Biomedical Signal Processing and Control 62 (2020) 102106

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

FPGA-based real-time epileptic seizure classification using Artificial Neural Network

Rijad Sarić [a,*], Dejan Jokić [a], Nejra Beganović [b], Lejla Gurbeta Pokvić [a,c], Almir Badnjević [a,c,d]

[a] International Burch University, Faculty of Engineering and Natural Sciences, Department of Electrical and Electronics Engineering, Sarajevo, Bosnia and Herzegovina
[b] Mid-Sweden University, Faculty of Science, Technology and Media, Department of Electronics Design, Sundsvall, Sweden
[c] Medical Device Inspection Laboratory VERLAB, Sarajevo, Bosnia and Herzegovina
[d] University of Sarajevo, Faculty of Pharmacy, Sarajevo, Bosnia and Herzegovina

ARTICLE INFO

Keywords:
ANN
FPGA
Biomedical signal processing
Real-time epilepsy classification

ABSTRACT

Epilepsy is a neurological disorder characterised by unusual brain activity widely known as seizure affecting 4-7% of the world's population. The diagnosis of this disorder is currently based on analysis of the electroencephalography (EEG) signals in the time-frequency domain. The analysis is performed applying various algorithms that yield high performance, however the challenge of effective real-time epilepsy diagnosis persists.

To address this, we have developed a Field Programmable Gate Array (FPGA) based solution for the classification of generalized and focal epileptic seizure types using a feed-forward multi-layer neural network architecture (MLP ANN).

The neural network algorithm is trained, validated and tested on 822 captured signals from Temple University Hospital Seizure Detection Corpus (TUH EEG Corpus) database. Inputs into the system were five main features obtained from EEG signals by time-frequency analysis followed by Continuous Wavelet Transform (CWT) and subsequent statistical analysis. Out of the total number of samples, 583 (70 %) of them were utilised during the system development in MATLAB and TensorFlow and 239 (30 %) samples were further used for subsequent testing of the model performance on the FPGA. Subsequently, the adequate parameters of the ANN model were determined by using k-Fold Cross-Validation. Finally, the best performing ANN model in terms of average validation data accuracy achieved during cross-validation was implemented on the FPGA for real-time seizure classification. The digital ANN solution was coded in Very High-Speed Integrated Circuit Hardware Description Language (VHDL) and tested on the FPGA using 30 % reaming data.

The results of this research demonstrate that epilepsy diagnosis with quite high accuracy (95.14 %) can be achieved with (5-12-3) MLP ANN implemented on FPGA. Also, the results show the steps towards appropriate implementation of ANN on the FPGA. These results can be utilised as the basis for the design of an application-specific integrated circuit (ASIC) allowing large serial production.

1. Introduction

Epilepsy is one of the most common neurological disorders affecting approximately 60 million people of all ages [1–3]. This disorder may disrupt sensory and motor functions, usually followed by repetitive epileptic seizures [4–6]. Although seizures generally last less than two minutes, they are manifested by the disturbance in the electrical activity of the brain neurons, which is life-threatening [7]. The recent findings reveal that the prevalence of epilepsy is nearly 80 % in low-income countries [8]. Almost 40 % of epilepsy patients do not obtain adequate medical treatment [9]. Besides, a vast majority of them are not adequately diagnosed since current diagnosis procedures require lengthy and exhausting medical examination.

Currently, in medical practice, epilepsy diagnosis is based on monitoring and analysis of electroencephalographic (EEG) recordings since the occurrence of epileptic seizures is noticeable in the frequency spectrum of EEG signals [10]. The main limitation of conventional intermittent EEG monitoring represents the difficulty in tracking electrocerebral activity during the hourly or daily diagnosis. Additionally, the interpretation of the clinical findings is complex, which may lead to incorrect diagnosis. This problem can be solved by using specific purpose biomedical device, based on machine learning algorithm capable of

* Corresponding author.
E-mail address: rijad.saric@stu.ibu.edu.ba (R. Sarić).

https://doi.org/10.1016/j.bspc.2020.102106
Received 5 March 2020; Received in revised form 9 July 2020; Accepted 21 July 2020
Available online 1 August 2020
1746-8094/

identifying suspicious brain activities in the real-time.

Artificial Neural Networks (ANNs) find significant application in various medical branches because of enabling detection of valuable artefacts as well as providing decision support that effectively reduces the workload for clinicians [11,12]. Application of ANNs in diagnosis of medical diseases, disorders and conditions has been proven to achieve excellent results [12–15]. ANNs are powerful tool in biomedical signal processing and analysis as well. For instance, Gabor et al. [16] utilize feed-forward ANN to analyse EEG signals and identify epileptic patterns in the recorded EEG data.

Numerous state-of-the-art ANNs are implemented directly onto the hardware to establish real-time diagnostic of different diseases. The most efficient way to implement prototype design of ANN is to use Field Programmable Gate Arrays (FPGAs) because of robust flexibility, reconfigurability and efficient parallel computing. However, FPGA can also bring some disadvantages such as circuit reprogramming, implementation complexity, high expenses, and lack of machine learning libraries [17]. For instance, Zairi et al. [18] propose an FPGA system for real-time diagnostic of arrythmia based on multilayer perceptron (MLP) ANN in combination with a discrete wavelet transform (DWT). Selvathi et al. [19] implement FPGA-based MLP ANN for diagnosis of benign or malignant breast cancer. Ahsan et al. [20] develop an efficient FPGA-based system to assist disabled or aged persons by making the hand gesture identification classifying electromyography (EMG) signals. Presently, extensive research has been conducted regarding analysis of EEG signals for diagnostic purposes. Karakaya et al. [21] utilize open-source Very High-Speed Integrated Circuit Hardware Description Language (VHDL) design of ANN to classify pre-processed EEG signals as normal or epileptic. As EEG signals belong to the group of nonlinear or nonstationary signals, Daoud et al. [22] take advantage of the Hilbert Transform (HT) to extract and select the most relevant features as inputs to the MLP ANN. Saleheen et al. [23] propose an embedded low-power hardware solution for automated real-time detection of epileptic seizures, combining ANN with a statistical function of variance, to extract relevant features of EEG signals. Rajaguru et al. [24] implement Wavelet Neural Network (WNN) on the FPGA to detect epileptic seizures using pre-processed EEG signals. Kueh et al. [25] demonstrate a feasibility study of how a simple ANN for classification of epileptic seizures could be effectively implemented on FPGA using massively parallel bit-serial processing. All highlighted research studies propose FPGA-based ANN classifier to considerably speed-up detection and classification of epileptic seizures. Certainly, medical device integrated with a well-trained ANN classifier could provide strong support to medical examiners during seizure monitoring and final diagnosis [26]. The advantage of such a device is its ability in determining correct seizure type observing raw EEG signals.

The aim of this research is to develop FPGA-based solution for classifying generalized and focal epileptic seizures including the case with no seizure occurrence using feed-forward multi-layer neural network known as MLP ANN. Finding the best performing ANN model and avoiding the potential bias, the k-Fold Cross-Validation is applied in TensorFlow framework where different ANN parameters are examined in order to obtain desirable data accuracy. The proposed FPGA seizure 5-12-3 ANN classifier differs from other existing solutions in terms of accuracy, hardware utilization and flexible design. The tested prototype design can be utilised as the basis for the implementation of application-specific integrated circuit (ASIC), allowing serial microchip production integrated with a pre-trained ANN seizure classifier.

2. Methodology

The workflow of the development of automated system for real-time diagnosis of epileptic seizures implemented on the FPGA board is illustrated in Fig. 1.

The development process consists of the following steps:

(1) Pre-processing of recorded EEG signals and extraction of valuable features
(2) Developing an automated system for diagnosis of epileptic seizures using MLP ANN with the best data accuracy developed in TensorFlow framework and further tested in MATLAB
(3) Implementation of developed automated system on the FPGA using VHDL
(4) Real-time validation of developed system
(5) Comparison between accuracies of training, validation and test data achieved in MATLAB and FPGA.

