ULECGNet: An Ultra-Lightweight End-to-End ECG Classification Neural Network

Jianbiao Xiao, Jiahao Liu, Huanqi Yang, Qingsong Liu, Ning Wang, Zhen Zhu, Yulong Chen, Yu Long, Liang Chang, Liang Zhou, and Jun Zhou, *Senior Member, IEEE*

***Abstract*—ECG classification is a key technology in intelligent electrocardiogram (ECG) monitoring. In the past, traditional machine learning methods such as support vector machine (SVM) and K-nearest neighbor (KNN) have been used for ECG classification, but with limited classification accuracy. Recently, the end-to-end neural network has been used for ECG classification and shows high classification accuracy. However, the end-to-end neural network has large computational complexity including a large number of parameters and operations. Although dedicated hardware such as field-programmable gate array (FPGA) and application-specific integrated circuit (ASIC) can be developed to accelerate the neural network, they result in large power consumption, large design cost, or limited flexibility. In this work, we have proposed an ultra-lightweight end-to-end ECG classification neural network that has extremely low computational complexity (∼8.2k parameters & ∼227k multiplication/addition operations) and can be squeezed into a low-cost microcontroller (MCU) such as MSP432 while achieving 99.1% overall classification accuracy. This outperforms the state-of-the-art ECG classification neural network. Implemented on MSP432, the proposed design consumes only 0.4 mJ and 3.1 mJ per heartbeat classification for normal and abnormal heartbeats respectively for real-time ECG classification.**

***Index Terms*—ECG, classification, lightweight, end-to-end, neural network, microcontroller, low power, low complexity.**

I. Introduction

HEART disease accounts for a large proportion of the world's disease-related deaths. According to the report released by the World Health Organization (WHO), the Cardiovascular Diseases (CVDs) are the leading causes of death worldwide [1]. Cardiac arrhythmias are the most important CVDs and have attracted wide attention from researchers and practitioners.

Manuscript received January 6, 2021; revised May 11, 2021; accepted June 10, 2021. Date of publication June 18, 2021; date of current version January 5, 2022. This work was supported in part by the National Key R&D Program of China under Grant 2019YFB2204500 and in part by the National Natural Science Foundation of China under Grant 62074026. *(Corresponding author: Jun Zhou.)*

The authors are with the University of Electronic Science and Technology of China, 611731 Sichuan, P. R. China (e-mail: jianbiao_x@std.uestc.edu.cn; jiahaoliu@std.uestc.edu.cn; 519458903yang@gmail.com; 201921010225@std.uestc.edu.cn; ning.wang@std.uestc.edu.cn; 201921010224@std.uestc.edu.cn; 16864723@qq.com; 995163042@qq.com; liangchang@uestc.edu.cn; zlzl@uestc.edu.cn; zhouj@uestc.edu.cn).

Digital Object Identifier 10.1109/JBHI.2021.3090421

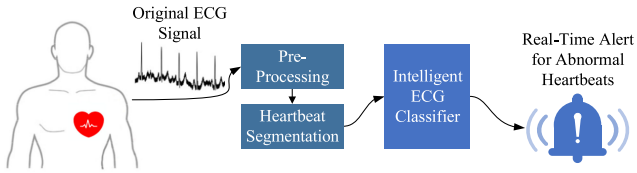

Fig. 1. Intelligent ECG monitoring.

Electrocardiogram (ECG), which records the electrical signals of the heart activity, is the most effective method for detecting cardiac arrhythmias [2].

ECGs are traditionally collected in a short period (e.g., 2–3 mins) using the bulky ECG equipment in hospitals then visually diagnosed by a cardiologist. However, the arrhythmias are difficult to detect in such a short time due to their intermittent occurrence, especially in the early stages of heart disease. Consequently, continuous and long-term ECG monitoring with arrhythmia detection capability is of great interest.

Recently, wearable intelligent ECG monitoring devices for long-term ECG monitoring have been proposed [3]–[5]. These devices integrate intelligent ECG classification to perform ECG monitoring and abnormal heartbeat detection. The original ECG signals are preprocessed and segmented into heartbeats, and then each heartbeat is sent to the intelligent ECG classifier for arrhythmias detection. When arrhythmias are detected, the user will be alerted to go to hospital for further diagnosis, as shown in Fig. 1. Several key features are required for these devices, including real-time, low power, and low cost.

There are two major methods for arrhythmia detection. One is a feature engineering-based ECG classification method, which is characterized by first performing feature extraction, and then sending the extracted features to a classifier for classification. However, the problem of this method is that its classification accuracy largely depends on the quality and quantity of the extracted features, which requires a lot of time and effort for feature engineering, and the accuracy is limited by human knowledge and experience. In addition, achieving a high classification accuracy requires a large number of artificial features, which significantly increase the computational complexity, leading to large processing delay and power consumption.

The other method is the end-to-end classification method, which directly sends the original signal to a neural network for classification. The advantage of this method is that the neural network can automatically learn the best features for classification through training. Therefore, this method is able to

address the limitation of human knowledge and experience, and thus obtain higher accuracy. However, the main disadvantage of the end-to-end neural network is its high computational complexity in terms of the number of parameters and the number of operations required. This makes it difficult to achieve real-time performance and low power consumption on a wearable device.

To accelerate the neural network computation, the graphics processing unit (GPU) and FPGA have been extensively used. The computation time can be significantly reduced by utilizing the parallel processing feature of GPU and FPGA [6]–[9]. However, the power consumption of GPU and FPGA is large, making them unsuitable for wearable ECG monitoring devices. To address this issue, ASICs dedicated to wearable intelligent ECG monitoring have been developed [10], [11]. By mapping the ECG processing algorithm to customized hardware, these designs are able to achieve real-time, low power, and low cost. But the performance is achieved by trading off flexibility. The customized hardware is difficult to catch up with the fast-changing algorithms. Although some ASIC designs incorporate a certain level of flexibility by providing configuration capability of some parameters, the flexibility is still very limited. Compared with ASIC, MCU offers much more flexibility through a fully programmable processor core and saves a significant amount of time and cost for integrated circuit (IC) design and fabrication. However, it is challenging to run the neural network-based ECG classification on low-cost MCUs with real-time performance and low power consumption due to high computational complexity.

In this work, we proposed an ultra-lightweight end-to-end ECG classification neural network and implemented it on a low-cost MCU. Several techniques have been proposed to greatly reduce the computational complexity of the ECG classification neural network while maintaining accuracy so that it can be implemented on the low-cost MCU to achieve low power consumption with real-time performance. The proposed solution can achieve a high classification accuracy of 99.1% with very low classification energy of 0.8mJ per heartbeat which to the best of our knowledge is the lowest among the existing non-ASIC-based ECG classification systems.

The rest of the paper is organized as follows: Section II reviews the related work. Section III describes the proposed ultra-lightweight end-to-end ECG classification neural network. Section IV shows the implementation details of the MCU-based ECG classification system using the proposed neural network. Section V presents the experimental results and related discussions. Section VI concludes the paper.

II. Related Work

The intelligent ECG classification is to automatically classify different types of heartbeats, including normal and abnormal heartbeats (e.g., arrhythmias). According to the common standard from the Association for the Advancement of Medical Instrumentation (AAMI), heartbeats can be divided into 'N' (normal or bundle branch block beats), 'S' (supraventricular ectopic beats), 'V' (ventricular ectopic beats), 'F' (fusion beats) and 'Q' (unknown beats) [12]. To achieve low-power and real-time intelligent ECG classification, both algorithm and hardware design need to be investigated.

One type of ECG classification algorithms is based on feature engineering, which extracts artificial features from the ECG signal and then sends the features to a classifier for ECG classification. For example, Faziludeen *et al.* [13] used discrete wavelet transform for signal pre-processing to obtain heart rate characteristics and then sent them to a support vector machine (SVM) for classification, which achieved better classification accuracy than other traditional classifiers such as K-nearest-neighbor (KNN) and decision-tree. Smíšek *et al.* [14] proposed a two-step SVM approach, which extracted global features first and then selects features via a genetic algorithm, and achieved higher accuracy than that of single-step SVM approaches. Jewajinda *et al.* [15] extracted the features of Hermite basis function coefficients, width parameter, and the time interval between two neighboring R-peaks, then fed them into block-based neural networks for classification to improve the accuracy. The feature engineering-based classification methods have some limitations. First, the effect of the selected features is limited by human knowledge and experience. Second, a large number of features are required in order to achieve high accuracy, resulting in high computational complexity. Finally, finding suitable features for classification takes a lot of time and effort.

