Received May 29, 2019, accepted June 7, 2019, date of publication June 10, 2019, date of current version June 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2921991

Automated Heartbeat Classification Using 3-D Inputs Based on Convolutional Neural Network With Multi-Fields of View

FEITENG LI[1], YIN XU[1], ZHIJIAN CHEN[1], AND ZHENYAN LIU[2]

[1]Institute of VLSI Design, Zhejiang University, Hangzhou 310027, China

[2]Chengdu tianfu new district people's hospital, Chengdu 610000, China

Corresponding author: Zhijian Chen (chenzj@vlsi.zju.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61801425.

ABSTRACT A high-performance method of automated heartbeat classification based on Convolutional Neural Network (CNN) is proposed in this paper. To make full use of the electrocardiogram information acquired from different parts of the human body, we present a novel 3-D data structure as the input of the CNN. The 3-D structure consists of multiple feature maps, each of which indicates the information collected from one lead and contains morphological characteristic, RR-interval, and beat-to-beat correlation feature. Besides, atrous spatial pyramid pooling (ASPP) module which uses filters with different resolutions is adopted to extract deep features in multi-fields of view. Validated on the MIT-BIH arrhythmia database, the proposed method yields an overall accuracy of 91.44% in the inter-patient practice. In particular, this method achieves 89.05% and 95.15% in the sensitivities of the supraventricular ectopic beat (SVEB) and ventricular ectopic beat (VEB) classes, respectively. With the high performance in detecting these two pathological classes, this method has potential clinical application.

INDEX TERMS Convolutional neural network (CNN), Electrocardiogram, heartbeat classification, inter-patient.

I. INTRODUCTION

Heartbeat classification is necessary in clinic for doctors to diagnose cardiac diseases. However, some non-lethal arrhythmias appear infrequently and analyzing long-term electrocardiogram (ECG) manually is time-consuming for doctors, making the automated ECG analysis useful. Nowadays, there are many methods for classifying heartbeats [1]–[22]. But due to the morphology similarity in heartbeats of different classes and the individual variation in P-QRS-T waveform, the performance in inter-patient practice is limited and needs to be further improved.

Philip de Chazal et al. [1] propose an automatic classification method using feature sets based on ECG morphology and heartbeat interval. First, twelve configurations of feature sets using ECG morphology, heartbeat interval, and RR-interval features are compared. Then they implement two linear discriminants (LDs) classifiers to process the features of different channels respectively and combine the outputs to form the final decision. This method achieves a sensitivity of 75.9%, a positive predictivity of 38.5% for the supraventricular ectopic beat (SVEB) class and a sensitivity of 77.7%, a positive predictivity of 81.8% for the ventricular ectopic beat (VEB) class, which represents the state-of-the-art performance of that time. However, significant time is needed to find the optimal features in this method, and LDs cannot further extract the depth features, which limits the performance potential.

Can Ye et al. [3] present a new approach for inter-patient heartbeat classification based on a combination of morphological and dynamic features. The morphological features consist of the wavelet features (approximation coefficients at level 4 and detail coefficients at levels 3 and 4) extracted by wavelet transform and the features extracted by independent component analysis (ICA, recovering independent source signals from a set of observed signals). The dynamic features of each heartbeat consist of four improved RR-interval related features, i.e., previous RR, post RR, local RR, and average RR interval. With these features, two individual support

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

vector machine (SVM) classifiers are trained separately to classify the heartbeats in different channels. They utilize a new probabilistic estimate-based approach to fuse the classification results from the two classifiers to make the final decision, which turns out to be effective in making a significant improvement over single-lead performance. However, because each channel needs to be processed separately, this approach is not suitable for ECG signals with more channels. It finally achieves an overall accuracy of 86.4%, a sensitivity of 60.8%, a positive predictivity of 56.4% for the SVEB class and a sensitivity of 81.5%, a positive predictivity of 63.1% for the VEB class.

Both of these papers argue that the ECG morphology and RR-interval are significant features, and introducing more channels is effective in improving the classification performance. There are also some papers containing similar features [4], [8], [13]–[15], [17], [19] and classifiers [8], [14], [15]. Some different features like statistics features [8], [12], sequential forward floating search [21], and classifiers like reservoir computing (RC) [6], weighted conditional random fields (CRF) [20] are utilized in others. Although there are various features and classifiers, all these papers first extract features manually and then classify heartbeats with traditional classifier(s). For higher accuracy, they spend much time to find and calculate the best features combination, which requires a certain amount of prior knowledge of the signals and is challenging for non-medical investigators.

Some researchers thereby start adopting neural network to extract deep features automatically and classify heartbeats with only a few superficial features as inputs [4], [9]–[13], [16], [18]. In [18], Xiaolong Zhai et al. propose a 2-D approach for patient-specific heartbeat classification based on convolutional neural network (CNN). A series of three adjacent beats are transformed into a 2-D coupling matrix inputs, which makes the convolutional filters readily capture the continuous waveforms and beat-to-beat correlation in adjacent heartbeats. They finally achieve a sensitivity of 76.8%, a positive predictivity of 74.0% for SVEB class and a sensitivity of 93.8%, a positive predictivity of 92.4% for VEB class. However, the patient-specific method has significant limitations in clinic, as it is time-consuming to acquire and label the typical ECG records of all patients and train the network with them.

This study aims to design and test an automated inter-patient heartbeat classification based on CNN. Every heartbeat is clustered to one of the five beat classes (normal beat (N), SVEB, VEB, fusion beat (F), or unclassifiable beat (Q)) recommended by ANSI/AAMI EC57:1998 standard [23]. Considering errors in the pathological classes (i.e., missing SVEBs or VEBs) can have dramatic consequences while errors in the normal class (i.e., incorrectly diagnosing a cardiac disease) are undesired but still not life-threatening [20], this study tends to achieve a great performance in pathological classes with an acceptable performance in detecting normal beats.

The main contributions of this paper are as follows:

1) We propose a novel 3-D data structure as the input of the CNN to help classify heartbeats better. The 3-D data structure is composed of features of morphology, RR-interval, and beat-to-beat correlation.
2) The atrous spatial pyramid pooling (ASPP) module is utilized in this paper to extract deep features in different fields of view, and the experimental results show that it is useful.
3) The proposed method achieves high performance in detecting the main pathological classes (SVEB and VEB classes) and an accepted performance in detecting N classes, which makes it possible for clinic use.

