Deep Multi-instance Networks for Bundle Branch Block Detection from Multi-lead ECG

Jing Hu[1], Wei Zhao[1], Dongya Jia[1], Cong Yan[1], Hongmei Wang[1], Zhenqi Li[1], Jiansheng Fang[1], Ming Yang[1]

Abstract— **Bundle branch block (BBB) is one of the most common cardiac disorder, and can be detected by electrocardiogram (ECG) signal in clinical practice. Conventional methods adopted some kinds of hand-craft features, whose discriminative power is relatively low. On the other hand, these methods were based on the supervised learning, which required the high cost heartbeat annotation in the training. In this paper, a novel end-to-end deep network was proposed to classify three types of heartbeat: right BBB (RBBB), left BBB (LBBB) and others with a multiple instance learning based training strategy. We trained the proposed method on the China Physiological Signal Challenge 2018 database (CPSC) and tested on the MIT-BIH Arrhythmia database (AR). The proposed method achieved an accuracy of 78.58%, and sensitivity of 84.78% (LBBB), 51.23% (others) and 99.72% (RBBB), better than the baseline methods. Experimental results show that our method would be a good choice for the BBB classification on the ECG dataset with record-level labels instead of heartbeat annotations.**

I. Introduction

Bundle branch block (BBB) is one of the most common heart diseases, and can be detected by electrocardiogram (ECG) signal. An ECG is a waveform record of bioelectrical changes produced by the heart muscle [1]. BBB can cause incomplete blood transfer from the atrium to the ventricle, with QRS interval greater than 120 ms and irregular heartbeat. Therefore, automatic classification of BBB on ECG signal is an essential function in ECG monitoring.

In recent years, many BBB detection methods have designed some hand-craft features based on the the duration and morphology of ECG waves and adopted machine learning techniques such as artificial neural network (ANN) and support vector machine (SVM) for the classification [2], [3], [4]. Our previous work [5] developed both deep features and hand-crafted features, and then the features were classified by the technique named ENCASE. These methods achieved good performance in BBB classification. However, these methods all extracted hand-crafted features, which relying on expert domain knowledge, and then classified with trained classifiers such as SVM, ANN and ENCASE. To improve the discriminative power of the features, we designed a deep neural networks to automatically learn pattern features for BBB classification.

On the other hand, previous BBB classification techniques require heartbeat annotations labeled by trained cardiologists. However, during continuous ECG monitoring, an ECG record is as long as 24 – 48h and contains 100,000 – 200,000 heartbeats, so cardiologists have to spend too much time for heartbeat annotation. Therefore, training classification algorithms on ECG recordings with record-level labels instead of heartbeat annotations is becoming an important issue in clinical practice.

The multi-instance learning (MIL) would be a good choice to solve this problem [6]. In the MIL, the classifier is trained on the labeled bags, and each bag contained several unlabeled instances. Sun [7] has applied MIL to cluster the features from the heartbeats without labels, and the labels of magnetic resonance images were assigned to the landmarks on the images in the brain disease diagnosis [8].

In this work, a novel method was proposed to classify three types of heartbeat: left BBB (LBBB), right BBB (RBBB) and others (abbreviated as L, R, O) based on an end-to-end deep networks with MIL-based training technique. The major contributions of this work include: 1. An end-to-end deep network is proposed for automatic BBB classification, which incorporates feature learning into the process of building models. 2. An attention-based MIL learning strategy is proposed for the training of the deep network. The proposed training technique does not require the heartbeats annotations for the training ECG records. Instead, it automatically discovers abnormal heartbeats through multi-instance learning. 3. We performed the cross-dataset evaluation on public ECG databases. In all experiments, our method shows superior performance than state-of-the-art methods.

II. Proposed Method

We proposed an end-to-end deep network with MIL-based training technique for BBB classification on ECG signal without the labeled heartbeats. The framework was shown in Fig. 1. Firstly, each ECG record was resampled to 250Hz, and the baseline drift was filtered by a zero-phase Butterworth high-pass filter (cutoff frequency 0.5Hz). Secondly, we segmented the heartbeat using the method as [9], Fig. 2 illustrates the generation of heartbeat segments. Thirdly, we designed a DNN network named RBNN to automatically classify L, R, O. The segmented heartbeats were used as input of the RBNN. Finally, The heartbeat type (L, R, O) was output.

