the fabrication procedure variations and environment temper-

atures. In fact, designing circuits that perform reliably under a vast range of extraneous factors is a challenging endeavor. As a result of this challenge, there is a dissonant condition between simulation results and the analog implementation. Furthermore, the reconfiguration of a very large scale integrated (VLSI) implementation is not achieved easily. Consequently, having a rapid prototyping platform for neural models with homogeneous flexibility in general purpose microprocessors seems essential. An FPGA is an ideal technology for this purpose. It is true that digital computation consumes more silicon area and power per operation than its analog counterpart; however, it affords extra merits. Having fascinating features such as low-cost, flexibility, reliability, and digital precision makes FPGAs popular as a promising choice over analog VLSI approaches for designing neuromorphic systems. A digital implementation of the spiking neural network is considered for its fast, high precision, and flexible storage structure. On the other hand, usually both analog and digital implementations are used. Analog implementations have more restrictions than digital ones. Using a reconfigurable and programmable device like FPGA can be an ideal option. The smart and small systems used in modern day-to-day applications and the possibility of their connection to the computer, require the implementation of neural network hardware in small volumes. In this structure, a large number of neurons are packed in order to implement the network at a huge scale. Based on the Euler recursive method, differential equations are solved for a neuron model.

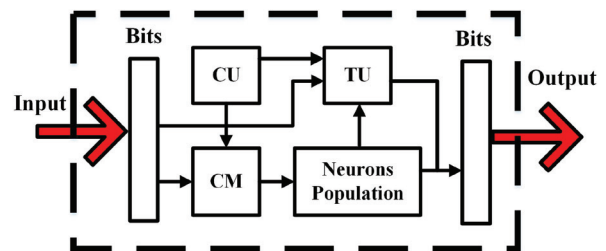


Figure 2: General structure of the proposed hardware

Fig. 2 presents an overview of the proposed hardware which consists of the training unit (TU), the coefficient matrix (CM), the control unit (CU), and the neural population. The TU deals with the process of training neurons based on their weights. The CM contains values of weights, parameters of neurons, and other network variables. CU produces the necessary control signals for the training of the proposed network and also controls the necessary conditions for passing the computational units. The process of training neuron weights is done in the training unit. Neuronal population consists of biochemical neurons that are used in all three layers of the network. This unit is employed for evaluation of the proposed CNN on an FPGA for tumor detection

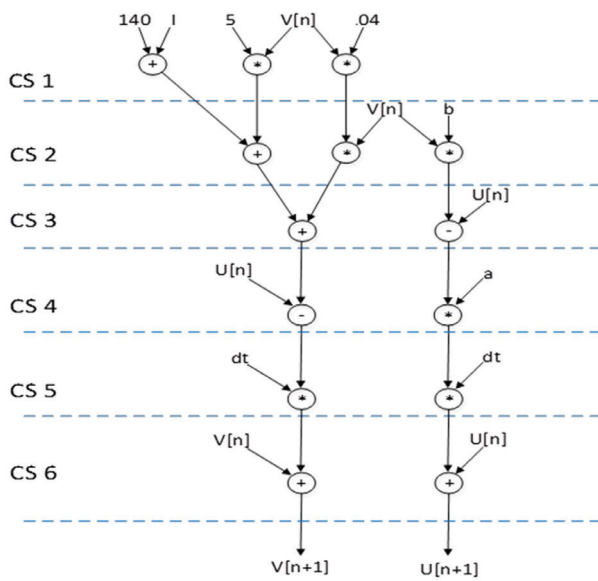


Figure 3: Scheduling diagram for the Izhikevich neuron

as a case study which will be discussed presently. As shown in Fig. 3, each Izhikevich neuron can be implemented at six different states by the presented scheduling diagram for describing in Verilog HDL as a hardware neuron unit. Also, Fig. 4 demonstrates the general structure of Izhikevich neuron by logical units. Accordingly, a hardware architecture is proposed based on combinational circuits such as Multiplexers, Multipliers, and arithmetic logic units (ALU). Each part of this architecture can perform on the base of the discretized the Izhikevich neuron in Euler method. Also, by using

functional units, the spiking neuron model can be described in HDLcode.

