




# ANNet: A Lightweight Neural Network for ECG Anomaly Detection in IoT Edge Sensors

Gawsalyan Sivapalan , Koushik Kumar Nundy, *Senior Member, IEEE*, Soumyabrata Dev, *Member, IEEE*, Barry Cardiff , *Senior Member, IEEE*, and Deepu John , *Senior Member, IEEE*

**Abstract**—In this paper, we propose a lightweight neural network for real-time electrocardiogram (ECG) anomaly detection and system level power reduction of wearable Internet of Things (IoT) Edge sensors. The proposed network utilizes a novel hybrid architecture consisting of Long Short Term Memory (LSTM) cells and Multi-Layer Perceptrons (MLP). The LSTM block takes a sequence of coefficients representing the morphology of ECG beats while the MLP input layer is fed with features derived from instantaneous heart rate. Simultaneous training of the blocks pushes the overall network to learn distinct features complementing each other for making decisions. The network was evaluated in terms of accuracy, computational complexity, and power consumption using data from the MIT-BIH arrhythmia database. To address the class imbalance in the dataset, we augmented the dataset using SMOTE algorithm for network training. The network achieved an average classification accuracy of 97% across several records in the database. Further, the network was mapped to a fixed point model, retrained in a bit accurate fixed-point environment to compensate for the quantization error, and ported to an ARM Cortex M4 based embedded platform. In laboratory testing, the overall system was successfully demonstrated, and a significant saving of  $\approx 50\%$  power was achieved by gating the wireless transmission using the classifier. Wireless transmission was enabled only to transmit the beats deemed anomalous by the classifier. The proposed technique compares favourably with current methods in terms of computational complexity and has the advantage of stand-alone operation in the edge node, without the need for always-on wireless connectivity making it ideal for IoT wearable devices.

**Index Terms**—Anomaly detection, edge computing, IoT sensors, LSTM, MLP, neural networks, power reduction.

## I. INTRODUCTION

CARDIOVASCULAR diseases (CVD) such as coronary heart disease (CHD), stroke, and other circulatory diseases

Manuscript received September 9, 2021; revised November 18, 2021; accepted December 17, 2021. Date of publication January 4, 2022; date of current version May 9, 2022. This work was supported in part by Microelectronic Circuit Centre Ireland under Grant MCCI-2018-03 and in part by Irish Research Council under Chist-Era Program. This paper was recommended by Associate Editor Sarah Ostadabbas. (*Corresponding author: Deepu John*).

Gawsalyan Sivapalan, Barry Cardiff, and Deepu John are with the School of Electrical and Electronic Engineering, University College Dublin, Dublin 4, Ireland (e-mail: gawsalyan.sivapalan@ucdconnect.ie; barry.cardiff@ucd.ie; deepu.john@ucd.ie).

Koushik Kumar Nundy is with the Think Biosolution Limited, D08 HKR9 Dublin, Ireland (e-mail: kknundy@thinkbiosolution.com).

Soumyabrata Dev is with the School of Computer Science, University College Dublin, Dublin 4, Ireland (e-mail: soumyabrata.dev@ucd.ie).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TBCAS.2021.3137646>.

Digital Object Identifier 10.1109/TBCAS.2021.3137646

account for roughly 30% of all global deaths in any given year. In addition, CVD is a leading cause of premature deaths and the primary driver of morbidity among all non-communicable diseases (NCD) [1]. It is estimated that CVDs cost the European Union, approximately € 169 billion annually, of which 62% is direct costs in the healthcare system and the remainder is productivity loss and informal care [2].

Continuous monitoring of physiological signals such as ECG using IoT enabled wearable devices is widely considered a solution to mitigate the costs and healthcare risks associated with CVDs [3]. Although the concept itself is not new, continuous monitoring of medical-grade electrocardiogram (ECG) is not yet a practical reality due to the large power consumption associated with constant wireless transmission. Analysing the data for detecting potential anomalies at the IoT sensor itself, can reduce the need for constant wireless transmission and thus reduce sensor power consumption.

There are several methods [4] reported in literature, for automatic multi-class ECG classification and simple anomaly detection using signal processing [5] and machine learning techniques [6]–[10]. However, many existing works often exhibit several flaws that make them non-ideal for real-world implementations. Several works attempt to classify ECG beats into multiple classes i.e. (Normal:N, Ventricular:V, Supra-Ventricular:S, Fusion:F, and Unclassified:Q). Attempting multi-class classification on edge sensors may not be ideal and mostly redundant, due to the computational complexity involved. In addition, from the perspective of the user of such a device, multi-class classification brings limited added value compared to simple anomaly detection. Some ([6], [11]) have implemented binary beat classification (Normal vs Abnormal), however, the computational complexity involved is still high and may result in large power consumption. Association for the Advancement of Medical Instrumentation (AAMI) [12] has established protocols to test medical instruments and yet several existing works ([13], [14]) have not taken these into account. Typical sampling rate of ECG in real-world sensors is around 250 Hz, while all the existing works use MIT-BIH records in its original, but rather unusual sampling rate of 360 Hz for performance evaluation and this results in unrealistic, but higher performance. MIT-BIH records are highly imbalanced in terms of class distribution, and these are not carefully handled by much existing work ([15], [16]). In real world scenarios, this will introduce sampling bias and overfitting issues etc to these reported works. Many works have failed to

produce a fixed-point model, which is the common and cost-effective environment in most IoT devices. The conversion of the floating point algorithms is subject to quantization errors and performance degradation, which isn't addressed well in existing works. All the above identified research gaps along with the issue of deteriorating performance of the existing approaches under unseen real-world conditions are addressed in our research.

This work aims to address the aforementioned problems by developing a low complexity machine learning algorithm for binary classification of the ECG signal that can be implemented locally on an IoT sensor. Only when an ECG beat is deemed anomalous by the classifier, wireless transmission will be enabled and thus sensor power consumption can be reduced. In addition, the issue of class imbalance in the MIT-BIH Arrhythmia database is addressed by augmenting the training data using Synthetic Minority Oversampling TEchnique (SMOTE) technique. This reduces disparities in real world performance of the proposed technique compared to the test data.

The proposed novel architecture embodies a Long Short Term Memory (LSTM) based recurrent block to identify the regularity of a typical time-series like data and simple Multi-Layer Perceptron (MLP) based block which learn the underlying relationship between the extracted features such as activation maps of Principal Component Analysis (PCA) coefficients of sequence of beats and ventricular rates. Our novel approach towards simultaneous training of all blocks will push the overall architecture to learn different properties of the sequence and complement each other while making a decision comparable to ensemble learning approach [11].

For this work, we also implemented floating and fixed-point versions of various machine learning building blocks and introduced fast approximate functions and their derivatives to facilitate model development in a floating-point environment and the subsequent mapping to a fixed-point implementation. Our approach is distinct from the widely used TensorFlow approach where there is usually the implementation loss (losses due to pruning, quantization, and look-up table based approximation), which remains in the network. The network proposed in this paper is significantly smaller in footprint (number of parameters and complexity) while achieving state-of-the-art performance. The fixed point model is deployed and tested in an ARM Cortex M4 based Nordic Semiconductor nRF52DK Bluetooth embedded development kit. A significant power saving of  $\simeq 50\%$  is achieved compared to sending every sample through a Bluetooth link. The low complexity of the proposed classifier, system level power reduction of the sensor and a reliable, real-world performance estimation using data augmentation makes the proposed approach a good choice for implementation in IoT edge applications.

The rest of the paper is organized as follows. Section II, explains the currently available solutions and algorithms for ECG classification and anomaly detection. Section III details the ECG dataset used in this work and any pre-processing performed on the data. Section IV provides details of the proposed neural network architecture and its fixed-point implementation. The performance analysis and the comparison of the proposed

method with previous approaches are evaluated and presented in Section V. The conclusions are presented in Section VI.

## II. RELATED WORKS

There are many approaches proposed in the literature for automatic detection and classification of cardiac arrhythmias from ECG signals. Arrhythmia classification is a pattern recognition task that can be done using syntactic or machine learning methods [17]. In traditional syntactic methods, ECG signal features are carefully extracted using signal processing and feature extraction methods such as frequency domain analysis, wavelet transform (WT), and morphological features after which hand-engineered algorithms and rules are applied to the extracted features to detect arrhythmia. Machine learning-based methods such as Decision Tree, Random forest, K-Nearest Neighbour, Support Vector Machine (SVM), Artificial Neural Network (ANN), Reservoir computing with logistic regression (RC), Linear discriminants (LD), Hidden Markov Models (HMM), hyper box classifiers, optimum-path forest, conditional random fields and rules-based models, and Bayesian models use a combination of signal features and morphologies as feature vectors to classify ECG signals [18]. However, the accuracy of these methods strongly depends on the selected learning technique and nature of training data; and the data is often limited with large variation in morphologies between patients.