II. DATASET

A. MATERIALS

The MIT-BIH arrhythmia database [24] used as the benchmark in many papers is presented to test the performance of the proposed method. To increase the signal-to-noise ratio (SNR) and be more realistic, all signals in these forty-eight recordings of the database are band-pass filtered at 0.1-100 Hz and then sampled at 360 Hz. Each of the records contains two 30-minute lead signals that are named as channel-one signal and channel-two signal respectively. The normal QRS complexes are usually prominent in the channel-one signals which are obtained from electrodes placed on the chest, while the ectopic beats often are more prominent in the channel-two signals which are obtained from electrodes placed near orthogonal to the mean cardiac electrical axis.

Philip de Chazal et al. [1] divide the records of the database into two sets with the same approximate proportion of beat types, i.e., DS1 and DS2. The paper concludes that to gain a more realistic evaluation of the real-world performance, DS1 and DS2 are used for training and testing separately. To compare the experimental results objectively and fairly, we follow the usage of this database in our work as many others do. Concretely, the DS1 set consists of all heartbeats from 22 records in MIT-BIH arrhythmia database (i.e., the records 101, 106, 108, 109, 112, 114, 115, 116, 118,119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223 and 230), and the DS2 set consists of all heartbeats from other 22 records (i.e., the records 100, 103, 105, 111, 113, 117, 121, 123,200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233, and 234).

Every heartbeat in DS1 or DS2 is clustered to one of the five beat classes, and because the abnormal heartbeats (i.e., SVEBs, VEBs, and F beats) appear infrequently [4], the distribution of heartbeat classes is extremely imbalanced. Table 1 shows the specific distribution of the MIT-BIH arrhythmia database. More than 89% of the heartbeats belong to N class with about 10% of the SVEBs and VEBs in total. Less than 1% of the beats belong to F class with only 15 Q beats in total.

TABLE 1. Summary of the distribution of heartbeats in DS1 AND DS2.

		N	SVEB	VEB	F	Q	Total
Full	Number	90030	2780	7004	802	15	100613
	Percentage of total	89.40%	2.76%	6.96%	0.80%	0.01%	100.00%
DS1	Number	45820	944	3785	414	8	50971
	Percentage of total	89.89%	1.85%	7.43%	0.81%	0.02%	100.00%
DS2	Number	44210	1836	3219	388	7	49660
	Percentage of total	89.03%	3.70%	6.48%	0.78%	0.01%	100.00%

Since CNN is a multi-classification algorithm based on balanced training data, the distribution of training data has a significant impact on the performance of CNN. The heavier the imbalance is, the worse the total performance [25] will be. As oversampling technique (duplicating instances of under-represented classes until a balanced dataset is created) is effective for improving the disadvantage [25], we extend the SVEB, VEB, and F classes for better performance (see section III for the detailed method). In consideration of the tiny minority of Q class in DS1 (only 7 beats), the extension of them will result in the network overfitting to these 7 beats, and due to the non-importance of Q beats in clinic, the Q beats are not extended in our work.

B. EVALUATION STRATEGY

Several metrics are defined to describe the performance of the proposed method and are used to compare with other works. The symbols in Table 2 are defined for the convenience of representing these metrics.

TABLE 2. Definition of basic statistical symbols.

		Algorithm label					
		n	sveb	veb	f	q	
Reference label	N	$C_{N,N}$	$C_{N,S}$	$C_{N,V}$	$C_{N,F}$	$C_{N,Q}$	$\sum(N)$
	S	$C_{S,N}$	$C_{S,S}$	$C_{S,V}$	$C_{S,F}$	$C_{S,Q}$	$\sum(S)$
	V	$C_{V,N}$	$C_{V,S}$	$C_{V,V}$	$C_{V,F}$	$C_{V,Q}$	$\sum(V)$
	F	$C_{F,N}$	$C_{F,S}$	$C_{F,V}$	$C_{F,F}$	$C_{F,Q}$	$\sum(F)$
	Q	$C_{Q,N}$	$C_{Q,S}$	$C_{Q,V}$	$C_{Q,F}$	$C_{Q,Q}$	$\sum(Q)$
		P_N	P_S	P_V	P_F	P_Q	$\sum(All)$

*$C_{I,J}$: sum of samples of Class I predicted as J(both I and J belong to five classes).
*$\sum(I)$: sum of samples of Class I.
*P_I: sum of samples predicted as Class I.
*$\sum(All)$: sum of all samples in DS2.

To calculate the metrics, we need to count the quantity of the correctly classified heartbeats of class I that is recorded as True positive of Class I (TP_I). As shown in Table 2, TP_I equals $C_{I,I}$.

The Whole Accuracy (Acc) standing for the overall performance of all classes is calculated in (1).

To evaluate the performance of each class, we define the Sensitivity (Sen) of Class I (Sen_I) as the proportion of correct classification in Class I heartbeats, which is shown in (2), and define the Positive predictivity (+P) of Class I ($+P_I$) as the probability of correct classification in heartbeats classified as Class, which is shown in (3).

$$Acc = (TP_N + TP_S + TP_V + TP_F + TP_Q)/(\sum(All)) \quad (1)$$

$$Sen_I = TP_I/\sum(I) \quad (2)$$

$$+P_I = TP_I/P_I \quad (3)$$

By using all the metrics mentioned above, we evaluate the performance of this method and compare it with other works.

III. METHOD

We propose four steps to build a complete model of the automated heartbeat classification method. First, the original ECG signals are preprocessed into normalized heartbeat signal segments. Second, we utilize the common features such as morphology, RR intervals and beat-to-beat correlation of the segments to form the inputs of the network as these features provide a more comprehensive description of the heartbeat. These kinds of features of all channels constitute the 3-D structure as the inputs, which is shown in the second block of Fig. 1. Third, a convolutional neural network with residual blocks, atrous spatial pyramid pooling (ASPP) module, global average pooling (GAP) and fully connected (FC) layer is established to extract deep multi-resolution features automatically and classify heartbeats. The outputs of the network are utilized as the prediction results. Finally, we utilize the mini-batch gradient descent (MBGD) method to train the CNN. Fig. 1 shows the flow of the process. In evaluation, the heartbeats are predicted one by one and the result of the current heartbeat is used to form the beat-to-beat correlation feature of the next heartbeat.

A. DATA PREPROCESSING

All original ECG signals in the MIT-BIH arrhythmia database are preprocessed in three stages to generate the normalized heartbeat segments which are used as the morphological features in this method. The algorithm 1 illustrates the preprocessing pseudo-code.