Figs. 5 and 6 represent the frequency movement for detecting cancer tumors at frequency 11 Hz and for non-cancerous tumors at frequency 80 Hz, respectively. Accordingly, the proposed hardware system is designed for recognizing the cancerous and non-cancerous tumors at two different frequencies. Fig. 7 shows how to change weights for 8 different neurons.

4 Brain tumors in MR images and the proposed convolutional neural network

A spiking convolutional neural networks (SCNN) was used for tumor detection as a signal processing application in magnetic resonance imaging (MRI). Based on Fig. 8, which illustrates the basic architecture, the images as inputs are transformed to spikes after pre-processing. The weights trained by a non-spiking CNN have been used in the spiking layers. The neuron with maximum activity (spike frequency) has been selected as the image's class. One of the most significant CNN-to-SNN approaches for energy efficient recognition is the structure displayed in Fig. 8 in which the weight normalization is employed to reduce the performance loss. The MRI images are employed for the SCNN by using STDP learning. There are 700 experimental MRI images as input, 80% of which are used for training

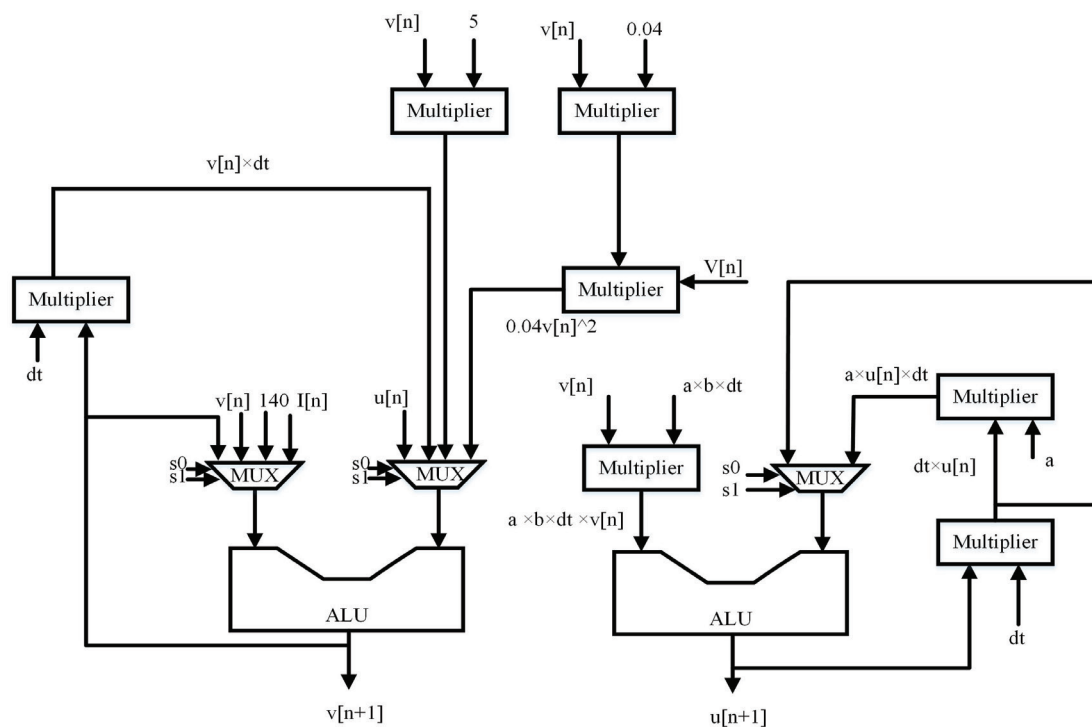


Figure 4: The general structure of the Izhikevich neuron and the used blocks

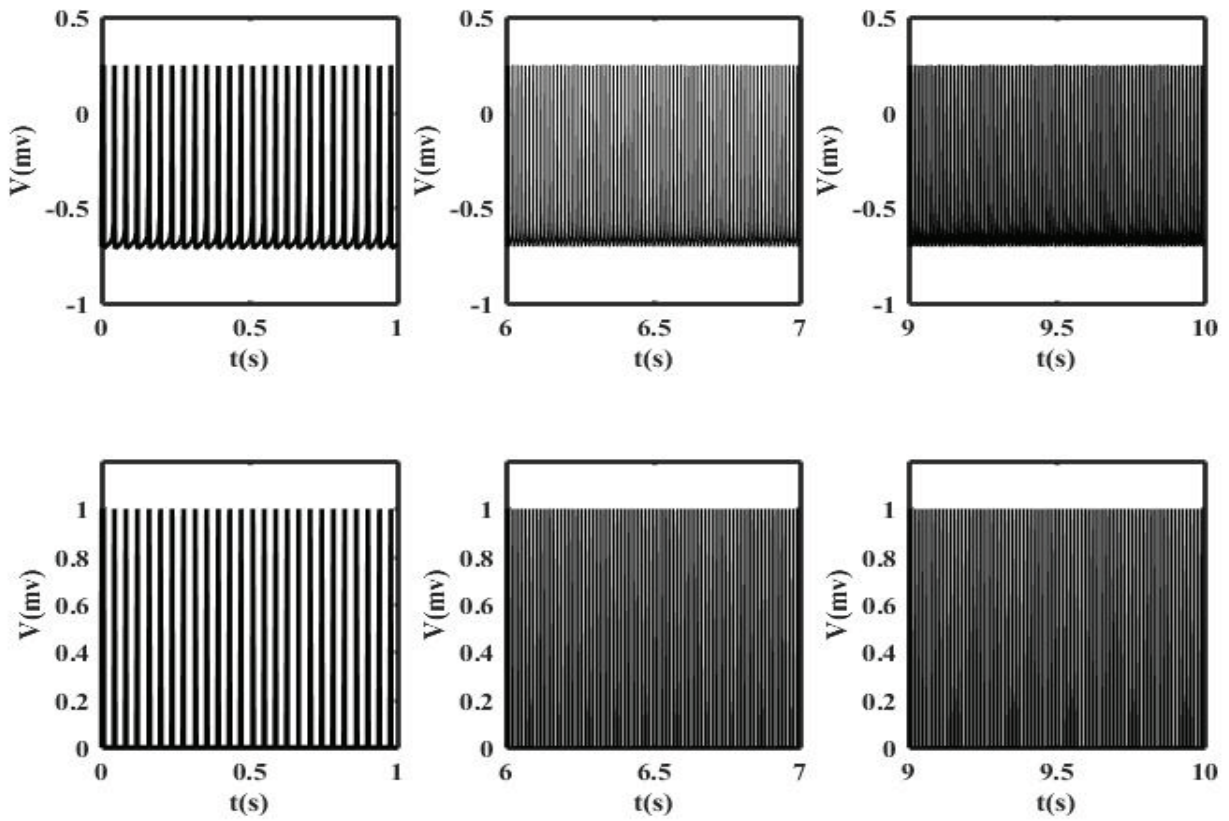


Figure 5: Convergence of the neuron of the network output layer to 80 Hz frequency

