IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 14 August 2022, accepted 11 October 2022, date of publication 19 October 2022, date of current version 27 October 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3215665

RESEARCH ARTICLE

Lightweight Shufflenet Based CNN for Arrhythmia Classification

HURUY TESFAI[1], HANI SALEH[1], (Senior Member, IEEE), MAHMOUD AL-QUTAYRI[1], (Senior Member, IEEE), MOATH B. MOHAMMAD[2], TEMESGHEN TEKESTE[1], AHSAN KHANDOKER[3], AND BAKER MOHAMMAD[1], (Senior Member, IEEE)

[1]Department of Electrical Engineering and Computer Science, System on Chip Center, Khalifa University, Abu Dhabi, UAE
[2]Texas College of Osteopathic Medicine, University of North Texas Health Science Center (UNTHSC), Fort Worth, TX 76107, USA
[3]Department of Biomedical Engineering, Healthcare Engineering Innovation Center, Khalifa University, Abu Dhabi, UAE

Corresponding author: Baker Mohammad (baker.mohammad@ku.ac.ae)

This work was supported in part by the Khalifa University Competitive Internal Research Award (CIRA) under Award CIRA-2020-053, and in part by the Alternative Numbering Systems and Fused Arithmetic for Deep Learning Hardware Implementation and System-on-Chip Center fund under Award RC2-2018-018.

ABSTRACT Recent advances in artificial intelligence (AI) and continuous monitoring of patients using wearable devices have enhanced the accuracy of diagnosing various arrhythmias, from the captured Electrocardiogram (ECG) signals. Achieving high accuracy when using Deep Neural Network (DNN) for ECG classification is accomplished at the cost of compute and memory intensive operations, thus limiting its deployment to devices with high computing capabilities, and makes it unsuitable for wearable edge devices. To facilitate the deployment of deep neural networks on wearable mobile edge devices with limited resources, a lightweight Convolution Neural Network (CNN) model based on the ShuffleNet architecture is proposed and implemented as a solution in this paper. A sliding window of variable stride is used to increase the number of under-represented classes in the database. Moreover, a novel encoding scheme is employed for labelling and training test set samples, allowing the model to detect multiple classes in one ECG segment. A loss function (Focal loss) that proved to be effective when applied for DNN training on an imbalanced dataset was also explored in this work. The proposed model outperformed traditional CNN with 9x less trainable parameters and improved the F1-score by 2%.

INDEX TERMS ECG, AI, health care, CNN, wearable electronics.

I. INTRODUCTION

The increase in the availability of wearable biomedical devices has facilitated our ability to monitor health conditions, outside clinical environments for extended periods of time. Healthcare workers can monitor their patients remotely and collect valuable clinical information about the patients' conditions. Cardiac arrhythmia, a condition in which the heart beats irregularly, is among the many conditions that require continuous monitoring and treatment, as it can lead to stroke or cardiac arrest [1]. Moreover, automated heart-beat classification of ECG signals collected from wearable devices is useful in medical practice [2]. Analyzing the massive amount data collected from these edge devices using artificial intelligence (AI) techniques enables a much better understanding of a patient's condition and hence improves the level of care.

Traditional arrhythmia classification techniques depend on ECG signal morphology (Fig 1) obtained from the feature extraction algorithms. As the QRS complex of ECG signals contain useful information that is required to classify cardiac arrhythmias, handcrafted features are extracted from the QRS complex [2]. However, ANNs (Artificial Neural Networks) are capable of learning features necessary for classification through training. In CNN, feature extraction and classification stages are merged to form one network with a CNN block responsible for feature extraction and a dense layer

The associate editor coordinating the review of this manuscript and approving it for publication was Anandakumar Haldorai.

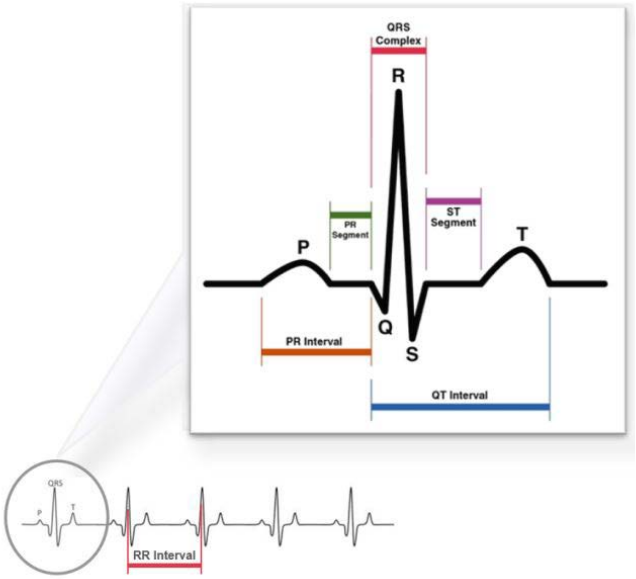


FIGURE 1. Typical ECG signal showing the main features PQRST of one beat.

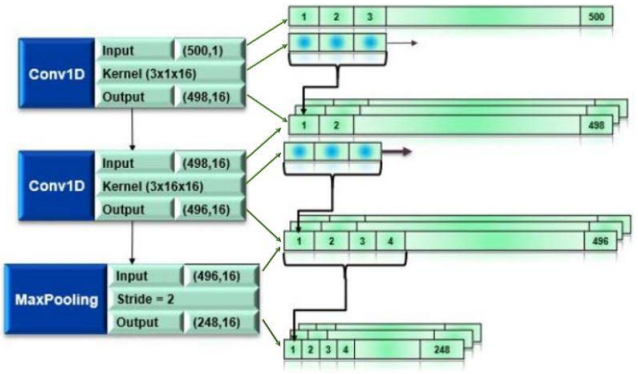


FIGURE 2. Typical one-dimensional CNN operation. Layer1: Filter (3 × 1 × 16) applied on input channel (500 × 1) gives feature map (498 × 16). Layer2: Filter (3 × 16 × 16) applied on feature map (498 × 16) gives feature map (496 × 16). Layer3: Down-sampling feature map (496 × 16) to output feature map (248 × 16).

for classification. An end-to-end classification system on raw ECG signal was proposed in [2]. Although there was no separate feature extraction stage in their work, location of the R peak was determined first before the input signal was segmented into beats of fixed length either by zero padding or truncating the beat whenever RR intervals were smaller or larger, respectively. A comprehensive review presented in [3] described Support Vector Machine (SVM) and ANN as the most popular classification algorithm for ECG signal processing with nearly half of all AI models being ANN mostly CNN in the past ten years. In the cardiologist level arrhythmia detection model reported in [4], a 34-layer CNN architecture was developed for an arbitrary length of ECG signal sampled at a frequency of 200Hz. Techniques that apply deep learning models have been very successful in arrhythmia classification and diagnosis of other heart diseases. However, the major challenge of running Deep Neural Network (DNN) inference on edge devices, such as health fit-bits and mobile phones, is the memory and power requirement due to its computational complexity. A Power Management Unit (PMU) architecture suitable for harvesting energy from multiple sources that is sufficient to power wearable devices was presented in [5]. An Efficient Thermal Energy Harvesting IC (EHIC) was also reported in [6] for microWatt system-on-chips without battery.

The two most commonly used types of neural network applied for ECG signal analysis in the literature are 1D CNN [7], [8], [9], [10], [11] and Long Short Term Memory (LSTM) recurrent neural network [12], [13], [14]. In [15], E. Essa et al. applied an ensemble learning with deep learning bagging models for ECG arrhythmia classification by combining CNN and classical feature extractors with LSTM. The benefit of using CNN (Fig 2) feature extraction over an ordinary fully connected neural network is achieved at the cost of employing many hidden layers. Most accurate CNN models reported in the literature have hundreds of hidden layers and thousands of filters [16], [17]. The computational complexity of these heavy networks leads to the need of executing billions of multiply-add operations [18].

