# PYNQ-Z2 Accelerated Cardiac Arrythmia Classifier – A Deep Learning based Approach

**Soumyashree Mangaraj & Abhyarthana Bisoyi**

School of Electronics Sciences
Odisha University of Technology and Research, Bhubaneswar, Odisha

**Abstract**

Electrocardiogram (ECG) signals are the primary characteristics of a physically fit human body; diagnosing cardiovascular diseases (CVDs) automatically using computer-aided tools has caught a significant attention in the current medical scenario. In recent times with the rapid growth of smart health-care system, IoT enabled edge devices make it possible for early diagnosis of diseases with resource constraint devices. PYNQ- a Python productivity on Xilinx platform, based hybrid convolution neural network (CNN) architecture has been proposed in this work for classifying arrythmia in reference to AAMI EC57 standard. A comparative investigation is conducted on volume of trainable parameters of the architecture, and accuracy of ECG classification. A customized FPGA IP for the proposed hybrid 1-D CNN architecture has been generated using Vitis High Level Synthesis (HLS) tool that would be implemented on PYNQ-Z2 board.

**Keywords:** ECG, CNN, PYNQ-Z2, Vitis HLS, IoT

## 1. Introduction

Ubiquitous networking with edge devices, capable of computing real-time data with reduced latency, low-power, and low-data rate builds the fundamental frame work of IoT infrastructure. A fast and reliable network connection helps in a consistent data transmission from the data source to the user with IoT-edge devices for further computation. The data collected from sensors and actuators are massive, and require highly efficient architecture to process the data. Data security and user's privacy will be the overriding concern in IoT enabled applications. In our article we will elucidate on IoMT: IoT for Medical application, in particular for cardiovascular disease (CVD) classification on PYNQ framework.

The study done by researchers are mostly based on software simulated deep learning (DL) models framed with CNN algorithm. Only a few studies in the literature have explored using resource-constrained FPGA acceleration for classifying ECG signals with CNNs. This approach focuses on detecting abnormalities in heartbeats efficiently under hardware limitations. This work will facilitate in developing an IoT enabled

healthcare system which will be deployable in remote areas. The proposed work aims to demonstrate that FPGA acceleration offers a state-of-the-art solution for implementing CNNs in heart disease classification. It highlights the potential of FPGAs to efficiently support deep learning models for medical diagnostics. PYNQ frame-work can be used as an IoT edge platform to communicate with cloud in conjunction with appropriate wireless connectivity module which will be an efficient architecture for IoMT application. This prototype will be used in remote locations as an edge device for medical applications.

## 2. Related Works

Electrocardiogram (ECG) signals are electrical voltages generated by the heart and recorded from the body's surface using an electrocardiograph. These signals reflect the heart's electrical activity and are crucial for diagnosing various cardiac conditions. The alarming numbers of CVD patients need to be diagnosed by exact abnormality finding from ECG signals. We have planned to carry our research work by developing a frame work to acquire real time ECG signal, accelerate it on hardware for classifying various arrythmia conditions at the edge node, and then transmitting it to some authorized medical experts securely via IoT empowered devices for further medication. Security becomes the prime challenge when the patient's data move through various healthcare professionals using inter-networked devices. In our architecture we will try to ensure secured user's information communication.

The acquired ECG signal need to be pre-processed for making it noise free by following necessary filtering steps. The filtered ECG signal is furthered analyzed to extract various morphological and temporal features to be used in machine learning (ML) based models. These physical feature extraction methods may lead to loss of a few salient information of ECG signal. The arrythmia classification can be made automatic and hand-crafted error free by employing deep learning (DL) algorithms-based models such as convolution neural network (CNN). In our research work we have proposed a hybrid CNN architecture which can be implementable on reconfigurable hardware platform- PYNQ, and further can be extended to be used for IoMT application. FPGA being a resource compelled target floor, researchers are working on resizing the huge data aiming to use less storage by introducing compressive sensing (CS) and sparsity by various pruning techniques. The complexity of deep neural network (DNN) is also getting modified to achieve efficient resource utilization with architectures like quantized CNN (QCNN) [1], Binarized Neural Networks (BNNs) [2], LUT Net [3], Ternary Weight Networks (TWNs) [4], and Trained Ternary Quantization (TTQ).