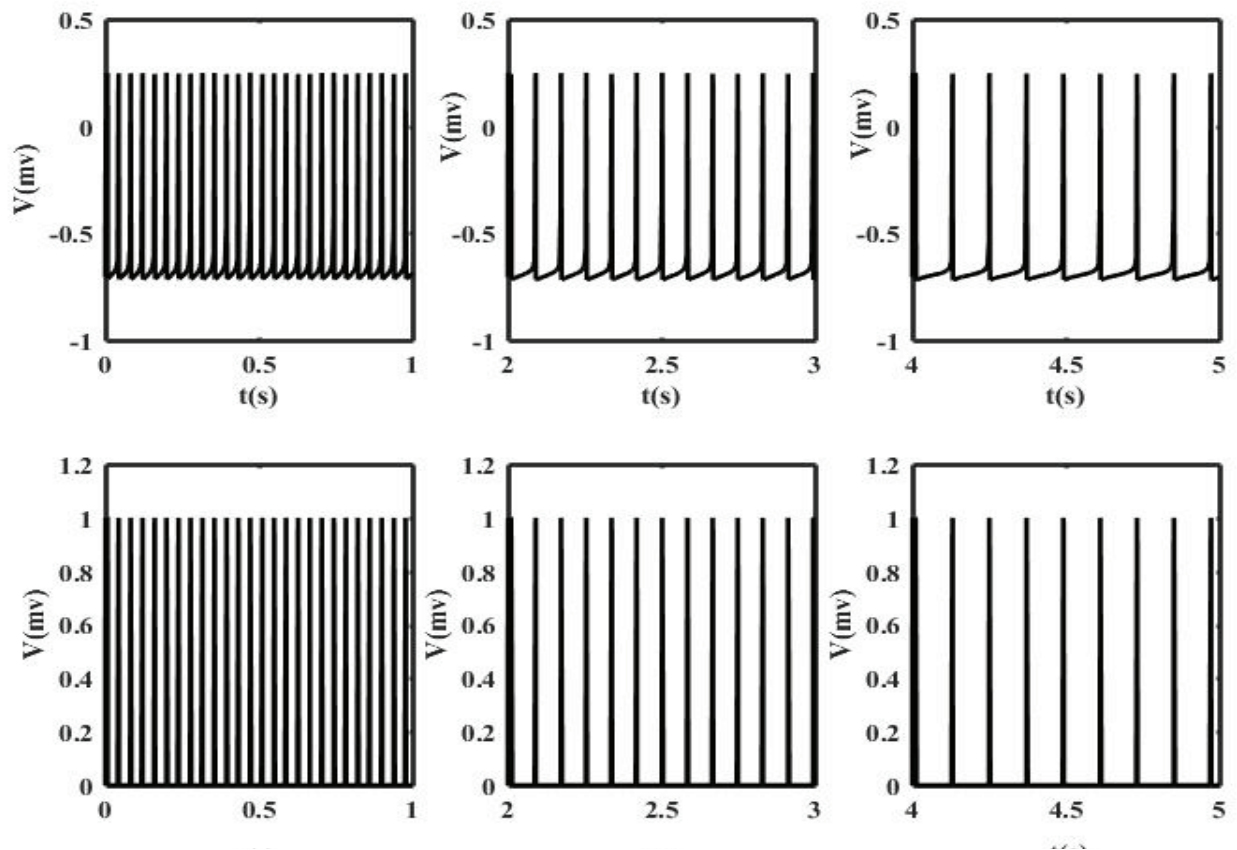


Figure 6: Convergence of the neuron of the network output layer to 11 Hz frequency