In Veeravalli *et al.* [17], Fast Dynamic Time Warping (FDTW) with a constraint window is used to formulate the cost feature matrix between the first 30 beats in a patient's record and K-means clustering is used to find the max cluster to nominate a beat as the global normal beat for that particular patient. Thereafter, DTW distance between all the incoming beats with respect to the selected global normal beat were computed. Further, anomalies in the data are detected using a Hampel filter. However, the approach fails to address the case when there are no multiple classes (i.e. Normal and Abnormal) present at the initial clustering phase and most occurring beats may also not always be the clinical normal beat. In addition, K-means clustering is an NP-hard problem and performance evaluation was done using only 15 records selected from the MIT-BIH arrhythmia database.

Zadeh *et al.* [13] uses a bandpass filter to do the pre-processing and an SVM classifier based on features from a Continuous Wavelet Transform is used. The approach has achieved 97% of Normal (N) vs Abnormal (S, V, F, and Q) test accuracy over 17,784 beats from a limited set of 8 selected patient records (118, 124, 207, 208, 209, 214, 222, and 223). Similarly in Jiang *et al.* [14], a block based neural network has been used with Hermite transform features over 49,600 selected beats to achieve 95.6% accuracy in identifying abnormal beats. However, the beat selection criteria used in this work were not specified.

Dan Li *et al.* [15] has introduced a 1D Convolution Neural Network (CNN) to classify ECG signals and achieved more than 98% test accuracy on selected 13,200 beats with a balanced down sampling of data to have an equal probability set of AAMI classes. Wavelet decomposition is used for pre-processing ECG signals and a SoftMax classifier is used in the neural network.

This approach is purely dependent on the local morphological information and ignores the simple and rich temporal features making it unfit for generalized tests. Kiranyaz *et al.* [16] have proposed an adaptive 1D CNN which is trained with both global and patient-specific data. The global part of the training set contains 245 representative beats, which includes 75 from each type: N, S, and V, and 13 from F and 7 from Q randomly sampled from the first 20 records (100 – 124). The first 5 min of data from each record is used for patient-specific training for that subject. For testing all 44 records except the paced records were considered and a 97% (N Vs A (S, V, F, and Q)) test accuracy is achieved. However, the approach heavily depends on the patient-specific training data and its characteristics during the first 5 min interval.

In [11], a Recurrent Neural Network (RNN) cell-based novel architecture is proposed to capture the temporal and spatial patterns of ECG signals. It uses ECG WT coefficients and RR interval properties in one RNN pipeline (Model Alpha) and PCA components from a concatenated vector of (WT, downsampled ECG beat and RR interval-based feature) in another RNN based pipeline (Model Beta). Each model has been trained individually and blended with the result of a new MLP network for better performance. A patient-specific training procedure has been followed along with global data collected from records 100-124 of MIT-BIH arrhythmia database. A total of 49,632 beats were tested to achieve 98% test accuracy and an F1 score of  $\sim 92\%$ . However, this method requires two leads of ECG signal which is difficult to acquire with a low profile wearable device [19]. Also, it is uncommon in that different portions of the concatenated feature vectors are presented to the RNN on each invocation whereas normally the RNN input vector contains the same set of features each time.

Das and Ari [20] have proposed a combined feature vector of 4 temporal features (pre-RR, post-RR, local-avg-RR, and global-avg-RR), 8 S-transform based features, and 20 wavelet decomposition based statistical features acquired from a single ECG beat. The architecture proposed uses an MLP based neural network to classify ECG beats. A six fold cross-validation test over 24 records from MIT-BIH arrhythmia database achieved 94.5% accuracy after patient specific training with 5 mins of data from respective ECG records and 200 global random training beats from the first 20 records.

Finally, in [6] a Two-Stage Neural Network (TSNN) that achieves 97.8% and 98.6% test accuracy over 48,310 individual beats, without and with biased training respectively is proposed. The first stage of the network takes an input of raw ECG beats and an MLP network classifies the beat into a Normal or Abnormal beat. The Abnormal beats from first stage will be fed to a second stage, where a CNN classifies these beats into AAMI classes N, S, V, F, and Q.

For performance evaluation of an ECG classifier, data from the same subject shouldn't be used for both training and testing. This is to ensure that the performance of the classifier on previously unseen records are evaluated. In addition, the maximum duration of a dataset used for training should follow the limits imposed by the AAMI protocol. Although many works in the literature strongly obey these tenets, only a few authors have

taken explicit precautions to follow the AAMI protocol precisely when reporting results. This can make fair comparisons between published works difficult.

### III. DATA SET AND PRE-PROCESSING

In this study, the MIT-BIH arrhythmia database [21] containing 48 ECG records excluding the paced<sup>1</sup> beat records are used for performance evaluation. Twenty-three of the records are intended to serve as a representative sample of routine clinical recordings and the remaining 25 records contain complex ventricular, junctional, and supraventricular arrhythmias. The records are bandpass filtered at 0.1-100 Hz and sampled at 360 Hz. There are over 100,000 labeled beats of 15 different heartbeat types. Each record has two ECG leads. The first lead is modified limb lead II (ML II) and the second lead is modified lead V1 or in some cases V2, V4, or V5. Two or more cardiologists independently annotated each record of 30-minute duration selected from 24-hour recordings [6], [14]. The database defines 15 types of beats. For the purpose of this work, we group the beats labeled as Supra Ventricular Ectopic beats (SVEB - S), Ventricular Ectopic Beats (VEB - V), Fusion Beat (F), and Unclassified Beat (Q) as 'Abnormal' and the remaining beats as 'Normal'. This categorization is consistent with AAMI standards.

ECG is usually affected by various noises like baseline wander (low-frequency noise in the range of 0-0.3 Hz), electrode contact noise, motion artifacts, power line interference (PLI) etc. which affects the efficacy of signal analysis [22]. Many real-world ECG devices perform baseline wander and PLI removal during acquisition and present a *clean* ECG signal at a typical sample rate of 250 Hz. To emulate this we perform the following data processing steps:

- Discrete Wavelet Transform (DWT) based denoising [18]
- PLI removal using a standard IIR notch filter at 60Hz
- Re-sampling from 360 Hz to 250 Hz.

An illustration of ECG noise removal using the above steps is shown in Fig. 1 .

Additionally, we chose to use R-peak location annotations from MIT-BIH database [21] directly instead of implementing our own R-peak detector. Since there are several existing works that achieve good accuracy for R-peak detection [23]–[25], we narrowed the scope of this work and focus exclusively on developing the classifier.

We have extracted a segment with samples extending from 250 ms before to 450 ms after the R-peak location. According to [18], this segment can sufficiently capture the entire beat (including P, T waves). At 250 Hz sampling rate, this segment corresponds to a vector of 175 samples which is the basic unit upon which our algorithm operates every time a R-peak is detected.

<sup>1</sup>Paced beats refer to ECG signals generated by the heart under the help of an external or implanted artificial pacemaker for the patients whose electrical conduction path of heart is blocked, or the native pacemaker is not functioning properly.

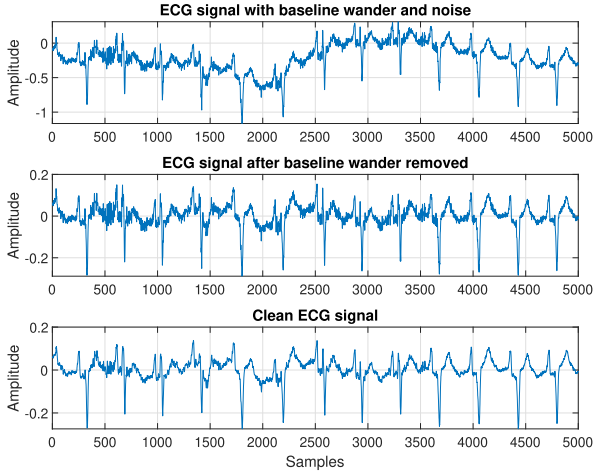


Fig. 1. ECG signal before and after denoising and notch filtering.

#### IV. PROPOSED ARCHITECTURE

This section provides the details of the proposed lightweight neural network, ANNet, starting with the feature vector formulation in Section IV-A, followed by the details of network architecture in Section IV-B. A fixed-point implementation of the proposed network is then discussed in Section IV-C.

