ASIC IMPLEMENTATION OF A PRE-TRAINED NEURAL NETWORK FOR ECG FEATURE EXTRACTION

Huruy Tekle Tefai, Hani Saleh, Temesghen Tekeste, Mahmoud Alqutayri, Baker Mohammad
(100049856, hani.saleh, temesghen.habte, mahmoud.alqutayri, baker.mohammad)@ku.ac.ae
System on Chip Center, Khalifa University

***Abstract*—The electrocardiogram signal (ECG), a record of electrical activity of the cardiac muscle, has been used in diagnosing many cardiopathies. Wearable devices equipped with readout sensors and circuits can be used to record and process weak ECG signals. In this paper, a pre-trained neural network was implemented for detecting the QRS feature of an ECG signal, which is crucial for auto-diagnostic of various cardiopathies. To take advantage of the fast evolution of artificial intelligence and its ability to find non-linear relationships, neural network based feature extraction of ECG signals for wearable devices was explored and tested using ASIC implementation flow. Firstly, a high-level simulation was carried out in MATLAB and verified with test data obtained from PhysioNET database. Recurrent neural network (RNN) MLP was created and trained using the data obtained from PhysioNET database. A high-level performance evaluation was carried out using the same network for P and T wave extraction. The weight and bias matrices obtained from the high-level trained network in MATLAB were used in the design of the hardware. An accuracy of 96.55% was achieved in the hardware implementation of the network.**

Keywords— ASIC, Hardware implementation, Neural Network, Multi-layer perceptron, ECG, Electrocardiogram.

***Index Terms*—QRS, ECG, Feature Extraction, RTL, Neural Network**

I. INTRODUCTION

Electrocardiogram (ECG) is a record of the electrical activity of the heart captured using different techniques. These waveforms are used in diagnosing different cardiopathies. Moreover, several applications have been proposed and implemented where ECG signals were used in emotion recognition and biometric identification. As it is impractical for a cardiologist to interpret very long sequence of data obtained over a long period of time, many systems were developed to automate the process of translating the waveforms to extract useful information. Several ECG signal acquisition techniques such as the standard 12 electrode ECG, Smart Chair using capacitively coupled electrodes [1], wearable devices, and RF based ones were used to capture the electrical activity of the heart. To track and monitor abnormal heart rhythms and cardiac symptoms over a longer period of time many wearable devices were developed. However, as the ECG signal captured from wrist tend to be weak, we need a powerful processing blocks to eliminate or suppress the noise originating from artifacts. For wearable devices, power efficient processors are also required to extract features that can be used for classification stage. In the process of ECG signal acquisition, different types of noise and artifact contaminate captured ECG signals. Various techniques are applied to reduce the effect of noise in the pre-processing stage of the system. Common types of noise are power line interference (a signal in the frequency of 50 or 60 Hz), baseline wonder noise (low-frequency 0.15 up to 0.3 Hz), electrode contact noise, electrode motion artifacts, muscle contractions, electrosurgical noise, and instrumentation noise. Noise and artifact suppression is performed with a series of sophisticated filtering blocks [2]. Once the signal's SNR is improved with the pre-processing stage a feature extraction technique is applied to obtain the distinctive ECG features such as the P-QRS-T complex features, statistical features, morphological features, and wavelet features. Among the most powerful algorithms for ECG QRS complex detection are Pan-Tompkins algorithm [3] [4] derivative [3] [5], digital filters [3], wavelet transform [5], Hilbert-Huang transform [7], neural networks [12]. The features extracted are then used in the classification stage of the system. The major categories of these classifiers are artificial neural networks (ANNs), LDA, k nearest neighbor (kNN), support vector machine (SVM), decision tree (DT), and Bayesian classifiers. To eliminate the feature extraction stage, a new End-to-End ECG Classification with Raw Signal Extraction and Deep Neural Networks (DNN) was proposed in [8], where DNN is used for both feature extraction and classification based on aligned heartbeats. That is raw ECG signal is fed to the proposed system and beat by beat classification decision is obtained. Though (ANN) is not a new concept, it has only started recently to drive the rapid development of many applications with emerging advancements in Graphical Processing Unit (GPU), availability of massive data and advanced chip manufacturing technologies. However, ANNs highly demanding computational complexity and large area/power requirement of its hardware implementation has made it impractical for low power real time applications [9]. Moreover, the extremely slow training phase is another major challenge with many applications of ANN especially in edge devices and mobile systems. Application of deep learning neural network for ECG feature extraction started growing recently in terms of accuracy and complexity with the increased availability of neural network hardware implementations [10]

978-1-7281-3320-1/20/$31.00 ©2020 IEEE

[11] . Two-level convolutional neural network (CNN) based QRS complex detection was proposed in [10]. ECG signal variations, avoids the need for hand-crafted features, improves accuracy in detecting features, and reduces computational cost. With the advancement of deep learning neural networks, subject independent ECG feature extraction and classification can be achieved with accuracy equivalent to state-of-the-art-techniques [12].