To reduce computational complexity, a significant amount of work has been devoted to developing hardware-friendly algorithms targeting edge devices with a limited power budget. Micro-controller class implementation of ultra-lightweight neural network for an end-to-end ECG classification was reported in [19] to reduce computational complexity while maintaining accuracy. Approximate computing is employed to reduce the precision of weights and activations in [20] and [21]. However, it is a challenge to maintain good accuracy with reduced model complexity. A deep compression technique in [22] employed pruning, trained quantization, and Huffman encoding. Moreover, some compact network architectures (i.e., compact DNNs) were explored in [23] and [24] to reduce the computational complexity of the overall design. Different lightweight CNN architectures which achieved state-of-the-art performance when applied to images and videos were also reported. Among the most popular light weight CNN architectures for image classification are MobileNet [25], ShuffleNet [18], MnasNet [26]. MobileNet and ShuffleNet use depthwise separable convolution which results in a smaller number of trainable parameters and computational cost. Most ECG classification techniques employ signal filtering to remove baseline wandering and noise artifacts to achieve high accuracy [27], [28], [29]. However, with advancements in deep learning techniques, high accuracy can be achieved even without filtering the signals. An accuracy of above 95%, 92.51% and 91.33% was reported in [30] for 13, 15, 17 classes, respectively, without applying any signal filtering. The accuracy obtained using DNN for noisy and

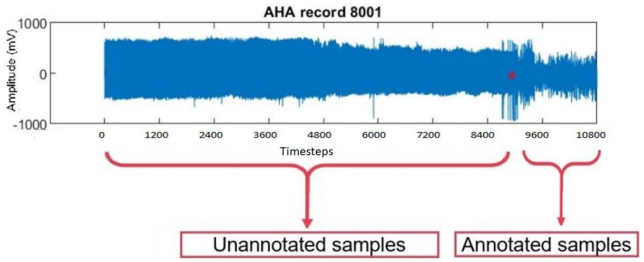


FIGURE 3. Three-hour sample record with the Last 30 minutes annotated.

noise-free ECG signals were compared in [9] and achieved comparable detection accuracy for both setups. Their experimental results demonstrated that noise removal is not necessary for ECG classification using deep learning techniques. A similar approach is followed in this work to remove the need for hardware for signal pre-processing and take advantage of the filtering feature of convolution neural networks.

In this work, a lightweight ShuffleNet based model is proposed and used for the classification of arrhythmias. The contribution of this paper could be summarized as:

- Develop a novel one dimensional ShuffleNet like network for light weight architecture targetting arrhythmia classification with 9x reduction factor in the number of trainable parameters.
- A novel encoding scheme is employed to label the training and test set samples, allowing the model to detect multiple classes in one sample.
- A sliding window of variable stride was used to increase the number of under represented classes in the database.
- The proposed model has outperformed traditional CNN by improving the F1-score by 2% with 9x less trainable parameters

The rest of this paper is organized as follows: Section II presents the dataset used and data preparation for the training of the proposed model, Section III introduces the proposed model and discusses the training methodology, Section IV presents the high-level performance result of this work, and finally, the paper is concluded in Section V.

A. ECG DATABASE

ECG signals obtained from the American Heart Association (AHA) ECG Database [31] were used in the analysis. The database consists of two series of beat-wise annotated ECG with 79 recordings of three-hour duration previously known as the public database and 75 recordings of three-hour duration previously known as the confidential database. Both series are broken down into eight arrhythmia classes with ten recordings per category. There are two channels in these recordings with a sampling rate of 250 samples per second per channel with twelve-bit resolution. Only the last 30 minutes of each three-hour-long recordings were annotated as shown in Fig 3.

TABLE 1. Number of annotated beats per arrhythmia.

Label	Description	Number of annotated beats
N	Beat of Non-Ventricular Origin	316609
V	Premature Ventricular Complex (PVC)	31187
E	Ventricular Escape	12
F	Fusion Beat	1266
R	R-on-T Beat	1552
P	Paced Beat	3171
Q	Questionable Beat - Indeterminate Origin	572
U	Unreadable	45
"[" or "]"	Ventricular Fibrillation or Flutter	3251

Table 1 summarizes the various cardiac arrhythmia diagnostic classes represented in the dataset, the associated number of annotated beats, and their legends used in the labels. The least represented category of the arrhythmia classes, Ventricular Escape, can cause an imbalance in the dataset and make the classification problem extremely difficult using deep learning models.

II. METHODOLOGY

A. TRAINING DATASET PREPARATION

Raw ECG signals were segmented into a fixed window of 2 seconds time segment, ensuring the presence of at least one beat and prepared for training without applying R peak detection algorithm. As shown in Figure 5, samples were obtained by segmenting the long sequence of ECG signal into fixed-length samples. However, this approach generated very few samples for the classes with smaller number of beats in the recordings.

B. LABEL ENCODING SCHEME

To label a window of fixed length having a combination of these arrhythmias, we came up with an encoding scheme similar to the embedded encoding that is commonly used in text samples preparation for deep learning text classification models. However, it can be seen from the plot in Fig 4 (second column) that the classes are not mutually exclusive, i.e. beats from different classes can exist in one segment. So softmax can not be used as an activation function in the output layer as it is only applicable when the classes are mutually exclusive. Other methods to tackle data imbalance were defined in [32] and [24] as Random Over-Sampling (ROS), Random Under-Sampling (RUS), and Synthetic Minority Over-Sampling Technique (SMOTE). There is no universal rule in the selection, all three should be carefully applied as ROS might cause overfitting leading to poor generalization performance, RUS might lead to loss of information when a significant portion of the majority class is removed and SMOTE might oversample uninformative samples or noisy samples.

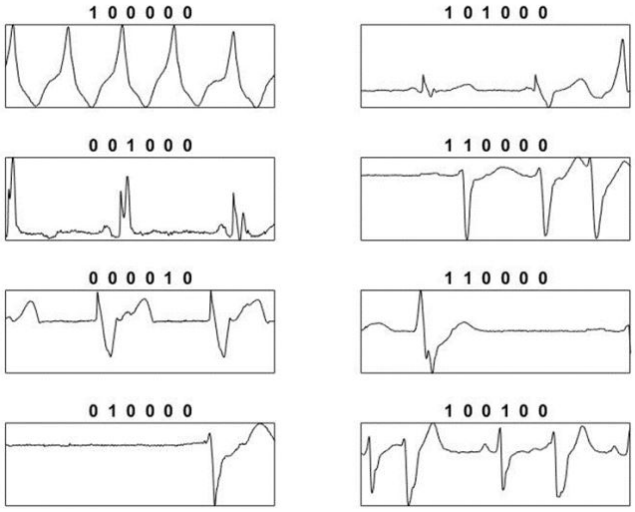


FIGURE 4. Label encoding vector of a 500 timesteps long samples [V E F R P VF] (samples from AHA database).

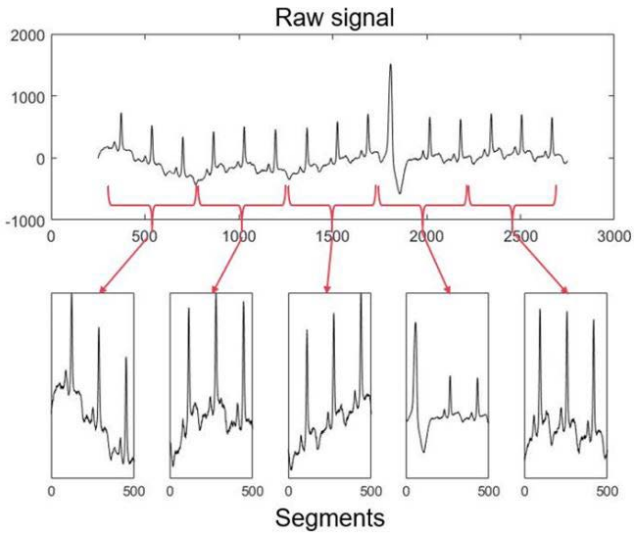


FIGURE 5. Samples generation with non-overlapping windows.