##### A. Feature Vectors

Two feature vectors are derived from the original ECG data samples, namely:

- $X$  : Input to the LSTM\_X Layer, and,
- $RR$  : Input to the MLP\_R Layer

1)  $X$ : For each beat, we then compute  $X_i$ , a vector of Principal Component Analysis (PCA) coefficients of length 6 (with respect to the principal Normal=N, RBBB=R, LBBB=L, Ventricular=V, Supra-Ventricular=S, and Fusion=F beats respectively), where  $i$  is the current beat index. An illustration of PCA feature vector construction for 2 random beats is shown in Fig. 4. The same process should be followed with the 5 beat window.

2)  $RR$ : The second feature vector is based on ECG RR interval information, and is defined as vector,  $[RR_i, RR_{i+1}, \overline{RR}_i, RR_{wSDNN_i}, RR_{Index_i}]$ , of length 5. The first two elements of this vector are the RR intervals just prior and after current ECG beat respectively. The third element is the average of 11 RR-intervals from  $RR_{i-9}$  to  $RR_{i+1}$ .  $RR_{wSDNN}$  and  $RR_{Index}$  are Heart Rate Variability (HRV) metrics<sup>2</sup> based on [26], which are defined as:

$$RR_{wSDNN_i} = \sqrt{\frac{1}{10} \sum_{j=-9}^{+1} W_j (RR_{i+j} - \overline{RR}_i)^2}$$

$$\text{where: } W_j = \begin{cases} 10 & j = 0 \\ 1 & \text{otherwise} \end{cases} \quad (1)$$

<sup>2</sup>Here the acronym ‘‘wSDNN’’ stands for ‘‘Weighted standard deviation’’ of the time between peaks. The  $RR_{Index}$  is a unitless ratio or index.

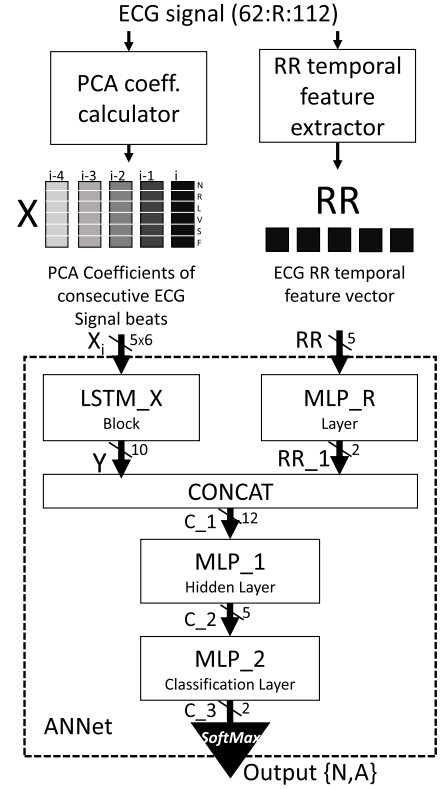


Fig. 2. Top level architecture of ANNet.

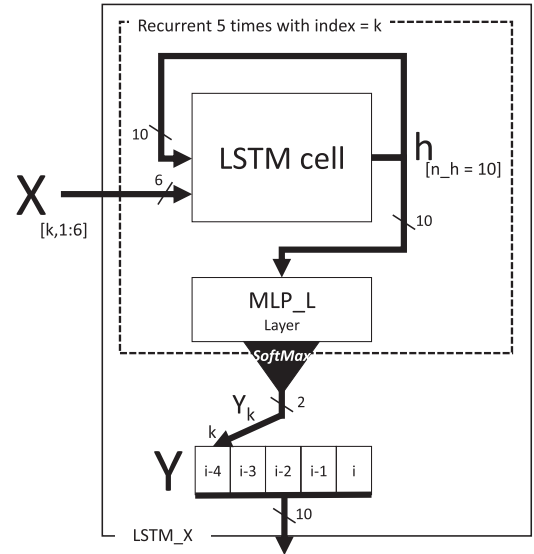


Fig. 3. LSTM cell pipeline as part of LSTM\_X (Fig. 2).

and

$$RR_{Index_i} = 2 \frac{RR_i - RR_{i-1}}{RR_i + RR_{i-1}} \quad (2)$$

##### B. Neural Network Architecture

The proposed network is composed of three main blocks: LSTM\_X, MLP\_R, and a blending block with MLP layers as

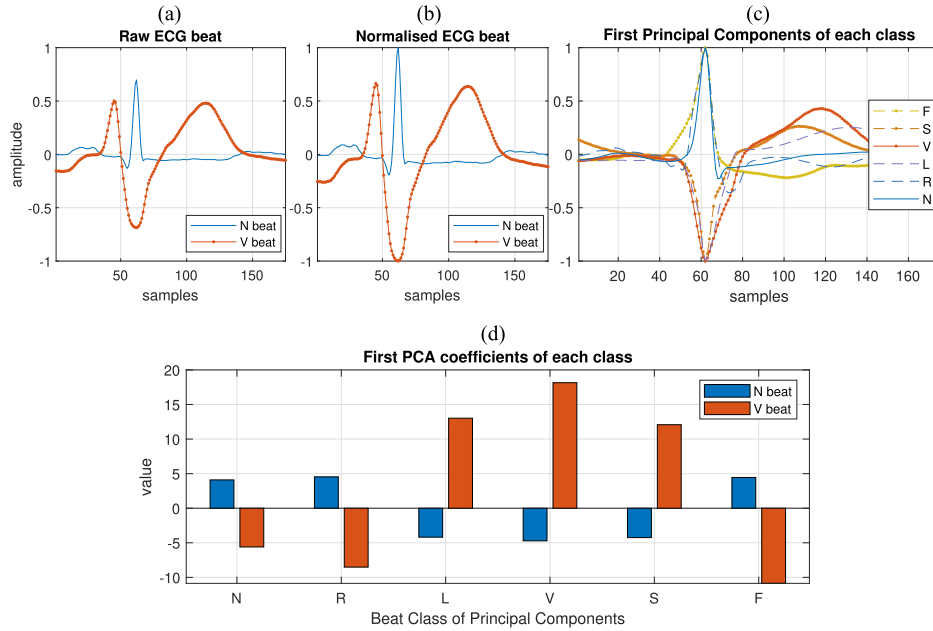


Fig. 4. Principal component analysis for the ECG records. (a) Original raw ECG signal randomly selected representing Normal and Ventricular beat. (b) Normalised ECG beat. (c) Compute the basis vectors (Principal Components) for each class. The most significant first basis vector of each class is shown in the plot. (d) Represent each beat as a coefficient with respect to each basis vector from every class.

seen in Fig. 2. The LSTM based recurrent block is selected to identify the regularity property of the typical time-series signal while all the other simple extracted features use MLP layers to learn the underlying relationship to predict abnormal beat. Simultaneous training of the blocks will result in complementary learning compared to ensemble learning of models [11], i.e. blocks will tend to learn different properties of the sequence and complement each other while making a decision.

The LSTM\_X block in Fig. 3 generates an attention map at its output ( $Y$ ) which keep track of past beats' properties, as in the work by Zhang *et al.* [27]. Fig. 3 shows the LSTM\_X block in more detail. The LSTM cell is executed 5 times for each ECG beat, with inputs being  $X_{i-4}$  through  $X_i$  in sequence. The output of the LSTM cell,  $h$  (of length 10) and internal state vectors are updated for each execution. Additionally, each output vector  $h$  is presented to the MLP\_L layer which generates a vector ( $Y_k$ ) of length two, as shown in Fig. 3. These two outputs are concatenated across the 5 execution cycles, forming the vector  $Y$  of length 10.

In parallel to the LSTM\_X block, there is an MLP\_R layer (Fig. 2) that takes the feature vector, RR, as input, and produces output, RR\_1, of length 2. This output is concatenated with the vector  $Y$  from the LSTM\_X block and is passed to an MLP network (MLP\_1, MLP\_2) with one hidden layer of 5 neurons and 2 outputs ( $C_3$ ). These two outputs are then passed to a SoftMax layer, which classifies the beat as Normal(N) or Abnormal(A).

The learning rate  $\varepsilon$  is set to 0.001 and the  $\beta$  value in the stochastic gradient descent algorithm is set to 0.9 for the network. A mini-batch size 128 is used and the error between the prediction and ground truth is used as the cost for back-propagation learning. More details about the parameters and

complexity of the network are analyzed and the results are presented in Section V.

### C. Fixed Point Implementation

ANNnet was initially implemented and trained in Matlab using floating point arithmetic (denoted as “Float” in Section V). Fixed point representation of the network reduces the complexity so that the model can be deployed on a cost effective low-profile embedded system. To port the algorithm to a fixed point embedded platform, we first replaced the activation functions in the model with fast approximate versions and then mapped these to a fixed point implementation. The model was implemented in C-language for an embedded platform. Also, a bit-accurate and fixed point version of the model was created in Matlab. This allowed us to take the previously trained coefficients from ANNnet's floating point version as a starting point and to *retrain* (in Matlab) the model using fixed point arithmetic. Note that this re-training step also requires derivatives of these fixed point activation functions for the back-propagation process; these are not implemented in the embedded environment, as they are only used during the retraining process. The details of the approximation functions and their mapping to fixed point arithmetic are given in Sections IV-C1–IV-C3.