In this work, hardware implementation of QRS detection was explored using shallow architecture neural network. To verify the functionality of the proposed system, a database of annotated ECG signals training data was obtained from PhysioNET database [13], a subset was used to train the network and another set was randomly selected for evaluation to generate the classification accuracy. The outline of the rest of the paper is as follows: section II presents the high-level design, section III discusses the hardware architecture, section IV presents the synthesis, place and route results, finally the paper is concluded in section V.

II. SYSTEM LEVEL DESIGN

To demonstrate that shallow neural networks can achieve accuracies comparable to those achieved by deep networks while learning complex relationships [14], a custom made shallow neural network was created in MATLAB with one recurrent layer and 200 inputs layer to capture a window of samples that is just enough to cover one beat at 250 Hz sampling frequency. The number of samples within the window is dependent on sampling frequency of the ADC and hence the size of the input of the network is set based on the sampling frequency. Moreover, a delayed version of the input sample and input to the recurrent layer were used to help the system to recognize past features. The architecture of the designed neural network is shown in Fig. 1.

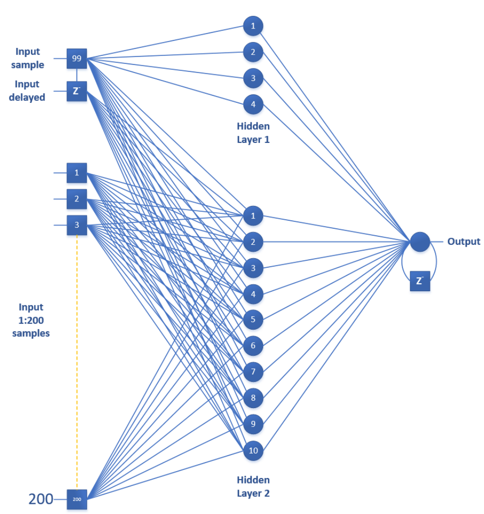


Fig. 1: Neural Network architecture used.

To train the network, labeled ECG data extracted from PhysioNET QT database [13] were used where each sample is labeled. The dataset is about 900 sec ECG recordings from a total of 105 patients. Every sample of the ECG signals in the database were labeled as QRS, P and T wave as shown in Fig. 2. The system was trained with ECG recording of those subjects whose modified limb lead II (ML II) reading was measured.

In this work, the network created was trained to identify the QRS samples based on the training of the labeled data. To distinguish the QRS samples from the rest of the other samples labeled P and T, a new vector was created with value '1' if corresponding label of a sample is QRS and '0' otherwise. To verify correct functionality of the network and compare the accuracy obtained, the records from those subjects with ML II lead were divided randomly into test and train data with training data being 70% of the available dataset and test data being 30%.

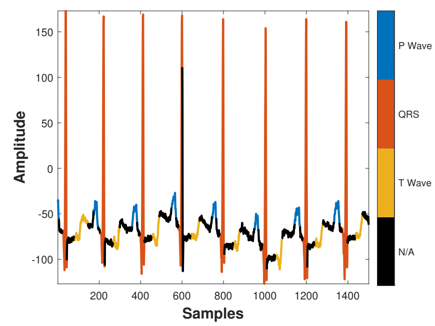


Fig. 2: Labeled ECG records from PhysioNET.

A Nguyen-Widrow layer initialization function was used to initialize the weights and biases during training phase. Moreover, the network was trained twice to improve the accuracy obtained in the first phase.

The signal was bandpass filtered before it was passed into the training phase to remove baseline wondering and high frequency interference.