2.1. Diagnosis of epilepsy

Diagnosis of particular epileptic seizure requires monitoring of EEG recordings to identify seizure type in the frequency spectrum of the signal [10]. Furthermore, a detailed description of events occurring before, during and after the seizure is necessary for a clinical presumptive epilepsy diagnosis. Impeccable epilepsy diagnosis [27,28] requires categorization of epilepsy as well as the type of seizure. Depending on a place of occurrence epileptic seizures are divided as follows: Generalized Non-Specific Seizure (GNS), Focal Non-Specific Seizure (FNS), combined Non-Specific Seizure and unknown epileptic seizure type.

Generalized epileptic seizures involve both brain hemispheres causing impaired awareness or loss of consciousness. The recorded EEG data of a patient with generalized epilepsy typically display spike-wave brain activities during clinical diagnosis. Focal epileptic seizures involve one brain hemisphere, usually causing weak awareness. The diagnosis of this type of epilepsy is supported by interictal EEG recordings.

Combined epilepsy involves both generalized and focal seizures including clinical diagnosis, whereas the unknown type of epilepsy indicates that a patient experiences a certain type of seizure. However, it is difficult to precisely determine the appropriate type as a result of insufficient EEG information.

The primary limitation of conventional EEG monitoring is the difficulty in tracking electrocerebral activity during the hourly or daily diagnosis process. Thereby, the interpretation of the collected results is difficult, leading to inaccurate diagnosis. The amplitude-integrated EEG (aEEG) technique replaces conventional EEG monitoring in many clinical practices. Recently, methods based on visual EEG information have offered significant benefits, however the video-EEG systems are usually not affordable [28].

2.2. Dataset description

EEG measurements used during the development of the automated system for the real-time classification of epileptic seizures are taken from Temple University Hospital Seizure Detection Corpus (TUH EEG Corpus) database [29]. This is an open-source database containing raw 16-channel EGG samples recorded using standardized 10-20 electrode configuration in Average Reference referential montage. All samples are stored in European Data Format (.edf) [30] together with a report written by the certified neurologist. The sampling frequency is 250 Hz, and it is also used in creating dataset samples. While developing the system, signals from the dataset are transformed into multiple samples, where each sample represents the corresponding EEG signal of 10 s duration. Any period below 10 s leads to poor classification since short-time periods do not generate informative features from considered EEG signals.

As it can be seen from Table 1, signals obtained from 315 patients are used for the system development. In total 247 samples contain focal epileptic seizure type, whereas 245 are with generalized epileptic seizure type and lastly 330 samples do not have epileptic seizure occurrence. The default distribution of TUH dataset including classes and number of patients is listed in Table 1.

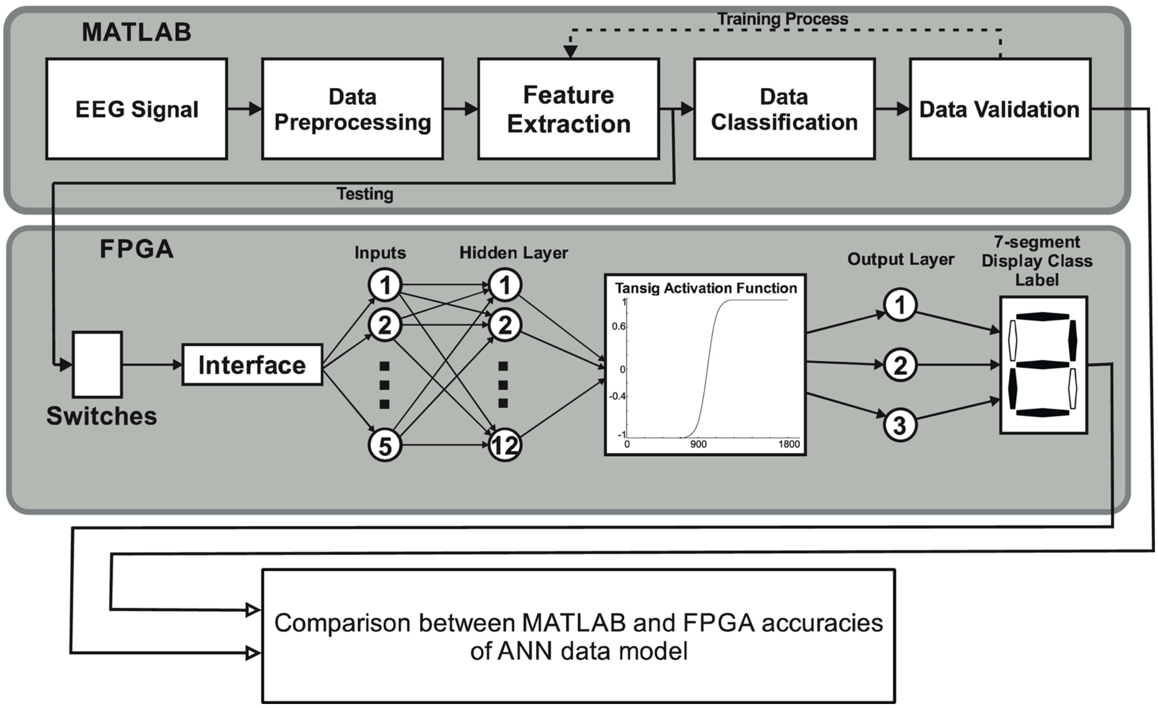


Fig. 1. The block diagram of a workflow towards the real-time implementation of FPGA-based ANN seizure classification system.

Table 1
Default TUH EEG Corpus dataset distribution for development of the ANN data model.

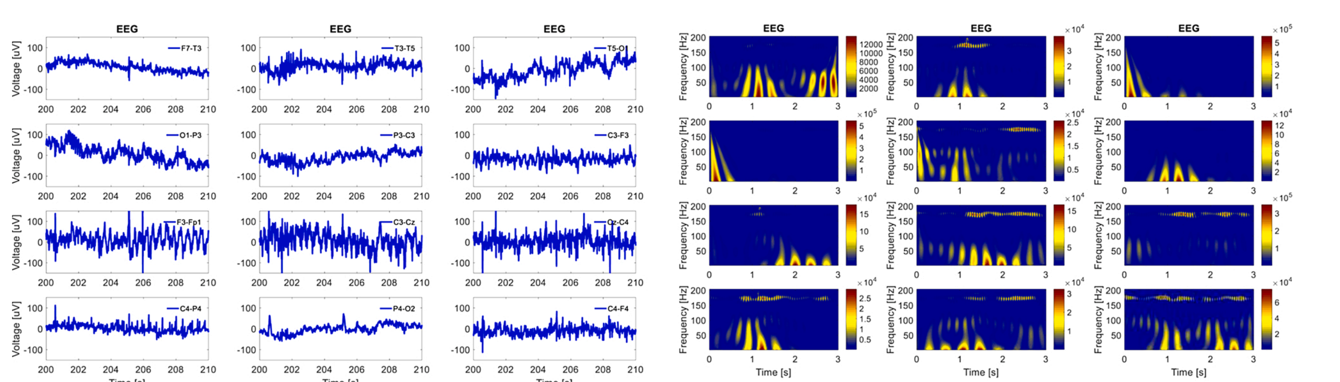
	Number of samples	Number of patients	Dataset etymology	Sample class distribution
Training	583 *(70%)*	265	49.5% male 50.5% female	FNS: 175 GNSZ: 174 NS: 234
Validation	239 *(30%)*	50	44.9% male 56.0% female	FNS: 72 GNSZ: 71 NS: 96
Total number	822	315		

2.3. Data processing and feature extraction

As EEG signals are non-linear in nature and contain artefacts from other processes in the human body, for accurate identification of epilepsy, pre-processing of EEG signals is required [31]. The first step in signal processing is to remove noise and artefacts that do not correspond to the EEG signal [31,32]. After removing noise from signal, feature extraction is performed to extract features that characterize the seizure [33–37]. Several pre-processing methods can be applied such as principle component analysis (PCA), continuous wavelet transformation (CWT), and higher-order spectral analysis (HOS) [36,37]. These pre-processing methods contribute to higher accuracy of epileptic seizure classification and therefore the abovementioned approach is applied in this research.