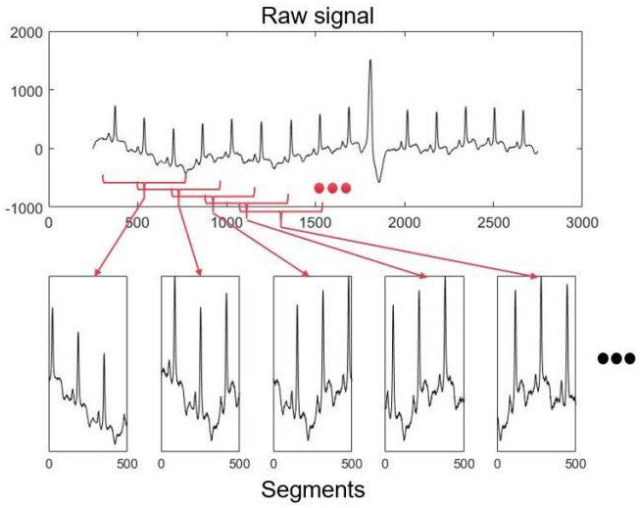


FIGURE 6. Samples generation with overlapping windows.

C. DATA IMBALANCE AND MITIGATION TECHNIQUES

1) DATA IMBALANCE

As shown in Table 1, the Ventricular Escape class for example has only 12 beats in the whole dataset and this causes data imbalance in the training set. A deep learning network not only requires a large corpus of training data to efficiently classify classes but it also depends heavily on a fairly balanced training dataset. To address data imbalance due to a small number of data samples from minor classes, many techniques are employed in training deep neural networks. To mitigate the class imbalance and improve the observability of the model for unusual heart activity, subjects with abnormal rhythms were considered in [4]. In [9] synthetic data was generated to mitigate data imbalance, resulting in higher performance than that reported in the model trained with an imbalanced dataset. Data augmentation, a technique used to artificially expand the size of the training dataset by generating a modified version of the original samples, is one of the widely used techniques in the literature [9]. The application of this technique is more commonly used in the computer vision community. However, some of the transformations such as rotation and flipping of the data are not applicable to ECG signals.

To increase the number of minor classes without duplicating the samples, a stepped window sampling technique is introduced where the long ECG signal series is sampled by striding a segment window by some offset. As presented in Fig. 6 the segments extracted share samples from neighboring segments. Through this, one beat from a minor class can appear in multiple windows, thereby increasing the number of training dataset for the underrepresented categories. The offset must be carefully selected, as it should not be close to zero to avoid duplicate segments and the offset should also not be equal to the size of the window as this will result in a small number of data from minor classes. If the offset value is close to zero, the segments obtained will be duplicates of the neighboring segment and may negatively impact the deep learning classification efficiency. Moreover, this overlapping window sampling technique must only be used on minor classes to balance it with those classes that are represented with a large number of beats in the dataset. For example, the offset that is used for Fusion beat with 12 samples needs smaller offset than the Paced beat with 3171 samples to balance the number of training samples.

These segments need to be labeled based on the beat classes within the window. Most prior efforts that use deep learning techniques to classify ECG signals employ R peak detection algorithm to segment ECG signal into unique beats [2], [9]. The beats are then labeled based on their annotations. However, due to heart rate variability, the size of each beat is different. Therefore, zero padding is applied for those beats that are shorter than a pre-specified window, and those beats that exceed a given window size are truncated. To circumvent

this requirement, ECG signal fragments of 2 and 3 seconds window size were analyzed in [33] assuming a unique class exists in each window. The authors in [30] also analyzed ten seconds segment with one dimensional CNN assuming one arrhythmia exists in addition to a normal beat within each window sample. For both cases, the network must be trained to detect only one condition in one window. In reality, different beats from any of the arrhythmia's may exist next to a normal beat or next to each other.

2) LOSS FUNCTION FOR IMBALANCED DATASET

A new class of loss function named focal loss was used in [34] to address the data imbalance problem for their dense object detector. A factor of $(1 - p_t)^{\gamma}$ is added to the traditional cross-entropy criterion to reduce the relative loss of well-classified samples.

$$\mathrm{FL}\,(p_{\mathrm{t}}) = -\alpha_{\mathrm{t}}(1 - p_{\mathrm{t}})^{\gamma} \log\,(p_{\mathrm{t}}) \tag{1}$$

where p_{t} and α_{t} are:

$$p_{\mathrm{t}}, \alpha_{\mathrm{t}} = \begin{cases} p, \alpha & \text{if } y = 1 \\ 1 - p, 1 - \alpha & \text{otherwise} \end{cases} \tag{2}$$

and $\alpha \in [0, 1]$ is the weighting factor commonly used in cross entropy to address the class imbalance. The effectiveness of this new loss function is studied and compared with the performance of the cross-entropy in the results section.

D. ONE DIMENSIONAL SHUFFLENET ARCHITECTURE AND MODEL TRAINING

Training of neural network models is done with the interaction of the following objects [35]:

- Layers, building blocks of the model
- The input data along with their corresponding labels
- The loss function, which is a measure of the success of the classification task
- The optimizer, which determines the speed at which the model is trained and how the parameters are updated

The dataset is split into 30% for testing and 70% for training. The training dataset is further split into training data and validation data where 20% of the training dataset is used for validation on every epoch throughout the training phase. The CNN model was developed and trained in Python using TensorFlow backend in Keras. Keras library was used to define the network architecture. The parameters chosen were obtained by exhaustive search starting from a heavy network with filters reaching 128 in some layers down to 16 filters to achieve the smallest possible architecture without compromising accuracy. This brute-force search technique was applied to ordinary one-dimensional CNN and a one-dimensional version of the ShuffleNet model in the search for optimal configuration. The selection of the optimizer and loss function in the compilation step depends on the classification task on hand [35]. RMSprop optimizer is recommended as it can be used for almost any classification

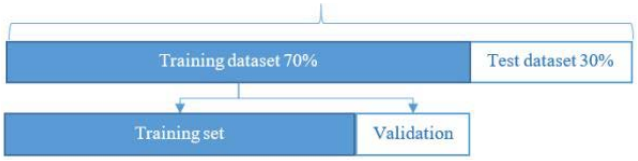


FIGURE 7. Test vs train dataset splitting.

TABLE 2. Recommended last later activation loss function.

Problem Type	Last Layer Activation	Loss Function
Binary classification	sigmoid	binary_crossentropy
Multiclass, single-label classification	softmax	categorical_crossentropy
Multiclass, multilabel classification	sigmoid	binary_crossentropy
Regression to arbitrary values	None	mse
Regression to values between 0 and 1	sigmoid	mse or binary_crossentropy

task [35]. Table 2 summarizes recommended last-layer activation and a loss function for the common classification and regression problems [35]. Dropout layer (50%) was added at the dense layer to prevent overfitting of the model to the major class. A learning rate of ''1e-4'' was used with the RMSprop optimizer and a loss function of cross-entropy. ReLu was applied at the output of all hidden layers and sigmoid is used at the output of the network. For multi-class problems, it is generally recommended to use softmax as the last activation function and categorical cross-entropy as the loss function instead of MSE. However, if the classes are not mutually exclusive with one sample belonging to multiple classes, sigmoids on the output layer and cross-entropy for the cost function results in better performance.

In this work, the application of both ordinary and light weight convolution neural network models in arrhythmia classification was explored. ShuffleNet [18] is a CNN architecture that proved to be highly computationally efficient. The distinct feature that makes ShuffleNet different from MobileNet is the channel shuffle operation for group convolution [18]. Group convolution means applying convolution to one input channel group at a time. Although this is also applied in Mobile-Net, each group is independently represented, and therefore the general representation of input channels is weak because data flow between channel groups is limited. Channel shuffle solves the problem of segregation between channel groups by shuffling the data in the input channels and, in turn, allowing each group's convolution output to fairly represent the input. This operation is what mainly differentiates the ShuffleNet from MobileNet. Another feature that makes these light-weight models hardware-friendly is the depthwise convolution. While capturing spatial and channel-wise information equivalent to the conventional

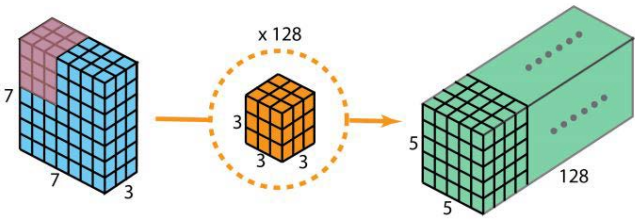


FIGURE 8. Ordinary convolution [36].