Recently, the end-to-end classification method has been proposed. In this type of method, the ECG signal after pre-processing and heartbeat segmentation is directly sent to the neural network for classification. Both feature extraction and classification are performed by the neural network automatically. For example, Kiranyaz *et al.* [16] segmented the ECG signal into one-beat & three-beat windows and fed them to an adaptive 1-D convolutional neural network (CNN) to automatically learn the features of the signal in order to improve accuracy. Xu *et al.* [17] segmented the ECG signal to heartbeats by leveraging the position information of R-peaks then used a CNN with two sub-sampling layers to classify the heartbeats. Zhai *et al.* [18] transformed three adjacent heartbeats in the ECG signal into a dual beat coupling matrix and then fed it into a 2-D CNN classifier, which achieved higher accuracy than 1-D CNN approaches. Compared with feature engineering-based classification methods, the end-to-end classification method achieves higher accuracy by automatically learning the most suitable features from the training data while reducing the effort of artificial feature selection. However, the neural network involves high computational complexity, resulting in high power consumption and large delay.

The hardware implementation of the neural network based ECG classification algorithm is another important aspect to be investigated. In the past, the GPU and FPGA had been extensively used for accelerating neural networks. For example, embedded artificial intelligence (AI) computing device such as Jetson TX2, which integrated 256 NVIDIA compute unified device architecture (CUDA) cores, has been developed for AI edge-computing and achieves 25× energy efficiency compared with central processing unit (CPU) [6]. Wess *et al.* [7] proposed an FPGA-based ECG classification system with piece-wise linear approximated transfer functions and fixed-point arithmetic to effectively reduce the hardware resources of the neural network. Zairi *et al.* [8] implemented a multi-layer perceptron (MLP) classifier with discrete wavelet transform (DWT)

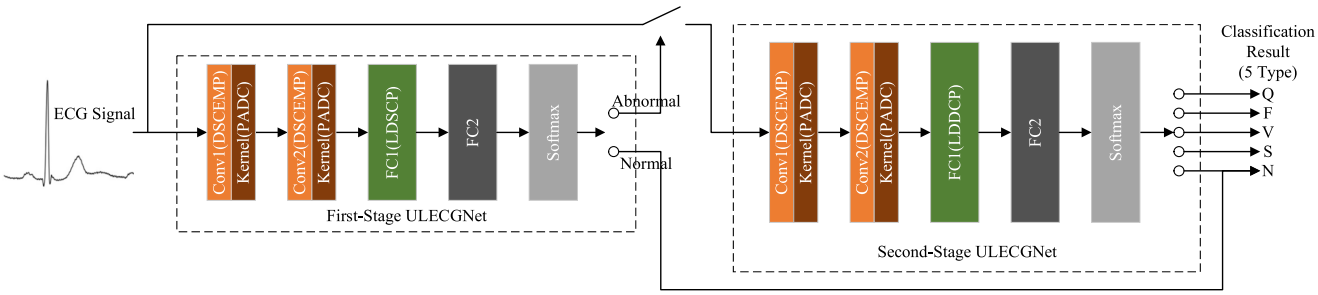

Fig. 2. Structure of the proposed ULECGNet for intelligent ECG classification.

on FPGA for real-time arrhythmia detection. Ponce *et al.* [9] proposed an FPGA-based parallel continuous neural network processor for automatic detection of arrhythmias. However, GPU and FPGA-based designs consume a significant amount of power and are not suitable for wearable devices. To optimize power consumption, ASICs dedicated to ECG classification have been developed. For example, Yin et. al [10] presented an ultra-low-power smart ECG processor in 65-nm CMOS that performed both ECG-based authentication and arrhythmia detection. Zhao *et al.* [11] presented a continuous-in-time discrete-in-amplitude (CTDA) event-driven patient-specific artificial neural network-cardiac arrhythmia classifier for wearable ECG sensors, which had been designed and implemented in 0.18 μm complementary metal-oxide-semiconductor (CMOS) process. But the ASIC-based designs rely on customized hardware with low flexibility which makes it difficult to catch up with the fast-evolving algorithms. Also, IC design takes a significant amount of design time, effort, and cost. In terms of flexibility and cost, MCU is a better solution and has been commonly used in commercial ECG monitoring devices [19]–[21]. However, as it is challenging to run neural networks on MCU, these products usually implemented the data collection on MCU and the classification on mobile phones, which consumed large power for wireless data transmission. Therefore, developing a lightweight ECG classification neural network is needed for realizing low power ECG monitoring devices.

III. Proposed Ultra-Lightweight End-to-End ECG Classification Neural Network

In this work, an ultra-lightweight end-to-end ECG classification neural network (ULECGNet) for MCU-based low power ECG classification system has been proposed. As shown in Fig. 2, four techniques have been proposed to greatly reduce computation complexity including the number of parameters and the number of operations while achieving high classification accuracy. Firstly, the Layer Decomposition with the Dual-Channel Pooling (LDDCP) technique is proposed to reduce the number of parameters in the fully-connected (FC) layer which dominates the number of parameters in the entire neural network. Secondly, the Depthwise Separable Convolution with Embedded Max-Pooling (DSCEMP) technique is proposed to reduce the number of operations in the convolution layer. Thirdly, the Pooling-Aware Dilated Convolution (PADC) technique is proposed to further reduce the number of operations in the convolution layer. Finally, a Revised Two-Stage Neural Network (RTSNN) structure is proposed to reduce the overall computational complexity of ECG classification. The detailed description of the proposed techniques is as follows.

A. Proposed LDDCP Technique

Compared with the MLP-based ECG classification neural network, the CNN-based ECG classification neural network can achieve higher accuracy as it extracts both global and local features. However, in the CNN-based ECG classification neural network, the fully-connected layers contain a large number of parameters. For example, in [2], [22], [23], the parameters of the fully-connected layers dominate the parameters of the entire neural network, especially the parameters of the first fully-connected layer account for more than 88% of that of the entire neural network. Simply cutting down the parameters will cause significant degradation of the classification accuracy.

In this work, we proposed a Layer Decomposition with the Dual-Channel Pooling (LDDCP) technique that adopts layer decomposition with dual-channel pooling in the first fully-connected layer to significantly reduce the number of parameters. Its details are shown in Fig. 3. The structure of the first fully-connected layer using the conventional method is shown on the left, and the structure of the first fully-connected layer using the proposed LDDCP technique is shown on the right.

In the conventional method, each neuron in the fully-connected layer has an independently trainable parameter matrix for the weighted fusion of all input feature maps, which causes a large number of parameters. In the LDDCP technique, we decompose the conventional fully-connected layer into a projection layer and a dual-channel pooling layer as shown in Fig. 3. The projection layer contains $m\times1\times1\times k$ convolution, which performs channel-to-channel correlation on the input feature maps and projects the feature vectors onto the higher dimension. The dual-channel pooling performs both max-pooling and weighted average pooling simultaneously, where the weighted average pooling integrates a convolution to achieve pooling with weighted accumulation. Compared with the single-channel pooling, the dual-channel pooling maintains more information from the input feature maps which help improve the classification accuracy. The projection layer together with the dual-channel pooling layer achieves equivalent operation as that of the fully-connected layer while significantly reducing the number of parameters.

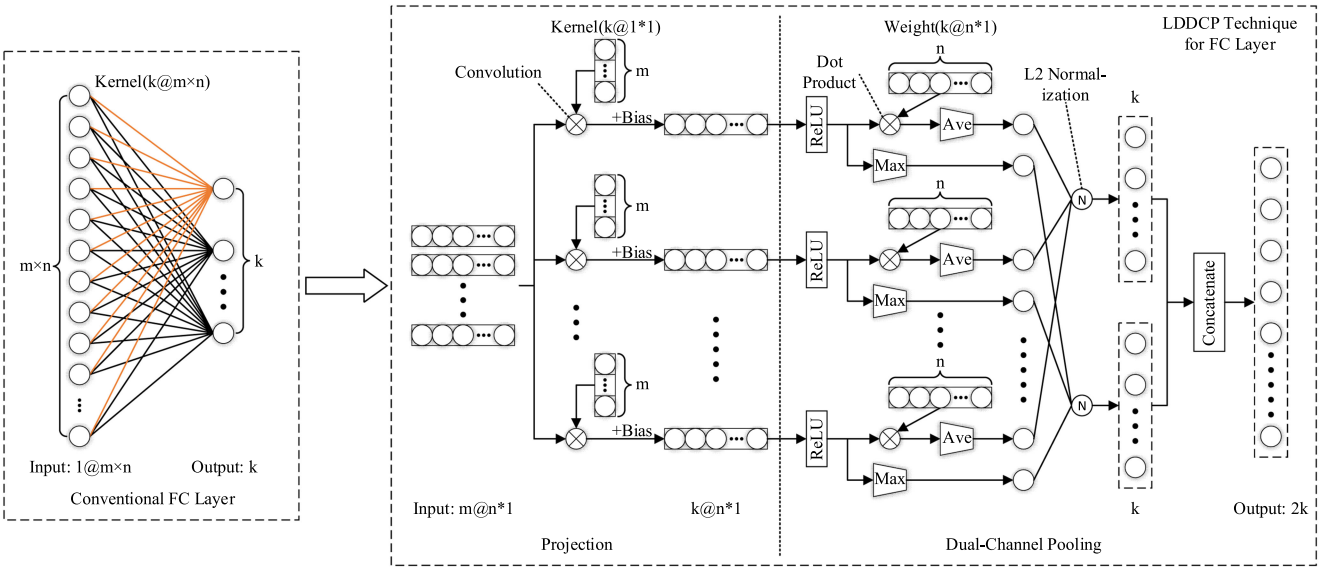

Fig. 3. The conventional fully-connected layer vs the proposed LDDCP technique (**"a@b∗c"**: the number of channels is **'a'** and the size is **'b∗c'**).