In the training process of the RBNN, the output of the softmax layer was ranked, and then the attention-based MIL pooling strategy was performed to calculate the loss.

A. Experimental Materials

The China Physiological Signal Challenge 2018 Database (CPSC) was used for training and validation. The well-known MIT-BIH Arrhythmia Database (AR) [10] was used

[1] J. Hu, W. Zhao, D Jia, C. Yan, H. Wang, Z. Li, J. Fang and M. Yang are with the CVTE Research, Guangzhou, China. {hujing, zhaowei, jiadongya, yancong, wanghongmei, fangjainsheng, yangming}@cvte.com

978-1-7281-1990-8/20/$31.00 ©2020 IEEE

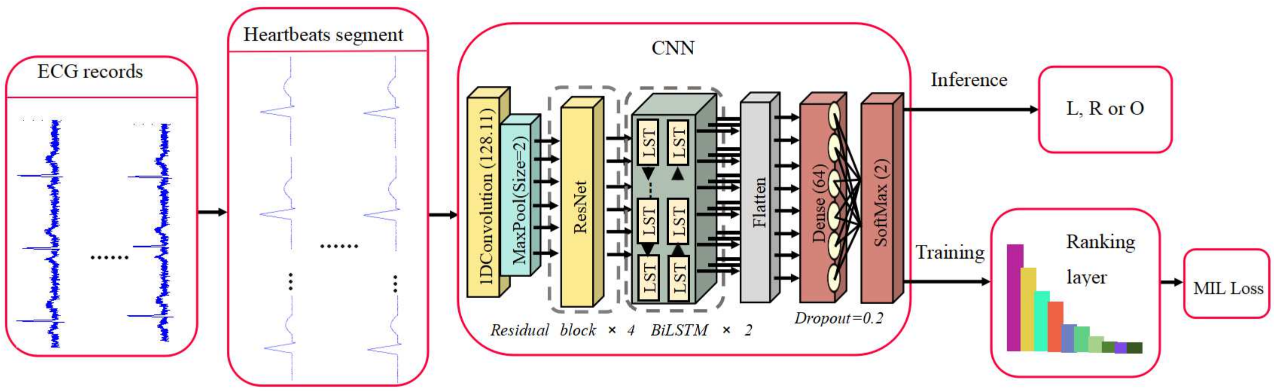

Fig. 1. Framework of proposed method

for test. The CPSC contains 6877 12-lead recordings digitized at 500Hz with record label, and the length of each recording was approximately $10 - 60$ seconds. The AR contains 48 two-lead ECG recordings with heartbeat annotations, sampling rate of which was 360Hz, and the length of each recording was approximately 30 minutes. For the CPSC, the heartbeat locations were detected by using the method [10] which dose not provide the beat annotation.

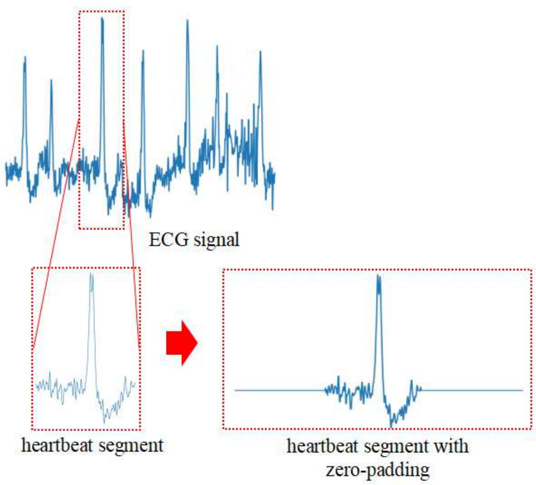

Fig. 2. Generation of beat segments

B. Deep RBNN Network

As shown in Fig. 3, we firstly built an end-to-end deep networks named RBNN for BBB classification. The proposed RBNN was composed of three concatenated modules: one convolutional layer with a Rectified Linear Unit (ReLU) plus a max-pooling layer, four residual convolution network (ResNet) blocks and two bidirectional long-short term memory network (BiLSTM) blocks. Then, the flatten layer was used to "flatten" the input, that is, one-dimensionalize the multi-dimensional input. The followed two fully-connected layers (dense layer and softmax layer) are used to calculate the probabilities of heartbeat belonging to L, R and O. The input of the RBNN model is the heartbeat segments of an ECG record with a size of 500 samples. The output is the class of the heartbeats (L, R or O).