The paper is organized in the following sequence. The proposed methodology is discussed in section II. It discusses the model proposed for arrythmia classification, and about its hardware implementation. The simulated results and discussion are presented in section III. Section IV briefs the conclusion and future research work directions.

## 3. Proposed Methodology

The research problem statement is classified into three sections: model preparation, hardware acceleration, and real-time data acquisition and computation. The processing of real time acquired ECG signal, and arrythmia classification will be done at the user end. The further medication will be done by the medical expert by sharing the data through cloud in a secured manner.

### A. Dataset used

The model proposed is trained with MIT-BIH Arrythmia Dataset [5,12], and will classify five types of heart beats: normal beat (N), supraventricular ectopic beat (S), ventricular ectopic beat (V), fusion beat (F) and unclassified beat (Q) as recommended by the Association for the Advancement of Medical Instrumentation (AAMI) [6]. The dataset used have total 55,126 beats with 256 normalized samples each. The data labels are encoded as, N- ‘0’, S- ‘1’, V- ‘2’, F- ‘3’ and Q- ‘4’ for five class classification. The encoded data are then split into train and test dataset in 85:15 ratio. The sample size of training and testing data are 46,857 and 8269 respectively. The model is validated by finding the validation accuracy and loss with 15% of the training dataset.

### B. Software Design

A hybrid CNN architecture is proposed by the authors which will be a light weight model to be accelerated on PYNQ-Z2 platform. The architecture is shown in Figure 1.

The input layer of the model has a dimension of 256×1. The bounded region with convolution followed by ReLU (Rectified Linear Unit) and max-pooling layers are iterated for 4 times. In each iteration the kernel size and number of filters are gradually decreasing aiming to reduce the size of weighted parameters, and it is tabulated in Table 1.