Removing noise from the signal is completed using a low-pass filter (cut-off frequency 50 Hz), since it is known that frequencies below 40 Hz are the most critical for the identification of epileptic seizures [4,38]. For filtering, Finite Impulse Response (FIR) filter type is used. Fig. 2 (a) shows recorded EEG signal with presence of GNSZ seizure displaying 12 out of 16 channels where all frequency components are available, while Fig. 2 (b) illustrates filtered EEG signal using Finite Impulse Response (FIR) filter type.

Afterward, for feature extraction Continuous Wavelet Transform (CWT) is applied [39]. Owing to the CWT, time-frequency description of signal is generated. Features calculated directly from CWT coefficients [40] are listed in Table 2.



(a) (b)

Fig. 2. EEG signal pre-processing. (a) case of GNSZ seizure occurrence (b) filtered GNSZ seizure after applying Finite Impulse Response (FIR) filter.

As a result of applying CWT on samples, each sample is represented with 160 features, but only 10 time-frequency based features per each channel are selected for further development since these are most relevant for epileptic seizure classification [32]. The final feature vector is represented as a 5 × 130 matrix, where individual feature vector elements (130 samples) are represented in Fig. 3. Each feature vector element represents signal energy content in a particular frequency bandwidth. Fig. 4 illustrates the output of the filtered signal after applying CWT in MATLAB.

2.4. Design of automated system for epileptic seizure diagnosis based on ANNs

For building system capable of performing diagnosis i.e. classification, several machine learning algorithms need to be assessed. In this research, ANNs are selected with respect to other classification algorithms such as Decision Trees (ID3), K-Nearest Neighbors (KNN), Random Forest (RF), Supported Vector Machine (SVM), and similar supervised learning algorithms. Neural Network is expected choice because other machine learning algorithms provide unsatisfactory performance, especially in the case of a short duration of seizure occurrence. The data model based on the SVM algorithm is evaluated with 86 % of validation accuracy, whereas the equivalent RF algorithm provides 79 % accuracy in predicting targeted classes. The main benefit of ANN in comparison to SVM is the fixed size of the model. In order words, ANN is parametric, while SVM is a non-parametric model. Additionally, RF could be an adequate option for training small datasets. However, it is less possible to improve the model performance bringing the new data samples.

A fundamental data model of feed-forward ANN is chosen to classify three seizure cases since such models are often used for the classification of processed or raw biomedical signals as reviewed in similar research studies [41,42]. The standard ANN architecture consists of input, hidden and output layer and it always goes in one direction. Considering, the relatively small volume of data the k-Fold Cross-Validation is the most decent way to determine the best performing ANN model. Hence, it is established by selecting five bins for default training data of 583 samples including uniform distribution of target classes. The cross-validation split ratio is 80 % for training and 20 % for validation data during five independent iterations where each time different set of bins are chosen for training and validation. After the five iterations, the average training and validation accuracy of the ANN model is calculated for particular hyper-parameters. Remaining 239 samples are used for testing the fine-tuned ANN model on the FPGA board after successful hardware implementation. The number of neurons in the hidden layer is determined by performance evaluation. Specifically, by gradually increasing the number of neurons and observing the performance of the ANN model in terms of average training and validation accuracy. Initially, model evaluation using cross-validation is conducted in MATLAB combining Backpropagation [43] as well as Quasi-Newton [44] training algorithms experimenting with activation functions. However, the model accuracy declines as the number of neurons increased, especially in the case of Sigmoid (Sig) and Rectified Linear Unit (ReLU) activation functions. Thus, ANN model optimization is required to solve the evident occurrence of overfitting. TensorFlow Python framework is applied to analyse

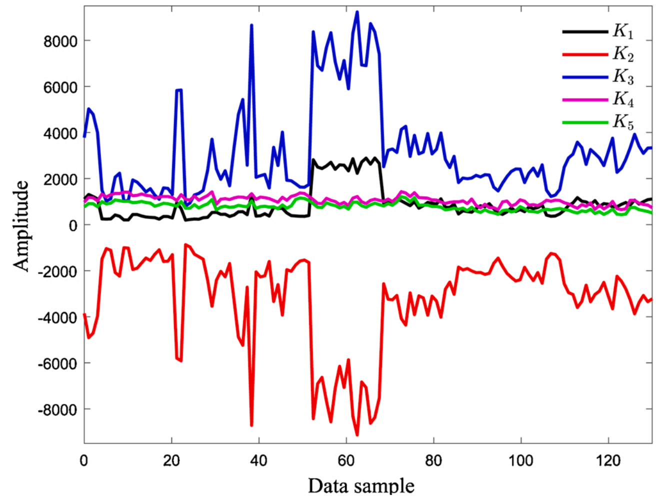


Fig. 3. Individual feature vector elements of arbitrary EEG signal sample described in the 10-seconds time interval during which the CWT coefficients and accompanying statistical variables are calculated.

and test different ANN models since it allows much higher flexibility than the MATLAB environment.

At first training data vector with processed EEG signals needs to be adequately formatted, making it suitable for feeding the ANN model. The useful technique to preserve overfitting in mini-batch training of ANN is data shuffling. In this uniform distribution of class labels or data vectors e.g. {1,1,1,2,2,2,2,3,3,3…} shuffling is applied to randomly change data positions, but the values remain the same. During the training process of shuffled data, it is less possible that the two or more training iterations (epochs) would be performed on the same data sequences. After that, the common Min-Max data normalization approach is applied by linearly transforming (scale) the data points between min and max values available in the considered dataset. The final step regarding data adjustment is to ensure that all categorical values in class label vector are converted into binary values since ANN models cannot operate with categorical data. Similarly, each ANN model designed in TensorFlow is evaluated using k-Fold Cross-Validation with the same data split during the five-independent training/test blocks. The models are developed using stochastic gradient descent (SGD) training algorithm with Adaptive Moment Estimation (Adam) [45] optimizer, which at the same time calculates the derivative and updates of training data instances. In fact, during the learning process of the ANN model, Adam optimizer maintains and adapts the learning rate for each parameter (weight and bias). The parameter-specific learning rates with Adaptive Gradient algorithm (AdaGrad) [46] is applied as the second approach. AdaGrad optimizer does adoption of learning rate to ANN parameters associated with frequently occurring features i.e. smaller updates (low learning rates) as well as higher updates (high learning rates) in case of infrequent features. Both methods are evaluated through cross-validation with a different number of neurons in the hidden layer (1) 4–60 neurons and (2) 6–80 neurons including different activation functions such as Sig, ReLU, Hyperbolic Tangent (Tanh) and Linear. As

Table 2
Time-frequency feature vector F1 obtained from EEG signal captured from 16-EEG scalp electrodes; F1 = [K1 … K5].

Feature vector element	Mathematical representation	Description
K_1	[avg(abs($X_{k\text{-}m}$ = 90−110))]	*$X_{k\text{-}m}$ - CWT coefficients between k and m CWT scales, mapped to corresponding frequency range and obtained from 16-EEG channels*
K_2	[min($X_{k\text{-}m}$ = 115−126)]	K_1 – individual feature is a vector which represents averaged
K_3	[max($X_{k\text{-}m}$ = 115−126)]	absolute values of CWT coefficients between the scales 90 and 110.
K_4	[sum(abs($X_{k\text{-}m}$ = 125−128))/ sum(abs($X_{k\text{-}m}$ = 60−128))]	K_5 – individual feature is a vector containing the percentage of
K_5	[sum(abs($X_{k\text{-}m}$ = 125−128))/ sum(abs($X_{k\text{-}m}$ = 1−128))]	overall signal energy presented in the last four CWT scales (lowest frequency bandwidth).