For the first convolutional layer, the kernel size and number of the convolutional kernels were 11 and 128 respectively. For the max-pooling layer, the pooling window size was set as 2. For the ResNet blocks, the kernel size and channel number of each convolutional layer were 11 and 64 respectively. And the number of neurons in the BiLSTM block and the dense layer were 32 and 64 respectively.

C. Deep Multi-instance Learning

The training of heartbeat-based BBB classification on the record with a short length could be treated as a typical MIL problem. If there exists a heartbeat on the multi-lead ECG record that is positive, the ECG record was labeled positive. And a multi-lead ECG record was labeled negative, if and only if all the heartbeats of which are negative.

Denote $\mathbf{X}$ as the input multi-lead ECG, C as the number of heart diseases (classes: L , R or O). Given a set of training multi-lead ECG records $T = \{\mathbf{X}_m, m = 1, ..., M\}$ (M is the number of the training ECG records), with corresponding labels l_m (classes: L, R, O), $l_m \in \{1, ..., C\}$. For the m-th training record $\mathbf{X}_m$, it is segmented into a set of heartbeats defined as $H(\mathbf{X}_m) = \{\mathbf{x}_{mn}, n = 1, ..., N\}$, which represents the n-th heartbeat on m-th ECG record). N is 50 in our training process. Zero padding is performed when the number of heartbeats in a ECG record is less than 50. These heartbeats were used as the training samples of the RBNN and their labels are inherited from the original ECG records. As shown by Eq.1, for the input m-th ECG record $\mathbf{X_m}$, the output of the RBNN is the matrix $\mathbf{R_m}$ with a size of $\Re^{C\times N}$, where the l_m-th value of the n-th column $\mathbf{r_{mn}}$ representing the probability of the heartbeat belonging to the class l_m, and $\mathbf{W}$ is the coefficient of the RBNN including weights vector and bias vectors.

$$\underset{l_m \in \{1,\ldots,C\}}{\mathbf{r}_{mn}} = \mathbf{P}(y = l_m | \mathbf{x}_{mn}; \mathbf{W}) \tag{1}$$

For all the heartbeats on m-th ECG record, The $\mathbf{R}_m$ can be flattened into a one-dimensional vector as $\mathbf{r}_m = (\mathbf{r}_{m1}, \mathbf{r}_{m2}, ..., \mathbf{r}_{mN})$ corresponding to flattened heartbeats

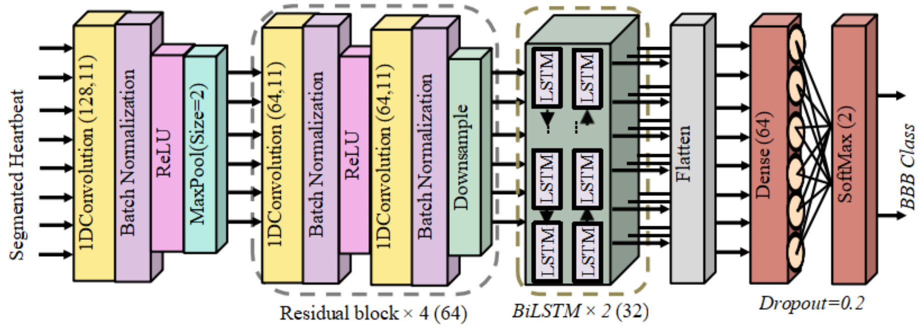

Fig. 3. Block diagram of the proposed RBNN model

$(\mathbf{x}_{m1}, \mathbf{x}_{m2}, ..., \mathbf{x}_{mN})$. We sort the value in the vector in descending order, as shown by Eq.2.