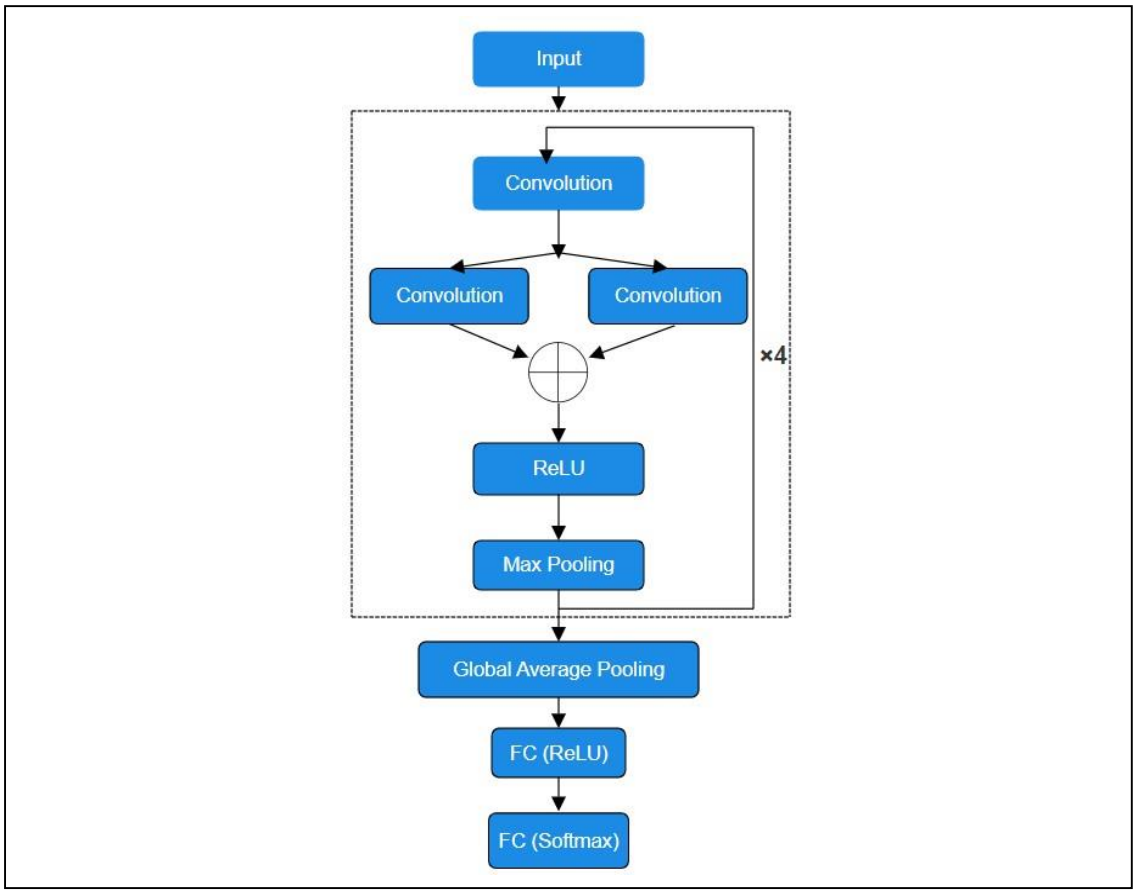


**Figure 1Proposed Hybrid CNN Model**

**Table 1 Filter summary of iterative convolution layers**

| Iteration | Kernel size | No. of filters | Stride |
|---|---|---|---|
| 1st | 7 | 32 | 1 |
| 2nd | 7 | 16 | 1 |
| 3rd | 5 | 8 | 1 |
| 4th | 3 | 4 | 1 |

The outputs of the first convolution layer are parallelly convoluted twice, and then the results are summed. We are proposing the parallel convolution which will be a faster, hardware efficient implementation as compared to the cascaded architecture of traditional CNN models. The summed value is moved through ReLU activation, and then subsampled in max-pooling layer with pooling size and stride as 2. The outputs after the 4th iteration, has been pooled through a global-average pooling (GAP) layer to reduce the input size before feeding to the fully-connected (FC) layers. There are two FC layers with 10 neurons in the first and 5 neurons in the second. The non-linear activation function used is ReLU in first FC layer to limit the maximum range of output, and soft-max in output layer to give

the probability of maximum classification accuracy. The numbers of weighted layers and their trainable parameters are summarized in Table 2.

| Layer type | Input | Output | Trainable Parameters |
|---|---|---|---|
| Convolution 1 | 256×1 | 256×32 | 256 |
| Convolution 2 & 3 | 256×32 | 256×32 | 7200 (for each layer) |
| Convolution 4 | 128×32 | 128×16 | 3600 |
| Convolution 5 & 6 | 128×16 | 128×16 | 1808 (for each layer) |
| Convolution 7 | 128×16 | 64×8 | 648 |
| Convolution 8 & 9 | 64×8 | 64×8 | 328 (for each layer) |
| Convolution 10 | 64×8 | 32×4 | 100 |
| Convolution 11 & 12 | 32×4 | 32×4 | 52 (for each layer) |
| Dense Layer1 | 4 | 10 | 40+10 (bias) |
| Dense Layer2 | 10 | 5 | 50+5 (bias) |

**Table 2 Weighted layers of proposed CNN Model**

## C. Hardware framework

PYNQ is an open-source project from AMD-XILINX. The target board for our proposed system will be the PYNQ- Z2 board which will be programmed through Jupyter Notebook. The CNN architecture is validated by doing C/C++ simulation in Vitis HLS tool, which further creates complex FPGA algorithms by synthesizing a C/C++ function into RTL. The PYNQ-Z2 board will be programmed through Python using hardware libraries and generated bit-stream file. The proposed hybrid CNN architecture is trained with Python language in Kaggle P100 GPU- 13GB RAM, and the model is stored in .h5 file format. This pretrained CNN model will be accelerated on PYNQ-Z2 framework through Vitis HLS, Vivado and Jupyter interface. The pre-trained 1-D CNN model is replicated in C++ for one layer of convolution, max-pooling, ReLU and output FC layer. The model replication in C++ is shown in Fig. 2. A FIFO queues the data samples, which are transmitted in sequential manner in to convolution layer. The kernel size is taken 7

```
1 # define FILTERS 32
2
3 void cnn
4 (float inp_data [ INP_ROWS ][ INP_COLS ],
5 float prediction [ Arrythmia_class ])
6 {
7 /* ******* Pre - processing the inp_data . ******* */
8 // Normalization and padding .
9
10 /* ******* Clone the normalized and padded image . ******* */
11 /*
12 * Clone the normalized and padded image in order to
13 * have an image for each parallel execution ( for each filter ).
14 */
15 /* ******* Parallel executions start here . ******** */
16
17 /* Dataflow section with streams used to transfer data between tasks :
18 * - convolution_layer ;
19 * - max_pooling_layer ;
20 * - dense_layer ;
21 * - dense_layer_softmax .
22 */
23 }
```