A. Preparing Input Signal

A matrix of two vectors, one for the ECG sample and another for the 200 samples window surrounding the input sample was created along with the corresponding annotated labels. This was fed into the network for training as input (See matrix below).

$$\begin{bmatrix} \text{ECG(i)} \\ \begin{bmatrix} \text{ECG(i-100)} & \ldots & \text{ECG(i)} & \ldots & \text{ECG(i+99)} \end{bmatrix} \end{bmatrix}$$

where i is mid sample index of the running window

A labels vector (Shown in blue in Fig. 3) was created with value '1' for indexes with 'QRS' labeled samples and '0' otherwise. This vector was then fed to the network as the desired output in the training phase.

The ECG samples vector and the 200 samples window array around every sample were fed to the network with the labeled output vector in order for the network to learn the pattern of the QRS complex samples. After the network was trained,

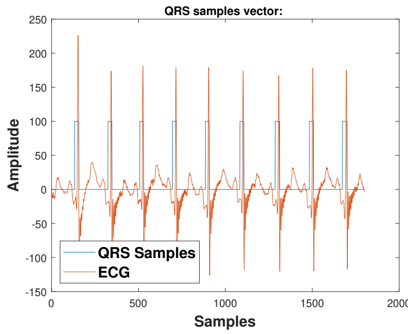


Fig. 3: Labels vector created to distinguish QRS samples

test data from different subjects were fed to the network to extract QRS samples as shown in Fig. 4 for one subject. It is clear from the plot that the network was able to detect QRS samples successfully from the test waveform fed. The accuracy varies from subject to subject ranging from 91% to 99%, but it achieved an average accuracy of 97%. The actual output of the network gives values in the range of 0 to 1 with those values close to 1 implying QRS sample. Fig. 5 shows output of the network with threshold set to differentiate QRS samples and improve accuracy. The accuracy obtained for one

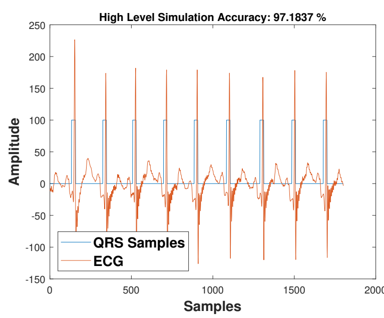


Fig. 4: High level accuracy 97.2%.

subject is shown in Fig. 4 where the QRS samples are plotted against the original ECG signal for one subject.

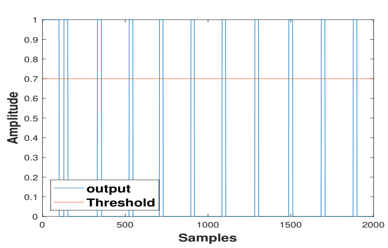


Fig. 5: Threshold.

Polynomial approximation was then used to evaluate the nonlinear sigmoid activation function (1) with degree two (2). A look up table could be used instead to reduce the computational complexity requirement of the polynomial approximation method.

$$y = \frac{1}{1+e^{-x}} \tag{1}$$

$$y) = \begin{cases} 0, & \text{if } x \leq -6 \\ 0.1998 + 0.0703 * x + 0.0063 * x^2, & \text{if } -6 < x \leq -4 \\ 0.3803 + 0.1756 * x + 0.0215 * x^2, & \text{if } -4 < x \leq -2 \\ 0.5020 + 0.2727 * x + 0.0406 * x^2, & \text{if } -2 < x \leq 0 \\ 0.4982 + 0.2723 * x - 0.0404 * x^2, & \text{if } 0 < x \leq 2 \\ 0.6179 + 0.1768 * x - 0.0217 * x^2, & \text{if } 2 < x \leq 4 \\ 0.7968 + 0.0718 * x - 0.0064 * x^2, & \text{if } 4 < x \leq 6 \\ 0.9923 + 0.0021 * x - 0.0001 * x^2, & \text{if } 6 < x \leq 8 \\ 1, & \text{if } x > 8 \end{cases} \tag{2}$$

B. High level Performance Result

In addition to extracting the QRS samples, the network was tested for accuracy (Acc), positive predictivity rate (PPR) and sensitivity (Sen) eqns (3) - (5) of extracting P and T wave features for one subject. A summary of the result is demonstrated in Table I, where TP, FP, FN stand for true positive, false positive and false negative.

$$Sen(\%) = \frac{TP}{TP+FN} * 100 \tag{3}$$

$$PPR(\%) = \frac{TP}{TP+FP} * 100 \tag{4}$$

$$Acc(\%) = \frac{TP}{TP+FN+FP} * 100 \tag{5}$$

TABLE I: High Level Performance Result

Percentage	**Train Data**			**Test Data)**		
	Acc	**Sen**	**PPR**	**Acc**	**Sen**	**PPR**
QRS	97.18	96.98	82.81	96.51	95.76	79.47
P wave	97.67	98.23	88.57	95.71	91.17	84.34
T wave	94.04	94.62	86.08	85.04	79.87	74.83

The drop in the P and T wave percentage of sensitivity, accuracy and positive predictivity shown in Table I is due to the distinct feature of QRS complex. Very high accuracy can be achieved with deeper network as demonstrated in [10] where the accuracy and sensitivity were greater than 99% for most of the records used from MIT-BIH-AR database. However, the hardware requirement of such algorithms will be extremely high.