$$\{\mathbf{r}'_{m1}, \mathbf{r}'_{m2}, ..., \mathbf{r}'_{mN}\} = sort\{\mathbf{r}_{m1}, \mathbf{r}_{m2}, ..., \mathbf{r}_{mN}\} \tag{2}$$

We assume that (1) heartbeats of the first k largest probabilities $\{\mathbf{r}'_{m1}, \mathbf{r}'_{m2}, ..., \mathbf{r}'_{mk}\}$ should be assigned with the same class label as that of whole ECG, and (2) the rest heartbeats should be labeled as negative, (3) each instance should be paid different attention (weight). Where, k is the coefficient, estimated by an adaptive way. To discover similarities among instances, the proposed attention based MIL loss function was designed as

$$L(\mathbf{W}) = \sum_{\mathbf{X}_m \in T} -\log\left(\sum_{i=1}^{k} a_i \mathbf{r}'_{mi}\right) \tag{3}$$

where, weight a_i is attention coefficient, determined by neural network.

III. Experimental Results

The sensitivity (Se) and overall accuracy (Acc) are calculated to measure the performance of the proposed method, which are defined as:

$$\begin{aligned} Se &= TP/(TP + FN) \times 100\% \\ Acc &= (TP + TN)/(TP + TN + FN + FP) \times 100\% \end{aligned} \tag{4}$$

where TP, TN, FN, FP represent the true positive, true negative, false negative and false positive respectively.

The algorithms were trained on CPSC and tested on AR. Three previous BBB classification techniques (ANN [3], SVM [4] and ENCASE [5]) were used for comparison. The performances of proposed method and baseline methods were compared in Table I. The McNemars test was used to calculate the P value (P).

The proposed method achieved the best results compared with the state-of-the-arts. For the classification of the BBB, the Se of the L, R, O were 84.78%, 99.72% and 51.23%, respectively.

We also compared the proposed method with three typical MIL algorithms (MI-SVM and mi-SVM [11], and SIL-SVM [12], as shown in Table II. The input of three typical MIL algorithms were two types: 1) hand-crafted features, the same as all the manual features from [5]; 2) deep features, the output of the flatten layer in the proposed method. Obviously, the performance of the proposed MIL method was significantly better than three typical MIL algorithms, and the performance of the MIL algorithms with deep features from the proposed method was better than those with hand-crafted features.

IV. Discussion

Previous hand-crafted feature-based methods required manual feature design, relying on expert knowledge and experience. To reduce the incompleteness caused by artificial design features and improve the discriminative power of the features, we proposed an en-to-end deep network to classify BBB (L, R, O). The proposed method can learn pattern features automatically and incorporates feature learning into the process of building models.

ResNet is a popular framework of ECG signal analysis, and BiLSTM constituted of forward LSTM and backward LSTM, can encode both front-to-back and back-to-front information. So deep features from the RBNN can indicate multi-view information. In this work, we concatenated ResNet framework and BiLSTM framework to extract the powerful features adaptively.

As shown in Table I, experimental results show that the proposed method with deep features achieved better performance than the baseline methods with hand-crafted features. Besides, as shown in Table II, deep features + MIL-SVM (MI-SVM, mi-SVM and SIL-SVM) all achieved better performance compared with hand-crafted features + MIL-SVM.

The proposed MIL training strategy with attention-based MIL pooling improved the sensitivity of 24.57% (L), 15.08% (O) and 0.58% (R) than the proposed method without MIL framework, as seen in Table I. The MIL algorithms basically perform better than supervised learning methods such as SVM, ANN, and ENCASE. The reason may be that for the supervised learning methods, it could be thought to annotate all instances in the positive bag as the positive. However, not all the heartbeats in the record from the patient suffered from the BBB were abnormal in the clinic practice.