Figure 2 Source code of model replication in C++

with 32 numbers of filters.

### D. System architecture

The system architecture with complete work flow is presented in Fig. 3. The proposed hybrid CNN model's acceleration on PYNQ-Z2 board, and its implementation for IoMT application is illustrated in the architecture. The real time ECG signal will be acquired on PYNQ board through an ADC, and will be pre-processed to generate a noise-free signal. The signal will be classified by the pre-trained model, and it will be communicated to the medical expert through PYNQ as an edge device.

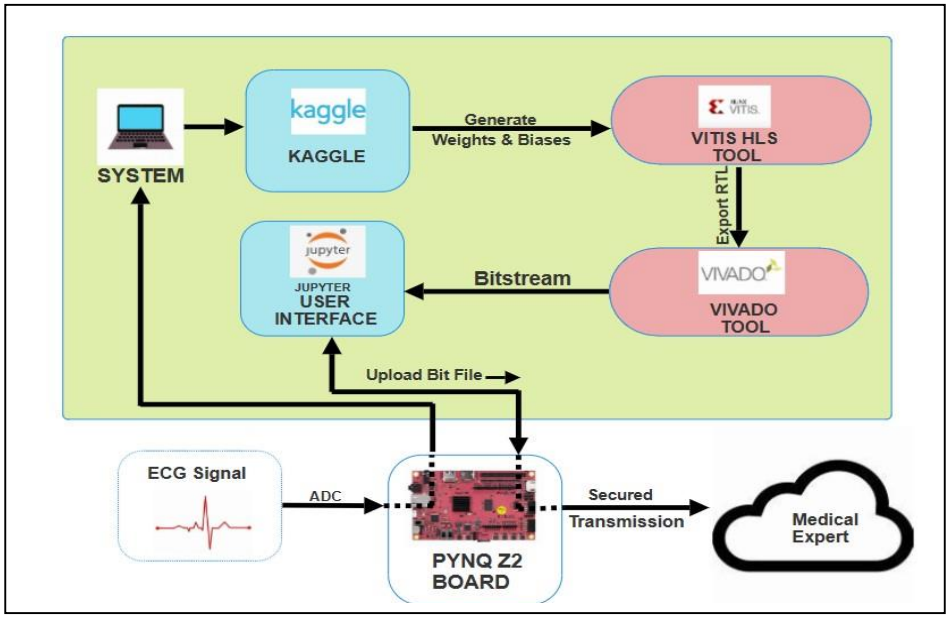


**Figure 2 System Architecture**

## 4. Simulation and Discussion

### A. Training and Evaluation Metrics

The MIT-BIH Arrythmia dataset is a highly imbalanced dataset, still without using data augmentation techniques the proposed architecture performs comparatively better as mentioned in Table 2. The proposed hybrid 1-D CNN structure achieves training and validation accuracies of 95.82% and 92.72%, respectively, after 200 epochs with a batch size of 50. These results are illustrated in the accuracy plot shown in Fig. 4. The categorical cross-entropy loss function computes the loss for multi-class single label ECG dataset, and is plotted in loss plot of Fig.4.