III. Hardware Implementation

The high-level model was modeled in register transfer level (RTL) code in Verilog for the inference implementation using weights and biases obtained in the high-level training. However, the conversion process of the weights and biases from floating-point arithmetic to a fixed precision binary needed for the hardware implementation caused a drop in the accuracy of the RTL model when the parameters were represented with 8 bits. To achieve better accuracy and improve the error introduced in the conversion process, 32bit fixed-point two's complement representation was used with 12 bits assigned for fractions and 20 bits for the integer part.

The overall block diagram of the hardware implementation of the work is shown in Fig. 6. The implementation has three modules for each layer and a shift register for the input samples. The memories for the weights and biases are encapsulated within each layer's module as shown in Fig. 6. The inference starts execution only when the first 200 samples are captured in the shift register and the infer_ready signal is set.

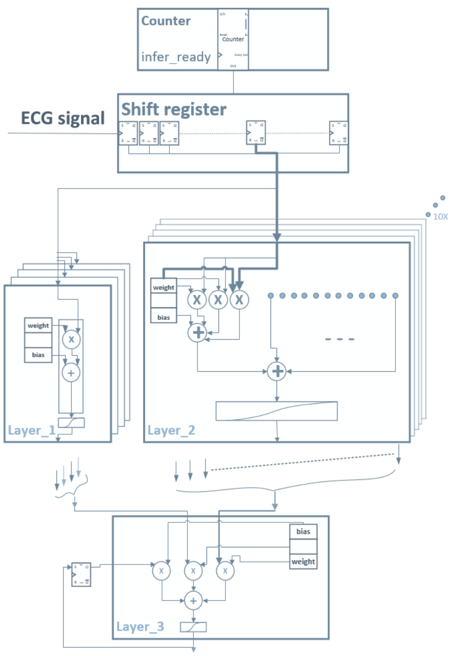


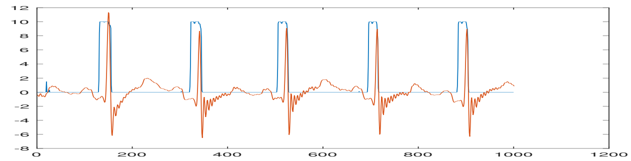
Fig. 6: HDL Block Diagram.

ECG samples were captured serially into the top module where a shift register is used to shift and store ECG vector of 200 samples with the middle element being the current sample. The weights and biases obtained from the training phase were converted to binary and stored in memory. Similarly, the coefficients of the polynomials used in the approximation of the activation function were also converted to binary.

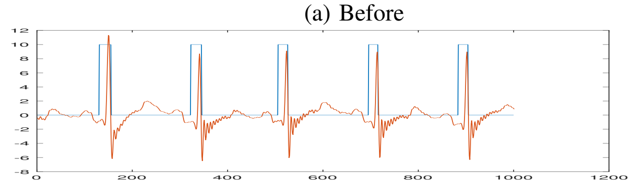
The RTL code was simulated for functional verification using a testbench. The output of the network was converted back to decimal to compare it with the original signal ECG signal in MATLAB. As shown in Fig. 7a and 7b before and after applying threshold respectively, the output is almost an exact duplicate of the QRS samples vector with accuracy of 96% for the subject considered for test. This high accuracy of the RTL design is achieved using one subject data randomly selected from the test data and with basic neural network architecture implying deeper networks can give improved accuracy but at a cost of computational complexity as it grows very fast with the depth of the network [15].

IV. SYNTHESIS, PLACE AND ROUTE

The RTL code was synthesized using Synopsys DC compiler and a gate level netlist was used for layout in Synopsys



(a) Before



(b) After

Fig. 7: Hardware Performance output before and after applying threshold

ICC compiler. The HDL design was synthesized and layout using a standard-cell library designed for Global Foundries 65nm technology low-power process. Hardware implementation of the design is summarized in Table II and III corresponding to power and area report respectively.