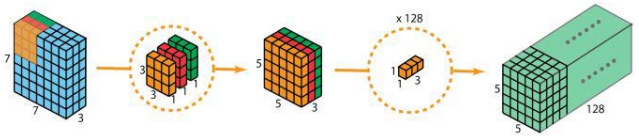


FIGURE 9. Depthwise separable convolution [36].

convolution, depthwise convolution results in fewer computations. In the conventional convolution shown above, a point-wise multiplication of 3×3 kernel for 3 channels is applied on all input channels. The result is stored in the 5×5 feature map and this will be applied 128 times for all output feature maps. Ignoring the biases and additions, the total number of computations is $3 \times 3 \times 3 \times 5 \times 5 \times 128 = 86{,}400$.

In order to utilize the effectiveness of one-dimensional CNN with sequential data and the hardware-friendly 2D ShuffleNet architecture, we propose a version of ShuffleNet with 1D architecture. In depthwise separable convolution shown above 3×3 filter is applied on each channel separately to give rise to three 5×5 out feature maps. To capture similar information along the depth of the channels as in ordinary convolution, 1×1 point-wise convolution is applied on the intermediate feature maps to obtain a single activation map. This is repeated 128 times for all output feature maps. Ignoring the biases and additions, total number of computations is $3 \times 3 \times 5 \times 5 \times 3 + 3 \times 5 \times 5 \times 128 = 10{,}275$. The difference in the number of computations between the regular convolution and the depthwise convolution becomes more significant as the network becomes deeper.

Appendix A presents the architecture of the proposed one dimensional ShuffleNet model and the shape of the kernels used in the model for each layer.

III. RESULTS AND PERFORMANCE EVALUATION

In order to verify the performance of our one-dimensional ShuffleNet model, we tested it with ECG arrhythmia classification task and achieved an overall accuracy of 92.23% and an F1-score of 97%. The high-level deep neural network model was developed, trained, and tested to classify six arrhythmia classes using the AHA dataset. Initially, the accuracy was steady at 75% and validation was sparsely distributed. After fine-tuning the model parameters and changing the loss function and the activation function of the output layer to cross-entropy and sigmoid respectively, the accuracy started to improve. This configuration is applicable to multi-class multi-label classification task. In addition to

TABLE 3. Confusion Matrix - ShuffleNet.

Using conventional cross entropy loss

Ann	V	E	F	R	P	VF
V	4388 94.94%	6 0.13%	98 2.12%	95 2.06%	28 0.61%	7 0.15%
E	9 1.11%	800 98.77%	0 0%	0 0%	0 0%	1 0.12%
F	133 9.49%	0 0%	1232 87.94%	11 0.79%	23 1.64%	2 0.14%
R	87 3.96%	0 0%	11 0.5%	2099 95.5%	0 0%	1 0.05%
P	17 0.92%	0 0%	18 0.98%	2 0.11%	1800 97.93%	10 0.05%
VF	19 0.65%	0 0%	2 0.07%	3 0.1%	0 0%	2901 99.18%

Using focal loss

Ann	V	E	F	R	P	VF
V	4354 94.2%	7 0.15%	126 2.73%	96 2.08%	38 0.82%	1 0.02%
E	4 0.49%	806 99.51%	0 0%	0 0%	0 0%	0 0%
F	120 8.57%	0 0%	1257 89.72%	4 0.29%	19 1.36%	1 0.07%
R	90 4.09%	0 0%	8 0.36%	2100 95.54%	0 0%	0 0%
P	12 0.65%	0 0%	9 0.49%	0 0%	1816 98.8%	1 0.05%
VF	7 0.24%	0 0%	0 0%	0 0%	7 .24%	2911 99.52%

the conventional cross-entropy, focal loss was also used to address the problem of data imbalance. Focal loss improves the performance of a DNN training when applied on an imbalanced dataset [34]. It is clear from the confusion matrices in Table 3 and Table 4 that the model focuses more on the underrepresented classes when focal loss is used in training. To compare the performance of both models for each class, a confusion matrix was computed and as displayed in the matrix, the ShuffleNet model outperforms the traditional CNN model with a fewer number of trainable parameters as shown in Table 3 and Table 4.

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = \frac{TP}{Total\ Actual\ Positive} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Precision = \frac{TP}{Total\ Predicted\ Positive}$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{5}$$

The result of the confusion matrices shows that the unique morphology of VF resulted in high classification accuracy in both models. However, a relatively low positive predictivity is observed with Fusion beats, which occur when a premature ventricular impulse is discharged late close to the time of the normal sinus rhythm. As the resulting waveform morphology combines normal beat and the PVC, both networks tend to predict fusion beat as Premature Ventricular Complex (PVC) [37]. Similarly, the prediction of Ventricular Escape 'E' class in both models was also negatively affected with

TABLE 4. Confusion matrix - traditional CNN model.

Using conventional cross entropy loss

Ann	V	E	F	R	P	VF
V	4030 87.19%	4 0.09%	165 3.57%	175 3.79%	240 5.19%	8 0.17%
E	5 0.62%	805 99.38%	0 0%	0 0%	0 0%	0 0%
F	185 13.2%	0 0%	1161 82.87%	18 1.28%	37 2.64%	0 0%
R	52 2.37%	0 0%	5 0.23%	2140 97.36%	1 0.05%	0 0%
P	10 0.54%	1 0.05%	11 0.6%	3 0.16%	1810 98.48%	3 0.16%
VF	4 0.14%	1 0.03%	2 0.07%	0 0%	5 5%	2913 99.59%

Using focal loss

Ann	V	E	F	R	P	VF
V	4177 90.37%	8 0.17%	94 2.03%	152 3.29%	184 3.98%	7 0.15%
E	0 0%	810 100%	0 0%	0 0%	0 0%	0 0%
F	268 19.13%	0 0%	1078 76.95%	8 0.57%	47 3.35%	0 0%
R	217 9.87%	0 0%	2 0.09%	1979 90.04%	0 0%	0 0%
P	34 1.85%	0 0%	12 0.65%	0 0%	1790 97.39%	2 0.11%
VF	13 0.44%	0 0%	2 0.07%	3 0.1%	1 0.03%	2906 99.35%

some of its beats classified as PVC. An escape ventricular beat or rhythm occurs due to failure of the sinus and atrioventricular (AV) node to stimulate an impulse resulting in the absence of a P wave and a widened QRS complex that resembles a premature ventricular complex/contraction. The morphological resemblance of this arrhythmia to PVC causes the model to predict a significant number of E beats as PVC. Ventricular Escape occurs after a pause of variable duration (but always greater than the normal sinus RR interval). Interpretation of performance measures used were F1-score, recall, and precision. If the dataset used in a classification task is balanced, accuracy and area under the receiver operating characteristic curve (ROC AUC) can be used. However, if the dataset is imbalanced, the precision and recall can be used. Moreover, if a balance between precision and recall is needed, F1-score a weighted average of Precision and Recall can be applied. Using accuracy as a performance measure alone is not enough especially when there is data imbalance as the model can over-fit to the major class.

Tables 5 and 6 summarizes the performance of both ShuffleNet and 1D CNN models on the test data reserved for evaluation in terms of precision, recall and F1-score. The parameters ''micro'', ''macro'', ''weighted'' and ''samples average'' are required for the multiclass/multilabel classification tasks. These parameters are defined in Keras library as [38]:

- ''micro'', calculate metrics globally by counting the total true positives, false negatives, and false positives;
- ''macro'', calculate metrics for each label, and find their unweighted mean. This does not take into account label imbalance

TABLE 5. Performance result - 1D ShuffleNet.