We used the notation “Qm.n” to denote our fixed point number formatting. Here it is assumed that all quantities are 2's complement signed numbers with “m” integer bits (including the sign) and “n” fractional bits. Accordingly the resolution is  $2^{-n}$  and the range is from  $-2^{m-1}$  to  $+2^{m-1} - 1$ . As we are targeting an embedded implementation, all variables will be either 16- or 32-bit wide. The number of fractional bits to use in the fixed point model of ANNnet was experimentally determined. We performed

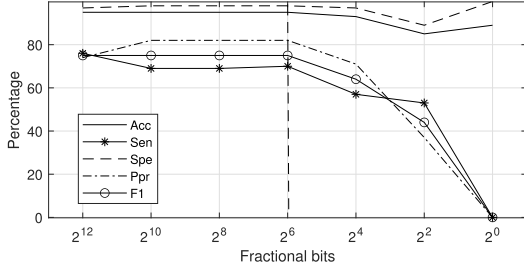


Fig. 5. Global training performance metrics of the proposed network with different level of quantization.

re-training of ANNet for various sizes of fractional bits and observed the performance as shown in Fig. 5. Based on this analysis, the number of fractional bits was chosen to be  $n = 6$ .

1) *Sigmoid Activation Function ( $\sigma$ )*: We used Sigmoid cell activation function for tuning the weights of neurons in ANNet. The below equations show the original sigmoid (3) and its derivative (4) used.

$$\sigma(x) = \frac{\exp(x)}{\exp(x) + 1} \quad (3)$$

$$\frac{\partial}{\partial x} \sigma(x) = \sigma(x)(1 - \sigma(x)) \quad (4)$$

Sigmoid function cannot be easily implemented in a fixed-point environment due to the presence of exponential functions. Therefore, it is translated to an approximate fast version of the function:  $\tilde{\sigma}(x)$ : Elliott Activation Function [28] (5), (6), to facilitate implementation on an embedded device:

$$\sigma(x) \simeq \tilde{\sigma}(x) = \frac{0.5x}{1 + |x|} + 0.5 \quad (5)$$

$$\Rightarrow \frac{\partial}{\partial x} \tilde{\sigma}(x) = \frac{0.5}{(1 + |x|)^2} \quad (6)$$

These can be implemented in fixed-point arithmetic with  $n$  fractional input,<sup>3</sup>  $\hat{x} = \lfloor 2^n x \rfloor$ , and output bits as follows:<sup>4</sup>

$$2^n \tilde{\sigma}(x) \simeq \hat{\sigma}(\hat{x}) \triangleq \left\lfloor \frac{2^{n-1} \hat{x}}{2^n + |\hat{x}|} \right\rfloor + 2^{n-1} \quad (7)$$

$$2^n \frac{\partial}{\partial x} \tilde{\sigma}(x) \simeq \partial \hat{\sigma}(\hat{x}) \triangleq \left\lfloor \frac{2^{3n-1}}{(2^n + |\hat{x}|)^2} \right\rfloor \quad (8)$$

2) *Tanh Activation Function*: It is mainly used inside the LSTM cell as the candidate gate  $C_t$  (Fig. 3) activation function. The original (9) and its derivative (10) are provided below along with its fast approximate fixed point implementation inspired from the work of Anguita *et al.* [29] is used in this work with modification to its coefficients considering bit level manipulation for efficiency.

$$\tanh(x) = \frac{(\exp(2x) - 1)}{(\exp(2x) + 1)} \quad (9)$$

<sup>3</sup>All rounding in this work is truncation.

<sup>4</sup>Note that multiplication by  $\cdot 2^k$ , and  $\cdot 2^{-k}$ , represents  $k$  bit shifts to left and right respectively. Also note that right bit shifts (negative  $k$ ) implicitly round towards zero and so this rounding will not be shown explicitly. The word sizes are chosen so as to mean that no overflows will occur when left shifting.

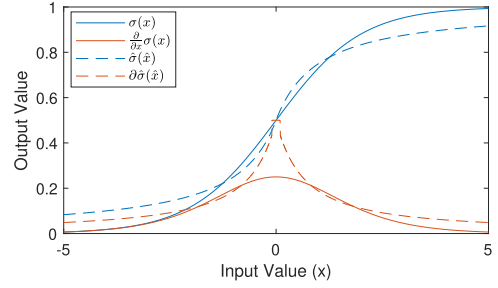


Fig. 6. Illustration of the sigmoid activation function,  $\sigma(x)$ , and its approximation  $\hat{\sigma}(\hat{x})$  and their respective derivatives as computed for  $n = 12$ .

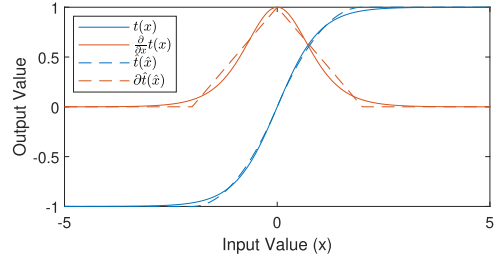


Fig. 7. Floating and fixed point approximation of tanh function.

$$\frac{\partial}{\partial x} \tanh(x) = 1 - \tanh(x)^2 \quad (10)$$

An approximate version of the tanh function is  $t(x)$ :

$$t(x) = \begin{cases} \text{sgn}(x) 1 & |x| \geq 2 \\ \text{sgn}(x) (1 - 0.25(|x| - 2)^2) & \text{otherwise} \end{cases} \quad (11)$$

$$\Rightarrow \frac{\partial}{\partial x} t(x) = \begin{cases} 0 & |x| \geq 2 \\ \text{sgn}(x) (1 - 0.5|x|) & \text{otherwise} \end{cases} \quad (12)$$

These can be implemented in fixed-point arithmetic with  $n$  fractional inputs and outputs as:

$$2^n \tilde{t}(\hat{x}) \simeq \hat{t}(\hat{x}) \triangleq \begin{cases} \text{sgn}(\hat{x}) 2^n & |\hat{x}| \geq 2^{n+1} \\ \text{sgn}(\hat{x}) (2^n - 2^{-(n+2)} (|\hat{x}| - 2^{n+1})^2) & \text{otherwise} \end{cases} \quad (13)$$

$$2^n \frac{\partial}{\partial x} t(x) \simeq \partial \hat{t}(\hat{x}) \triangleq \begin{cases} 0 & |\hat{x}| \geq 2^{n+1} \\ \text{sgn}(\hat{x}) (2^n - 2^{-1} |\hat{x}|) & \text{otherwise} \end{cases} \quad (14)$$

An illustration of the original and approximate versions for sigmoid and tanh functions are given in Fig. 6 and Fig. 7 respectively.

3) *SoftMax Function*: This is primarily used in the classification layers. It is not related to a single-neuron output rather it computes a normalised vector,  $\mathbf{s}$  (15), based on the vector,  $\mathbf{x}$ , of outputs from the last fully connected layer:

$$s_i(\mathbf{x}) = \frac{\exp(x_i)}{\sum_k \exp(x_k)} \quad \forall 0 \leq i < |\mathbf{x}| \quad (15)$$

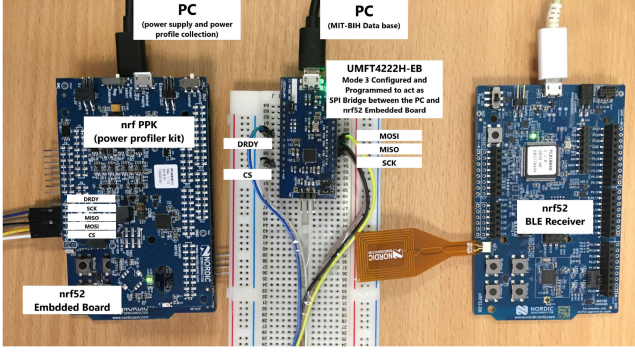


Fig. 8. Test setup used for experimental verification.

$$\Rightarrow \frac{\partial}{\partial x_i} s_i(\mathbf{x}) = s_i(\mathbf{x})(1 - s_i(\mathbf{x})) \forall 0 \leq i < |\mathbf{x}| \quad (16)$$

where  $|\mathbf{x}|$  is the cardinality of the vector  $\mathbf{x}$ , which for our binary classifier we have  $|\mathbf{x}| = 2$ .