To ensure this, the parameter matrix of the fully-connected layer ($W_{m\times n}$) is split into two vectors ($V_{m\times 1}$ and $V_{1\times n}$) according to (1):

$$W_{m\times n} \rightarrow V_{m\times 1} V_{1\times n} \tag{1}$$

where $V_{m\times 1}$ and $V_{1\times n}$ are the parameter vectors of projection and dual-channel pooling obtained from the neural network training.

In the proposed LDDCP technique, a ReLU activation layer is inserted between the projection layer and the dual pooling layer to improve the non-linearity. An L2 normalization layer, which normalizes the feature vectors from the max-pooling and average pooling, is added after the dual-channel pooling layer. The L2 normalization is expressed as below (2):

$$y_i = \frac{x_i}{\sqrt{\sum_{j=1}^{d} x_j^2}} \tag{2}$$

where $\{x_i\}$ denotes the features vector, which has d elements in total.

Compared with the conventional fully-connected layer, the proposed LDDCP technique is able to significantly reduce the number of parameters from $m \times n \times k$ to $(m+n) \times k$, as shown in Fig. 3. The reduction percentage (RP_p) is expressed as below (3):

$$RP_p = \left[1 - \left(\frac{1}{m} + \frac{1}{n}\right)\right] \times 100\% \tag{3}$$

The m and n (e.g., 50×32) are usually large for the first fully-connected layer and small (e.g., 1×64) for the second fully-connected layer. This is the reason why we apply the LDDCP technique only to the first fully-connected layer. In the first fully-connected layer, for the typical values of $m = 50$, $n = 32$, and $k = 64$, the percentage of reduction for the number of parameters is 94.9%. With the proposed LDDCP technique, the number of multiplications and additions is slightly increased as shown in Table I. For the same typical value of m, n and k, the increased percentages of multiplications and additions are only 2.13% and 2.06% respectively, which are negligible.

TABLE I
COMPARISON OF PARAMETERS AND OPERATIONS BETWEEN THE CONVENTIONAL METHOD AND THE PROPOSED LDDCP TECHNIQUE

	Conventional	with LDDCP
No. Parameter	$m\times n\times k$	$(m+n)\times k$
No. Multiplication	$m\times n\times k$	$m\times n\times k+(n+2)\times k$
No. Addition	$m\times n\times k$	$m\times n\times k+(n+1)\times k$

'*m*': the number of input feature map channels, **'*n*'**: the size of the input feature maps, **'*k*'**: the number of output channels in the projection layer.

B. Proposed DSCEMP Technique

Another issue of CNN-based ECG classification neural network is that the number of operations (i.e., multiplications and additions) is large, which is mainly caused by the convolutional layers. In this work, a Depthwise Separable Convolution with Embedded Max-Pooling (DSCEMP) technique is proposed to reduce the number of operations.

As shown in Fig. 4, a depthwise separable convolution (DSC), which consists of a depthwise convolution layer and a pointwise convolution layer, is applied to replace the conventional convolution. The depthwise separable convolution was first introduced in image recognition applications [24], [25]. Compared with the conventional convolution, it can largely reduce the number of operations by transforming the convolution into two convolution layers along with space and depth separately.

In the DSC followed by the max-pooling, the pointwise convolution layer is followed by a max-pooling layer to reduce the dimension of the output feature map. This indicates that the operation in the pointwise layer actually contains large redundancy, as the computed values are directly cut down by the max-pooling layer. Motivated by this observation, we proposed to remove the max-pooling layer and insert an embedded

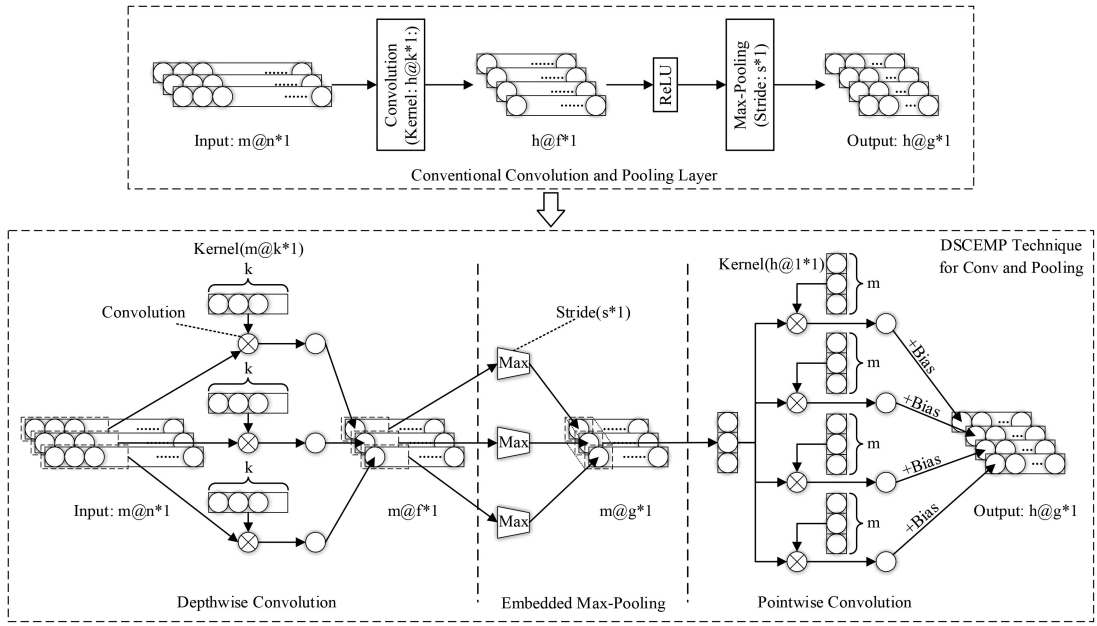

Fig. 4. The conventional convolution & pooling layer vs the proposed DSCEMP technique (**"a@b∗c"**: the number of channels is **'a'** and the size is **'b∗c'**).

TABLE II
COMPARISON OF COMPUTATIONS BETWEEN CONVENTIONAL METHOD AND DSCEMP TECHNIQUE IN THE CONVOLUTION AND POOLING LAYERS

	No. Multiplication	No. Addition
Conventional	$m\times f\times k\times h$	$m\times f\times k\times h$
with DSC	$m\times f\times(k+h)$	$m\times f\times(k-1)+m\times f\times h$
with DSCEMP	$m\times(f\times k+g\times h)$	$m\times f\times(k-1)+m\times g\times h$

'*m*': the number of input channels for depthwise convolution, **'*f*'**: the output feature maps length of the depthwise convolution, **'*k*'**: the kernel length of the depthwise convolution, **'*g*'**: the output feature maps length of the pointwise convolution, **'*h*'**: the number of output channel for the pointwise convolution.

max-pooling layer between the depthwise convolution layer and the pointwise convolution layer to form the DSCEMP, as shown in Fig. 4. This significantly reduces the number of operations in the pointwise convolution layer. As shown in Table II, with the proposed DSCEMP technique, the number of multiplication and addition is compressed and the reduction percentage ($RP_{\times}$, RP_{+}) can be expressed as below (4)(5). In the convolution layer, for the typical values of $h = 32, k = 9, f = 88$, and $g = 44$, the percentage of reduction for multiplications and additions are 91.3% and 91.7%, respectively.

$$RP_{\times} = \left[1 - \left(\frac{1}{h} + \frac{1}{k} \times \frac{g}{f}\right)\right] \times 100\% \tag{4}$$

$$RP_{+} = \left[1 - \left(\frac{\mathrm{k}-1}{h \times k} + \frac{1}{k} \times \frac{g}{f}\right)\right] \times 100\% \tag{5}$$

We have also analyzed the impact of the proposed DSCEMP technique on classification accuracy. Firstly, as shown in Fig. 5, in the conventional DSC of 1×4 vector and 3 channels, the pointwise convolution transforms the 3 channels into 1 channel. Here for the simplification of illustration, the weights of the kernels are all assumed to be '1', and the biases are all assumed