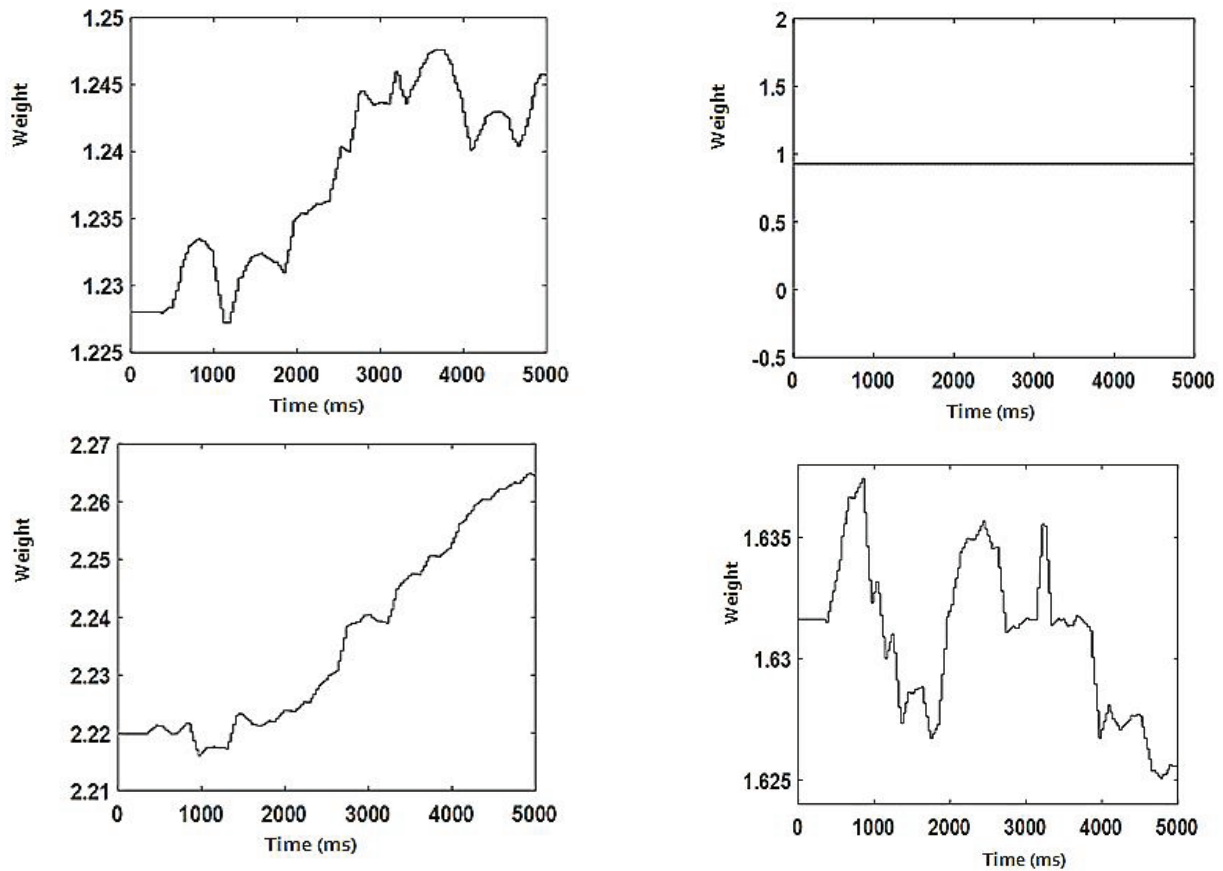


Figure 7: Changing weights during the training process

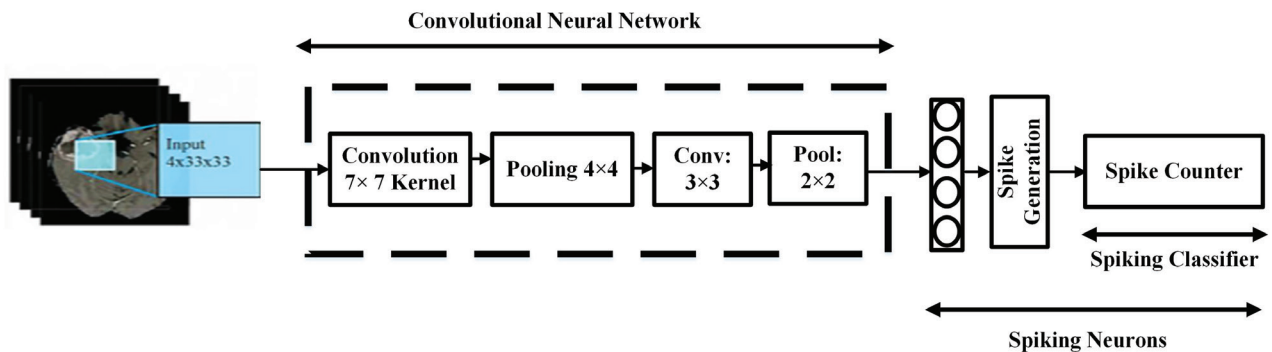


Figure 8: Architecture - The CNN is a two-way path in which the rotation on the input piece is performed in two ways.

and 20% are applied for test. All simulations have been done in MATLAB.

Fig. 9 provides a better picture of the performance of the proposed architecture during shaping features which suggests that it is an improvement on the basis of the Dice criteria in all areas of the tumor.

5 Physical Implementation

The proposed SCNN is used for recognizing the tumors in MRI images which are categorized based on the

presence or absence of tumors as shown in Fig. 10. On the base of this figure, a physical implementation can be proposed for detecting the differences in these two categories. Fig. 11 depicts the final results of physically-implemented network by which the images with and without tumors (displayed in Fig. 10) can be recognized at 11 Hz and 80 Hz frequencies, respectively. Furthermore, the outcomes of the physical implementation are presented as experimental results, verifying and validating the accuracy of the proposed method. Also, the hardware utilization summary results are presented in Table 1 in which an efficient implementation has been obtained. Table 2 compares the synthesis results