We used normalisation with offset '1' as approximate version of the SoftMax function,  $\tilde{s}(\mathbf{x})$ :

$$\tilde{s}_i(\mathbf{x}) = \frac{1 + x_i}{|\mathbf{x}| + \sum_k x_k} \quad \forall 0 \leq i < |\mathbf{x}| \quad (17)$$

$$\Rightarrow \frac{\partial}{\partial x_i} s_i(\mathbf{x}) = \frac{1}{1 + x_i} s_i(\mathbf{x})(1 - s_i(\mathbf{x})) \quad \forall 0 \leq i < |\mathbf{x}| \quad (18)$$

These can be implemented in fixed point arithmetic with  $n$  fractional inputs and outputs as before:

$$2^n \tilde{s}_i(\hat{\mathbf{x}}) \simeq \hat{s}_i(\hat{\mathbf{x}}) \triangleq \left\lfloor \frac{2^n(2^n + \hat{x}_i)}{2^n|\hat{\mathbf{x}}| + \sum_k \hat{x}_k} \right\rfloor \quad (19)$$

$$2^n \frac{\partial}{\partial x_i} s_i(\mathbf{x}) \simeq \partial \hat{s}_i(\hat{\mathbf{x}}) \triangleq \left\lfloor \frac{\hat{s}_i(\hat{\mathbf{x}})(2^n - \hat{s}_i(\hat{\mathbf{x}}))}{2^n + \hat{x}_i} \right\rfloor \quad (20)$$

#### D. Embedded System

The nRF52DK development kit from Nordic Semiconductor with Segger Embedded Studio is used to deploy and test ANNet in real-time. The experimental setup is shown in Fig. 8. Nordic nRF PPK (Power Profiler Kit) shield along with the development kit is connected to a PC and is used to measure the power consumption using the Nordic Power Profiler Software. An FTDI FT4222H SPI bridge is configured to act as an ECG sensor by transferring signals from MIT-BIH arrhythmia test records stored in the PC for fair emulation. The ECG signal is pre-processed offline as per Section III creating a 250 Hz sampled signal and quantized to 16-bit signed fixed point numbers, composed of 4 integer bits<sup>5</sup> followed by 12 fractional bits (Q4.12 format<sup>6</sup>) along with a flag indicating R-peak locations. This is the input to our embedded, fixed-point ANNet classifier. The overall flow diagram is provided in Fig. 9.

<sup>5</sup>We avoid overflows by truncating  $\pm(2^3 - 2^{12})$ , which was verified to be sufficient to capture the entire MIT database.

<sup>6</sup>i.e. 16 bit numbers on the range  $[-8, +8]$  with resolution  $2^{-12}$

1) *Preprocessor and Feature Extraction*: The input, output, and internal calculations within these blocks are implemented using 16 bit, Q4.12 binary fixed-point arithmetic. A ring buffer (length 1000) is maintained to store incoming samples and another ring buffer (length 11) is used to record the R peak locations with the sample buffer. For each beat detected, a window of 175 samples (62:R:112) is taken and normalized, and then the PCA coefficients are calculated using the stored principal components. Additionally, the RR features (Section IV-A2) are calculated from the R peak locations. The combined feature vector is then fed to the ANNet module for classification.

2) *ANNet*: The ANNet module uses 32 bit arithmetic represented in Q26.6 fixed-point format corresponding to a resolution or  $2^{-6}$ . The large word size is used to avoid overflows in intermediate results throughout the ANN. The classification results are then sent back to the PC for verification of the results and to obtain performance metrics. The detailed operation of this block was described in Section IV-B.

## V. EXPERIMENTAL SETUP & RESULTS

For evaluating the proposed network, we have used MIT-BIH arrhythmia database [30]. In total, 100,661 ECG beats from the database were used for this study. In accordance with De Chazal *et al.* [31]'s recommendation, we have split the database into two distinct sets namely DS1 and DS2 for evaluation purposes. DS1 consists of 22 ECG records with patient numbers 101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223 and 230 while DS2 consists of the remaining 22 ECG records with patient numbers 100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233 and 234. DS1 and DS2 include a blend of representative samples of routine clinical recordings and uncommon but clinically significant arrhythmias. This split allows both training and testing sets to have the same approximate proportion of beat classes.

#### A. Augmentation of Imbalanced Dataset

The MIT-BIH dataset is severely imbalanced with more than 90% beats of the type 'Normal' class. This can affect the real-world performance and therefore, it is beneficial to balance the training data so that the underlying proportion of beat type in the training set will not have a major impact while tuning the network parameters. Some of the previous approaches use the Conditional Data Grouping, Biased Training, or downsampling method to address the imbalance and identity issue of the training database [7]. However, it is visible that there is a significant variation among beats, and selecting very few to balance the classes will limit the learning of the network to a particular subspace of available data representation. Hence, we used Synthetic Minority Oversampling TEchnique (SMOTE [32], [33]) for augmenting the dataset by creating synthetic data of the minority class [34].

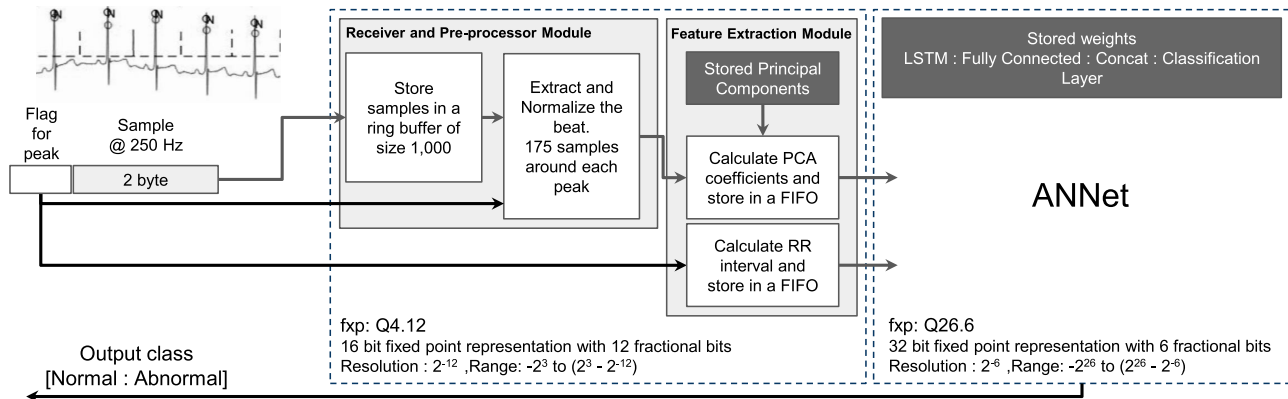


Fig. 9. Flow diagram of the implementation of the algorithm on the embedded environment.

TABLE I  
PERFORMANCE OF THE PROPOSED FLOATING POINT AND FIXED POINT NETWORKS AT EACH TRAINING PHASES

Format	Train. Scope	TN	FN	TP	FP	Acc	Sen	Spe	Ppr	F1	G
Float	5 min - Global - SMOTE	29,367	317	4,269	7,527	0.81	0.93	0.80	0.36	0.52	0.58
Float	5 min - Global	34,194	712	3,874	2,700	0.92	0.84	0.93	0.59	0.69	0.71
Float	5 min - Pat.Spe	35,952	496	4,090	942	0.97	0.89	0.97	0.81	0.85	0.85
Fixed	Converted from 5 min Global Float	33,105	2,994	1,592	3,789	0.84	0.35	0.90	0.30	0.32	0.32
Fixed	5 min - Global	35,694	1,099	3,487	1,200	0.94	0.76	0.97	0.74	0.75	0.75
Fixed	5 min - Pat.Spe	36,088	800	3,786	806	0.96	0.83	0.98	0.82	0.83	0.83
<b>Considering 70% of ECG beats for training and remaining 30% as test set</b>											
Fixed	70% - Global	12,917	395	1,184	417	0.95	0.75	0.97	0.74	0.74	0.74
Fixed	70% - Pat.Spe	13,123	212	1,367	211	0.97	0.87	0.98	0.87	0.87	0.87

TN: True Negative, FN: False Negative, TP: True Positive, FP: False Positive.

Acc: Accuracy, Sen: Sensitivity, Spe: Specificity, Ppr: Positive Prediction, F1: F1-Score, G: G-Score.

## B. Evaluation Method

This section describes the steps which we have followed to train and evaluate the performance of ANNet. Table I, illustrates the performance at each step from floating point global level training to fixed point patient specific training.

1) *Global Training*: We used DS1 from the MIT-BIH database, which consists of 50,982 beats, for training our network. Many prior works have used DS1 as the global training set and hence no performance test is conducted over this subset of data.