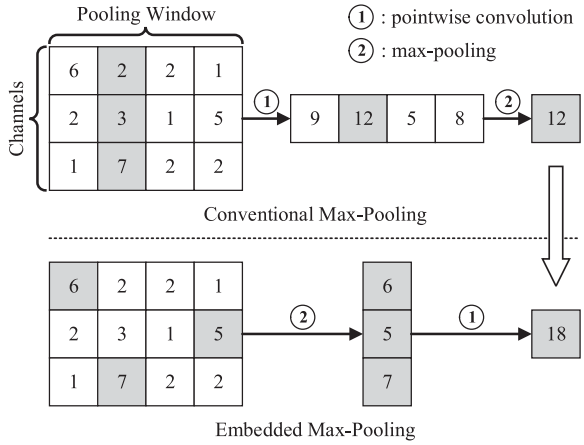

Fig. 5. The conventional max-pooling vs the proposed embedded max-pooling.

to be '0'. After that, the max-pooling is performed which selects the max value from the 1×4 vector, as shown in Fig. 5. In the proposed DSCEMP technique, as shown in Fig. 5, one max value is selected from each of the 4 channels. Then the 4 channels are transformed into 1 channel by the pointwise convolution. The effect of the DSCEMP is similar to that of the DSC as both of them transform the three 1×4 vectors into one value, therefore it has a negligible impact on the classification accuracy. Moreover, the effect of DSCEMP is even better in terms of keeping diverse information in the input vectors as the points at a different position in each channel are selected to generate the output value as shown in Fig. 5, while in the DSC only the points at the same positions are selected to generate the output value as shown in Fig. 5. This can be seen from the experimental results from later sections.

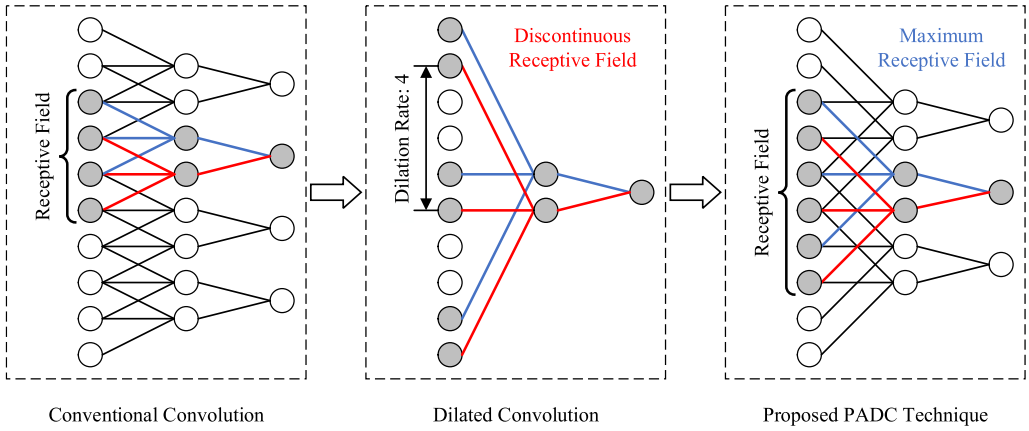

Fig. 6. The conventional convolution, the dilated convolution and the proposed pooling-aware dilated convolution technique.

C. Proposed Pooling-Aware Dilated Conv Technique

In the ULECGNet, each convolutional layer is followed by a pooling layer to reduce the dimension and increase the receptive field. The size of the receptive field is important for classification accuracy. On the one hand, small receptive fields will limit the classification performance. On the other hand, large receptive fields will result in high computational complexity. In image recognition applications, the dilated convolution has been proposed to achieve large receptive fields without increasing the computational complexity by skipping connections during the convolution [26].

In this work, we proposed a Pooling-Aware Dilated Convolution (PADC) technique to reduce computational complexity while maintaining sufficient receptive fields. Different from the conventional dilated convolution, in the proposed PADC technique, the way of skipping connections is dependent on the size of subsequent pooling. As shown in Fig. 6, in the conventional convolution, the neighboring 4 neurons in the input layer are connected to a neuron in the layer 1. To increase the receptive field without increasing the number of connections, the dilated convolution can be adopted by skipping connections during the convolution. If K represents the number of connections skipped, the dilation rate is defined as $K + 1$. A common practice to apply the dilated convolution is to follow a binary fashion for the dilation rate for a group of convolution layers (e.g., dilation rate of 4, 8, 16 for three consecutive convolution layers) [26], [27], [28], as shown in Fig. 6. This will quickly enlarge the receptive field and reduce computation complexity. However, we observed that a dilation rate larger than the pooling size will result in discontinuous receptive fields, as shown in Fig. 6. This will degrade the classification accuracy as what can be seen from the section of experimental results. Therefore, in the proposed PADC, the dilation rate is set to the same as the size of the subsequent pooling as shown in Fig. 6. This will help maintain a high classification accuracy at certain cost of computation complexity.

D. Proposed Revised Two-Stage NN Structure

In our previous work [29], we have proposed to reduce the power consumption of the ECG classification neural network by a two-stage neural network. The first stage of the neural network classifies the heartbeat into normal and abnormal. If it is abnormal, the second stage of the neural network will classify it into 'S', 'V', 'F', 'Q' and 'N', according to the AAMI standard [12]. If the heartbeat is normal, the second stage will not be activated which saves the power consumption significantly as the normal case occurs much more frequently than the abnormal case. In this work, we have also adopted a Revised Two-Stage NN (RTSNN) with some revisions to further optimize the computational complexity.

In the previous work, the first stage of the neural network is an MLP stage and the second stage is a CNN stage. In this work, we have applied all of the aforementioned techniques to the CNN stage to reduce the computational complexity while maintaining accuracy, as shown in Fig. 7. In the meanwhile, considering that MLP has a large number of parameters, we have also changed the first stage from MLP to CNN and applied our proposed techniques. This helps to reduce the number of parameters of the first-stage neural network by 33.7×. When applying the proposed LDDCP technique to the first stage, we used a slightly different strategy – layer decomposition with the single-channel pooling (LDSCP). Instead of using dual-channel pooling, we used single-channel pooling to further reduce the computational complexity. Theoretically, this will also reduce the classification accuracy, however, as the task of the first stage is relatively simple (i.e., only classify normal and abnormal), the impact of this revision on the accuracy is negligible.

IV. MCU Implementation

The proposed ULECGNet is an ultra-lightweight neural network that is able to be implemented on a low-cost MCU. Compared with ASIC design, it significantly reduces the design time and cost while increasing the flexibility for algorithm update. In this work, we have implemented the ULECGNet on a commonly used low-cost MCU (i.e., MSP432) from TI. Fig. 8 shows the MSP-EXP432P401R MCU Launchpad Evaluation Board that was used. The MSP432 MCU contains 256 kB embedded flash in which 32 kB is used for storing the parameters of the ULECGNet and 17 kB is used for storing the system code. It also contains 64 kB SRAM in which 35 kB is used for storing the

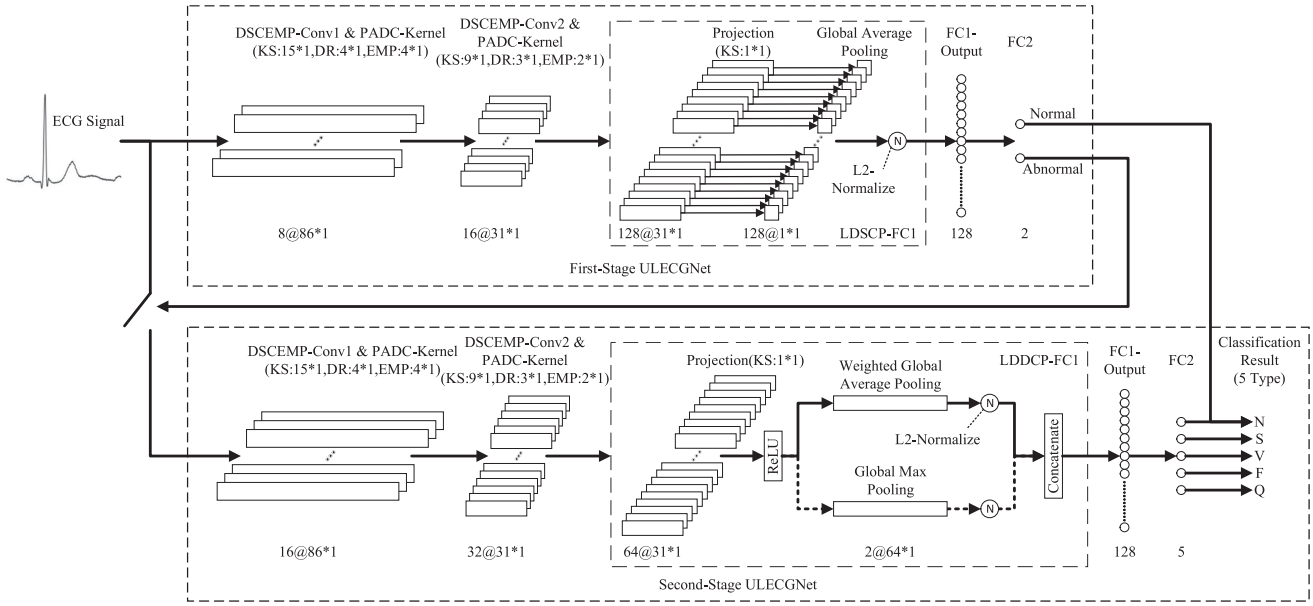

Fig. 7. The proposed revised two-stage NN structure (**"KS":** convolution kernel size, **"DR"**: dilation rate, **"EMP"**: size of embedded max-pooling).