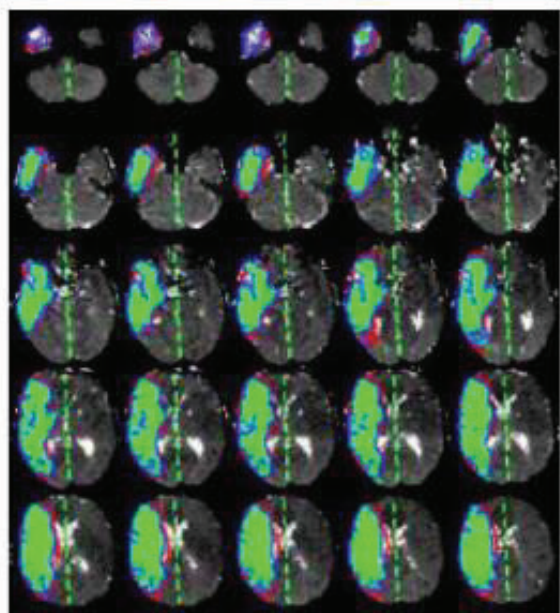


Figure 9: Results of applying the proposed method based on the criterion on one of the database images

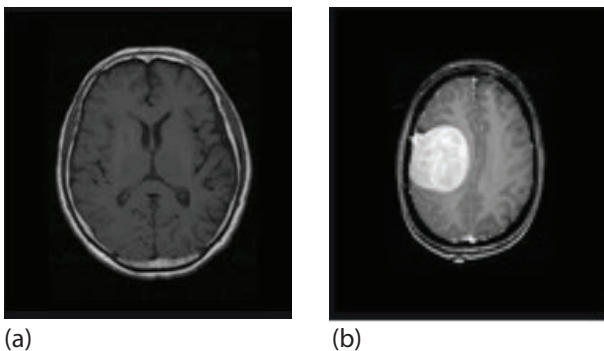


Figure 10: Two MRI images a) without and b) with tumors”

of the present study with those of two other studies reported in the literature, which suggests the lower cost of SCNN as compared with CNN on a Virtex-6 FPGA ML605 board. The merits of SCNN as an ideal choice for hardware implementation purposes is evident.

Table 1: Device utilization summary

Logic Utilization	Used	Available	Utilization
Slice registers	6775	301440	2%
Slice LUTs	24115	150720	16%
Bonded I/Os	9	600	1%
Blocks	1	416	0%
DSP48E1S	20	768	2%

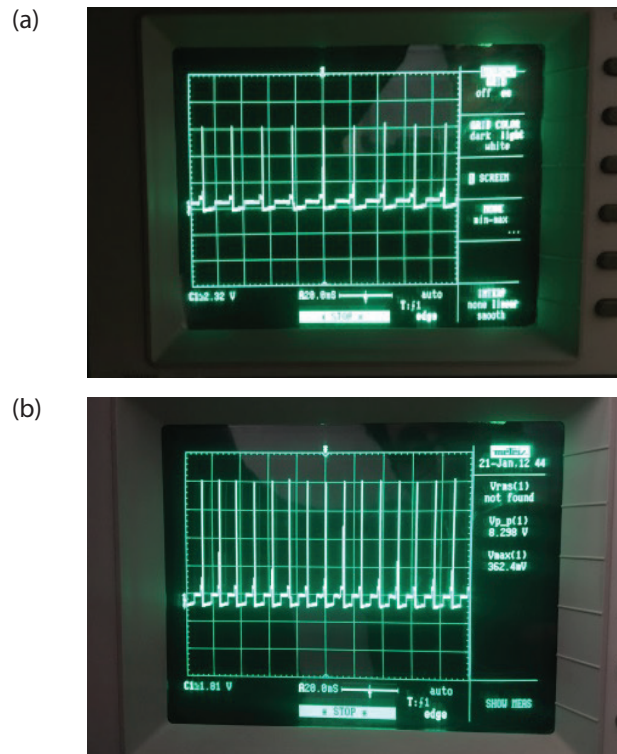


Figure 11: Results of the physical implementation of the proposed SNN: (a) for images with tumor at frequency 11 Hz (b) for images without tumor at frequency 80 Hz

Table 2: Comparison of the synthesis results of the study with the previously published works

Ref.	Slice registers	Slice LUTs
[3]	33.87%	61.3%
[26]	18.15%	32.82%
This work	2%	16%