The Function Segment is employed to divide the signals into segments and guarantee each of them contains only one heartbeat (the locations of R-peaks exist in the database annotation). Each segment starts at the 0.14s (50 points at a sampling rate of 360) after the previous R-peak (P1), rather than the fixed time before current R-peak, and ends at the 0.28s (100 points at a sampling rate of 360) after current R-peak (P2). The approach helps the segments to contain more comprehensive morphological features, as the heart rate is always changing and the fixed start point may lose information (heartbeats with wide P-QRS-T complexes) or bring in interference information (heartbeats with short RR-intervals).

Such segmentation way will result in different widths of segments, which fails to satisfy the fixed inputs shape requirement of the CNN. Therefore, with resampling method, all the split heartbeat segments are scaled into the same width 128 in Function Scale.

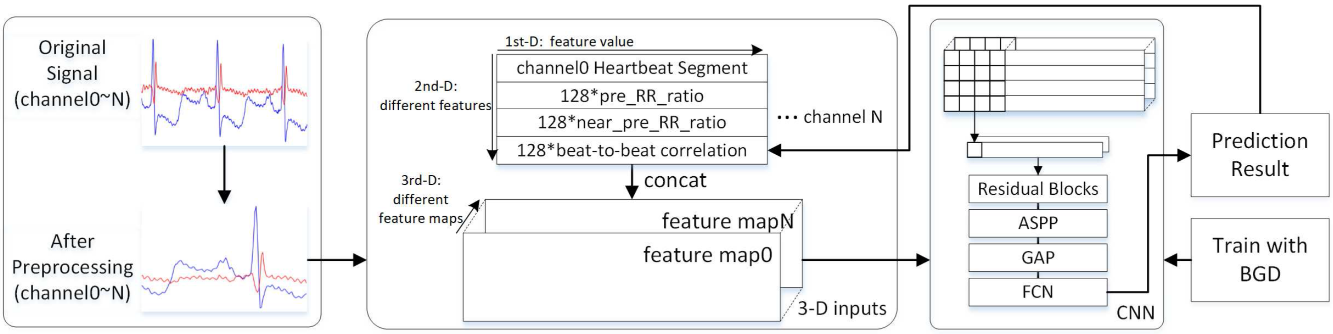

FIGURE 1. Schematic representation of the proposed CNN-based method for Heartbeat Classification.

Besides, the baseline wander commonly found in ECG signals [1] will cause covariate shift [26]. In Function SuppressDC, we suppress the baseline wander with a simple method (i.e., subtract the average of a segment from each point), and output the normalized heartbeat segment with an average of 0. This approach locks all segments into similar locations and thus advances the ability of the neural network to be spatially invariant to the input data, which can improve the performance of the network and accelerate the training [27]. Because CNN naturally has a certain noise tolerance, other sophisticated noise reduction methods are not adopted in our work.

Algorithm 1 Signal Preprocessing

Define:

sig: original ECG signal;
final_seg: preprocessed heartbeat segments;
Segment: divide signal to heartbeat segments;
Scale: resample signal at given frequency;
SuppressDC: suppress direct current component;

$seg\_queue = Segment(sig, P1, P2, annotation)$
for each *seg* in *seg_queue* **do**
  $resamp\_seg = Scale(seg, width)$
  $final\_seg = SuppressDC(resamp\_seg)$
end for

B. CNN INPUTS GENERATING

The proposed 3-D inputs structure supports multiple feature maps, each of which consists of features extracted from a different ECG channel, which is shown in the second block of Fig. 1. Although more channels will bring more different information, restricted by the MIT-BIH arrhythmia database, two channel signals are used in our experiments.

Because some heartbeats of different classes ((N vs. SVEB), (VEB vs. F) and (N vs. F)) naturally characterize with considerable morphological similarity of the heartbeat waveforms, the RR interval and beat-to-beat correlation features which have additional potential to reveal differences of beat classes are involved to help improve the classification performance.

The interval of current and previous R-peak namely pre-RR interval is generally different between normal and arrhythmic heartbeats of a person [15]. However, for the patient population, differences in the individual basic heart rate can cause overlapping of the pre-RR interval distributions between normal and arrhythmic heartbeats. The pre-RR ratio [14] (the ratio of the current pre-RR interval to the average of all pre-RR intervals of the corresponding record) is presented to overcome the deficiency. The approach of calculating the pre-RR ratio is similar to unifying the basic heart rate of everyone, and it helps to eliminate the overlap. Since it is not possible to obtain the RR-intervals behind the current heartbeat during prediction, all pre-RR intervals before the current heartbeat rather than all pre-RR intervals of the record are used to calculate the ratio. As the basic heart rate of a person varies with mood and movement state [8], we present the near-pre-RR ratio (the ratio of the current pre-RR interval to the average of ten pre-RR intervals before the heartbeat) to adapt to it. These two ratios injected into the 3-D inputs help the neural network to increase adaption of inter-individual and intra-individual heart rate changes.

Given that the feature of beat-to-beat correlation can assist in boosting the classification performance [18], we add it to the inputs of CNN. Table 3 illustrates that SVEBs and VEBs are always adjacent to the beat of the same class or N class and the obvious class-correlation between adjacent beats has additional potential to help distinguish heartbeats when the heartbeats are similar to more than one class in morphological and RR-interval features. As the beat-to-beat correlation cannot be directly obtained in inference, the prediction result of the previous heartbeat will be used as the feature of the current beat.

The most direct way to build the inputs of CNN is to splice these scalar features and the normalized heartbeat segments into longer 1-D feature maps and then compose the 1-D inputs (only feature value dimension) or 2-D inputs (feature value and channel dimensions). However, with low occupancy in the inputs, these scalar features are difficult to learn by the network. Thus we propose the 3-D inputs structure as the second block of Fig. 1 shows. First, all the scalar features are duplicated to vectors of the same width as

TABLE 3. Neighboring heartbeats statistics in MIT-BIH arrhythmia database.