Using conventional cross entropy loss

Class	Annotation	Precision	Recall	F1-Score	Support
0	V	0.95	0.95	0.95	4622
1	F	1.00	1.00	1.00	1854
2	E	0.90	0.87	0.89	1705
3	R	0.96	0.96	0.96	2300
4	P	0.99	0.99	0.99	2352
5	VF	1.00	0.99	0.99	2926
micro avg		0.97	0.96	0.96	15759
macro avg		0.97	0.96	0.96	15759
weighted avg		0.97	0.96	0.96	15759
samples avg		0.97	0.97	0.97	15759

Using focal loss

Class	Annotation	Precision	Recall	F1-Score	Support
0	V	0.96	0.94	0.95	4622
1	F	1.00	1.00	1.00	1854
2	E	0.92	0.89	0.90	1705
3	R	0.96	0.95	0.96	2300
4	P	0.99	0.99	0.99	2352
5	VF	1.00	1.00	1.00	2926
micro avg		0.97	0.96	0.97	15759
macro avg		0.97	0.96	0.96	15759
weighted avg		0.97	0.96	0.97	15759
samples avg		0.97	0.97	0.97	15759

TABLE 6. Performance result - traditional CNN model.

Using conventional cross entropy loss

Class	Annotation	Precision	Recall	F1-Score	Support
0	V	0.94	0.87	0.90	4622
1	F	1.00	1.00	1.00	1854
2	E	0.90	0.77	0.83	1705
3	R	0.94	0.97	0.95	2300
4	P	0.98	0.99	0.99	2352
5	VF	1.00	1.00	1.00	2926
micro avg		0.96	0.93	0.95	15759
macro avg		0.96	0.93	0.94	15759
weighted avg		0.96	0.93	0.94	15759
samples avg		0.96	0.94	0.95	15759

Using focal loss

Class	Annotation	Precision	Recall	F1-Score	Support
0	V	0.95	0.85	0.89	4622
1	F	1.00	1.00	1.00	1854
2	E	0.94	0.66	0.77	1705
3	R	0.95	0.88	0.92	2300
4	P	0.99	0.98	0.99	2352
5	VF	1.00	0.99	1.00	2926
micro avg		0.97	0.90	0.93	15759
macro avg		0.97	0.89	0.93	15759
weighted avg		0.97	0.90	0.93	15759
samples avg		0.93	0.91	0.92	15759

- ''weighted'', calculate metrics for each label, and find their average weighted by support (the number of true instances for each label). This alters 'macro' to account for label imbalance; it can result in an F1-score that is not between precision and recall
- ''samples'', calculate metrics for each instance, and find their average (only meaningful for multilabel classification where this differs from function accuracy score)

The highest score in all performance measures obtained was for VF classes due to its unique morphology of the waveform.

To demonstrate the effectiveness of our approach, we trained our model using the publicly available MIT-BIH

TABLE 7. Performance result - 1D ShuffleNet using MIT-BIH dataset.

Using conventional cross entropy loss					
Class	**Annotation**	**Precision**	**Recall**	**F1-Score**	**Support**
0	N	1.00	0.99	1.00	54281
1	S	0.96	0.91	0.93	6727
2	V	0.99	0.99	0.99	36530
3	F	0.97	0.96	0.97	9791
micro avg		0.99	0.98	0.99	107329
macro avg		0.98	0.96	0.97	107329
weighted avg		0.99	0.96	0.99	107329
samples avg		0.98	0.98	0.98	107329

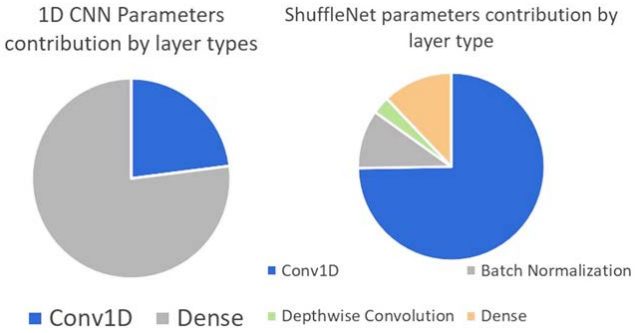


FIGURE 10. Parameters contribution by different layer types.

Arrhythmia database and compared it with prior work. The classes considered in this work are [Normal (N), Superventricular Ectopic Beat (SVEB), Ventricular Ectopic (VEB), Fusion (F)] as defined in AAMI standard. As the ECG records were sampled at 360 Hz, we segment the waveform into a window of 720 timesteps to include at least two beats in one window. These 2s segments were annotated using the labeling scheme shown in Figure 4 into four classes as the Q (Unknown beat) class was not considered in the classification task. To increase the number of those classes which are not well represented in the dataset, we apply flexible stride when generating samples following similar technique used in Figure 6. That is smaller stride size is used for those minority classes and higher stride for those classes which are well represented in the dataset. This improves the class imbalance in the dataset. The dataset is split into 70% training and 30% test data with 20% of the training data being used for validation during the training phase. The summary of the performance evaluation of the proposed model using the MIT-BIH dataset is presented in Table 7.

Figure 10 shows the contribution of the layers to the total number of trainable parameters. While the conventional one-dimensional CNN dense layer requires more than 75% of the overall memory of the model, the convolution operation contributes around 75% trainable parameters in the ShuffleNet model. The dense layer in the ShuffleNet model has relatively fewer number of parameters due to the application of Global MaxPooling operation at the final convolution layer of the model. In a traditional convolutional neural network, feature maps of the last convolution layer are flattened and fed to a dense fully connected layer acting as a bridge between

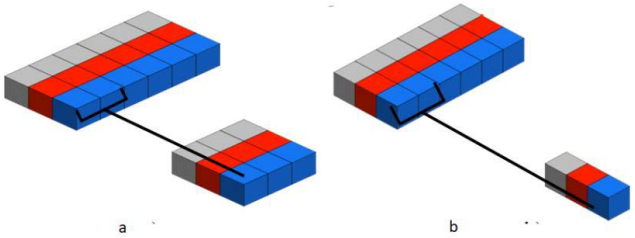


FIGURE 11. MaxPooling (a) vs global MaxPooling (b).

TABLE 8. Performance comparison 1D ShuffleNet Vs 1D.

Model	Parameters	No of Layers	Training time (s)	Inference time
1D ShuffleNet	9,510	42*	6,901.89	12.39
1D CNN	79,088	23*	6,733.52	16.40

* Including MaxPooling layer

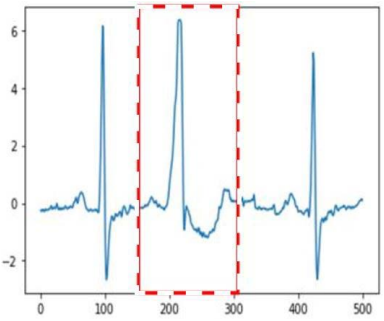


FIGURE 12. Sample input with PVC and normal beat.

the feature extractor convolutional layers and classifier network. Unlike in MaxPooling, where elements from a fixed size pool are compared for maximum value, the entire input block is considered as one pool in Global MaxPooling. As a result, the size of the last feature map of the convolution block is very small compared to the model without Global MaxPooling. In [28], they replaced a fully connected network with a Global average pooling and claimed it can act as a structural regularizer enforcing feature maps to be confidence maps concepts. The application of both Global max and average pooling in our one-dimensional ShuffleNet has resulted in a small number of parameters in the classifier network.