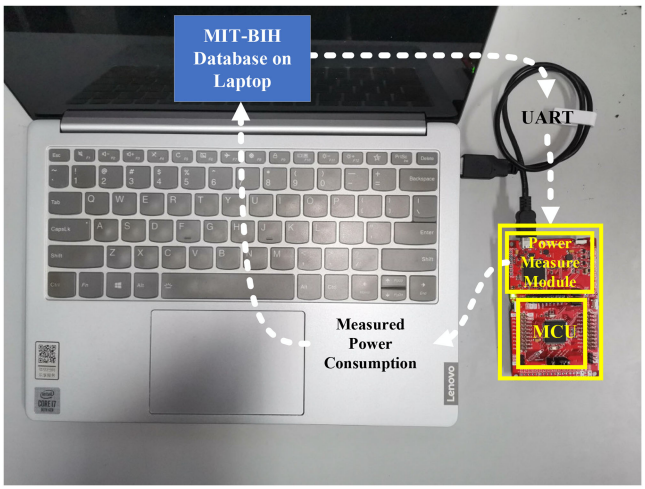

Fig. 8. Experimental platform based on MSP-EXP432P401R MCU launchpad evaluation board.

intermediate data such as the feature maps of the ULECGNet. The 32-bit floating point unit (FPU) is used for computing the multiply-accumulate (MAC) operations of the ULECGNet.

During the implementation, as much existing work [30]–[32], we have adopted the Pan-Tompkins algorithm for the R-peak detection and heartbeat segmentation. Each heartbeat segment window is $[-133, 266]$ around the R-peak point, which contains 400 points. The segmented heartbeats are sent to the ULECGNet for end-to-end classification.

In the implementation, we also adopted some optimization techniques to reduce redundant operations. For the implementation of the neural network, we reduced the amount of computation by adopting a multi-stage address calculation method. The implementation of CNN on MSP432 is usually realized by many layers of embedded loops and the computation (e.g., MAC operation) & control instructions (e.g., data address calculation) are usually placed in the innermost loop. This causes a significant amount of computation in which some are actually redundant. To address this issue, we adopted a multi-stage address calculation method, in which the address is calculated stage by stage from the outermost loop to the innermost loop instead of being calculated only in the innermost loop, as shown in Table III. As the innermost loop is executed most frequently, this significantly reduces the amount of computation for the address calculation.

V. Experimental Results and Analysis

In this section, the performance of the proposed ULECGNet is evaluated, including classification accuracy, computation complexity (No. of parameters & No. of operations), and energy consumption.

A. Database

In this work, we used the ECG data from the MIT-BIH arrhythmia database [33] to evaluate the proposed methods. The database includes 48 records, and each record includes 30 minutes of ECG signals selected from 24-hour recordings of 47 individuals (25 males aged 32 to 89 and 22 females aged 23 to 89). The ECG signals are obtained by the Beth Israel Hospital Arrhythmia Laboratory containing more than 4000 long-term Holter recordings obtained between 1975 and 1979. The database contains 23 individuals' data chosen at random from this set, and 24 individuals' data selected from the same set to include a variety of rare but clinically important phenomena that would not be well-represented by a small random sample of Holter recordings. The original ECG signals are filtered using a band-pass filter of 0.1-100 Hz and then digitized at 360 Hz. The database contains annotation for both timing information and heartbeat type information verified by doctors.

TABLE III
IMPLEMENTATION OF CONVOLUTION ON MCU

Conventional Method	**Our Method**
Required: Input feature maps *x*, weights of kernel *w*, bias of kernel *b*, number of input channel *ic*, number of output channel *oc*, length of input feature maps *xl*, length of kernel *kl*. **Required**: Calculate the length of the output feature map *yl* according to *xl* and *kl* and malloc memory of length *ks* to the output feature map *y* with zero initialization. **for** *i* in 1 to *oc* **do**: Calculate the memory address *ba* of *b* on the *i*-th output channel. **for** *j* in 1 to *yl* **do**: **for** *k* in 1 to *ic* **do**: **for** *m* in 1 to *kl* **do**: Calculate the memory address *wa* of the *m*-th kernel weight on the *i*-th output channel and the *k*-th input channel. Calculate the memory address *xa* of the *m*-th value of *x* on the *k*-th input channel which corresponds to the *j*-th value of *y* on the *i*-th output channel Calculate the product of *x*[*xa*] and *w*[*wa*] and add it to *y*[*j*]. **end for** **end for** Add *b*[*ba*] to *y*[*j*]. **end for** **end for**	**Required**: Input feature maps *x*, weights of kernel *w*, bias of kernel *b*, number of input channel *ic*, number of output channel *oc*, length of input feature maps *xl*, length of kernel *kl*. **Required**: Calculate the length of the output feature map *yl* according to *xl* and *kl* and malloc memory of length *ks* to the output feature map *y* with zero initialization. **for** *i* in 1 to *oc* **do**: Calculate the memory address *ba* of *b* on the *i*-th output channel. Calculate the memory address offset of the *w* and *x* caused by the *i*-th output channel and assign them to *wa_i* and *xa_i* respectively. **for** *j* in 1 to *yl* **do**: Calculate the memory address offset of *x* caused by the *j*-th value of *y* and add it to *xa_i* then assign result to *xa_j*. **for** *k* in 1 to *ic* **do**: Calculate the memory address offset of *w* and *x* caused by the *k*-th input channel and add them to *wa_i* and *xa_j* and assign results to *wa_k* and *xa_k* respectively. **for** *m* in 1 to *kl* **do**: Calculate the memory address offset of *w* and *x* caused by the *m*-th kernel weight and add them to *wa_k* and *xa_k* and assign results to *wa* and *xa* respectively. Calculate the product of *x*[*xa*] and *w*[*wa*] and add it to *y*[*j*]. **end for** **end for** Add *b*[*ba*] to *y*[*j*]. **end for** **end for**

TABLE IV
THE DETAILED STRUCTURE OF THE PROPOSED ULECGNET

Layer		Size	D.R.	Ch.	Pa.
Input		1*400	-	-	-
First stage:					
Conv1 (DSCEMP+ PADC)	Depthwise Conv1	1×15	4	1	15
	EMP1	1×4	-	1	-
	Pointwise Conv1	1×1	-	8	16
Conv2 (DSCEMP+ PADC)	Depthwise Conv2	1×9	3	8	72
	EMP2	1×2	-	8	-
	Pointwise Conv2	1×1	-	16	144
FC1 (LDSCP)	Projection	1×1	-	128	2176
	GAP	1×31	-	128	-
FC2	Fully-Connected	128×2	-	-	258
Second stage:					
Conv1 (DSCEMP+ PADC)	Depthwise Conv1	1×15	4	1	15
	EMP1	1×4	-	1	-
	Pointwise Conv1	1×1	-	16	32
Conv2 (DSCEMP+ PADC)	Depthwise Conv2	1×9	3	16	144
	EMP2	1×2	-	16	-
	Pointwise Conv2	1×1	-	32	544
FC1 (LDDCP)	Projection	1×1	-	64	2112
	WGAP	1×31	-	64	1984
	GMP	1×31	-	64	-
FC2	Fully-Connected	128×5	-	-	645

"EMP": Embedded Max-Pooling, **"GAP"**: Global Average Pooling, **"GMP"**: Global Max-Pooling, **"WGAP"**: Weighted GAP, **"D.R."**: dilation rate, **"Ch."**: the number of output channels, **"Pa."**: the number of parameters.

B. Detailed Structure of the Proposed ULECGNet

The detailed structure of the proposed ULECGNet is shown in Table IV. The ULECGNet consists of two stages.