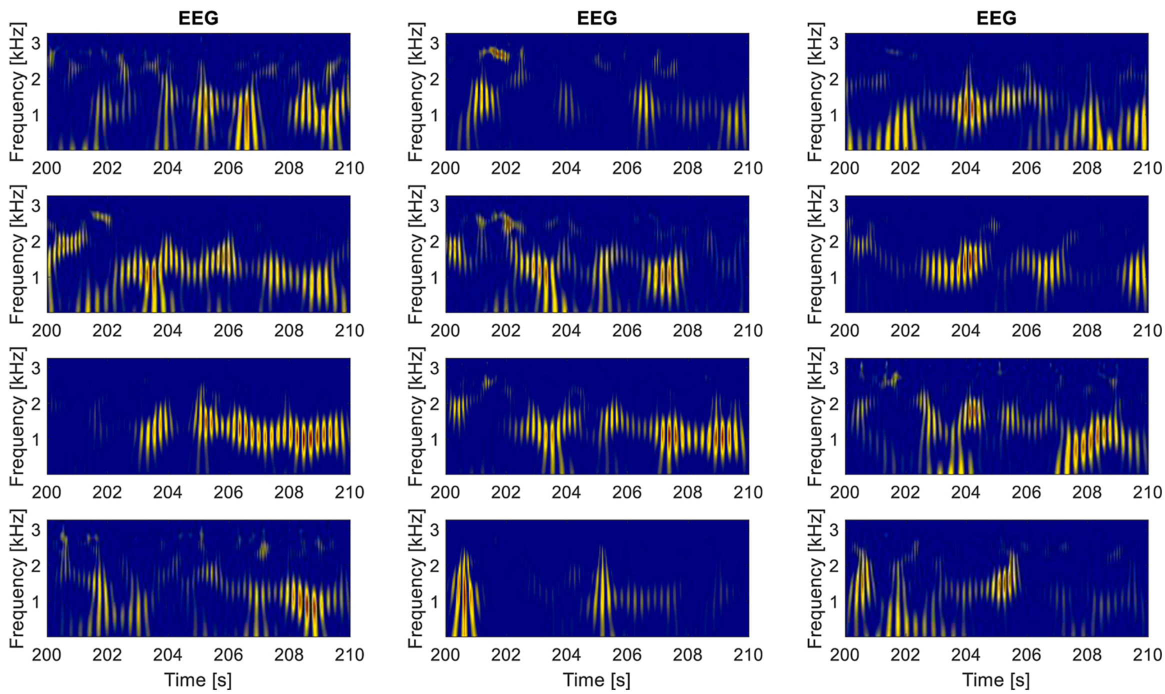


Fig. 4. EEG signal features recorded using 10-seconds interval and selected by applying Continuous Wavelet Transform (CWT).

the dataset contains more than two classes, for the optimization of the model two multi-class loss (cost) functions namely Categorical Cross-Entropy and Sparse Cross-Entropy are used. The purpose of these functions is to compute the desired quantity that ANN model seeks to minimize during a training process. Lastly, the number of epochs is chosen to be 80, including a fixed batch size as well as drop out value. This can help in avoiding overfitting and the problem know as Early Stopping where ANN model does not go over all available training data instances. In Tables 3 and 4, the achieved average training and validation accuracies of ANN models with different hyper-parameters and two learning algorithms are listed. It is noticeable that the best average validation accuracy of 96.7 % is achieved using SGD training of ANN model with sigmoid activation function, 46 neurons in the hidden layer and categorical cross-entropy loss function. On the other side, the highest validation accuracy of 94.8 % for parametric specific ANN model is reached with the same activation function but 80 neurons in the hidden layer including sparse cross-entropy loss function.

ANN model based on SGD training algorithm and Adam optimizer with 12 neurons, Sig-activation function and categorical cross-entropy is selected for FPGA-based implementation. The average validation accuracy of this model is 95.5 %. Parameters and accuracy of this ANN model need to be observed in MATLAB as well. According to [47] considering that date size (n) is sufficiently large, the time complexity of SGD algorithm is expressed as $O(p/\varepsilon)$, where p indicates the number of features and ε accuracy to be achieved for the objective function (excess test error). Essentially, the time complexity of this algorithm is smaller than $O(p^3)$, assuming that the number of iterations is less than $O(p^2)$. Standard gradient descent (GD) has higher total time complexity compared to SGD which requires the time of $O(1/\varepsilon)$ per each epoch. Instead of computing the sum of all gradients before updating the weights as GD, SGD tends to decrease computational cost per each epoch by updating the weights after analysis of every training sample. Thus, the computational cost per iteration of SGD is $O(d)$ where d stands for parameter dimension. The total computational cost of SGD algorithm is computed as $O(d/\varepsilon)$.

The highest validation accuracy of 97.4 % achieved in the third iteration of cross-validation for ANN model trained with SGD, Sig activation function, 55 neurons in the hidden layer, categorical CE and 80 epochs is illustrated in Fig. 5.

Extracted features obtained from the EEG signals are inputs to the developed network. They are fed into the (5-12-3) ANN model and normalized using offset and gain values generated during the training phase. Each neuron calculates a weighted sum of inputs whose result is sent to the non-linear Sig activation function. The selection of the sigmoid activation function is especially important for creating this ANN model that predicts whether a seizure has occurred or not, as well as

Table 3
Average of training/validation accuracies calculated after performing k-Fold Cross-Validation on stochastic gradient descent (Adam) ANN models by varying neurons in the hidden layer, activation functions, and loss functions in 80 epochs.

Training algorithm	Loss function (Multi-Class)	Hidden layer neurons	Average Training and Validation Model Accuracy after k-Fold Cross-Validation (k = 5) & Epochs = 80			
			Sigmoid (%)	ReLu (%)	Tanh (%)	Linear (%)
Adam – Stochastic Gradient Descent Method	Categorical Cross-Entropy	4	94.2 | 92.1	85.6 | 82.3	84.5 | 79.6	89.2 | 85.0
		12	97.5 | 95.5	92.6 | 89.1	87.4 | 86.5	93.1 | 89.2
		24	98.3 | 96.2	94.2 | 88.8	94.2 | 91.3	93.8 | 92.1
		46	98.6 | 96.7	94.8 | 91.4	94.7 | 92.1	94.6 | 92.7
	Sparse Cross-Entropy	5	94.6 | 94.0	89.6 | 87.7	94.2 | 92.9	87.3 | 82.2
		15	94.5 | 92.8	92.8 | 89.0	95.3 | 91.2	90.0 | 89.3
		30	94.6 | 93.2	94.6 | 92.0	95.1 | 92.6	92.1 | 88.7
		60	95.3 | 94.1	95.1 | 92.8	95.2 | 93.0	94.2 | 89.6

Table 4
Average of training/validation accuracies calculated after performing k-Fold Cross-Validation on adaptive gradient (AdaGrad) ANN models by varying neurons in the hidden layer, activation functions, and loss functions in 80 epochs.

Training algorithm	Loss function (Multi-Class)	Hidden layer neurons	Average Training and Validation Model Accuracy after *k*-Fold Cross-Validation (k = 5) & Epochs = 80			
			Sigmoid (%)	ReLu (%)	Tanh (%)	Linear (%)
AdaGrad – Parameter Specific Learning Method	Categorical Cross-Entropy	6	82.8 | 78.3	88.5 | 84.3	82.6 | 77.3	88.3 | 84.2
		15	87.5 | 85.0	93.1 | 90.1	85.7 | 82.3	91.7 | 88.3
		35	88.3 | 86.4	93.8 | 90.2	91.7 | 89.2	92.3 | 91.1
		55	90.7 | 88.2	95.0 | 93.3	93.1 | 92.0	93.4 | 92.5
	Sparse Cross-Entropy	8	86.4 | 82.3	87.4 | 81.3	80.5 | 76.4	84.3 | 78.2
		20	93.1 | 88.3	91.3 | 89.7	87.8 | 85.3	89.6 | 84.6
		40	94.2 | 93.4	93.2 | 91.0	91.3 | 89.2	91.4 | 88.7
		80	95.7 | 94.8	94.8 | 91.3	93.4 | 90.7	91.9 | 89.8

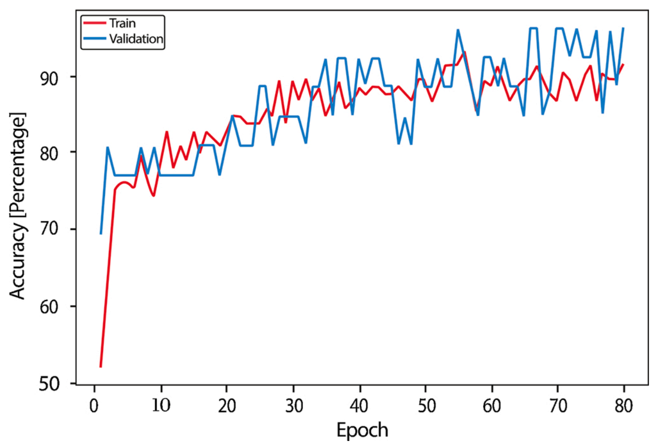


Fig. 5. The max validation accuracy achieved in third iteration of cross-validation data split using ANN trained with SGD and Sigmoid activation function.

classify the correct type of seizure namely GNSZ or FNSZ available in the TUHH EEG dataset [29]. The sigmoid function or "*S*" shaped curve is generally expressed as

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

After finding the ANN model with adequate accuracy including not high complexity in terms of FPGA implementation, an entire network is coded and executed in MATLAB to approve obtain results in TensorFlow. The ANN model yields the same training and validation performance of the network. Finally, The ANN with 5 inputs, 12 neurons in the hidden layer, 3 neurons in the output layer and predicted class label at the output is chosen for diagnosis of epilepsy seizure. The block diagram of the developed multi-layer feed-forward neural network is depicted in Fig. 6.