It is often said that it is difficult to visualize representations learned by CNN models that are readable to humans and the network is considered as a black box. However, several techniques have been developed to visualize and interpret these representations [35]. Intermediate activation maps were plotted in Figure 13 and 14 to visualize the response of the filters. Some of these activation maps were zero as they are responsive to specific inputs only; this could also be a symptom of a high learning rate. Activation maps provide useful information related to a representation that the model learns. As demonstrated in Figure 13 and 14, deeper layers tend to encode more meaningful patterns that is in

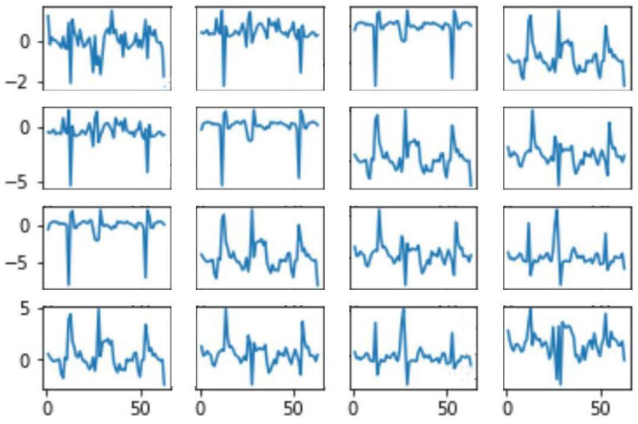


FIGURE 13. Layer-16 activation features visualization (ShuffleNet model): Magnitude (y-axis) of each neuron of the 16 convnet outputs represented in each timestep (x-axis).

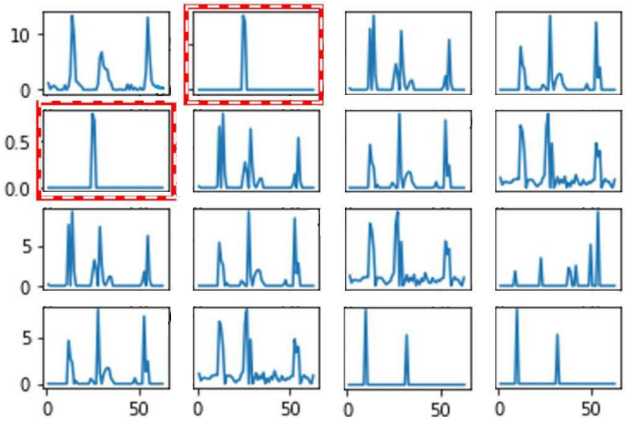


FIGURE 14. Layer-36 activation features visualization (ShuffleNet model): PVC beat detected marked with red dots.

TABLE 9. Performance comparison ShuffleNet model.

Reference	Dataset	Network	F1-Score	Parameters	# classes
This work	AHA	1D Shuf-fleNet	97%	9,718	6
Traditional 1D CNN	AHA	1D CNN	95%	79,088	6
This work	MIT-BIH	1D Shuf-fleNet	98%	12,788	4
Acharya et al. [9]	MIT-BIH	1D CNN	76.14%*	19,715*	5
O. Yıldırım et al. [30]	MIT-BIH	1D CNN	92.45%†	3,054,656†	5
Chen et al. [39]	MIT-BIH	1D CNN	96.4	-	13
Romdhane et al. [40]	MIT-BIH	1D CNN	98.38%‡	-	5

* Not reported, the figures presented are calculated
† Considered 13 classes and used large kernel size
‡ Larger network with preprocessing for R peak detection used

a form comprehensible by humans than the shallow ones. The input sample plotted in Figure 12 portrays two convnet outputs of layer-36 (marked with a red dotted rectangle in 14) that are responsive to patterns from PVC beat in the input signal.

Table 8 presents the effectiveness of our light weight architecture as it has shorter inference time and comparable training time with 9x less number of parameters. Moreover, as seen in the results presented in Table 9, our light weight model has proven to classify sequential data (signals in this case) efficiently and with better accuracy. It should be noted that the comparison made is with prior works which applied one dimensional CNN but for multi-class one label classification as we could not find an arrhythmia classification experiment with multi-class multi-label classification task. It can be seen from Table 9 that the ordinary CNN model developed for comparison achieved less accuracy compared to all other models that were trained to predict only one arrhythmia for a given input sample.

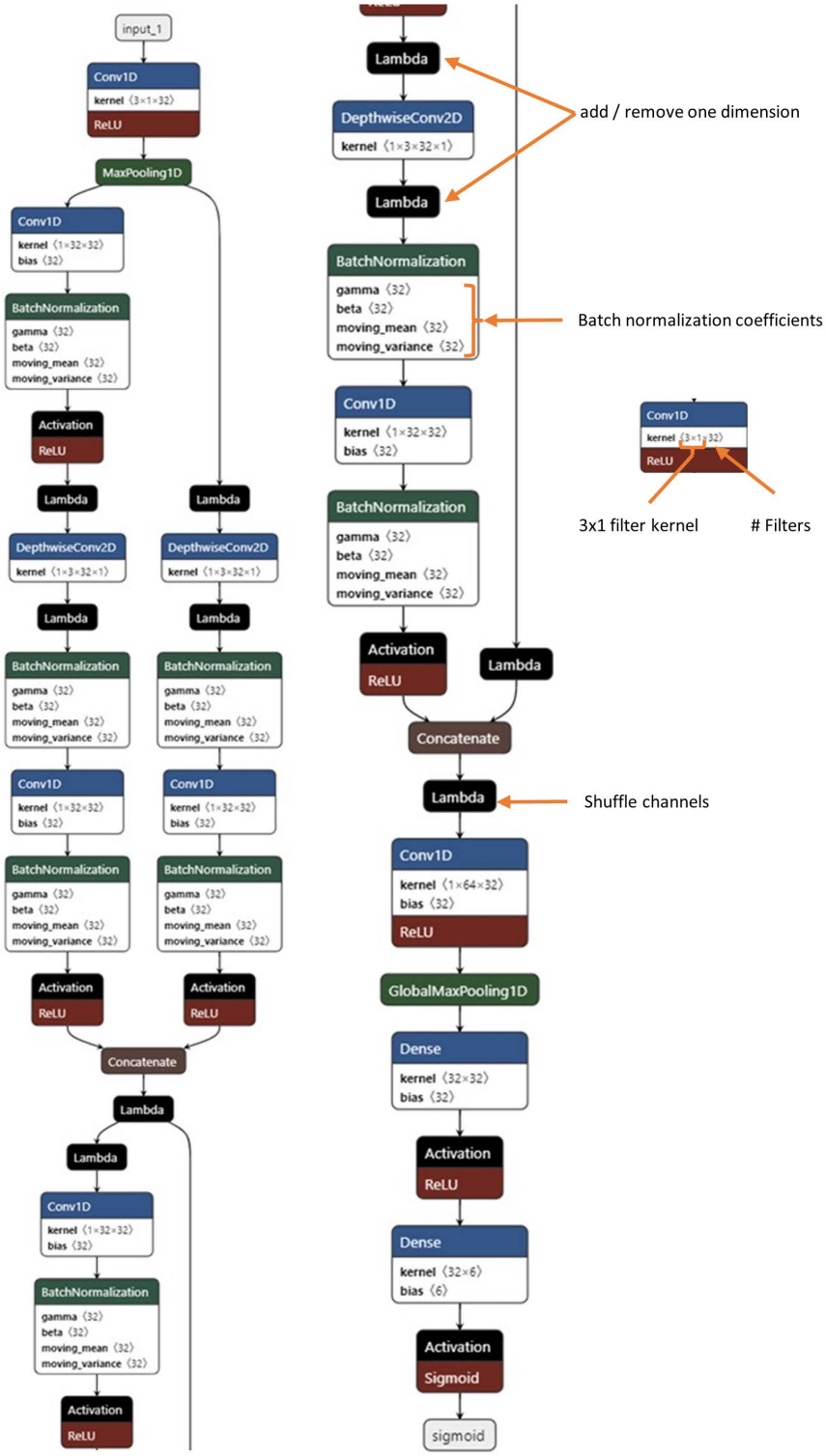
IV. CONCLUSION AND FUTURE WORK

This paper demonstrated the application of hardware-friendly deep learning neural networks for arrhythmia classification tasks. We presented the design and implementation of a lightweight one-dimensional CNN model based on ShuffleNet architecture that requires a lower number of trainable parameters. The network was developed and tested using a dataset obtained from the AHA. The lightweight model presented has 9x fewer parameters than the traditional implementation and, hence, less memory requirement than conventional CNN architectures. The Focal loss was used in place of the standard cross-entropy loss function to improve the performance. This work proved that lightweight neural networks based on depth-wise and point-wise convolution outperform conventional CNN models in classification accuracy. While maintaining 9x less trainable parameters, our model has outperformed traditional CNN by 2%, achieving an F1-score of 97%. Moreover, the encoding approach for training and test set samples enables the model to detect multiple classes from one sample of ECG signal.