Previous Beat	Current Beat	Percentage of Total Current Class Beat
SVEB	SVEB	58.78%
N	SVEB	40.38%
VEB	SVEB	0.70%
F	SVEB	0.14%
N	VEB	81.44%
VEB	VEB	13.98%
F	VEB	4.33%
SVEB	VEB	0.25%
SVEB	SVEB	60.56%
SVEB	N	38.64%
SVEB	VEB	0.68%
SVEB	F	0.11%
VEB	N	85.56%
VEB	VEB	13.99%
VEB	SVEB	0.27%
VEB	F	0.19%

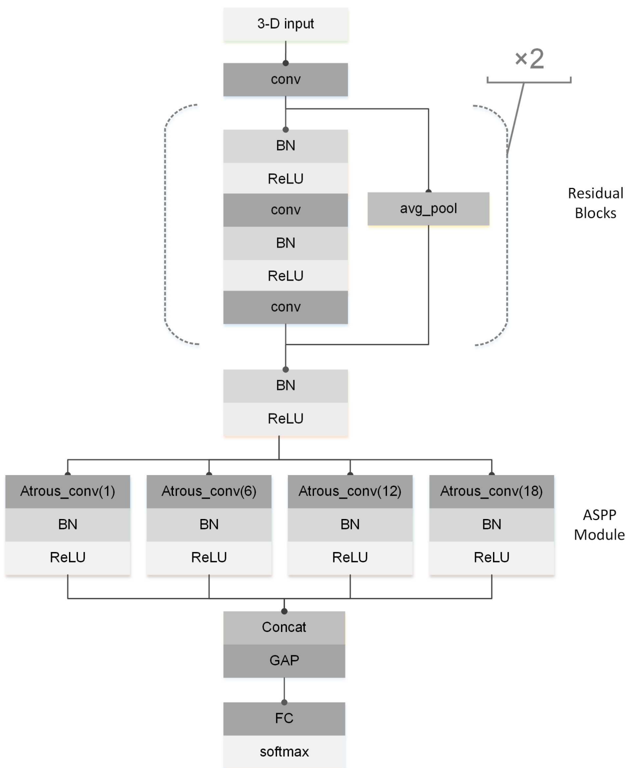

FIGURE 2. High-level architecture of the proposed CNN. The neural network contains one first convolution layer, two residual blocks, one ASPP module, one GAP layer, and one FC layer.

the normalized heartbeat segments (128). Then, these vectors and the heartbeat segments in each channel are stacked in the height dimension. Finally, feature maps are stacked in channel dimension to form the inputs.

C. CNN

The proposed CNN is responsible for inferring the classes of input samples. As depicted in Fig. 2, the network contains 10 convolution layers (conv) and one FC layer in total. The output of the FC layer is fed to a softmax which produces the prediction result of the network.

The micro-architectures of each layer of the network are shown in Table 4. To compress the size of the network, in the first convolution layer, the 4×4 filters with valid padding is presented to convolute the $4 \times 128 \times 2$ inputs to $1 \times 125 \times 16$ tensor. In all subsequent layers, all filters have the same shape of 1×3.

TABLE 4. Micro-architectures for the proposed CNN. The super parameters in each convolution layer.

Layer Name	Output Size	Kernel Size	Padding	Stride	Rate
3-D inputs	[4x128x2]				
First Layer	[1x125x16]	[4x4]	VALID	1	1
Res B1 Layer1	[1x125x16]	[1x3]	SAME	1	1
Res B1 Layer2	[1x125x16]	[1x3]	SAME	1	1
Res B2 Layer1	[1x63x64]	[1x3]	SAME	2	1
Res B2 Layer2	[1x63x64]	[1x3]	SAME	1	1
ASPP Layer1	[1x61x32]	[1x3]	VALID	1	1
ASPP Layer2	[1x61x64]	[1x3]	VALID	1	6
ASPP Layer3	[1x61x64]	[1x3]	VALID	1	12
ASPP Layer4	[1x61x64]	[1x3]	VALID	1	18
Concat	[1x61x224]				
GAP	[224]				
FC	[5]				
Parameters	63.3k				
FLOPs	3.89M				

*First Layer: the fisrt conv layer in the network
*Res B(x) Layer(y): conv layer y in the Residual Block x
*ASPP Layer(x): Atrous_conv layer x in the ASPP module

To make the network easier to optimize, and to gain accuracy from the increase of the convolution layer, following the first convolution layer, two residual blocks with average pooling (avg_pool) shortcuts are joined up [28] in cascade.

The ASPP module connected to the residual blocks enables the CNN to own multi-fields of view to exploit multi-scale features [29], [30]. Although the atrous convolution is proposed to solve the problem of image segmentation, the idea of extracting features with different resolutions is very suitable for the analysis of temporal signals. Four parallel atrous convolution layers (Atrous_conv(x)) with different atrous rates are applied in the module. Considering the basic time characteristics of P-QRS-T waves (variation range of time), we set the atrous rate values to 1, 6, 12, and 18 to extract both detailed and macroscopic features of heartbeats. The output feature maps of ASPP module are concatenated in channel dimension (Concat) and injected into the GAP layer which has no parameters to lead to overfitting. Furthermore, GAP greatly reduces parameters of the network and is more robust to spatial translations of the inputs [31]. The concrete implementations of ASPP module and GAP layer are shown in Fig. 3.

Batch normalization (BN) [26] technique is ubiquitous in all convolution layers, as experiments show that it helps the CNN with ASPP module to be trained better [29]. The output of every BN layer is activated with the Rectified Linear

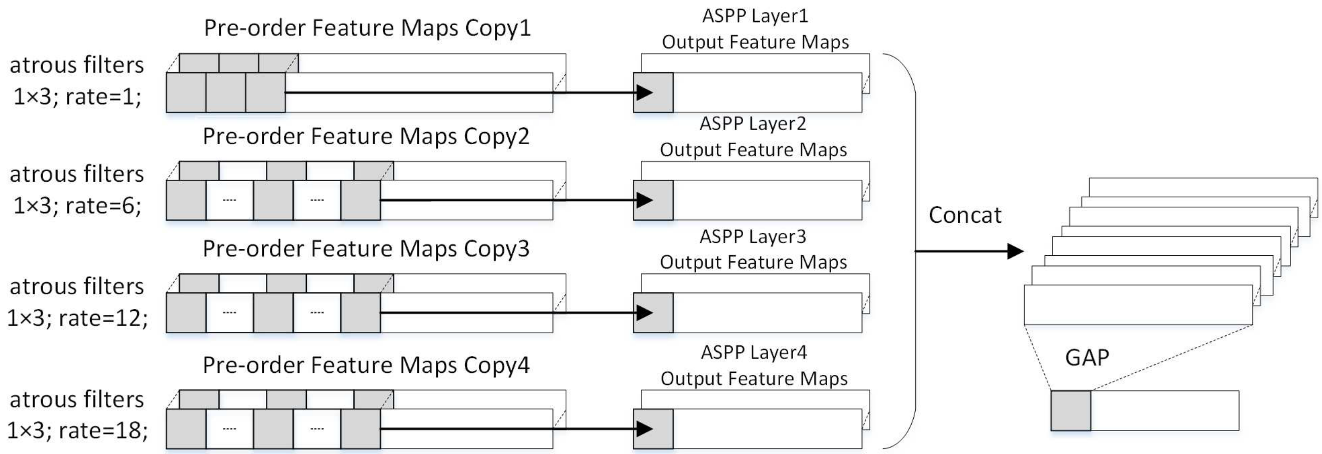

FIGURE 3. ASPP module and GAP layer in the proposed CNN. All the output feature maps of the ASPP module are merged in channel dimension.