2.5. Implementation of developed neural network on FPGA

The FPGA Altera D2-115 Cyclone IV E series board is used for the hardware implementation of the developed seizure classification system. The reprogrammable architecture of FPGAs makes the implementation of complex digital circuits simpler because of configurable logic blocks linked through programmable interconnections. The FPGA provides an opportunity of implementing feasible, optimized and fully parallel ANN architecture. With respect to microcontrollers [48], which are also programmable integrated circuit, FPGAs provide better on-chip learning properties. Moreover, FPGA-based implementations can be easily mapped onto new improved FPGAs, thus providing more cost-effective chip development in larger serial ASIC production for biomedical devices suitable for real-time diagnosis and classification of epileptic seizures.

FPGA-based design is fully coded in VHDL. The use of direct VHDL programming enables easy access to internal variables and VHDL subcomponents leading at the same time to increased reliability. In this way, it is possible to design special-purpose embedded hardware modules with greater control over constituent mathematical operations of every ANNs such as multiplication, addition, subtraction, and division.

To implement the developed system on the FPGA, all ANN parameters such as weights, biases, nonlinear activation function, accompanying operations are represented as fixed-point variables with three decimals because of easier hardware-based description compared to standard floating-point representation [49–51]. The fixed-point notation (18-bit radix point) is used as the 5 × 13-bit format or precisely 5-bits for integer and 13-bits for the fraction part. This type of optimization makes embedded coding much easier as well as efficient especially in case of heavy mathematical operations of ANN models for seizure classification [52–54]. Twelve neurons in the hidden layer accept five inputs and one output, where each input is assigned through switches available on the FPGA board or more practically through the interface that contains decimal to binary conversion and transfers the data to neuron blocks using buses. From the Register Transfer Level (RTL) point of view, each neuron in the hidden layer contains five logical multiplier blocks as well as adders, where 18-bit values are extracted from the resulting 36-bit product and then added together. However, the most difficult part is associated with the approximation of the nonlinear sigmoid activation function on the FPGA-based chip using direct VHDL coding. The approximation of the sigmoid function is achieved using a customized lookup table. All values coming from neurons in the hidden layer are scaled in order to make it suitable for calculating the correct index of the lookup table. In addition, scaling is usually undertaken by combining multiplication as well as shifting operations at the lowest level. The trick of the multiplication in case of fixed-point binary notation is that the radix point moves depending on where the original points are. Therefore, the customized lookup table is implemented together with an arithmetic rounding scheme where each index is described as 11-bit unsigned binary number data type.

Lastly, the class label is outputted on the 7-segment display available on the FPGA board. RTL block of designed look-up table contains two multipliers, four adders, MUX, comparator, and synchronized RAM which serves as a binary array where all binary values are stored. The comparator is used to compare some of the obtained indexes with actual index values to prevent inaccurate calculations if they accidentally occur.

3. Results

The developed MLP ANN is subsequently tested based on 239 EEG signals acquired from 50 patients that are taken from the TUH EEG Corpus database [29]. The obtained testing accuracy is 95.14 % after implementing the best-performing ANN model on the FPGA using VHDL coding. Following, the FPGA-based design is tested through functional

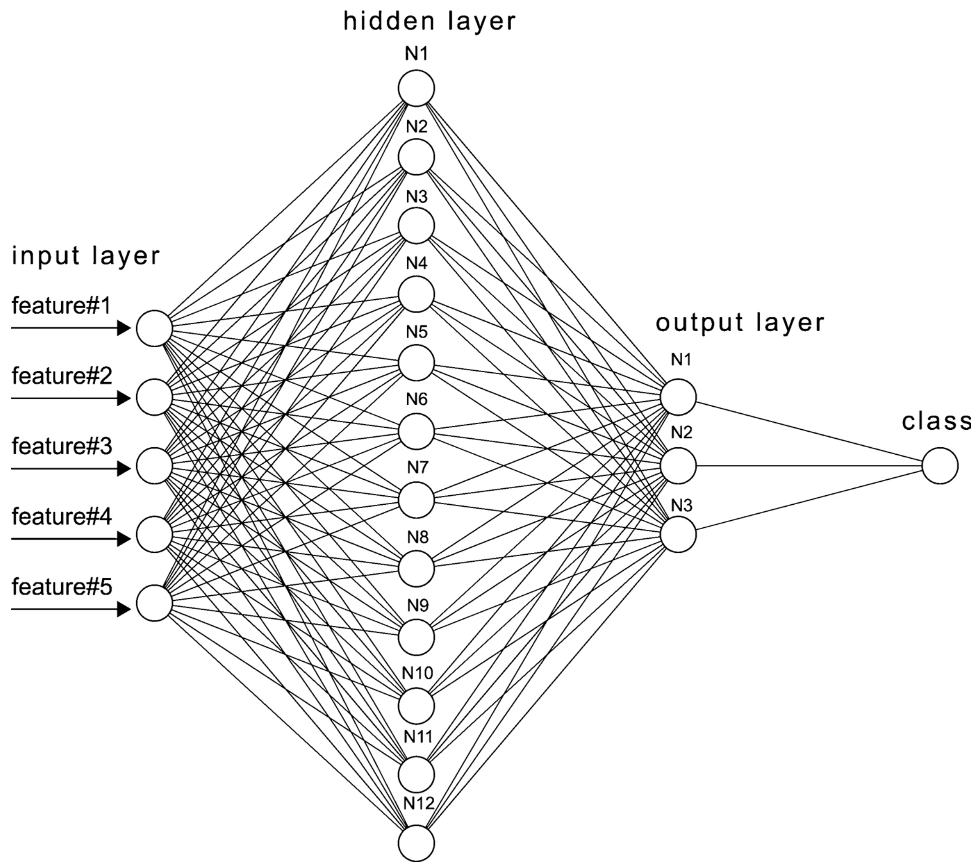


Fig. 6. The 5-12-3 structure of the selected feed-forward MLP ANN model for FPGA-based implementation.

and timing simulation using Intel ModelSim software tool.

The adequate implementation of feed-forward MLP ANN on the FPGA-based embedded device should cover three necessary steps including data representation, weight-bias precisions, and lastly approximation of activation function. During FPGA implementation of developed high-accuracy MLP ANN, 18-bit fixed-point representation is used following the conclusion of Holt et al. [54] that 16-bit fixed point is the minimum acceptable precision that does not affect the performance of the ANN data model on FPGA chip especially in the case of the classification process. The implementation of neurons in the hidden layer on the FPGA-based embedded system is shown in Fig. 7, while the implementation of the nonlinear sigmoid activation function is depicted in Fig. 8. Depending on the ANN application, there are two the most common ways to implement the activation function, namely piece-wise linear approximation, or lookup tables [55]. As already mentioned, in this research, the activation function is modelled using lookup tables.

Upon the FPGA implementation, the overall system performance is scrutinized. For testing purpose, the new 239 samples not seen by the developed network are used to evaluate the performance of the model on the hardware. The system embedded on the FPGA provides an accuracy of 95.14 %. The degradation of performance of only 0.26 % is noticed in Fig. 9 illustrating the difference between accuracies obtained during the development of ANN in MATLAB and TensorFlow as well as its actual implementation on the FPGA. This slight decrease of accuracy is expected as a result of the bit-stream arithmetic applied, and it evidently affects the network performance. However, the reduction to 18-bit fixed-point data is considered insignificant.