In future, a different network containing convolution and classifier blocks could be trained with prealigned ECG segments based on R peak locations instead of a fixed window. This could potentially improve feature extraction capabilities of the convolution blocks. Then, the classifier network could also be replaced with a new dense layer trained with the pre-trained convolution blocks frozen using fixed windowed ECG samples as input instead of R peak location based segmented ECG signals.

APPENDIX A
ONE DIMENSIONAL SHUFFLENET ARCHITECTURE

Lambda blocks are inserted before and after the depthwise convolution to add and remove one dimension respectively to the one-dimensional arrays of activation maps as there is no builtin depthwise convolution in keras for such activation maps.



REFERENCES

[1] A. B. de Luna, P. Coumel, and J. F. Leclercq, "Ambulatory sudden cardiac death: Mechanisms of production of fatal arrhythmia on the basis of data from 157 cases," *Amer. Heart J.*, vol. 117, no. 1, pp. 151–159, 1989.

[2] 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. Health Inform.*, vol. 23, no. 4, pp. 1574–1584, Jul. 2019.

[3] Y. Wei, J. Zhou, Y. Wang, Y. Liu, Q. Liu, J. Luo, C. Wang, F. Ren, and L. Huang, "A review of algorithm & hardware design for AI-based biomedical applications," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 2, pp. 145–163, Apr. 2020.

[4] Y. H. Awni, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Med.*, vol. 25, pp. 65–69, Jan. 2019.

[5] M. Alhawari, T. Tekeste, B. Mohammad, H. Saleh, and M. Ismail, "Power management unit for multi-source energy harvesting in wearable electronics," in *Proc. IEEE 59th Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Oct. 2016, pp. 1–4.
[6] M. Alhawari, D. Kilani, B. Mohammad, H. Saleh, and M. Ismail, "An efficient thermal energy harvesting and power management for μWatt wearable BioChips," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2016, pp. 2258–2261.
[7] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mech. Syst. Signal Process.*, vol. 151, Apr. 2021, Art. no. 107398.
[8] M. Sarlija, F. Jurisic, and S. Popovic, "A convolutional neural network based approach to QRS detection," in *Proc. 10th Int. Symp. Image Signal Process. Anal.*, Sep. 2017, pp. 121–125.
[9] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych, and R. S. Tan, "A deep convolutional neural network model to classify heartbeats," *Comput. Biol. Med.*, vol. 89, pp. 389–396, Oct. 2017.
[10] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, and M. Adam, "Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals," *Inf. Sci.*, vol. 415, pp. 190–198, Nov. 2017.
[11] O. Yildirim, T. R. San, and U. R. Acharya, "An efficient compression of ECG signals using deep convolutional autoencoders," *Cognit. Syst. Res.*, vol. 52, pp. 198–211, Dec. 2018.
[12] O. Faust, A. Shenfield, M. Kareem, T. R. San, H. Fujita, and U. R. Acharya, "Automated detection of atrial fibrillation using long short-term memory network with RR interval signals," *Comput. Biol. Med.*, vol. 102, pp. 327–335, Nov. 2018.
[13] Ö. Yildirim, "A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification," *Comput. Biol. Med.*, vol. 96, pp. 189–202, Jul. 2018.
[14] B. Ganguly, A. Ghosal, A. Das, D. Das, D. Chatterjee, and D. Rakshit, "Automated detection and classification of arrhythmia from ECG signals using feature-induced long short-term memory network," *IEEE Sensors Lett.*, vol. 4, no. 8, pp. 1–4, Aug. 2020.
[15] E. Essa and X. Xie, "An ensemble of deep learning-based multi-model for ECG heartbeats arrhythmia classification," *IEEE Access*, vol. 9, pp. 103452–103464, 2021.
[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[17] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.
[18] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6848–6856.
[19] J. Xiao, J. Liu, H. Yang, Q. Liu, N. Wang, Z. Zhu, Y. Chen, Y. Long, L. Chang, L. Zhou, and J. Zhou, "ULECGNet: An ultra-lightweight end-to-end ECG classification neural network," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 1, pp. 206–217, Jan. 2022.
[20] E. H. Lee, D. Miyashita, E. Chai, B. Murmann, and S. Simon Wong, "LogNet: Energy-efficient neural networks using logarithmic computation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 5900–5904.
[21] B. Moons, B. De Brabandere, L. Van Gool, and M. Verhelst, "Energy-efficient ConvNets through approximate computing," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2016, pp. 1–8.
[22] S. Han, H. Mao, and W. J. Dally, "Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding," 2015, *arXiv:1510.00149*.
[23] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
[24] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size," 2016, *arXiv:1602.07360*.
[25] A. Howard, M. Sandler, B. Chen, W. Wang, L.-C. Chen, M. Tan, G. Chu, V. Vasudevan, Y. Zhu, R. Pang, H. Adam, and Q. Le, "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Nov. 2019, pp. 1314–1324.
[26] M. Tan, B. Chen, R. Pang, V. Vasudevan, M. Sandler, A. Howard, andQ. V. Le, "MnasNet: Platform-aware neural architecture search for mobile," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 2820–2828.
[27] C. Haritha, M. Ganesan, and E. P. Sumesh, "A survey on modern trends in ECG noise removal techniques," in *Proc. Int. Conf. Circuit, Power Comput. Technol. (ICCPCT)*, Mar. 2016, pp. 1–7.
[28] P. Singh, S. Shahnawazuddin, and G. Pradhan, "Significance of modified empirical mode decomposition for ECG denoising," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 2956–2959.
[29] M. M. Butt, U. Akram, and S. A. Khan, "Denoising practices for electrocardiographic (ECG) signals: A survey," in *Proc. Int. Conf. Comput., Commun., Control Technol. (ICT)*, Apr. 2015, pp. 264–268.
[30] Ö. Yildirim, P. Plawiak, R.-S. Tan, and U. R. Acharya, "Arrhythmia detection using deep convolutional neural network with long duration ECG signals," *Comput. Biol. Med.*, vol. 102, no. 1, pp. 411–420, Nov. 2018.
[31] *American Heart Association ECG Database USB*. Accessed: Mar. 15, 2021. [Online]. Available: https://www.ecri.org/american-heart-association-ecg-database-usb
[32] J. L. Leevy, T. M. Khoshgoftaar, R. A. Bauder, and N. Seliya, "A survey on addressing high-class imbalance in big data," *J. Big Data*, vol. 5, no. 1, pp. 1–30, 2018.
[33] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, and M. Adam, "Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network," *Inf. Sci.*, vol. 405, no. 1, pp. 81–90, Sep. 2017.
[34] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2980–2988.
[35] F. Chollet, *Deep Learning With Python*, vol. 361. Shelter Island, NY, USA: Manning, 2018.
[36] K. Bai. (Feb. 2019). *A Comprehensive Introduction to Different Types of Convolutions in Deep Learning*. [Online]. Available: https://towardsdatascience.com/a-comprehensive-introduction-to-different-types-of-convolutions-in-deep-learning-669281e58215
[37] M. Gertsch, *The ECG Manual: An Evidence-Based Approach*. Berlin, Germany: Springer, 2008.
[38] *sklearn.metrics.precision_recall_fscore_support-Scikit-Learn 0.24.1 Documentation*. Accessed: Mar. 15, 2021. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision_recall_fscore_support.html
[39] A. Chen, F. Wang, W. Liu, S. Chang, H. Wang, J. He, and Q. Huang, "Multi-information fusion neural networks for arrhythmia automatic detection," *Comput. Methods Programs Biomed.*, vol. 193, Sep. 2020, Art. no. 105479.
[40] T. F. Romdhane and M. A. Pr, "Electrocardiogram heartbeat classification based on a deep convolutional neural network and focal loss," *Comput. Biol. Med.*, vol. 123, Aug. 2020, Art. no. 103866.