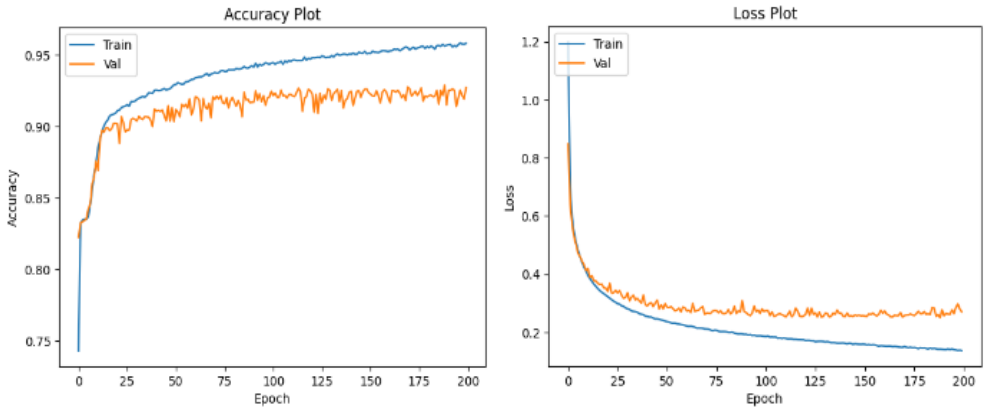


**Figure 4 1-D CNN model plots of accuracy and loss**

The confusion matrix for heartbeat classification on the test-set is presented in Fig. 5. The beats S, V and F classification accuracy can be improved by improving the size of data samples. We will introduce the other dataset available to handle the mis-classification in future.

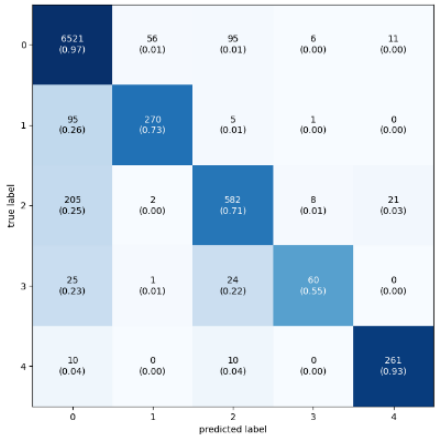


**Figure 5 Confusion Matrix on test set**

The number of trainable weighted parameters of the proposed model are 23,725, and it is less as compared to 29, 263 [10], hence convincing as a light weight CNN model. The accuracy comparison is presented in Table 3.

**Table 3 Accuracy comparison of arrythmia classification**

| Work | Method | Accuracy (%) |
|---|---|---|
| Kachuee et al. [7] | Deep residual CNN | 93.4 |
| Aphale et al. [8] | ArrhyNet- 1-D CNN | 92.73 |
| HA et al. [9] | CNN | 90.4 |
| Tiwari et al. [10] | Aug+Deep Residual CNN | 95.6 |
| Proposed work | Hybrid CNN | 95.82 |

## B. Hardware acceleration and Implementation Results

The weights and biases for convolution and dense layers are extracted from .h5 file of the trained hybrid 1-D CNN model, and saved in the header files in 32-bit floating-point format to create C++ project in Vitis HLS tool. The header files and .cpp file are written for convolution, pooling, ReLU activation and dense layers. The synthesis is done by interfacing the directives in s_axilite mode. The C-simulation is done with in.dat and out.dat testbench files. The RTL file will be generated after successful completion of synthesis and simulation in Vitis HLS tool. In Fig. 6 the IP module for the proposed 1-D CNN model is presented. The Vivado generated bit-stream file and the hardware files are transported to Jupyter interface by Overlay function for implementing the model on PYNZ-Z2 board.