The first stage includes two DSCEMP+PADC convolution layers, an LDSCP fully-connected layer, and a conventional fully-connected layer. The DSCEMP+PADC convolution layer includes a depthwise convolution layer with dilated kernels, an embedded max-pooling layer, and a pointwise convolution layer. The LDSCP fully-connected layer includes a projection layer and a global average pooling layer.

The second stage includes two DSCEMP+PADC convolution layers with the number of output channels doubled compared to the first stage, an LDDCP fully-connected layer, and a conventional fully-connected layer. The LDDCP fully-connected layer includes a projection layer, a weighted global average pooling layer, and a global max-pooling layer.

C. Evaluation Method

We have reviewed some state-of-the-art ECG classification methods based on neural network [16], [29], [31], [34]–[39] and found that they adopt different evaluation methods, mainly including three different experimental protocols as shown in Table V. Therefore, for fair comparison, we have adopted all the three experimental protocols (i.e., Exp.1, Exp.2, and Exp.3) when comparing our work with the state-of-the-art work. In addition, we adopted Exp.1 for the evaluation of the impact of different proposed technique in our method.

As other existing work, five metrics including Acc, Spe, Sen, Ppr and $F1$ are used for the evaluation of the classification accuracy. In (6)(7)(8)(9), TP refers to true positive, FN refers to false negative, FP refers to false positive, TN refers to true negative, Acc refers to accuracy, Spe refers to specificity, Sen refers to sensitivity, Ppr refers to predictability, and $F1$ refer to F1 score. The four metrics (Acc, Spe, Sen, and Ppr) for supraventricular ectopic beat (SVEB) and ventricular ectopic beat (VEB) are calculated and analyzed as suggested by AAMI

TABLE V
EXPERIMENTAL PROTOCOLS

	Ref. Work	Description
Exp.1	[29][34] [35][36] [39]	50% of the heartbeat samples of the 48 records are randomly picked and used for training. The remaining 50% are used for testing.
Exp.2	[31][37]	22 (known as DS1) of the 44 records (102, 104, 107, and 217 are removed) are used for global training (80% for training and 20% for validation). In the remaining 22 records (known as DS2*), the first 5 minutes of each record are used for local training (to obtain a patient-specific model as different records are from different patients), and the remaining minutes of the same record are used for testing.
Exp.3	[16][38]	20 (known as DS100) of the 44 records (102, 104, 107, and 217 are removed) are used for global training (80% for training and 20% for validation). In the remaining 24 records (known as DS200**), the first 5 minutes of each record is used for local training and the remaining minutes of the same record are used for testing.

*: DS2, Record 100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233, and 234.
**: DS200, Record 200, 201, 202, 203, 205, 207, 208, 209, 210, 212, 213, 214, 215, 219, 220, 221, 222, 223, 228, 230, 231, 232, 233, and 234.

standard [12], where the SVEB and VEB are the two most common arrhythmias. In addition, the confusion matrix and the overall accuracy which is given in (9) are also evaluated (ϕ_i refers to the number of samples of the class i that are correctly predicted), as in other work [29], [34], [35], [39].

$$N = TP + TN + FP + FN \tag{6}$$

$$Acc = \frac{TP + TN}{N} \quad Spe = \frac{TN}{TN + FP} \tag{7}$$

$$Sen = \frac{TP}{TP + FN} \quad Ppr = \frac{TP}{TP + FP} \tag{8}$$

$$F1 = \frac{2 \times Sen \times Ppr}{Sen + Ppr} \quad Over\ Acc = \frac{\sum \phi_i}{N} \tag{9}$$

To evaluate the computational complexity of the proposed techniques, both the number of parameters and the number of operations are calculated. As the proposed ULECGNet is a two-stage neural network as [29], to calculate the effective number of operations, we first calculated the number of operations for the two stages respectively, then multiplied that of the second-stage by the ratio of abnormal heartbeats, and added it to that of the first stage.

In addition, the processing time and energy consumption of the implemented ULECGNet on the MSP432 have been measured using the time & energy measurement tool from TI during its real-time running.

D. Classification Accuracy

According to the AAMI standard [12], each heartbeat can be classified into five classes, including 'N', 'S', 'V', 'F', and 'Q'. Table VI shows the confusion matrix of the proposed ULECGNet combining all the proposed techniques. From Table VI, the classification accuracy of VEB, SVEB, and the overall classification accuracy can be calculated, as shown in Table VII. The overall accuracy is 99.1%.

Fig. 9 shows the contribution of each proposed technique on the reduction of the number of parameters. Here the baseline is a single-stage end-to-end neural network including two convolution layers, two max-pooling layers, and two fully-connected layers that are able to achieve an overall classification accuracy

TABLE VI
CONFUSION MATRIX OF EXP. 1

Ground Truth	Classification Result				
	N	*S*	*V*	*F*	*Q*
N	44122	29	34	12	9
S	160	1109	7	0	4
V	89	0	3326	8	0
F	54	0	43	308	2
Q	36	2	3	0	4011

TABLE VII
CLASSIFICATION ACCURACY OF EXP. 1

		Baseline	L.	L.+D.	L.+D. +P.	L.+D.+ P.+R.
VEB	*Acc*	99.8	99.7	99.7	99.7	99.7
	Sen	97.6	97.3	97.2	97.5	97.2
	Spe	99.9	99.9	99.9	99.8	99.8
	Ppr	98.8	98.0	97.9	97.2	97.5
	F1	98.2	97.6	97.5	97.3	97.3
SVEB	*Acc*	99.7	99.6	99.6	99.6	99.6
	Sen	89.5	87.8	85.8	87.3	86.6
	Spe	99.9	99.9	99.9	99.9	99.9
	Ppr	97.2	95.2	96.7	96.6	97.3
	F1	93.2	91.4	90.9	91.7	91.6
Overall Accuracy		99.3	99.1	99.0	99.1	99.1

'L.': the proposed LDDCP technique, **'D.'**: the proposed DSCEMP technique, **'P.'**: the PADC technique, **'R.'**: the proposed RTSNN structure.

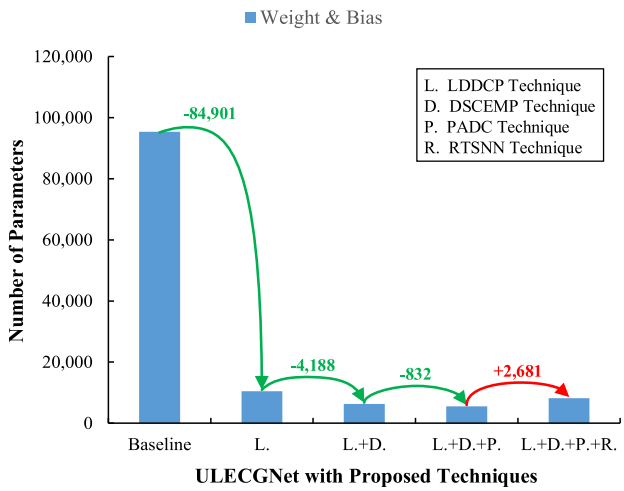

Fig. 9. Number of parameters of ULECGNet with proposed techniques.

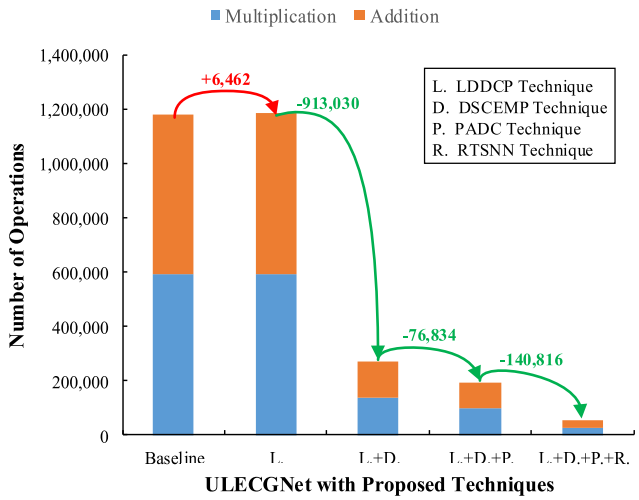

Fig. 10. Number of operations of ULECGNet with proposed techniques.