HURUY TESFAI received the B.S. degree in electrical and electronics engineering from the Eritrea Institute of Technology, in 2012, and the M.S. degree in electrical and computer engineering from the Khalifa University of Science and Technology, Abu Dhabi, UAE, in 2020, where he is currently pursuing the Ph.D. degree with focus on hardware friendly deep neural network architectures. His current research interest includes energy-efficient hardware implementation of deep neural network accelerators.

HANI SALEH (Senior Member, IEEE) received the Bachelor of Science degree in electrical engineering from the University of Jordan, the Master of Science degree in electrical engineering from the University of Texas at San Antonio, and the Ph.D. degree in computer engineering from the University of Texas at Austin.

He has been an Associate Professor of electronic engineering at Khalifa University, since 2012. He is a Co-Founder and an Active Researcher at the Khalifa University Research Center (KSRC) and the System-on-Chip Research Center (SOCC), where he led the multiple IoT projects for the development of wearable blood glucose monitoring SOC and a mobile surveillance SOC. He has a total of 19 years of industrial experience in ASIC chip design, microprocessor design, DSP core design, graphics core design, and embedded system design. Prior to joining Khalifa University, he worked for many leading semiconductor design companies including a Senior Chip Designer (Technical Lead) at Apple Inc., Intel (ATOM mobile microprocessor design), AMD (Bobcat mobile microprocessor design), Qualcomm (QDSP DSP core design for mobile SOC's), Synopsys (designed the I2C DW IP included in Synopys DesignWare library), Fujitsu (SPARC compatible high performance microprocessor design), and Motorola Australia. He has 21 issued U.S. patents, and a number of pending patent application, and over 120 articles published in peer reviewed conferences and journals in the areas of AI accelerators, artificial intelligence, digital system design, computer architecture, DSP, and computer arithmetic. His research interests include the IoT design, deep learning, AI hardware design, DSP algorithms design, DSP hardware design, computer architecture, computer arithmetic, SOC design, ASIC chip design, FPGA design, and automatic computer recognition.

MAHMOUD AL-QUTAYRI (Senior Member, IEEE) is a Professor of electrical and computer engineering and the Associate Dean for Graduate Studies at the College of Engineering, Khalifa University (KU), UAE. He is also affiliated with the KU System-on-Chip Center, which focuses on the design of high performance and energy efficient electronic devices and systems for a wide range of applications. Prior to joining Khalifa University, he worked at De Montfort University, U.K., and the University of Bath, U.K. He also worked at Philips Semiconductors, Southampton, U.K. He has authored or coauthored numerous technical papers in peer-reviewed journals and international conferences. He has also coauthored a book titled *Digital Phase Lock Loops: Architectures and Applications* and edited a book titled *Smart Home Systems*. This is in addition to a number of book chapters and patents. His current research interests include embedded systems, in-memory computing, emerging memory technologies, the energy efficient IoT systems, efficient edge computing, artificial intelligence hardware implementation, wireless sensor networks, and cognitive wireless networks.

He received a number of awards during his undergraduate and graduate education and professional career. His professional services include serving on the editorial board of some journals and membership of the steering, organizing, and technical program committees of many international conferences.

MOATH B. MOHAMMAD graduated the Bachelor of Science degree *(magna cum laude)* in biology from the Texas A&M University. He is currently pursuing the Medical degree with the UNTHSC–Texas College of Osteopathic Medicine. During his time at Texas A&M, he worked as an EMT with Texas A&M EMS. His current research interests include cardiology, orthopedics, and surgery.

TEMESGHEN TEKESTE received the B.S. degree in electrical engineering from Asmara University, Eritrea, in 2007, the M.S. degree in microsystems engineering from the Masdar Institute of Science and Technology, Abu Dhabi, in 2012, and the Ph.D. degree in electrical and computer engineering from Khalifa University, in 2017. In 2006, he joined the Eritrea Institute of Technology, Asmara, as a Teaching Assistant and Lab Instructor. In 2010, he joined the Masdar Institute, as an M.Sc. Student and his research was focused on low-power, mixed signal integrated circuit design that involved schematic design, simulation, layout, and verification of an ultra-low power clock generator for heart rate sensor IC. His Ph.D. research was on ultra-low power digital circuit design for biomedical application. After his Ph.D., he worked as a Postdoctoral Researcher at Khalifa University, where he was investigating efficient neural network architectures for low-power devices. He is currently working as a Research Specialist at Tawazun Technology and Innovation. He holds two issued U.S. patents, and has authored journal and conference papers. His research interests include DSP hardware design, computer architecture, ASIC chip design, and FPGA design.

AHSAN KHANDOKER received the Ph.D. degree in electronics and biomedical engineering from the Muroran Institute of Technology, Japan, in 2004. He is currently a Professor of biomedical engineering at Khalifa University, Abu Dhabi, UAE. He has multidisciplinary research accomplishments in the area of sleep, diabetes, fetal medicine, psychiatry, biomechanics, bioinstrumentation, bio-signal processing and circuits, and nonlinear modeling. His research projects are funded by the Abu Dhabi Department of Education and Knowledge; the Bill and Melinda Gates Foundation; an Australian Research Council; and Khalifa University Internal Funds in cardiac and mental health monitoring research area in collaboration with Cleveland Clinic, Abu Dhabi; and several key international medical research facilities in Australia, Germany, and Japan. He is a Theme Leader at the Healthcare Engineering Innovation Center (HEIC), Khalifa University. He has published over 100 journal articles and 140 conference papers. A number of ideas proposed in his work have influenced the efforts of the bio-signal processing platforms developed by companies such as ResMed Sydney; Compumedics, Melbourne, Australia; Atom Medical Company, Tokyo; and the start-up company MARP Abu Dhabi (Twinkle Heart Fetal Phonogram device). He received an Australian Research Council Fellowship from the Department of Electrical and Electronic Engineering, University of Melbourne, Australia.

BAKER MOHAMMAD (Senior Member, IEEE) received the B.S. degree from the University of New Mexico, Albuquerque, the M.S. degree from Arizona State University, Tempe, and the Ph.D. degree from the University of Texas at Austin, in 2008, all in ECE. He is the Director of System-on-Chip Center and a Professor of EECS at Khalifa University. Prior to joining Khalifa University, he was a Senior Staff Engineer/the Manager at Qualcomm, Austin, USA, for six years, where he was engaged in designing high performance and low-power DSP processors used for communication multi-media application. Before joining Qualcomm, he worked for ten years at Intel Corporation on a wide range of microprocessors design from high performance, server chips > 100Watt (IA-64), to mobile embedded processor low power sub 1 watt (xscale). He has over 16 years of industrial experience in microprocessor design, emphasizing memory, low power circuit, and physical design. He is also engaged in a microwatt range computing platform for wearable electronics and WSN focusing on energy harvesting, power management, and power conversion, including efficient dc–dc, ac–dc converters. He authored/coauthored over 100 refereed journals and conference proceedings, three books, 18 U.S. patents, multiple invited seminars/panelists, and the presenter of three conference tutorials, including one tutorial on Energy Harvesting and Power Management for WSN at the 2015 (ISCAS). His research interests include VLSI, power-efficient computing, high yield embedded memory, emerging technology, such as Memristor, STTRAM, and In-Memory-Computing, hardware accelerators for Cyber-Physical Systems. He has received several awards, including the KUSTAR Staff Excellence Award in intellectual property creation, the IEEE TVLSI Best Paper Award, the 2016 IEEE MWSCAS Myrill B. Reed Best Paper Award, the Qualcomm Qstar Award for excellence on performance and leadership, the SRC Techon Best Session Papers for 2016 and 2017, the 2009 Best Paper Award for Qualcomm Qtech Conference, and the Intel Involve in the Community Award for volunteer and impact on the community. He participates in many technical committees at IEEE conferences and reviews for TVLSI, IEEE Circuits and Systems journals. He is an Associate Editor of IEEE Access and IEEE Transactions on Very Large Scale Integration (TVLSI) journals.