6 Conclusions

In this paper, an automated fragmentation of brain lesions was presented based on deep convolution neural networks. The proposed architecture is a great improvement in the voxel-based classification with regard to neighboring information and a number of features. In addition to the ability to apply MR images, the proposed method can also be applied to enhanced-contrast scan images. Moreover, by appropriate training of this learning method, a wide range of medical images taken with different devices can also be covered. A spiking neural network was used to detect benign and malignant tumors. Furthermore, hardware design and digital implementation on the FPGA framework improved the speed, accuracy, and flexibility of

the proposed spiking network, which uses a combination of the Izhikevich and the STDP learning model. The experimental test results confirmed the accuracy of the proposed model. To the best of our knowledge, while there are some research studies on medical image processing by deep learning, particularly convolutional neural networks, no instance of hardware implementation of SCNN was found to use medical images as a dataset. Also, our dataset is a special one based on experimental work and it is not a general dataset. Our hardware implementation based on a Virtex-6 FPGA ML605 board, presents a hardware module which is appropriate for MRI-embedded devices and this reduces human mistakes. Also, based on the results of the study, the possibility of physical implementation is recommended.

7 References

1. A. Putnam et al., "A reconfigurable fabric for accelerating large-scale datacenter services," ACM SIGARCH Computer Architecture News, vol. 42, no. 3, pp. 13–24, 2014.
<https://doi.org/10.1145/2678373.2665678>
2. A. Pullini, et al., "A heterogeneous multi-core system-on-chip for energy efficient brain inspired computing," IEEE Transactions on Circuits and Systems II (TCAS-II): Express Briefs, 2017.
<https://doi.org/10.1109/TCSII.2017.2652982>
3. C. Zhang et al., "Optimizing fpga-based accelerator design for deep convolutional neural networks," Proceedings of the ACM/SIGDA Int Symp on Field-Programmable Gate Arrays, pp. 161-170, 2015.
<https://doi.org/10.1145/2684746.2689060>
4. M. Motamedi, et al., "Design space exploration of fpga-based deep convolutional neural networks," proceedings of Asia and South Pacific Design Automation Conference (ASP-DAC), pp. 575-580, 2016.
<https://doi.org/10.1109/ASPDAC.2016.7428073>
5. A. Rahman, J. Lee, and K. Choi. "Efficient FPGA acceleration of Convolutional Neural Networks using logical-3D compute array," proceedings of Design, Automation and Test in Europe (DATE), pp. 1393-1398, 2016.
<https://doi.org/978-3-9815370-7-9/DATE16>
6. F. Ponulak and A. Kasinski, "Introduction to spiking neural networks: Information processing, learning and applications," Acta Neurobiologiae Experimentalis, vol. 71, no. 4, pp. 409–433, 2011.
<https://www.ncbi.nlm.nih.gov/pubmed/22237491>
7. A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," Journal of Physiol, vol. 117, no. 4, pp. 500–544, Aug. 1952.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1392413/>
8. E. M. Izhikevich, "Which model to use for cortical spiking neurons?" IEEE Trans. Neural Netw., vol. 15, no. 5, pp. 1063-1070, Sep. 2004.
<https://doi.org/10.1109/TNN.2004.832719>
9. E. M. Izhikevich, "Simple model of spiking neurons," IEEE Trans. Neural Netw., vol. 14, no. 6, pp. 1569-1572, Nov. 2003.
<https://doi.org/10.1109/TNN.2003.820440>
10. R. FitzHugh, "Impulses and physiological states in theoretical models of nerve membrane," Biophysical Journal, vol. 1, no. 6, pp. 445– 466, Jul. 1961.
[https://doi.org/10.1016/S0006-3495\(61\)86902-6](https://doi.org/10.1016/S0006-3495(61)86902-6)
11. J. Nagumo, S. Arimoto and S. Yoshizawa, "An active pulse transmission line simulating nerve axon," Proc Inst. Radio Eng, vol. 50, no. 10, pp. 2061–2070, Oct. 1962.
<https://doi.org/10.1109/JRPROC.1962.288235>
12. R. M. Rose and J. L. Hindmarsh, "The assembly of ionic currents in a thalamic neuron I. The three-dimensional model," Proceeding of the Royal Society, vol. 237, no. 1288, pp. 267–288, Aug. 1989.
<https://doi.org/10.1098/rspb.1989.0049>
13. H. Paugam-Moisy and S. Bohte, "Computing with spiking neuron networks," In Handbook of natural computing, Springer Berlin Heidelberg, pp. 335-376, 2012.
https://doi.org/10.1007/978-3-540-92910-9_10
14. G. Indiveri, E. Chicca, and R. Douglas, "A VLSI array of low-power spiking neurons and bistable synapses with spike-timing dependent plasticity," IEEE Trans. Neural Netw., vol. 17, no. 1, pp. 211–221, Jan. 2006.
<https://doi.org/10.1109/TNN.2005.860850>
15. R. Tapiador-Morales, A. Linares-Barranco, A. Jimenez-Fernandez, G. Jimenez-Moreno, "Neuromorphic lif row-by-row multiconvolution processor for fpga", IEEE Transactions on Biomedical Circuits and Systems, vol. 13, no. 1, pp. 159-169, Feb 2019.
<https://doi.org/10.1109/TBCAS.2018.2880012>
16. T. Naka and H. Torikai, "A Novel Generalized Hardware-Efficient Neuron Model Based on Asynchronous CA Dynamics and Its Biologically Plausible On-FPGA Learnings" IEEE Trans. Circuits Syst. II, Express Briefs, vol. 66, no. 7, pp. 1247–1251, July. 2019.
<https://doi.org/10.1109/TCSII.2018.2876974>
17. G. Karimi, M. Gholami, and E. Z. Farsa, "Digital implementation of biologically inspired Wilson model, population behavior, and learning," *Int. J. Circuit Theory Appl.*, vol. 46, no. 4, pp. 965–977, Apr. 2018.
<https://doi.org/10.1002/cta.2457>
18. E. Z. Farsa, S. Nazari, and M. Gholami, "Function approximation by hardware spiking neural network," *J. Comput. Electron.*, vol. 14, no. 3, pp. 707–716, Sep. 2015.
<https://doi.org/10.1007/s10825-015-0709-x>