TABLE II: Power Report

$Category$	$Value$
$InternalPower(mW)$	14.45
$NetSwitchingPower(mW)$	1.58
$LeakagePower(\mu W)$	184.71
Total Power(mW)	16.03

TABLE III: Area Report

$Category$	$Value$
$Combinational Area\ (\mu m)^2$	11,789,489.24
$Non-Combinational Area\ (\mu m)^2$	540,115.55
$Buf/Invarea\ (\mu m)^2$	278,356.33
Total area$(\mu m)^2$	12,329,604.80

V. CONCLUSION

In this work a neural network architecture for ECG signal QRS complex detection was created and trained with focus on its hardware implementation. The pre-trained neural network was implemented in hardware with coefficients for biases and weights obtained from MATLAB high level model. It was demonstrated that shallow networks can achieve accuracy comparable to state-of-the-art work. In addition to QRS complex detection, a high-level performance evaluation was performed to test the designed network for P and T wave features extraction capability. The RTL design functionality was verified with a testbench.

REFERENCES

[1] A. Aleksandrowicz and S. Leonhardt, “Wireless and Non-contact ECG Measurement System - the Aachen SmartChair,” Acta Polytech. J. Adv. Eng., vol. 47, no. 4–5, pp. 68–71, 2007.

[2] S. Kaplan Berkaya, A. K. Uysal, E. Sora Gunal, S. Ergin, S. Gunal, and M. B. Gulmezoglu, “A survey on ECG analysis,” Biomed. Signal Process. Control, vol. 43, pp. 216–235, 2018.

[3] M. Darweesh, T. Habte, H. Saleh, B. Mohammad, and M. Ismail, “Novel logarithmic ECG feature extraction algorithm based on pan and tompkins,” Midwest Symp. Circuits Syst., no. October, pp. 1–4, 2017.

[4] J. Pan and W. J. Tompkins, “A Real-Time QRS Detection Algorithm,” IEEE Trans. Biomed. Eng., vol. BME-32, no. 3, pp. 230–236, 2007.

[5] N. M. Arzeno, Z. De Deng, and C. S. Poon, “Analysis of first-derivative based QRS detection algorithms,” IEEE Trans. Biomed. Eng., vol. 55, no. 2, pp. 478–484, 2008.

[6] T. Tekeste et al., “Adaptive ECG interval extraction,” Proc. - IEEE Int. Symp. Circuits Syst., vol. 2015-July, no. 1, pp. 998–1001, 2015.

[7] B. Lenka, “Time-frequency analysis of non-stationary electrocardiogram signals using Hilbert-Huang Transform,” 2015 Int. Conf. Commun. Signal Process. ICCSP 2015, pp. 1156–1159, 2015.

[8] S. S. Xu, M. W. Mak, and C. C. Cheung, “Towards End-to-End ECG Classification with Raw Signal Extraction and Deep Neural Networks,” IEEE J. Biomed. Heal. Informatics, vol. PP, no. 8, p. 1, 2018.

[9] D. A. Nguyen, H. H. Ho, D. H. Bui, and X. T. Tran, “An Efficient Hardware Implementation of Artificial Neural Network based on Stochastic Computing,” NICS 2018 - Proc. 2018 5th NAFOSTED Conf. Inf. Comput. Sci., pp. 237–242, 2019.

[10] Y. Xiang, Z. Lin, and J. Meng, “Automatic QRS complex detection using two-level convolutional neural network,” Biomed. Eng. Online, vol. 17, no. 1, pp. 1–18, 2018.

[11] C. M. A. Hedayat, H. Chia, X. Zhou, C. Richard., “Supervised ECG Interval Segmentation Using LSTM Neural Network,” Int’l Conf. Bioinforma. Comput. Biol. BIOCOMP’18, p. 71, 2018.

[12] M. Sarlija, F. Jurisic, and S. Popovic, “A convolutional neural network based approach to QRS detection,” Int. Symp. Image Signal Process. Anal. ISPA, no. Ispa, pp. 121–125, 2017.

[13] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals (2003).

[14] Ba, Jimmy, and Rich Caruana. ”Do deep nets really need to be deep?.” Advances in neural information processing systems. 2014.

[15] Livni, Roi, Shai Shalev-Shwartz, and Ohad Shamir. ”On the computational efficiency of training neural networks.” Advances in neural information processing systems. 2014.