Unit (ReLU) since CNN with ReLUs trains several times faster than their equivalents with hyperbolic tangent (tanh) units [32].

There are about 63.3k parameters in the network, and one inference takes about 3.89M floating point operations (multiply-adds).

D. PARAMETERS TRAINING

All about 49660 different training samples in DS1 are utilized to train the proposed neural network. As mentioned above, the training data is seriously imbalanced, which has significant negative effects on the performance of the proposed method. To overcome the inherent defect and achieve a higher and more balanced performance, especially for abnormal classes, we duplicate SVEB, VEB, and F beat samples and make the quantity of each of them similar to the quantity of N beat samples, i.e., duplicate SVEB, VEB, and F beats 24, 13 and 113 times respectively.

The network is trained by the MBGD method with a mini-batch size of 128. To balance each mini-batch and balance the training number of each sample, we randomly divide the whole extended training dataset into batches, and each time sequentially select one batch as the input mini-batch to train the network (the parameters will be updated for each mini-batch). After all batches have been traversed, keep repeating the process until the loss converges.

IV. RESULTS

This section presents the classification results of all records in DS2 and the comparisons with other papers. The samples in evaluation dataset are tested one by one to meet the actual scenario in clinic and satisfy the 3-D inputs structure.

A. INDIVIDUAL CLASSIFICATION

In inter-patient practice, there are few methods that could achieve great performance in all classes. In general, it is hard to attain high performance in both N and SVEB classes at the same time. Most papers tend to achieve a relatively high Sen_N (over 90%) and a relatively low Sen_S (less than 80%) that is more likely to cause the problem of missing detection of diseases. The group of F, VEB and N classes faces the same dilemma. The proposed method attempts to balance the performance of all normal and abnormal classes as much as possible, and achieves a more comprehensive performance in most records. Table 6 shows the detailed classification results of each record in DS2.

TABLE 5. Confusion matrices of the proposed method.

		Algorithm label					
		n	sveb	veb	f	q	Sen
Reference label	N	40587	2938	232	453	0	91.81%
	SVEB	140	1635	55	6	0	89.05%
	VEB	82	44	3063	30	0	95.15%
	F	218	0	45	125	0	32.22%
	Q	3	0	4	0	0	0.00%

Over 80% of the records get an Acc of higher than 90% which is strongly related to Sen_N as the proportion of normal beats is close to 90%. In Sen_N, nearly two-thirds of records exceed 96%, and over 80% of the records achieve more than 91%. At least 80% of the records surpasses 98% in $+P_N$.

As suggested by the ANSI/AAMI EC57 standard [23], we focus on evaluating the performance of two majority arrhythmia classes (SVEB and VEB). Record 232 with over 77% SVEBs reaches a very high Sen_S of 96.09% and a $+P_S$ of 99.92%. The records with 1% to 10% SVEBs attain an acceptable average Sen_S of 84.52% and an average $+P_S$ of 52.56%, while the rest records with less than 1% SVEBs achieve a low Sen_S and $+P_S$.

In Sen_V, 63% of the records exceed 97%, and seven-eights of the records attain more than 85%. The $+P_V$ of nearly two-thirds of records surpasses 90%.

TABLE 6. Detailed classification results of every record in DS2.

Record	Number of Beats					N		SVEB		VEB		F		Acc
	All	N	SVEB	VEB	F	Sen	+P	Sen	+P	Sen	+P	Sen	+P	
100	2270	2236	33	1	0	98.79%	100.00%	100.00%	55.00%	100.00%	100.00%	-	-	98.81%
103	2082	2080	2	0	0	99.13%	99.90%	0.00%	0.00%	-	-	-	-	99.04%
105	2570	2524	0	41	0	96.79%	99.88%	-	-	92.68%	36.89%	-	-	96.54%
111	2122	2121	0	1	0	99.86%	100.00%	-	-	100.00%	50.00%	-	-	99.86%
113	1792	1786	6	0	0	99.78%	99.94%	0.00%	0.00%	-	-	-	-	99.44%
117	1532	1531	1	0	0	99.87%	99.93%	0.00%	0.00%	-	-	-	-	99.80%
121	1861	1859	1	1	0	99.84%	99.95%	0.00%	0.00%	100.00%	33.33%	-	-	99.79%
123	1516	1513	0	3	0	85.72%	100.00%	-	-	100.00%	100.00%	-	-	85.75%
200	2598	1741	30	825	2	94.20%	97.91%	73.33%	15.28%	92.85%	98.33%	0.00%	0.00%	93.46%
202	2134	2059	55	19	1	96.36%	99.10%	81.82%	44.55%	36.84%	50.00%	0.00%	0.00%	95.41%
210	2647	2420	22	195	10	91.07%	98.88%	59.09%	25.00%	88.21%	92.97%	10.00%	0.55%	90.29%
212	2745	2745	0	0	0	98.98%	100.00%	-	-	-	-	-	-	98.98%
213	3248	2638	28	220	362	99.92%	92.49%	0.00%	0.00%	86.36%	78.19%	33.43%	78.51%	90.73%
214	2260	2001	0	256	1	99.75%	99.80%	-	-	99.22%	99.22%	0.00%	0.00%	99.56%
219	2152	2080	7	64	1	61.39%	99.53%	14.29%	0.14%	98.44%	40.91%	0.00%	0.00%	62.31%
221	2425	2029	0	396	0	68.75%	100.00%	-	-	99.75%	100.00%	-	-	73.81%
222	2481	2272	209	0	0	91.90%	97.89%	67.46%	47.96%	-	-	-	-	89.84%
228	2051	1686	3	362	0	99.17%	99.52%	33.33%	100.00%	98.34%	97.27%	-	-	98.93%
231	1569	1566	1	2	0	36.08%	99.65%	0.00%	0.00%	50.00%	100.00%	-	-	36.07%
232	1778	397	1381	0	0	93.70%	87.53%	96.09%	99.92%	-	-	-	-	95.56%
233	3076	2228	7	830	11	91.88%	98.84%	28.57%	66.67%	97.83%	99.27%	27.27%	1.63%	93.11%
234	2751	2698	50	3	0	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	-	-	100.00%
Total	49660	44210	1836	3219	388	91.81%	98.92%	89.05%	35.41%	95.15%	90.11%	32.22%	20.36%	91.44%

Record 213 with more than 11% of F beats achieves a Sen_F of 33.43% and a $+\text{P}_F$ of 78.51%. Other records with less than 1% F beats attain an average Sen_F of 6.21% and an average $+\text{P}_F$ of 0.36%.