19. A. Jiménez-Fernández et al., "A binaural neuromorphic auditory sensor for FPGA: A spike signal processing approach," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 4, pp. 804–818, Apr. 2016. <https://doi.org/10.1109/TNNLS.2016.2583223>
20. S. Yang et al., "Digital implementations of thalamocortical neuron models and its application in thalamocortical control using FPGA for Parkinson's disease", *Neurocomputing*, vol. 177, pp. 274-289, 2016. <https://doi.org/10.1016/j.neucom.2015.11.026>
21. E. Zaman Farsa, A. Ahmadi, M. A. Maleki, M. Gholami, and H. Nikafshan Rad, "A low-cost high-speed neuromorphic hardware based on spiking neural network," *IEEE Trans. Circuits Syst. II: Express. Briefs*, published, Jan. 2019. <https://doi.org/10.1109/TCSII.2019.2890846>
22. E. I. Guerra-Hernandez, A. Espinal, P. Batres-Mendoza, C. Garcia-Capulin, R. D. J. Romero-Troncoso, H. Rostro-Gonzalez, "A FPGA-based neuromorphic locomotion system for multi-legged robots", *IEEE Access*, vol. 5, pp. 8301-8312, Apr. 2017. <https://doi.org/10.1109/ACCESS.2017.2696985>
23. Michael Pfeiffer, Thomas Pfeil, "Deep learning with spiking neurons: Opportunities & challenges", *Frontiers in Neuroscience*, vol. 12, pp. 774, 2018. <https://doi.org/10.3389/fnins.2018.00774>
24. H. Markram, W. Gerstner, P. J. Sjstrm, "A history of spike-timing-dependent plasticity", *Front. Synap. Neurosci.*, vol. 3, 2011. <https://doi.org/10.3389/fnsyn.2011.00004>
25. M. R. Azghadi, et al. "Efficient design of triplet based spike-timing dependent plasticity," *IEEE International Joint Conference on Neural Networks (IJCNN)*, 2012. <https://doi.org/10.1109/IJCNN.2012.6252820>
26. A. Azarmi Gilan, et al. "FPGA-based Implementation of a Real-Time Object Recognition System using Convolutional Neural Network", *IEEE Transactions on Circuits and Systems II: Express Briefs*. 2019. <https://doi.org/10.1109/TCSII.2019.2922372>



Copyright © 2019 by the Authors.

This is an open access article distributed under the Creative Commons Attribution (CC BY) License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Arrived: 27. 09. 2019

Accepted: 17. 12. 2019