The SMOTE algorithm is used to address the class imbalance in DS1 and the total number of beats have been increased to 91,729 after augmentation. After global training with this augmented DS1 dataset, the network achieved  $\approx 81\%$  of accuracy with  $\approx 52\%$  of F1 score over the DS2 test sets. This allows the network to learn without getting biased due to the large presence of the normal(N) beats. Once the network is trained without any influence from the imbalanced data, the SMOTE synthetic data is removed from the training set and the network is trained globally with the original training set. This step improves the performance of the network to  $\approx 92\%$  of accuracy with  $\approx 69\%$  of F1 score.

The model is converted to a fixed point version which introduces approximation errors due to the use of our fast-approximation functions (Section IV-C) and quantization errors. This results in a performance drop from 92% to 84% in accuracy,

as seen in Table I. Using the model as a starting point, the fixed-point model was re-trained in a fixed-point bit-accurate environment to optimise performance. This step improves the fixed-point model global performance to 94%.

2) *Patient Specific Training*: The patient-specific training and testing have been performed using the DS2 dataset and the results of individual records are presented in Table II. The training set is created with the first 5 min data from individual records in DS2 according to AAMI recommendations. The network performance is evaluated among the other approaches and presented below in Section V-C. During this training phase, the learning rate  $\epsilon$  is set to a low value so that the layer weights will not deviate much from prior learning with augmented DS1 dataset. In comparison to the previous floating-point global training, this step improves the accuracy to  $\approx 97\%$  with an F1 score of  $\approx 85\%$ . Similarly, the fixed-point patient specific training from the global fixed point model as starting point, improved the overall accuracy to 96%.

A total of 41,480 beats from DS2 is used for unseen-testing and this excludes the 5-minute patient specific training set. The total number of True Positive (TP), False Negative (FN), False Positive (FP), True Negative (TN) are given in Table II. Further, training and testing with a 70:30 split also conducted on the proposed model and the results are provided in Table I only for study purposes.



TABLE II  
PATIENT SPECIFIC TRAINING SUMMARY OF FLOATING POINT NETWORK FOR  
THE DS2: MIT-BIH ARRHYTHMIA DATASET

Record	TN	FN	TP	FP	Test Acc.
100	1,857	4	26	13	0.99
103	1,725	1	1	1	1.00
105	2,094	6	28	26	0.99
111	1,765	0	1	9	0.99
113	1,498	0	5	2	1.00
117	1,282	0	1	0	1.00
121	1,434	0	2	123	0.92
123	1,261	0	3	3	1.00
200	1,423	27	703	14	0.98
202	1,671	22	49	128	0.92
210	1,964	60	133	44	0.95
212	2,283	0	0	1	1.00
213	2,099	151	339	109	0.90
214	1,621	8	207	41	0.97
219	1,710	17	41	4	0.99
221	1,668	15	301	35	0.98
222	1,626	49	160	279	0.84
228	1,392	4	301	35	0.98
231	1,264	0	0	10	0.99
232	298	39	1,127	20	0.96
233	1,781	63	639	75	0.95
234	2,236	30	23	0	0.99
Gross	35,952	496	4,090	942	0.97

### C. Performance

The proposed implementation achieved higher performance compared to many other works while maintaining lower complexity despite all the others using higher sampling rate (360 Hz) and floating point implementations. [13], [17] and [15] has reported performance, which is on par or marginally ( $\sim 1\%$ ) higher than ours, however, the training and testing do not follow AAMI recommendations [12] and consist of less than half the numbers of test beats for evaluation. Similarly [11], which reports marginally higher accuracy, involves irregular use of LSTM cells and uses two lead ECG data which is difficult to acquire in a wearable device. [6] and [16] exclusively depends on the local morphology of a single beat ECG segment and uses a CNN based architecture. These approaches wouldn't be able to extract and make use of the RR interval information and isn't a recommended practice as the morphologies vary a lot across patients and may result in poor performance under a new unseen environment. In addition, [11], [16], [35] and [6] consume more than a million instruction cycle to classify single beat to achieve that marginal performance while [35]'s 1D-CNN architecture is used irregularly over temporal and morphology feature extraction. MIT-BIH dataset, which is used for performance evaluation by most of the above approaches, is an unbalanced dataset. Training with an unbalanced dataset is not properly addressed in these works, which makes the prediction by these methodologies subject to sampling bias. On the flip side, the test accuracy also depends on the selection of test records and their properties, therefore, achieving marginal performance does not signify the effort and guarantees the best model. For example, [20] architecture which requires very low ( $\sim 2.5 K$ )

instruction cycles for beat classification performs a 6 fold test and reports the average performance which makes the results less reliable. In essence, most of the existing works achieve comparatively lower performance than ours and show several implementation drawbacks as discussed above.

The proposed network achieves a higher level of accuracy in routine clinical records, while the overall performance ( $\simeq 97\%$ ) in the DS2 recordings has not surpassed the level of state-of-the-art mainly due to the presence of regular abnormal signal. It is observed that the signals which belong to the normal category sometimes do have different morphology, which introduces significant changes in the PCA coefficients and thus affecting the performance.

According to the MIT-BIH arrhythmia database [30], it is expressed that records belonging to 200, 203, 214, and 222 are corrupted with the occasional burst of noise and artifacts. In 214, 215, and 228, there exist few episodes of tape slippage, and 219 and 232 are with long pauses. There are occurrences of axis shifts in records 203 and 223. The records 201, 212, 213, and 223 show either abnormally high or slow cycles for the relevant beat type, making it more complex to analyze. Records 203 and 207 are included in the global training set, despite their classification being onerous even for experienced cardiologists to manually annotate. This results in accuracy of the model appearing lower than in typical conditions.

The records 121, 202, 213, and 222 are the worst performing among the DS2 dataset. In 222, the incorrectly classified beats belong to PAC - Premature Atrial Contractions (Physionet-A) which is not a dangerous arrhythmia class. In records 213, the beats which are classified as normal belong to the Fusion category and these Fusion PVC beats are almost identical to normal in morphology. In addition, records 103, 105, 121, 123, 212, 222, and 234 do not have any abnormal beat sample to represent in the first 5 min training episodes and therefore, it became almost impossible to do the patient-specific training with the first 5 min episode as suggested by AAMI.

### D. Complexity Analysis

Referring to ANNet description in Section IV-B, we have enumerated the input vector sizes and the number of trained coefficients in each of the layers of our proposed ANN architecture in Table IV below. In total, we have 791 parameters which is comparatively very low with respect to many previous approaches in Table III.

Table III applies various assumptions in calculating the number of parameters and instructions per classification for those algorithms which lacks detailed architectural information. The values are calculated based on the confusion matrix provided in the related publications and converting it to  $N(N, L, R, j, e)$  versus A classes. The other parameters and instructions were also calculated based on the revealed architecture for classification of N versus A only. The input feature vector extraction and other operations such as filtering are not considered for complexity calculations since it is considered a separate block in most cases and most models are not fully dependent on specific features. To compute the complexity in terms of instruction per classification,

TABLE III  
PERFORMANCE COMPARISON OF THE PROPOSED METHOD WITH OTHER APPROACHES

Method	Feature	Algorithm	# Test Beats	Test Accuracy	# Parameter	# Multiplication	# Addition	# Division & Other Op.	# Instruction per classification
Das and Ari [20]	RR interval / WT / S Transform	WT / ST / MLP-NN	24*25min	0.945	2,277*	1,184*	1,216*	84*	<b>2,484*</b>
Kiranyaz et al. [16]	Raw ECG beat samples	Adaptive 1D-CNN	49,557	0.97	1,809*	929,650 <sup>[6]</sup>	929,650 <sup>[6]</sup>	20*	1,859,320*
Veeravalli et al. [17]	DTW Cost w.r.t N beat	DTW / K-Mean / Hampel filter	n/a	n/a	<b>262</b>	76,116	76,134	18	152,268
Dan Li et al. [15]	Raw ECG beat samples	1D-CNN	13,200	0.98	66,029	198,180	198,180	20,376*	416,736*
Xia and Xie [35]	Raw ECG beat samples / RR interval	1D-CNN	58,584	0.99	206,020	1,289,312 <sup>[6]</sup>	1,289,312 <sup>[6]</sup>	16*	2,578,640*
Saadatnejad et al. [11]	Wavelet Transform / RR interval	WT / PCA / RNN	49,632	0.977	45,991	448,040	670,640	593,468*	1,712,148*
Wang et al. (w/o BT) [6]	Raw ECG beat samples	MLP / CNN	48,310	0.978					180,208* (Stage1)
Wang et al. (w BT) [6]	Raw ECG beat samples	MLP / CNN	48,310	0.986	90,352	90,100	90,100	8*	$\gg 1.3\text{Mn}$ (TSNN)
This Work (Proposed)	PCA / RR interval	LSTM / MLP-NN	22*25min	0.97	<b>791</b>	5,190	7,340	7,016*	<b>19,546*</b>

The Table applies various assumptions in calculating the value for Accuracy, No of Parameters, No of Operations (Total No of Additions and Multiplication similar to [6], Division, and other operation such as sampling and SoftMax calculations) and No of Instructions per classification for the algorithm in the literature which has minimum available information to do the same as described in Section V-D.