There are still deficiencies in the method for a small part of records. A certain number of normal beats and SVEBs are misclassified to each other, as they have similar signal segments. Many F beats, and N or VEB beats are misclassified to each other, as the morphology of the F beats varies from almost N beats to almost VEB beats.

It is unlikely to train a perfect model for all classes of the heartbeat with the limited training dataset. In this paper, SVEBs and VEBs, the main abnormal beats, are given primary consideration, ensuring an acceptable performance in N (more than 90%) and a certain performance in F beats. Table 5 shows the confusion matrices of the five-class heartbeats classification results. It indicates that this method has potential clinical application.

B. PERFORMANCE COMPARISON

Table 8 shows the performance comparisons between the proposed method and several other methods that follow the guidelines of the AAMI. With the recommendation of AAMI, we focus on comparing the performance in detecting SVEBs and VEBs. As Sensitivity which stands for the proportion of true positive samples is particularly significant for clinical use [6], the comparisons focus on Sensitivity with consideration of positive predictivity.

Our work adopts the evaluation scheme proposed by Philip de Chazal et al. [1] as [3], [6], [8] and [22] do. These papers utilize different classifiers as Table 8 shows. It can be seen that the proposed method greatly improves the performance

TABLE 7. Effects of segment width.

SW	Sen_N	$+\text{P}_N$	Sen_S	$+\text{P}_S$	Sen_V	$+\text{P}_V$	Sen_F	$+\text{P}_F$
64[a]	92.80%	98.68%	84.86%	37.19%	91.46%	90.42%	37.11%	23.08%
96	93.94%	98.83%	89.49%	43.19%	96.05%	91.56%	1.80%	1.54%
128	91.81%	98.92%	89.05%	35.41%	95.15%	90.11%	32.22%	20.36%
196	92.48%	98.90%	89.60%	44.86%	95.31%	91.07%	30.41%	9.22%
256	93.17%	98.62%	86.27%	41.29%	91.86%	84.29%	36.86%	26.24%

[a] The atrous rate values are settled to 1, 6, 9, and 12, as the kernels with 18 rate exceed the width of the feature maps in ASPP module.

of Sen_S and achieves a similar Sen_V compared with the top performance work. Like the other methods, the $+\text{P}_S$ of the proposed method is low, but the $+\text{P}_V$ of this method is up to 90%, which is better than most other methods. Besides, this method achieves an acceptable performance in N class. Compared with [6], the proposed method achieves a higher average accuracy in detecting the main pathological beats (SVEBs and VEBs) (92.10% vs. 87.72%) and a higher average accuracy in detecting the normal and main pathological beats (92.00% vs. 90.75%). In general, the performance advantage of the method, especially in abnormal classes, demonstrates its clinical application value.

A state-of-the-art patient-specific method based on 2-D CNN is presented in [18]. In order to compare performance with it, we add some extra heartbeats to the training set as they do. In the patient-specific practice, our method achieves an obvious advantage in abnormal classes and a similar performance in N class as Table 8 shows.

In general, the proposed method achieves a special advantage in detecting SVEBs and VEBs which is of great importance in clinic, and also gains an acceptable performance in N and F classes.

TABLE 8. Comparisons of the classification performance with the previous works.

Paper	Classifier Paradigm	Sen_N	$+P_N$	Sen_S	$+P_S$	Sen_V	$+P_V$	Sen_F	$+P_F$
Philip de Chazal et al. [1]	LDs	86.86%	99.00%	75.94%	38.50%	77.74%	81.59%	89.43%	8.57%
Mar Tanis et al. [8]	MLP	89.64%	99.12%	83.22%	33.53%	86.75%	75.89%	61.08%	16.57%
Can Ye et al. [3]	SVM	88.51%	97.54%	60.86%	52.34%	81.49%	61.38%	19.59%	2.50%
Escalona-Moran et al. [6]	RC	96.82%	98.90%	79.37%	49.80%	96.06%	99.49%	92.26%	95.40%
Janne Takalo-Mattila et al. [22]	CNN	91.89%	97.00%	62.49%	55.86%	89.23%	50.85%	0.52%	2.63%
Zhai Xiaolong et al. [18][a]	CNN	97.64%	98.51%	76.79%	74.04%	93.80%	92.36%	79.58%	62.36%
proposed	CNN	91.80%	98.92%	89.05%	35.41%	95.15%	90.11%	32.22%	20.36%
proposed[a]	CNN	96.96%	99.07%	86.19%	72.49%	92.16%	88.46%	83.80%	59.81%

*MLP: multilayer perceptron
[a]Patient-specific practice: The heartbeats in the first 5 minutes of the testing records are added to the DS1 for training, the rest of heartbeats in DS2 are used for testing.
[a]DS1: records 100, 101, 103, 105, 106, 108, 109, 111, 112, 113, 114, 115, 116, 117, 118, 119, 121, 122, 123 and 124
[a]DS2: records 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

TABLE 9. Comparisions of results using different input forms.

Experiments	Sen_N	$+P_N$	Sen_S	$+P_S$	Sen_V	$+P_V$
Original	91.81%	98.92%	89.05%	35.41%	95.15%	90.11%
No. 1	84.76%	98.26%	74.40%	37.42%	94.19%	63.51%
No. 2	82.11%	99.36%	93.08%	26.84%	92.11%	70.11%

*Original: result of the proposed method with proposed 3-D Inputs
*No. 1: result of the proposed method with segments that start from the 0.28s before the current R-peak and end with the 0.56s after the current R-peak [3]
*No. 2: result of the proposed method with segments that start from the 0.20s before the current R-peak and end with the 0.28s after the current R-peak [6]

V. DISCUSSION

The results show that the proposed method based on CNN achieves a competitive classification performance. In this section, we will discuss the influence of the 3-D inputs and multi-fields of view on the performance.