The utilization of resources on FPGA chip after successful implementation of the developed ANN model needs to be considered since high utilization on FPGA chip could have several issues regarding interconnect capacity, performance, clock distribution and I/O capacity. The 11 % of hardware footprint of presented ANN implementation means that this way of implementation can be applied for designing specific purpose biomedical device. As such means, device would have much lower hardware cost and it can be widely deployed as an assistive medical tool. The summary of synthesized resources usage is listed in Table 5. The 11 % of hardware resources is a pretty low number for such a massively parallel implementation.

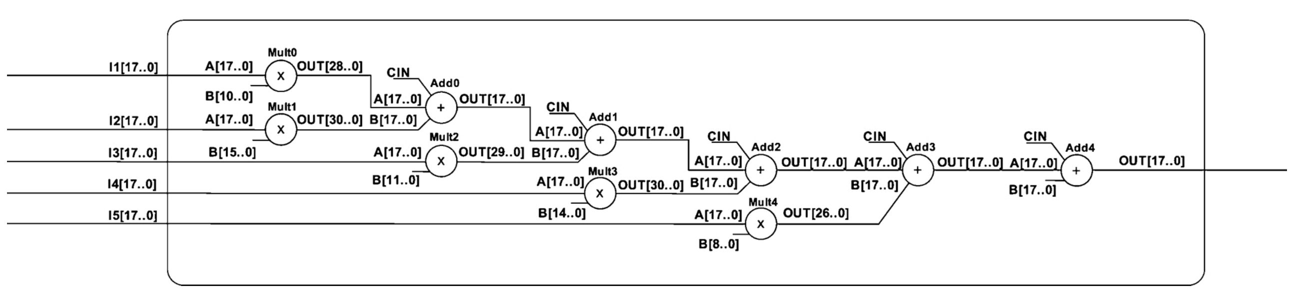


Fig. 7. RTL diagram of the single neuron in the hidden layer of ANN classifier implemented on the FPGA-based embedded system.

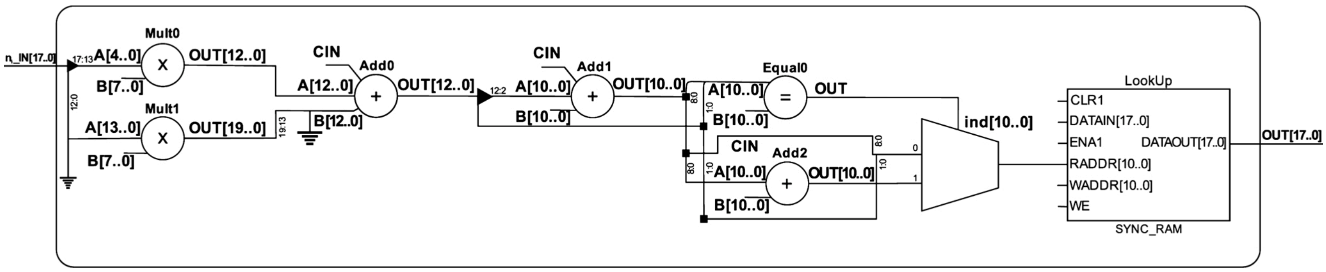


Fig. 8. RTL diagram of the nonlinear sigmoid activation function implemented on the FPGA-based embedded system.

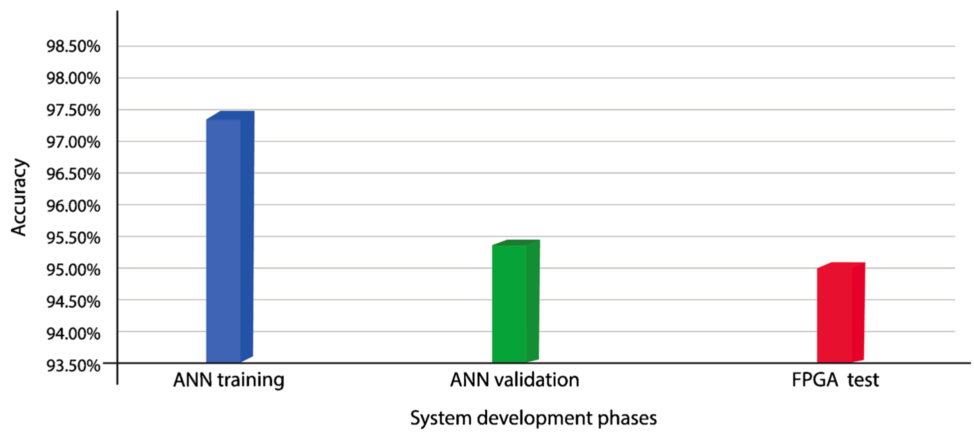


Fig. 9. The comparison between accuracies obtained during training and validation in MATLAB and TensorFlow including data testing on the FPGA board.

Execution of the developed ANN model on the FPGA Altera D2-115 Cyclone IV E board is extremely fast. All processes and calculations are executed in parallel or precisely in 20 nanoseconds (ns) because the base clock frequency is 50 MHz. Execution goes without any delays during the computation on main logical elements like adders and multipliers. Each generated logic circuit performs in the one clock cycle that is 20 ns.

4. Discussion

There has been a significant increase in research aiming to develop smart biomedical devices for personalized, faster, and more accurate disease diagnosis, treatment, and monitoring. Management of chronic diseases and disorders is one of the top priorities in modern healthcare.

One of the widespread chronic disorders is epilepsy that affects 50-70 million people worldwide and accounts for 0.75 % of the global burden of disease [1–3]. Many people with epilepsy have more than one type of seizure and may have other symptoms of neurological problems as well. It is not rare that EEG testing, clinical history, family history and outlook are similar among a group of people with epilepsy. In these situations, their condition can be defined as a specific epilepsy syndrome. Having seizures including epilepsy can affect one's safety, relationships, work, driving, and other normal activities. That is why there is a need for a biomedical device capable of diagnosis and classification of this disorder.

Table 5
The utilization of hardware resources on the FPGA Altera D2-115 board after implementation of best-performing ANN model.

Resource	Utilization [no. of logic elements]	Percentage [%]
Combinational ALUT	12,971	17%
Dedicated Registers	114	12%
I/O Pins	51	10%
9-bit Multipliers	116	22%
Total Logic Elements	13,014	11%

Recent approaches in diagnosis and classification of epileptic seizure have seen the application of ANNs and other machine learning algorithms [56–60]. Over the years, various ANN classifiers based on different features of EEG signal have been developed [23–25]. These classification algorithms provide high accuracy performance, but their implementation has been almost strictly in the software environment, thus not so portable, scalable, and usable in the real-time and busy clinical environment. Also, the performance of software-based ANN has a drawback concerning execution time compared to hardware-based ANN that could provide high speed for the purpose of real-time applications [17].

The hardware implementation of automated intelligent systems, based on ANNs or other machine learning algorithms makes these systems portable, scalable and therefore more usable not only in the real-time clinical environment but in remote areas as well where the presence of medical specialist is in deficit. The major disadvantage of hardware-based platforms is related to obtaining high-level of data accuracy as a result of the large hardware costs required for specific computations [61]. However, this research suggests that ANN trained with SGD algorithm and high data accuracy can indeed be implemented in hardware with high-level of accuracy and low hardware cost, namely in FPGA-based systems. The achieved accuracy of the FPGA-based system is 95.14 % with respect to Karakaya et al. [21] where the model performance is 86.7 %. The accuracy is higher in terms of classification of a different epileptic seizure, compared to Geethu et al. [62] which implemented the SVM classifier on FPGA to detect whether a patient suffers epilepsy or not. Subasi [63] in his research study used Discrete Wavelet Transform (DWT) and the mixture of expert models which yields accuracy of 95 %. However, Particle Swarm Optimization and Morlet activation function can be used to improve the accuracy [24] but hardware resources increase as well. Accuracies near 100 % have been achieved previously [23,64] but in respect to them, in this research ANN model is trained using the larger number of features related to the epileptic seizure and validated on the larger dataset.