\* This value is an estimate based on assumptions over the operations involved and architectures unveiled.

TABLE IV  
NEURAL NETWORK PARAMETER COUNT

Layer	Size	No of Parameter
<b>LSTM_X pipeline</b>		
Input	6 * 5	n/a
LSTM	10	$4 * 10 * (6 + 10 + 1) = 680$
MLP	2	$2 * (10 + 1) = 22$
Output	2 * 5	n/a
<b>MLP_R pipeline</b>		
Input	5	n/a
MLP	2	$2 * (5 + 1) = 12$
Output	2	n/a
<b>Blending Block</b>		
CONCAT	10 + 2	n/a
MLP	5	$5 * (12 + 1) = 65$
MLP	2	$2 * (5 + 1) = 12$
Output	2	n/a
<b>Total</b>		<b>791</b>

we have reasonably assumed that for addition and multiplication, it will cost one instruction cycle and for division, it costs 2 instruction cycles based on the Cortex M4 Technical Reference Manual [36]. The instruction for all activation functions is considered with one division and one addition operation only, even though different activation functions behave differently within different regions of the input value. These assumptions are uniformly applied across all the other methodologies despite the fact that others are using floating point architecture and therefore

consume more clock cycles or instruction cycles, particularly in activation and SoftMax classification functions with exponential value calculations.

For example, in [6], a TSNN was proposed where the first stage does binary Normal / Abnormal classification similar to our design, but the parameter count is significantly higher and the MLP based network used has a complexity of approximately 180 k instructions cycles per classification whereas the network proposed in this work consumes less than 20 k instruction cycles, a 9-fold decrease. More comprehensive comparisons are provided in Table III.

### E. Power Consumption Analysis

The current consumption of the proposed system on nRF52 DK is measured in real-time using an nRF PPK shield and the results are shown in Fig. 10. For a baseline reference, the average current consumption for sending all the samples through BLE(Bluetooth Low Energy) without executing the ANNet ('NO ANN' in Fig. 10) is  $112.68\mu\text{A}$  for a 30 minute record sampled at 250 Hz; this can be seen as a flat reference line in Fig. 10.

Operation of the ANN increases the current consumption expended on microprocessor computations which is expected to be dependent on the number of beats per second. However, this is offset by the mechanism of only triggering Bluetooth transfers when anomalous beats are detected (both true and false positives). These two effects can be clearly seen in Fig. 10 where it is evident that average current is correlated positively with both

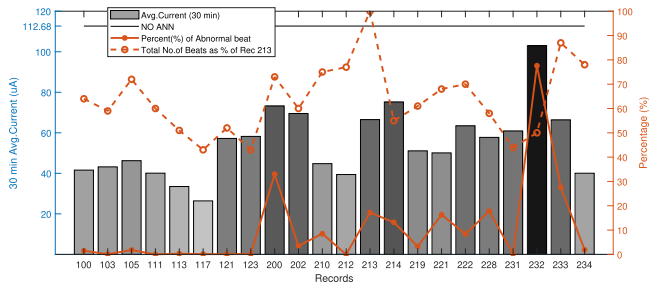


Fig. 10. Average current consumption in  $\mu A$  over 30 minutes of DS2 records. We can see that record 213 has the highest average heart beat rate (scaled to 100%). The records with the largest number of abnormal beats (e.g., 232) show the highest current.

the average number of beats per second as expected and with the percentage of abnormal beats.

For example, record 117 has the lowest number of beats (1,534) and a very low abnormal beat rate consumes only  $26.4\mu A$  on average. However, record 232 has 77.6% of its beats being classified as abnormal and consumes the highest high average current of  $103\mu A$ . Another example is record 213, with 3,549 total beats with 17.2% abnormality consumes approximately  $66.5\mu A$  on average.

The test results prove the significant power efficiency, or the order of  $\simeq 50\%$ , are typically achieved across many of the records compared to the alternative of continuously wireless transmission of the ECG signal. In real-life personal health monitoring, the occurrence of arrhythmias is extremely sporadic and power savings are expected to be higher than 50% ensuring the longevity of the battery.

## VI. CONCLUSION

In this study, we have proposed a lightweight neural network to classify ECG into Normal and Abnormal beats. The network takes an input feature vector created from the coefficients of the PCA using 5 consecutive beats and a temporal feature vector created from the ventricular R-R interval rate. The proposed method is able to achieve low complexity with higher anomalous signal detection accuracy in routine clinical recordings and reasonable accuracy in complex records. The algorithm was ported to an embedded platform by replacing various activation functions with approximations and the mapping to fixed point after retraining resulting in very little implementation loss and a design having the lowest computationally complexity with respect to the state of the art.

Compared to continuous data transmission, we demonstrated that gating the wireless transmission using a binary classifier so that only anomalous beats are transmitted, can significantly reduce the overall system power consumption

## REFERENCES

- [1] W. A. Zoghbi *et al.*, "Sustainable development goals and the future of cardiovascular health: A statement from the global cardiovascular disease taskforce," *J. Amer. Heart Assoc.*, vol. 3, no. 5, 2014, Art. no. e000504.
- [2] J. Leal *et al.*, "Economic burden of cardiovascular diseases in the enlarged European union," *Eur. Heart J.*, vol. 27, no. 13, pp. 1610–1619, Feb. 2006.

- [3] D. L. T. Wong *et al.*, "An integrated wearable wireless vital signs biosensor for continuous inpatient monitoring," *IEEE Sensors J.*, vol. 20, no. 1, pp. 448–462, Jan. 2020.
- [4] Y. Wei *et al.*, "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.
- [5] E. J. da *et al.*, "ECG-based heartbeat classification for arrhythmia detection: A survey," *Comput. Methods Programs Biomed.*, vol. 127, pp. 144–164, 2016.
- [6] N. Wang, J. Zhou, G. Dai, J. Huang, and Y. Xie, "Energy-efficient intelligent ECG monitoring for wearable devices," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 5, pp. 1112–1121, Oct. 2019.
- [7] Y. Zhao, Z. Shang, and Y. Lian, "A  $13.34\mu W$  event-driven patient-specific ANN cardiac arrhythmia classifier for wearable ECG sensors," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 2, pp. 186–197, Apr. 2020.
- [8] D. L. T. Wong, Y. Li, D. John, W. K. Ho, and C. H. Heng, "Resource and energy efficient implementation of ECG classifier using binarized CNN for edge AI devices," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2021, pp. 1–5.
- [9] M. Saeed *et al.*, "Evaluation of level-crossing ADCs for event-driven ECG classification," *IEEE Trans. Biomed. Circuits Syst.*, to be published, doi: [10.1109/TBCAS.2021.3136206](https://doi.org/10.1109/TBCAS.2021.3136206).
- [10] J. Huang, S. Lin, N. Wang, G. Dai, Y. Xie, and J. Zhou, "TSE-CNN: A two-stage end-to-end CNN for human activity recognition," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 1, pp. 292–299, Jan. 2020.
- [11] S. Saadatnejad, M. Oveisi, and M. Hashemi, "LSTM-based ECG classification for continuous monitoring on personal wearable devices," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 2, pp. 515–523, Feb. 2020.
- [12] A.-A. EC57, *Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms*. Arlington, VA, USA: Association for the Advancement of Medical Instrumentation, 1998.
- [13] A. E. Zadeh and A. Khazaei, "High efficient system for automatic classification of the electrocardiogram beats," *Ann. Biomed. Eng.*, vol. 39, no. 3, pp. 996–1011, 2011.
- [14] W. Jiang and S. G. Kong, "Block-based neural networks for personalized ECG signal classification," *IEEE Trans. Neural Netw.*, vol. 18, no. 6, pp. 1750–1761, Nov. 2007.
- [15] D. Li, J. Zhang, Q. Zhang, and X. Wei, "Classification of ECG signals based on 1D convolution neural network," in *Proc. IEEE 19th Int. Conf. e-Health Netw. Appl. Services (Healthcom)*, 2017, pp. 1–6.
- [16] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, Mar. 2016.
- [17] B. Veeravalli, C. J. Deepu, and D. Ngo, "Real-time, personalized anomaly detection in streaming data for wearable healthcare devices," in *Handbook of Large-Scale Distributed Computing in Smart Healthcare*, Berlin, Germany: Springer, 2017, pp. 403–426.
- [18] E. J. d. S. Luz *et al.*, "ECG-based heartbeat classification for arrhythmia detection: A Survey," *Comput. Methods Programs Biomed.*, vol. 127, pp. 144–164, 2016.
- [19] C. J. Deepu, X. Y. Xu, D. L. T. Wong, C. H. Heng, and Y. Lian, "A  $2.3\mu W$  ECG-on-chip for wireless wearable sensors," *IEEE Trans. Circuits Syst. II: Exp. Briefs*, vol. 65, no. 10, pp. 1385–1389, Jul. 2018.
- [20] M. K. Das and S. Ari, "ECG beats classification using mixture of features," *Int. Scholarly Res. Notices*, vol. 2014, 2014, Art. no. 178436. [Online]. Available: <https://doi.org/10.1155/2014/178436>
- [21] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, May/June 2001.
- [22] S. M. Anwar, M. Gul, M. Majid, and M. Alnowami, "Arrhythmia classification of ECG signals using hybrid features," *Comput. Math. Methods Med.*, vol. 2018, pp. 1–8, 2018.
- [23] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985.
- [24] J. Li, A. Ashraf, B. Cardiff, R. C. Panicker, Y. Lian, and D. John, "Low power optimisations for IoT wearable sensors based on evaluation of nine QRS detection algorithms," *IEEE Open J. Circuits Syst.*, vol. 1, pp. 115–123, 2020, doi: [10.1109/OJCS.2020.3009822](https://doi.org/10.1109/OJCS.2020.3009822).
- [25] C. J. Deepu, X. Zhang, C. H. Heng, and Y. Lian, "A 3-lead ECG-on-chip with QRS detection and lossless compression for wireless sensors," *IEEE Trans. Circuits Syst. II: Exp. Briefs*, vol. 63, no. 12, pp. 1151–1155, Dec. 2016.
- [26] F. Shaffer and J. Ginsberg, "An overview of heart rate variability metrics and norms," *Front. Public Health*, vol. 5, p. 258, 2017. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fpubh.2017.00258/text>