A. EFFECTS OF SEGMENT WIDTH

To determine the effect of segment width (SW) on the performance, we experiment with 5 different widths (64, 96, 128, 196, and 256). Table 7 shows that the experiments with 96, 128, and 196 SWs achieve similar performance in N, SVEB and VEB classes. However, the Sen_F of the experiment with 96 SW is really low (1.80%), and the 196 SW does not result in additional performance improvements, so the SW is eventually set to 128 in this study.

B. EFFECTS OF INPUTS FORM

The proposed 3-D inputs consist of several particular parts, including the normalized segments, two types of pre-RR ratio, and the feature of beat-to-beat correlation. Experiments show that without the pre-RR ratios, the CNN can only gain at most 60% of Sen of Class N and SVEB in average, which is not necessary to be discussed with tables. The effects of importing the normalized segments and beat-to-beat correlation are discussed below.

1) EFFECTS OF THE PROPOSED NORMALIZED SEGMENT

Different from the heartbeat segmentation with a fixed start point before current R-peak [3], [6], the proposed segment starts from an adaptive point to accommodate different heart rates. We utilize the same additional features (features of RR-interval and beat-to-beat correlation) and the three kinds of heartbeat segments (the proposed segment and the segments mentioned in [3] and [6]) respectively to build the inputs and train the proposed CNN with them. The results shown in Table 9 illustrate that the proposed segment with more complete heartbeat information helps the CNN classifying heartbeats better.

2) EFFECTS OF IMPORTING THE BEAT-TO-BEAT CORRELATION

We set up four extra experiments with controlling variables method to explore the influence of importing the feature of beat-to-beat correlation, as shown in Table 10. To quantify the holistic effects of importing the previous heartbeat class that stands for the feature of beat-to-beat correlation, we set up the No. 1 experiment, i.e., the CNN is trained and evaluated using the 3-D inputs without importing the previous heartbeat class. It can be seen that adding the previous heartbeat class as the feature of beat-to-beat correlation will increase the Sen_S by over 10% without performance loss in N and VEB classes. In No. 2 experiment, we adopt the adjusted beat-to-beat correlation, i.e., if there are 20 consecutive predictions that are abnormal, we will initialize the current classification result to 0 which stand for the basic heartbeat (N class), which is reasonable because abnormal heartbeats appear infrequently and there are inevitably misclassifications. Comparison experiments show that this method achieves about 2% of increment in Sen_N with tiny decrement in Sen_S (0.55%) and Sen_V (0.05%). To determine the influence of misclassifications, we use the exact previous heartbeats classes instead of the predictions in No. 3 experiment. It shows that there is an overall improvement in all N, SVEB and VEB beats, i.e., more accurate predictions in the past will result in more accurate predictions in the coming.

TABLE 10. Effects of importing the beat-to-beat correlation.

Experiments	Sen_N	$+\text{P}_N$	Sen_S	$+\text{P}_S$	Sen_V	$+\text{P}_V$
Original	91.81%	98.92%	89.05%	35.41%	95.15%	90.11%
No. 1	91.89%	98.64%	80.50%	41.17%	94.87%	93.05%
No. 2	93.61%	98.92%	88.56%	42.69%	95.10%	90.06%
No. 3	95.36%	99.06%	91.45%	55.80%	95.25%	89.44%

*Original: the trained CNN with the original last predictions
*No. 1: the retrained CNN without the information of the past signal
*No. 2: the trained CNN with the adjusted last predictions
*No. 3: the trained CNN with the real last heartbeats classes

TABLE 11. Effects of importing multi-fields of view.

Experiments	Sen_N	$+\text{P}_N$	Sen_S	$+\text{P}_S$	Sen_V	$+\text{P}_V$
Original	91.81%	98.92%	89.05%	35.41%	95.15%	90.11%
No. 1	87.23%	98.33%	88.18%	43.74%	92.30%	88.00%

*Original: the method with multi-fields of view
*No. 1: the method without multi-fields of view

C. EFFECTS OF MULTI-FIELDS OF VIEW

With four different parallel atrous convolution layers, the ASPP module enables the neural network to own multi-fields of view which helps to exploit multi-scale features. To analyze the specific strength of the multi-fields of view, we set up an extra experiment without it, i.e., use the filters with same rate in all four parallel convolution layers. Table 11 shows the results of these two comparison experiments. Compared with the method without multi-fields of view, the proposed method achieves an advantage in all N, SVEB and VEB beats. Concretely, there is a 5.25% increment in Sen_N, a 0.97% increment in Sen_S, and a 3.09% increment in Sen_V.

VI. CONCLUSION

In this paper, a method based on CNN for automated inter-patient heartbeat classification is proposed. In this method, we present a 3-D inputs and a 10 convolution layer neural network with multi-fields of view to acquire a high performance. The 3-D inputs structure supports multiple feature maps and each of them contains the morphological characteristic, RR-interval features, and beat-to-beat correlation extracted from different leads respectively. Filters of different rates in the ASPP module enable the CNN to have multi-filed of view to exploit the multi-scale features. Based on the benchmark MIT-BIH

arrhythmia database and the practice recommended by AAMI, the proposed method achieves an overall accuracy of 91.44%. In terms of single class performance, it yields a normal class sensitivity of 91.81%, an SVEB class sensitivity of 89.05% and a VEB class sensitivity of 95.15%. Compared with the existing works, this method achieves the highest average sensitivity of these three classes beats and has a great advantage in detecting abnormal heartbeats, which is pivotal in clinic. The results show that the method is of positive significance in clinical detection of heart diseases.

REFERENCES

[1] P. De Chazal, M. O.'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1196–1206, Jul. 2004.

[2] A. B. Rad, T. Eftestøl, K. Engan, U. Irusta, J. T. Kvaløy, J. Kramer-Johansen, L. Wik, and A. K. Katsaggelos, "ECG-based classification of resuscitation cardiac rhythms for retrospective data analysis," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 10, pp. 2411–2418, Oct. 2017.

[3] C. Ye, B. V. K. V. Kumar, and M. T. Coimbra, "Heartbeat classification using morphological and dynamic features of ECG signals," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 10, pp. 2930–2941, Oct. 2012.

[4] P. Li, Y. Wang, J. He, L. Wang, Y. Tian, T.-S. Zhou, T. Li, and J.-S. Li, "High-performance personalized heartbeat classification model for long-term ECG signal," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 78–86, Jan. 2017.