The overall hardware footprint of ANN implementation in this research is 11 % with respect to Daoud et al. [22] utilizing 19.3 % of total hardware resources. This is a significant decrease in hardware footprint in comparison to the work by Saleheen et al. [23] where 44 % of the FPGA resources utilization is considered as moderate hardware footprint.

This research study has demonstrated that neural network with a larger number of neurons can be implemented on the ASIC while achieving higher performance speed and lower cost of implementation. Such FPGA solution based on MLP ANN are scalable and portable, so they are very valuable in the real-time diagnosis of the epileptic seizures in a clinical and non-clinical environment.

5. Conclusion

Systems based on artificial intelligence can significantly contribute to solving the problem of effective real-time epilepsy diagnosis. The main goal of this paper was to describe the development of a high accuracy automated system for classification of epileptic seizures. The system was developed by utilising 822 samples of EEG signals acquired from Temple University Hospital Seizure Detection Corpus (TUH EEG Corpus) database.

The results of this paper confirm that automated system based on machine learning algorithm with high accuracy can be successfully implemented on FPGA chips which make them portable and easy to use in everyday diagnosis situation in healthcare institutions. The high-performance automated system yielded accuracy of 95.14 % while tested on FPGA board.

Application-specific integrated circuit (ASIC) such as FPGA prototype presented in this paper, is the next step in developing expert/intelligent systems for disease classification. In this way, these expert systems become scalable and portable which makes their application in clinical and non-clinical environment easier.

***The innovation** in the paper represents concept of implementation of application-specific integrated circuit (ASIC) based on neural network with a larger number of neurons while achieving higher performance speed and lower cost of implementation. Such FPGA solution based on MLP ANN is scalable and portable, so they are very valuable in real-time diagnosis of the epileptic seizures in clinical and non-clinical environment.*

Credit author statement

Rijad Sarić: Full FPGA-based implementation of ANN model using VHDL coding, Implementation of k-Fold Cross-Validation algorithm in TensorFlow for finding the best performing ANN model, Validation of selected ANN model in MATLAB, Writing and Editing. **Dejan Jokić**: Supervision during the implementation of FPGA-based embedded system. **Nejra Beganović**: Finding the TUH EEG Corpus dataset intended for epileptic seizure classification, Establishing raw EEG signal processing and feature extraction/selection in MATLAB, Writing and Editing. **Lejla Gurbeta Pokvić**: Supervision during the development and testing phase regarding the FPGA-based system for seizure classification, Writing-Reviewing and Editing. **Almir Badnjević**: Supervision.

Declaration of Competing Interest

The authors hereby declare to have no conflict of interest.

References

[1] S.S. Alam, M.I.H. Bhuiyan, Detection of seizure and epilepsy using higher order statistics in the EMD domain, IEEE J. Biomed. Health Inform. 17 (2) (2013) 312–318.
[2] J. Liu, Z. Liu, T. Chen, R. Xu, Treatment of epilepsy in China: formal or informal? Neural Regen. Res. 8 (35) (2013) 3316–3324.
[3] M. Jane England, C.T. Liverman, A.M. Schultz, L. Strawbridge, Epilepsy across the spectrum: promoting health and understanding, Epilepsy Curr. Am. Epilepsy Soc. 12 (2012) 245–253.
[4] L. Feldman, B. Lapin, R.M. Busch, J.F. Bautista, Evaluating subjective cognitive impairment in the adult epilepsy clinic: effects of depression, number of antiepileptic medications, and seizure frequency, Epilepsy Behav. 81 (2018) 18–24.
[5] D.M. Hilkman, W.N. van Mook, W.H. Mess, V.H. van Kranen-Mastenbroek, The use of continuous EEG monitoring in intensive care units in the Netherlands: a national survey, Neurocrit. Care (2018) 1–8.
[6] S. Grossberg, Towards solving the hard problem of consciousness: The varieties of brain resonances and the conscious experiences that they support, Neural Netw. 87 (2017) 38–95.
[7] S.A. Gacio, Amplitude-integrated electroencephalography for neonatal seizure detection an electrophysiological point of view, Arq. Neuropsiquiatr. 77 (2019) 122–130.
[8] A. Ngugi, C. Bottomley, I. Kleinschmidt, L. Sander, C. Newton, Estimation of burden of active and life-time epilepsy: a metanalytic approach, Epilepsia 51 (2010) 883–890.
[9] J. French, Refractory epilepsy: clinical overview, Epilepsia 48 (Suppl. 1) (2007) 3–7.
[10] D. Chadwick, Diagnosis of epilepsy, Lancet 336 (1990) 291–295.
[11] J.L. Patel, R. Goyal, Applications of artificial neural networks in medical science, Curr. Clin. Pharmacol. 2 (2007) 217–226.
[12] P. Lisboa, A. Taktak, The use of artificial neural networks in decision support in cancer: a systematic review, Neural Netw. 19 (2006) 408–415.
[13] A. Catic, L. Gurbeta, A. Kurtovic-Kozaric, S. Mehmedbasic, A. Badnjevic, Application of Neural Networks for classification of Patau, Edwards, Down, Turner and Klinefelter Syndrome based on first trimester maternal serum screening data, ultrasonographic findings and patient demographics, BMC Med. Genomics (2018) 11–19.
[14] L. Gurbeta, M. Maksimovic, E. Omanovic-Miklicanin, A. Badnjevic, E. Sejdic, A telehealth system for automated diagnosis of asthma and chronical obstructive pulmonary disease, J. Am. Med. Inform. Assoc. (2018).
[15] A. Badnjevic, L. Gurbeta, E. Custovic, An expert diagnostic system to automatically identify asthma and chronic obstructive pulmonary disease in clinical settings, Sci. Rep. 8 (1) (2018) 11645.
[16] A.J. Gabor, M. Seyal, Automated interictal EEG spike detection using artificial neural networks, Electroencephalogr. Clin. Neurophysiol. 83 (5) (1992) 271–280.
[17] F. Morgan, S. Cawley, B. Mc Ginley, S. Pande, L.J. Mc Daid, B. Glackin, J. Maher, J. Harkin, Exploring the evolution of noc-based spiking neural networks on FPGAs, in: International Conference on Field-Programmable Technology, IEEE, 2009, pp. 300–303.
[18] H. Zairi, M. Kedir Talha, K. Meddah, S. Ould Slimane, FPGA-based system for artificial neural network arrhythmia classification, Neural Comput. Appl. 32 (2020) 4105–4120.
[19] D. Selvathi, D. Nayagam, FPGA implementation of on-chip ANN for breast cancer diagnosis, Intell. Decis. Technol. 10 (2016) 1–12.
[20] M. Ahsan, M. Ibrahimy, O. Khalifa, M. Ullah, VHDL modeling of EMG signal classification using artificial neural network, J. Appl. Sci. 12 (3) (2012) 244–253.
[21] B. Karakaya, T. Kaya, A. Gulten, FPGA-based ANN design for detecting epileptic seizure in EEG signal, Balk. J. Electr. Comput. Eng. 6 (2) (2018) 15–19.
[22] H.G. Daoud, A.M. Abdelhameed, M. Bayoumi, FPGA implementation of high accuracy automatic epileptic seizure detection system, 61st IEEE International Midwest Symposium on Circuits and Systems (MWSCAS) (2018) 407–410.
[23] M.U. Saleheen, H. Alemzadeh, A.M. Cheriyan, Z. Kalbarczyk, R.K. Iyer, An efficient embedded hardware for high accuracy detection of epileptic seizures, in: 3rd International Conference on Biomedical Engineering and Informatics, 5, 2010, pp. 1889–1896.
[24] H. Rajaguru, S. Prabhakar, FPGA implementation of a wavelet neural network with particle swarm optimization learning for epileptic seizure detection, Int. J. Mech. Eng. Technol. 9 (2018) 1141–1154.
[25] S.M. Kueh, T.J. Kazmierski, Massively Parallel Bit-Serial Neural Networks for Fast Epilepsy Diagnosis: A Feasibility Study, Vol. 10, 2016.
[26] C. Elger, C. Hoppe, Diagnostic challenges in epilepsy: seizure underreporting and seizure detection, Lancet Neurol. 17 (2018) 211–212.
[27] I.E. Scheffer, S. Berkovic, G. Capovilla, M.B. Connolly, J. French, L. Guilhoto, E. Hirsch, S. Jain, G.W. Mathern, S.L. Moshé, D.R. Nordli, E. Perucca, T. Tomson, S. Wiebe, Y.-H. Zhang, S.M. Zuberi, ILAE classification of the epilepsies: position paper of the ILAE commission for classification and terminology, Epilepsia 58 (2017) 512–521.
[28] S.A. Gacio, Amplitude-integrated electroencephalography for neonatal seizure detection. an electrophysiological point of view, Arq. Neuropsiquiatr. 77 (2019) 122–130.
[29] V. Shah, E. von Weltin, S. Lopez, J.R. McHugh, L. Veloso, M. Golmohammadi, I. Obeid, J. Picone, The temple university hospital seizure detection Corpus, Front. Neuroinform. 12 (2018) 1–8.
[30] R. Kennett, Modern electroencephalography, J. Neurol. 259 (2012) 783–789.
[31] S. Ramgopal, S. Thome-Souza, M. Jackson, et al., Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy, Epilepsy Behav. 37 (2014) 291–307.
[32] L. Boubchir, B. Daachi, V. Pangracious, A review of feature extraction for EEG epileptic seizure detection and classification, 2017 40th International Conference on Telecommunications and Signal Processing (TSP) (2017) 456–460.