- [27] Z. Zhang, Z. Lv, C. Gan, and Q. Zhu, "Human action recognition using convolutional LSTM and fully-connected LSTM with different attentions," *Neurocomputing*, vol. 410, pp. 304–316, 2020.
- [28] D. Elliott and D. L. Elliott, "A better activation function for artificial neural networks," Univ. Maryland. ISR Tech. Rep. TR 93-8, 1993.
- [29] D. Anguita, G. Parodi, and R. Zunino, "Speed improvement of the back-propagation on current-generation workstations," in *Proc. Portland: World Congr. Neural Netw.*, Portland, Oregon, Jul. 1993, vol. 1, pp. 11–15.
- [30] A. L. Goldberger *et al.*, "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [31] P. deChazal *et al.*, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1196–1206, Jul. 2004.
- [32] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.
- [33] Oversampling imbalanced data: Smote related algorithms. Accessed: Oct. 12, 2020. [Online]. Available: <https://github.com/minoue-xx/Oversampling-Imbalanced-Data/releases/tag/1.0.1>
- [34] L. Xiaolin, B. Cardiff, and D. John, "A 1D convolutional neural network for heartbeat classification from single lead ECG," in *Proc. 27th IEEE Int. Conf. Electron. Circuits Syst.*, 2020, pp. 1–2.
- [35] Y. Xia and Y. Xie, "A novel wearable electrocardiogram classification system using convolutional neural networks and active learning," *IEEE Access*, vol. 7, pp. 7989–8001, 2019.
- [36] Cortex-m4 technical reference manual. Accessed: Oct. 2021. [Online]. Available: <https://developer.arm.com/documentation/ddi0439/b/Floating-Point-Unit/FPU-Functional-Description/FPU-instruction-set>



**Gawsalyan Sivapalan** received the B.Sc. degree in electronics and telecommunication engineering (Hons.) from the University of Moratuwa, Moratuwa, Sri Lanka, in 2016. He has been a Qualified Member of the Chartered Institute of Management Accountants - U.K., since 2016. He was a Research Intern with the Singapore University of Technology and Design, Singapore, from October 2014 to March 2015. During 2017–2018, he was a Senior Executive with Corporate Planning and Strategy development, Dialog Axiata PLC, a super brand and a blue-chip telecommunication arm of Axiata Group Berhad in Sri Lanka. He is currently working toward the Master by Research degree with the School of Electrical and Electronic Engineering, University College Dublin, Dublin, Ireland. His research interests include the development of low complexity, power-efficient machine learning solutions for wearable devices. He co-founded Kairos Sensing, a university spin-off providing wearable-based high-performance capturing solutions to elite clients like Sri Lanka National Cricket in 2016/17 and the winner of Disrupt 2.0 - FutureX technopreneur challenge in 2016.



**Koushik Kumar Nundy** (Senior Member, IEEE) received the Bachelor of Technology degree in electronics and communication engineering from the National Institute of Technology, Durgapur, India, in 2010, and the M.Sc. degree in telecommunication from the Hong Kong University of Science and Technology, Hong Kong, in 2011. He is the CTO and Co-Founder of Think Biosolution Limited, a medical devices and healthcare IT company in Dublin, Ireland. He was with multiple research and academic institutes, including the National University of Singapore, Singapore, and the Indian Institute of Science, Bangalore, India, and has presented his work in multiple conferences and journals. He was also with the R&D team at Altai Technologies Ltd, a Pioneer in wireless communication systems. He is the Ireland Area Chair of the IEEE Young Professionals. He received the NTSE Scholarship in 2004 and NGS Scholarship in 2013. He was the recipient of multiple awards, including Roche Unicorn Champion 2020, Innovation of the Year 2017, Start-up Weekend 2015, and has been featured in publications such as the Irish Times, Rochester Business Journal, and Silicon Republic.



**Soumyabrata Dev** (Member, IEEE) received the B.Tech. degree in electronics and communications (*summa cum laude*) from the National Institute of Technology, Silchar, India, in 2010, and the Ph.D. degree in electrical and electronics from Nanyang Technological University, Nanyang, Singapore, in 2017. He is currently an Assistant Professor with the School of Computer Science, University College Dublin, Dublin, Ireland. He was a Postdoctoral Researcher with the ADAPT SFI Research Centre, Dublin. From August to December 2015, he was a Visiting Student with the Audiovisual Communication Laboratory, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. He has authored or coauthored more than 60 papers. His research interests include remote sensing, statistical image processing, machine learning, and deep learning.



**Barry Cardiff** (Senior Member, IEEE) received the B.Eng., M.Eng. Sc., and Ph.D. degrees in electronic engineering from University College Dublin, Ireland, in 1992, 1995, and 2011, respectively. He was a Senior Design Engineer or Systems Architect for Nokia from 1993 to 2001, moving to Silicon & Software Systems (S3 group) thereafter as a Systems Architect in their R&D division focused on wireless communications and digitally assisted circuit design. Since 2013, he has been an Assistant Professor with University College Dublin, Dublin, Ireland. His research interests include digitally assisted circuit design and signal processing for wireless and optical communication systems. He holds several U.S. patents related to wireless communication.



**Deepu John** (Senior Member, IEEE) received the B.Tech. degree in electronics and communication engineering from the University of Kerala, Thiruvananthapuram, India, in 2002, and the M.Sc. and Ph.D. degrees in electrical engineering from National University Singapore, Singapore, in 2008 and 2014, respectively. He is currently an Assistant Professor with the School of Electrical and Electronics Engineering, University College Dublin, Dublin, Ireland. From 2014 to 2017, he was a Postdoctoral Researcher with the Bio-Electronics Lab, National University Singapore. Previously, he was a Senior Engineer with Sanyo Semiconductors, Gifu, Japan. He was a Member of the organizing committee or technical program committee for several IEEE conferences, including TENCON, ASICON, IS-CAS, BioCAS, and ICTA. He was the Guest Editor of IEEE TRANSACTIONS ON CIRCUITS & SYSTEMS-I and IEEE OPEN JOURNAL OF CIRCUITS & SYSTEMS. He is currently an Associate Editor for IEEE TRANSACTIONS ON BIOMEDICAL CIRCUITS & SYSTEMS, IEEE TRANSACTIONS ON CIRCUITS & SYSTEMS- II and *Wiley International Journal of Circuit Theory & Applications*. His research interests include low-power biomedical circuit design, energy-efficient signal processing, and edge computing. He is the recipient of the Institution of Engineers Singapore Prestigious Engineering Achievement Award in 2011, Best Design Award at the Asian Solid-State Circuit Conference in 2013, and IEEE Young Professionals, Region 10 Individual Award in 2013.