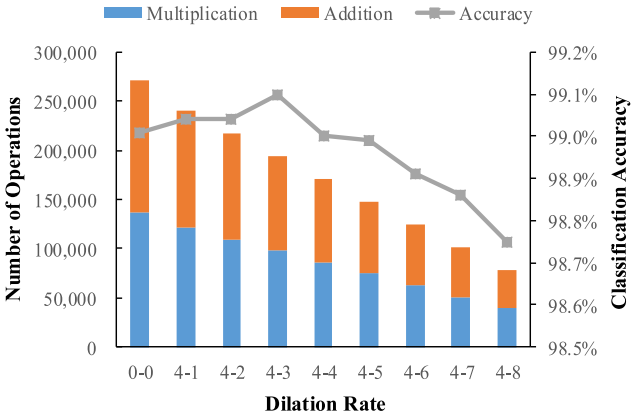

Fig. 11. Number of operations and classification accuracy with different dilation rate.

of 99.3% with computational complexity similar to the baseline in [29]. It can be seen that with the proposed LDDCP, DSCEMP, and PADC techniques added one by one, the number of parameters is reduced by 84901, 4188, and 832, respectively. With the proposed Revised Two-Stage NN technique, the number of parameters is increased to 8k which is still only 8.6% of the baseline, but the number of operations is significantly reduced as shown in Fig. 10.

Fig. 10 shows the contribution of each proposed technique to the reduction of the number of operations including multiplication and addition. It can be seen that with the proposed DSCEMP, PADC, and RTSNN techniques, the number of operations is reduced by 913030, 76834, and 140816. For the proposed LDDCP, the number of operations is slightly increased by 6462, but it leads to a significant reduction in the number of parameters as shown in Fig. 9.

Fig. 11 shows the impact of the dilation rate on the number of operations for the proposed PADC technique. Here M-N means that the dilation rate of the first depthwise convolution layer is M×1 and that of the second depthwise convolution layer is N×1. The classification accuracy is also shown in Fig. 11. It can be seen that as the dilation rate increases, the number of operations decreases, but the classification accuracy first increases and then decreases. The classification accuracy reaches the highest for dilation rate of 4-3, and therefore we adopted this dilation rate in the proposed PADC technique.

TABLE VIII
CONFUSION MATRIX OF THE PROPOSED FIRST-STAGE NN AND PREVIOUS WORK [29]

Ground Truth	Classification Result	
	Normal	*Abnormal*
Normal	42382, 95.9% (43577, 98.6% [29])	1824, 4.1% (629, 1.4% [29])
Abnormal	**138, 1.5%*** **(455, 5.0%*** **[29])**	9024, 98.5% (8707, 95.0% [29])

∗ The calculated false negative rate is 1.51% (this work) and 4.97% ([29]).

Table VIII compares the confusion matrix of the proposed first-stage ULECGNet and the MLP of previous work [29]. Compared to MLP, the proposed ULECGNet achieves significantly improved the false negative rate (FNR), it can be seen that the FNR is significantly reduced from 4.97% to 1.51%. It may be critical in real-world applications that lower FNR means fewer detrimental missed diagnosis. In addition, it is noted that we trained the first-stage NN using a biased training scheme as previous work [29] in which the loss function is modified to suppress the FNR of the first-stage NN even though the false positive rate (FPR) is slightly increased. By this way, more heartbeats will be classified as abnormal and passed to the second-stage NN for further classifications. The FNR is effectively reduced and the increased FPR can be dealt with by the second-stage NN.

Table IX compares the classification accuracy of the proposed ULECGNet and the state-of-the-art work. It can be seen that for Exp.1 and Exp.2 the proposed ULECGNet has the highest overall accuracy (99.1% and 98.6%). For Exp.3, the overall accuracy (97.3%) of the proposed method is the second highest which is only 0.3% lower than the highest (97.6%). To analyze more comprehensively, we also compared other key metrics including sensitivity and F1 score, it can be seen that our work achieves better performance in almost metrics. More importantly, the computational complexity of the proposed method is significantly lower than the existing work, as will be shown in the subsequent section.

E. Computational Complexity

As mentioned previously, the proposed ULECGNet features ultra-low computational complexity while maintaining high classification accuracy. Therefore, we have also compared the computational complexity of the proposed method with the state-of-the-art work with similar classification accuracy [16], [29], [31], [35], [36], [38], [39]. It is noted that [37] is not compared here as it adopts a spiking neural network involving differential and exponential computation, which is difficult to estimate the computational complexity. Also, the spiking neural network has lower classification accuracy as shown in Table IX. As the work [29] and the proposed work adopt a two-stage neural network, for them we have calculated the computational complexity with and without activating the second stage. It can

TABLE IX
COMPARISON OF CLASSIFICATION ACCURACY BETWEEN THE PROPOSED ULECGNET AND THE STATE-OF-THE-ART METHODS

		TBME 2016 [16]	TBME 2017 [34]	MEMEA 2018 [35]	EMBC 2018 [36]	TBioCAS 2019 [37]	TBioCAS 2019 [29]	JBHI 2020 [38]	JBHI 2020 [31]	TBME 2020 [39]	Proposed		
Experimental Protocol		Exp.3	Exp.1	Exp.1	Exp.1	Exp.2	Exp.1	Exp.3	Exp.2	Exp.1	Exp.1	Exp.2	Exp.3
VEB	*Acc*	98.4	97.4	N/A	99.6	97.9	99.1	98.9	99.6	N/A	**99.7**	**99.6**	**99.4**
	Sen	95.0	92.1	97.0	**98.8**	80.2	91.8	93.1	95.8	N/A	97.2	**96.9**	**95.9**
	Spe	98.7	98.1	N/A	99.6	99.8	99.6	99.5	**99.9**	N/A	**99.8**	99.8	**99.8**
	Ppr	88.1	88.2	97.0	95.5	97.3	95.3	95.6	**97.8**	N/A	**97.5**	97.1	**98.3**
	F1	91.4	90.1	97.0	97.1	88.0	93.5	94.3	96.8	N/A	**97.3**	**97.0**	**97.1**
SVEB	*Acc*	96.6	98.4	N/A	99.1	N/A	99.4	96.7	99.0	N/A	**99.6**	**99.1**	**98.0**
	Sen	64.6	88.6	**96.0**	92.7	N/A	79.5	**78.3**	75.6	N/A	86.6	**77.7**	59.5
	Spe	98.1	99.1	N/A	99.3	N/A	99.9	99.7	99.9	N/A	**99.9**	**99.9**	**99.9**
	Ppr	62.0	88.8	75.0	80.2	N/A	96.3	92.5	**98.9**	N/A	**97.3**	97.7	**95.9**
	F1	63.3	88.7	84.0	86.0	N/A	87.2	**84.8**	85.7	N/A	**91.7**	**86.6**	73.4
Overall Accuracy		95.1	94.8	97.6	N/A	97.3	98.4	**97.6**	N/A	96.3	**99.1**	**98.6**	97.3

"N/A": The values are not applicable which are not provided in the corresponding papers [31], [35]–[37], [39].

TABLE X
COMPARISON OF THE COMPUTATIONAL COMPLEXITY BETWEEN THE PROPOSED ULECGNET AND THE STATE-OF-THE-ART METHODS

	TBME 2016 [16]	MEMEA 2018 [35]	EMBC 2018 [36]	TBioCAS2019 [29]		JBHI 2020 [38]	JBHI 2020 [31]	TBME 2020 [39]	Proposed	
				100%A	90%N+ 10%A				100%A	90%N+ 10%A
No. of Mul*	929,650	788,124,928	3,244,023,812	749,620	156,052	62,200	302,390	4,416,700	114,784	**26,966**
No. of Add*	929,650	788,241,792	3,244,023,812	749,620	156,052	62,200	307,601	4,416,700	112,525	**25,968**
No. of Parameters*	10,993	35,550,147	3,028,740	198,037		62,505	198,771	1,597,732	**8,157**	

∗: The values are calculated based on the structure of neural network provided in the papers.
For two-stage neural network, **'N'**: normal heartbeats, **'A'**: abnormal heartbeats, **"90%N+10%A"**: 90% normal and 10% abnormal heartbeats.

TABLE XI
COMPARISON OF ENERGY CONSUMPTION ON MCU

	Type	Platform	Classification Energy Per Heartbeat (mJ)			Classification Time Per Heartbeat (ms)		
			100%N	100%A	90%N+10%A	100%N	100%A	90%N+10%A
TBioCAS2019 [29] (re-implemented on the same MCU)	CNN & MLP	MSP432	2.24	21.06	4.12	71	640	128
Proposed	CNN	MSP432	0.44	3.10	0.71	18	106	27

For two-stage neural network, **'N'**: normal heartbeats, **'A'**: abnormal heartbeats, **"90%N+10%A"**: 90% normal and 10% abnormal heartbeats.

be seen from Table X that the proposed work has the lowest computational complexity in terms of both the number of operations (multiplication & addition) and the number of parameters. Compared to other work, the reduction of parameters and the number of multiplications & additions are significant.