[33] P.N. Kumar, H. Kareemullah, EEG signal with feature extraction using SVM and ica classifiers, International Conference on Information Communication and Embedded Systems (ICICES2014) (2014) 1–7.
[34] Y. Li, W.-G. Cui, H. Huang, Y.-Z. Guo, K. Li, T. Tan, Epileptic seizure detection in EEG signals using sparse multiscale radial basis function networks and the fisher vector approach, Knowledge Based Syst. 164 (2019) 96–106.
[35] M. Nova, L. Vyslouzilova, Z. Vojtech, O. Stepankova, Towards computer supported search for semiological features in epilepsy seizure classification. World Congress on Medical Physics and Biomedical Engineering 2018, Springer, 2019, pp. 363–366.
[36] L. Wang, X. Long, R.M. Aarts, J.P. van Dijk, J.B. Arends, EEG-based seizure detection in patients with intellectual disability: which EEG and clinical factors are important? Biomed. Signal Process. Control 49 (2019) 404–418.
[37] Y. Li, X.-D. Wang, M.-L. Luo, K. Li, X.-F. Yang, Q. Guo, Epileptic seizure classification of EEGs using time-frequency analysis based multiscale radial basis functions, IEEE J. Biomed. Health Inform. 22 (2) (2018) 386–397.
[38] N.D. Schartz, S.K. Wyatt-Johnson, L.R. Price, S.A. Colin, A.L. Brewster, Status epilepticus triggers long-lasting activation of complement C1Q-C3 signaling in the hippocampus that correlates with seizure frequency in experimental epilepsy, Neurobiol. Dis. 109 (2018) 163–173.
[39] J. Peng, H. Wang, A. Jiang, Q. Xie, H. Wang, Quantitative calibration model for infrared spectroscopy using continuous wavelet transform combined with genetic algorithm, International Conference on Wavelet Analysis and Pattern Recognition (2012) 269–273.
[40] N. Beganovic, J. Kevric, D. Jokic, Identification of diagnostic-related features applicable to EEG signal analysis, Annual Conference of the Prognostics and Health Management Society (2018).
[41] M.G. Ruano, A.E. Ruano, On the use of artificial neural networks for biomedical applications, in: V. Balas, J. Fodor, A. Várkonyi-Kóczy, J. Dombi, L. Jain (Eds.), Soft Computing Applications – Advances in Intelligent Systems and Computing, vol. 195, Springer, Berlin, 2013.
[42] M. Akay, Neural network in processing and analysis of biomedical signals, in: Nonlinear Biomedical Signal Processing, Fuzzy Logic, Neural Networks, IEEE, 2000, pp. 69–97.
[43] F. Fan, W. Cong, G. Wang, General backpropagation algorithm for training second-order neural networks, Int. J. Numer. Method. Biomed. Eng. 34 (2017).
[44] A. Likas, A. Stafylopatis, Training the random neural network using Quasi-Newton methods, Eur. J. Oper. Res. 126 (2000) 331–339.
[45] S.J. Reddi, S. Kale, S. Kumar, On the convergence of Adam and beyond, Proceedings of International Conference on Learning Representation (2018).
[46] J. Duchi, E. Hazan, Y. Singer, Adaptive subgradient methods for online learning and stochastic optimization, J. Mach. Learn. Res. 12 (2011) 2121–2159.
[47] L. Bottou, Large-scale machine learning with stochastic gradient descent, in: Y. Lechevallier, G. Saporta (Eds.), Proceedings of COMPSTAT, 2010.
[48] H. Wang, W. Shi, C. Choy, Hardware design of real time epileptic seizure detection based on STFT and SVM, IEEE Access 6 (2018) 67277–67290.
[49] V. Geethu, S. Santhoshkumar, An efficient FPGA realization of seizure detection from EEG signal using wavelet transform and statistical features, IETE J. Res. (2018) 1–11.
[50] A. Saidi, S. Ben Othman, W. Kacem, S. Ben Saoud, FPGA implementation of EEG signal analysis system for the detection of epileptic seizure, International Conference on Advanced Systems and Electric Technologies (ICASET) (2018) 415–420.
[51] L. Marni, M. Hosseini, J. Hopp, P. Mohseni, T. Mohsenin, A real-time wearable FPGA-based seizure detection processor using MCMC, IEEE International Symposium on Circuits and Systems (ISCAS) (2018) 1–4.
[52] R. Jacob, K.P. Menon, Implementation of EEG feature extractor and classifier for seizure detection on FPGA, International Conference on Intelligent Computing and Control Systems (ICICCS) (2017) 307–310.
[53] J. Gotman, Automatic seizure detection: improvements and evaluation, Electroencephalogr. Clin. Neurophysiol. 76 (4) (1990) 317–324.
[54] J.L. Holt, T.E. Baker, Back propagation simulations using limited precision calculations, in: Seattle, USASeattle International Joint Conference on Neural Networks, Vol. 2, 1991, pp. 121–126. IEEE.
[55] D.F. Wolf, R.A. Romero, E. Marques, Using embedded processors in hardware models of artificial neural networks, in: V Simposio Brasileiro De Automação Inteligente, Vol. 9, 2001. Brazil.
[56] J. Zhu, P. Sutton, FPGA implementations of neural networks-a survey of a decade of progress. International Conference on Field Programmable Logic and Applications, Springer, 2003, pp. 1062–1066.
[57] M. Marchesi, G. Orlandi, F. Piazza, A. Uncini, Fast neural networks without multipliers, IEEE Trans. Neural Netw. 4 (1) (1993) 53–62.
[58] K.R. Nichols, M.A. Moussa, S.M. Areibi, Feasibility of floating-point arithmetic in FPGA based artificial neural networks, in: Proceedings of the 15th International Conference on Computer Applications in Industry and Engineering, San Diego, California, USA, Citeseer, 2002.
[59] L. Reyneri, Theoretical and implementation aspects of pulse streams: an overview, in: Proceedings of the Seventh International Conference on Microelectronics for Neural, Fuzzy and Bio-Inspired Systems, IEEE, 1999, pp. 78–89.
[60] H. Hikawa, A new digital pulse-mode neuron with adjustable activation function, IEEE Trans. Neural Netw. 14 (1) (2003) 236–242.
[61] A.G. Blaiech, K.B. Khalifa, M. Boubaker, M.H. Bedoui, LVQ neural network optimized implementation on FPGA devices with multiple-wordlength operations for real-time systems, Neural Comput. Appl. 29 (2) (2018) 509–528.
[62] V. Geethu, S. Santhoshkumar, An efficient FPGA realization of seizure detection from EEG signal using wavelet transform and statistical features, IETE J. Res. 66 (3) (2020) 315–325.
[63] A. Subasi, EEG signal classification using wavelet feature extraction and a mixture of expert model, Expert Syst. Appl. 32 (4) (2007) 1084–1093.
[64] S. Draghici, On the capabilities of neural networks using limited precision weights, Neural Netw. 15 (3) (2002) 395–414.