[5] J. Oster, J. Behar, O. Sayadi, S. Nemati, A. E. W. Johnson, and G. D. Clifford, "Semisupervised ECG ventricular beat classification with novelty detection based on switching Kalman filters," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 9, pp. 2125–2134, Sep. 2015.

[6] M. A. Escalona-Morán, M. C. Soriano, I. Fischer, and C. R. Mirasso, "Electrocardiogram classification using reservoir computing with logistic regression," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 3, pp. 892–898, May 2015.

[7] U. Satija, B. Ramkumar, and M. S. Manikandan, "Automated ECG noise detection and classification system for unsupervised healthcare monitoring," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 3, pp. 722–732, May 2018.

[8] T. Mar, S. Zaunseder, J. P. Martínez, M. Llamedo, and R. Poll, "Optimization of ECG classification by means of feature selection," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 8, pp. 2168–2177, Aug. 2011.

[9] E. Choi, A. Schuetz, W. F. Stewart, and J. Sun, "Using recurrent neural network models for early detection of heart failure onset," *J. Amer. Med. Informat. Assoc.*, vol. 24, no. 2, pp. 361–370, Mar. 2017.

[10] Z. Xiong, M. K. Stiles, and J. Zhao, "Robust ECG signal classification for detection of atrial fibrillation using a novel neural network," in *Proc. Comput. Cardiol. (CinC)*, vol. 44, Sep. 2017, pp. 1–4.

[11] M. Zubair, J. Kim, and C. Yoon, "An automated ECG beat classification system using convolutional neural networks," in *Proc. ICITCS*, Sep. 2016, pp. 1–5.

[12] S. Hong, M. Wu, Y. Zhou, Q. Wang, J. Shang, H. Li, and J. Xie, "ENCASE: An ENsemble ClASsifiEr for ECG classification using expert features and deep neural networks," in *Proc. Comput. Cardiol. (CinC)*, Sep. 2017, pp. 1–4.

[13] 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)*, Oct. 2017, pp. 1–6.

[14] H. Huang, J. Liu, Q. Zhu, R. Wang, and G. Hu, "A new hierarchical method for inter-patient heartbeat classification using random projections and RR intervals," *Biomed. Eng. Online*, vol. 13, no. 1, p. 90, Dec. 2014.

[15] C.-C. Lin and C.-M. Yang, "Heartbeat classification using normalized RR intervals and morphological features," *Math. Problems Eng.*, vol. 2014, May 2014, Art. no. 712474.

[16] M. Sadrawi, C.-H. Lin, Y.-T. Lin, Y. Hsieh, C.-C. Kuo, J. C. Chien, K. Haraikawa, M. F. Abbod, and J.-S. Shieh, "Arrhythmia evaluation in wearable ECG devices," *Sensors*, vol. 17, no. 11, p. 2445, Oct. 2017.

[17] M. L. Soria and J. Martínez, "Analysis of multidomain features for ECG classification," in *Proc. 36th Annu. Comput. Cardiol.*, Sep. 2009, pp. 561–564.

[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] C. Ye, B. V. K. V. Kumar, and M. T. Coimbra, "Combining general multi-class and specific two-class classifiers for improved customized ECG heartbeat classification," in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR)*, Nov. 2012, pp. 2428–2431.

[20] G. De Lannoy, D. Francois, J. Delbeke, and M. Verleysen, "Weighted conditional random fields for supervised interpatient heartbeat classification," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 1, pp. 241–247, Jan. 2012.

[21] Y. Bazi, N. Alajlan, H. AlHichri, and S. Malek, "Domain adaptation methods for ECG classification," in *Proc. Int. Conf. Comput. Med. Appl. (ICCMA)*, Jan. 2013, pp. 1–4.

[22] J. Takalo-Mattila, J. Kiljander, and J.-P. Soininen, "Inter-patient ECG classification using deep convolutional neural networks," in *Proc. 21st Euromicro Conf. Digit. Syst. Design (DSD)*, Aug. 2018, pp. 421–425.
[23] AAMI, *AAMI EC57:2012 Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms*. Arlington, VA, USA: AAMI, 1998.
[24] 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/Jun. 2001.
[25] D. Masko and P. Hensman, "The impact of imbalanced training data for convolutional neural networks," Ph.D. dissertation, School Comput. Sci. Commun., KTH Roy. Inst. Technol., Stockholm, Sweden, 2015.
[26] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: https://arxiv.org/abs/1502.03167
[27] M. Jaderberg, K. Simonyan, K. Kavukcuoglu, and A. Zisserman, "Spatial transformer networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 2017–2025.
[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[29] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*. [Online]. Available: https://arxiv.org/abs/1706.05587
[30] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2017.
[31] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, *arXiv:1312.4400*. [Online]. Available: https://arxiv.org/abs/1312.4400
[32] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

FEITENG LI was born in Shijiazhuang, Hebei, China, in 1992. He received the B.S. degree in the Internet of Things engineering from the School of Information Science and Technology, Southwest Jiaotong University, Chengdu, China, in 2014. He is currently pursuing the Ph.D degree with the College of Electrical Engineering, Zhejiang University, Hangzhou, China. His research interests include physiological signals processing with machine learning and ultra-lower-power neural network accelerator.

YIN XU was born in Ningbo, Zhejiang, China, in 1990. She received the B.S. degree in electronic and information engineering from Zhejiang University, in 2013, and the Ph.D. degree from the College of Electrical Engineering, Zhejiang University, Hangzhou, China, in 2018. Her research interests include machine learning, and ultra-lower-power neural network accelerator.

ZHIJIAN CHEN received the B.S. degree from the College of Electrical Engineering, Zhejiang University, Hangzhou, China, in 2006, and the Ph.D. degree from the College of Electrical Engineering, Zhejiang University, Hangzhou, China, in 2011. From 2011 to 2013, he was a postdoctor in the College of Electrical Engineering, Zhejiang University. Since 2013, he has been a Lecturer in the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China. His research interest includes ultra-low-power physiological signal processor design.

ZHENYAN LIU was born in Yantai, Shandong, China in 1991. She received the B.S. degree in traditional Chinese medicine and the M.S. degree in acupuncture and moxibustion from the Chengdu University of Traditional Chinese Medicine, Chengdu, China, in 2014 and 2017, respectively. Since 2017, she has been a clinical doctor in Chengdu Tianfu New District People's Hospital. Her research interests include the mechanism of acupuncture in treating acute pancreatitis, and the relationship between physiological signals and diseases.