F. Energy Consumption

For the ULECGNet implemented on the MSP432 MCU, the classification energy consumption can be measured by using the TI's EnergyTrace Plus Tool. As the proposed ULECGNet contains two stages, we have measured the classification energy consumption for activating one stage and two stages by feeding normal heartbeats and abnormal heartbeats to the network. For normal heartbeats, only the first stage of the ULECGNet will be activated, and the measured classification energy consumption is 0.44 mJ. For abnormal heartbeats, both stages of the ULECGNet will be activated, and the measured classification energy consumption is 3.1 mJ. In the MIT-BIH database, 90% of recorded heartbeats are normal heartbeats [33], therefore the average energy consumption can be calculated and is 0.71 mJ for the proposed ULECGNet.

Table XI shows the classification energy consumption of the proposed work when implemented on the MCU. We have also implemented our previous work [29] (which has the closest computational complexity to the proposed method) on the same MSP432 MCU for comparison. However, as the number of parameters of the neural network in [29] still exceeds the memory

size of MSP432, we had to optimize it to reduce the number of parameters by network pruning. As shown in Table XI, the proposed ULECGNet has much lower classification energy than [29]. The classification time of the proposed ULECGNet is 18 ms and 106 ms for normal and abnormal heartbeats respectively, which is sufficient for real-time ECG classification.

VI. Conclusion

In this work, an ultra-lightweight end-to-end ECG classification neural network (i.e., ULECGNet) has been proposed. Several techniques have been proposed to reduce the computational complexity of the neural network while maintaining high classification accuracy. Validated using the MIT-BIH database, the ULECGNet achieves an overall classification accuracy of 99.1% with only ∼8.2k parameters & ∼227k multiplication/addition operations. This outperforms the state-of-the-art ECG classification neural network. The proposed ULECGNet has been implemented on a low-cost MCU MSP432 and consumes only 0.4 mJ per heartbeat classification for normal heartbeats and 3.1 mJ for abnormal heartbeats for real-time ECG classification.

References

[1] "Cardiovascular diseases (CVDs)," World Health Organization, Geneva, Switzerland, 2019. [Online]. Available: https://www.who. int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[2] W. Sun, N. Zeng, and Y. He, "Morphological arrhythmia automated diagnosis method using gray-level co-occurrence matrix enhanced convolutional neural network," *IEEE Access*, vol. 7, pp. 67123–67129, 2019.

[3] S. M. Abubakar, W. Saadeh, and M. A. B. Altaf, "A wearable long-term single-lead ECG processor for early detection of cardiac arrhythmia," in *Proc. Des. Autom. Test Europe Conf. Exhib.*, Dresden, Germany, Mar. 2018, pp. 961–966.

[4] Z. Zhao *et al.*, "Noise rejection for wearable ECGs using modified frequency slice wavelet transform and convolutional neural networks," *IEEE Access*, vol. 7, pp. 34060–34067, 2019.

[5] H. Ozkan, O. Ozhan, Y. Karadana, M. Gulcu, S. Macit, and F. Husain, "A portable wearable tele-ECG monitoring system," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 1, pp. 173–182, Jan. 2020.

[6] NVIDIA JETSON TX2. Accessed: Feb. 28, 2021. [Online]. Available: http://www.nvidia.cn/autonomous-machines/embedded-systems/jetson-tx2/

[7] M. Wess, P. D. Sai Manoj, and A. Jantsch, "Neural network based ECG anomaly detection on FPGA and trade-off analysis," in *Proc. IEEE Int. Symp. Circuits Syst.*, Baltimore, MD, May 2017, pp. 1–4.

[8] H. Zairi, M. Kedir Talha, K. Meddah, and S. Ould Slimane, "FPGA-based system for artificial neural network arrhythmia classification," *Neural Comput. Appl.*, vol. 32, no. 8, pp. 4105–4120, Apr. 2020.

[9] M. Alfaro-Ponce, I. Chairez, and R. Etienne-Cummings, "Automatic detection of electrocardiographic arrhythmias by parallel continuous neural networks implemented in FPGA," *Neural Comput. Appl.*, vol. 31, no. 2, pp. 363–375, Feb. 2019.

[10] S. Yin *et al.*, "A 1.06-uW smart ECG processor in 65-nm CMOS for real-time biometric authentication and personal cardiac monitoring," *IEEE J. Solid-State Circuits*, vol. 54, no. 8, pp. 2316–2326, Aug. 2019.

[11] Y. Zhao, Z. Shang, and Y. Lian, "A 13.34 μW 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.

[12] ANSI/AAMI "Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms," *Assoc. Adv. Med. Instrument*, no. EC57, 1998.

[13] S. Faziludeen and P. V. Sabiq, "ECG beat classification using wavelets and SVM," in *Proc. IEEE Conf. Inf. Commun. Technol.*, Thuckalay, Tamil Nadu, India, Apr. 2013, pp. 815–818.

[14] R. Smíšek *et al.*, "SVM based ECG classification using rhythm and morphology features, cluster analysis and multilevel noise estimation," in *Proc. Comput. Cardiol.*, Rennes, France, Sep. 2017, pp. 1–4.

[15] Y. Jewajinda and P. Chongstitvatana, "FPGA-based online-learning using parallel genetic algorithm and neural network for ECG signal classification," in *Proc. Int. Conf. Electr. Eng./Electron. Comput. Telecommun. Inf. Techn.*, Chiang Mai, Thailand, 2010, pp. 1050–1054.

[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] X. Xu and H. Liu, "ECG heartbeat classification using convolutional neural networks," *IEEE Access*, vol. 8, pp. 8614–8619, 2020.

[18] X. Zhai and C. Tin, "Automated ECG classification using dual heartbeat coupling based on convolutional neural network," *IEEE Access*, vol. 6, pp. 27465–27472, 2018.

[19] Heal Force: Easy ECG Monitor – Prince-180D (3 Lead). Accessed: Feb. 28, 2021. [Online]. Available: http://www.healforce.com/en/html/products/portableecgmon itors/healthcare-portable-ECG-monitors-Prince-180D.html

[20] SOMA Technology: Philips Intellivue MX40. Accessed: Feb. 28, 2021. [Online]. Available: https://www.somatechnology.com/Patient-Monitors/philips-intellivue-mx40.aspx

[21] Heal Force: Easy ECG Monitor – PC-80B. Accessed: Feb. 28, 2021. [Online]. Available: http://www.healforce.com/cn/index.php?ac=article&at=read&did=394

[22] J. Huang, B. Chen, B. Yao, and W. He, "ECG arrhythmia classification using STFT-based spectrogram and convolutional neural network," *IEEE Access*, vol. 7, pp. 92871–92880, 2019.

[23] 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.

[24] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," Apr. 2017. [Online]. Available: http://arxiv.org/abs/1704.04861

[25] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, Jul. 2017, pp. 1800–1807.

[26] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. Int. Conf. Learn. Represent*, San Juan, Puerto rico, May 2016.

[27] N. Kalchbrenner *et al.*, "Neural machine translation in linear time," Mar. 2017. [Online]. Available: http://arxiv.org/abs/1610.10099

[28] A. Paszke *et al.*, "ENet: A deep neural network architecture for real-time semantic segmentation," Jun. 2016. [Online]. Available: http://arxiv.org/abs/1606.02147

[29] 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.

[30] J. Park, K. Lee, and K. Kang, "Intelligent electrocardiogram monitoring system for early arrhythmia detection," in *Proc. Int. Conf. Adv. Inf. Netw. Appl.*, Victoria, BC, Canada, May 2014, pp. 1105–1110.

[31] 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.

[32] P. Cheng and X. Dong, "Life-threatening ventricular arrhythmia detection with personalized features," *IEEE Access*, vol. 5, pp. 14195–14203, 2017.

[33] 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, Jun. 2001.

[34] P. Li *et al.*, "High-performance personalized heartbeat classification model for long-term ECG signal," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 78–86, Jan. 2017.

[35] B. Murugesan *et al.*, "ECGNet: Deep network for arrhythmia classification," in *Proc. IEEE Int. Symp. Med. Meas. Appl.*, Rome, Italy, Jun. 2018, pp. 1–6.

[36] P. Xie *et al.*, "Bidirectional recurrent neural network and convolutional neural network (BiRCNN) for ECG beat classification," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Honolulu, HI, USA, Jul. 2018, pp. 2555–2558.

[37] A. Amirshahi and M. Hashemi, "ECG classification algorithm based on STDP and R-STDP neural networks for real-time monitoring on ultra low-power personal wearable devices," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 6, pp. 1483–1493, Dec. 2019.

[38] S. S. Xu, M. -W. Mak, and C. -C. Cheung, "I-vector-based patient adaptation of deep neural networks for automatic heartbeat classification," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 3, pp. 717–727, Mar. 2020.

[39] J. Wang, R. Li, R. Li, B. Fu, C. Xiao, and D. Z. Chen, "Towards interpretable arrhythmia classification with human-machine collaborative knowledge representation," *IEEE Trans. Biomed. Eng.*, to be published, doi: 10.1109/TBME.2020.3024970.