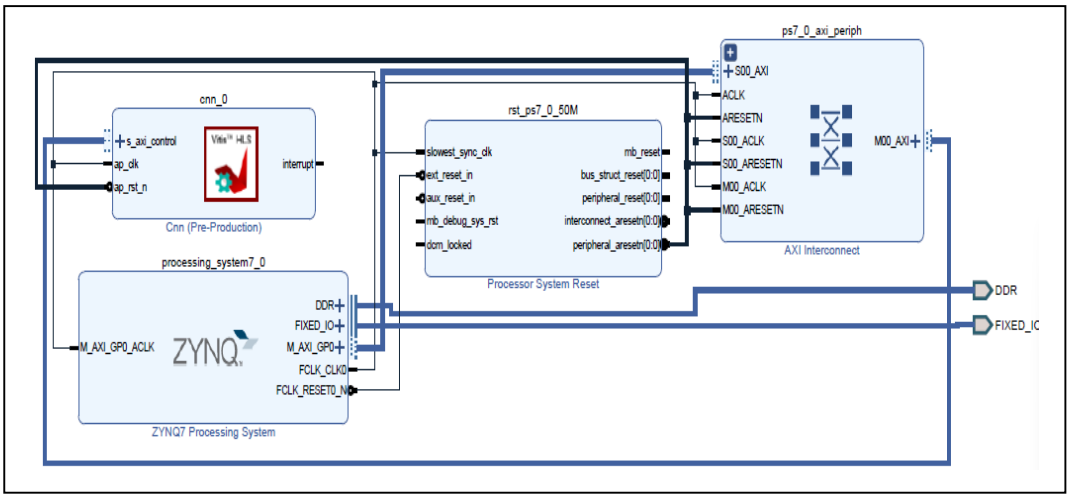


*Figure 6 FPGA IP of 1-D CNN*

## 5. Conclusion

In this work an efficient light weight 1-D CNN architecture to classify five ECG beats is proposed. The DL architecture design is implemented on hardware platform like PYNQ framework. The proposed architecture has a smaller number of trainable parameters, that will improve the memory requirement of storing weights and biases. The authors would introduce pruning to prepare a quantized CNN architecture such that the hybrid architecture will be implemented on IoT edge device to serve IoMT application in future.

## REFERENCES


[1] Z.Qi, J.Cao,Y.Zhang, S. Zhang and Q. Zhang, "FPGA Implementation of Quantized Convolutional Neural Networks," in IEEE 19th International Conference on Communication Technology, 2019.

[2] I. Hubara, M. Courbariaux, D. Soudry, R. El-Yaniv and Y. Bengio, "Binarized Neural Networks," in Advances in Neural Information Processing Systems 29 (NIPS), 2016.

[3] E. Wang, J. J. Davis, P. . Y. K. Cheung and G. A. Constantinides, "LUTNet: Rethinking Inference in FPGA Soft Logic," in IEEE 27th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM), 2019.

[4] F. Li and B. Liu, "Ternary weight networks," in arXiv preprint arXiv:1605.04711, 2016.

[5] "MIT-BIH Supraventricular Arrhythmia Database," Available online at https://physionet.org/content/svdb/1.0.0/.

[6] A. J. Prakash and S. Ari, "AAMI Standard Cardiac Arrhythmia Detection with Random Forest Using Mixed Features," in IEEE 16th India Council International Conference (INDICON), 2019.

[7] M. Kachuee, S. Fazeli and M. Sarrafzadeh, "ECG Heartbeat Classification: A Deep Transferable Representation," in IEEE International Conference on Healthcare Informatics, 2018.

[8] A. S. Siddhasanjay, E. John and . T. Banerjee, "ArrhyNet: A High Accuracy Arrhythmia Classification Convolutional Neural Network," in IEEE International Midwest Symposium on Circuits and Systems (MWSCAS), 2021.

[9] H. A. Deepak and T. Vijaykumar, "ECG Beat Classification using CNN," in IEEE International Conference on Data Science and Information System (ICDSIS), 2022.

[10] S. Tiwari and P. R. Muduli , "Convolutional Neural Network-based ECG Classification on PYNQ-Z2 Framework," in IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), 2022.

[11] Mohammad Kachuee, Shayan Fazeli, and Majid Sarrafzadeh, ECG Heartbeat Classification: A Deep Transferable Representation. In 2018 IEEE International

[12] Conference on Healthcare Informatics

G.B. Moody and R.G. Mark. 2001. The impact of the MIT-BIH Arrhythmia Database. IEEE Engineering in Medicine and Biology Magazine 20, 3 (2001), 45–50.