TABLE I

COMPARISON OF THE PROPOSED METHOD AND BASELINE METHODS WITH THE SUBJECT-ORIENTED PATTERN

Method	$Acc(\%)$	$Se_O(\%)$	$Se_L(\%)$	$Se_R(\%)$	P
ANN [3]	33.62	8.97	27.31	64.57	<0.001
SVM [4]	36.74	11.78	31.24	67.21	<0.001
ENCESE [5]	63.89	37.38	55.97	98.31	<0.001
ours without MIL	65.17	36.15	60.21	99.14	<0.001
ours with max-pooling-based MIL	74.30	47.31	76.18	99.41	<0.001
ours	78.58	51.23	84.78	99.72	-

TABLE II

COMPARISON OF THE PROPOSED METHOD AND THREE TYPICAL MIL ALGORITHMS

Method	$Acc(\%)$	$Se_O(\%)$	$Se_L(\%)$	$Se_R(\%)$	P
hand-crafted fetures + MI-SVM	53.92	16.54	41.38	56.29	<0.001
hand-crafted fetures + mi-SVM	68.71	24.87	47.54	67.63	<0.001
hand-crafted fetures + SIL-SVM	72.92	29.22	61.34	84.08	<0.001
deep fetures + MI-SVM	53.92	24.85	57.24	79.68	<0.001
deep fetures + mi-SVM	68.71	38.27	72.49	95.36	<0.001
deep fetures + SIL-SVM	72.92	44.13	75.46	99.18	<0.001
ours	78.58	51.23	84.78	99.72	-

The proposed deep MIL method also obtained better performance than the typical MIL algorithms with the same deep features, as shown in Table II. This shows the superiority of our end-to-end trained deep MIL for BBB classification on ECG signal. Our methods can train end-to-end, that is, the parameter of the feature extractor and classifier can updated together. While the MIL-SVM required extracting features first and then train the classifier.

As shown in Table II, we also compared the proposed MIL learning method with the max-pooling MIL based method with the standard MIL assumption. It's clearly that the perfomance of our method is better. Max-pooling MIL based method only considers one heartbeat of the positive bags, leading to unstable responses on heartbeats and not making full use of the others of the positive bags. While our method employs more heartbeat of each ECG record and is more efficient than max-pooling MIL based method.

V. CONCLUSION

In this paper, a novel deep MIL learning framework was developed to classify the BBB without heartbeats annotation on multi-lead ECG signals. We first processed the ECG signals and segmented the heartbeat. Then, we built RBNN to automatically extract heartbeats information and discover the abnormal heartbeats via multi-instance learning. The proposed method was trained on CPSC and tested on AR, with an accuracy of 78.58% ,and sensitivity of 84.78% (L), 51.23% (O) and 99.72% (R) in BBB classification on ECG signal. Experimental results showed that our method obtained a clear improvements compared with state-of-the-art methods. In the future, we will explore some sophisticated algorithms to further improve ECG classification performance and extend the proposed method to classify the more kinds of abnormal rhythm on ECG.

REFERENCES

[1] C. Ye, B. V. K. Vijaya Kumar, and M. T. Coimbra. Heartbeat classification using morphological and dynamic features of ecg signals. *IEEE Transactions on Biomedical Engineering*, 59(10):2930–2941, Oct 2012.

[2] M. Llamedo and J. P. Martinez. Heartbeat classification using feature selection driven by database generalization criteria. *IEEE Transactions on Biomedical Engineering*, 58(3):616–625, March 2011.

[3] R. Allami, A. Stranieri, V. Balasubramanian, and H. F. Jelinek. A genetic algorithm-neural network wrapper approach for bundle branch block detection. In *2016 Computing in Cardiology Conference (CinC)*, pages 461–464, Sep. 2016.

[4] N. Sultana and Y. Kamatham. Msvm-based classifier for cardiac arrhythmia detection. In *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 1314–1318, Sep. 2016.

[5] J. Hu, W. Zhao, D. Jia, C. Yan, H. Wang, Z. Li, and T. You. A novel detection method of bundle branch block from multi-lead ecg. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 79–82, July 2019.

[6] Thomas G. Dietterich, Richard H. Lathrop, and Toms Lozano-Prez. Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence*, 89(1):31 – 71, 1997.

[7] L. Sun, Y. Lu, K. Yang, and S. Li. Ecg analysis using multiple instance learning for myocardial infarction detection. *IEEE Transactions on Biomedical Engineering*, 59(12):3348–3356, Dec 2012.

[8] Mingxia Liu, Jun Zhang, Ehsan Adeli, and Dinggang Shen. Landmark-based deep multi-instance learning for brain disease diagnosis. *Medical Image Analysis*, 43:157 – 168, 2018.

[9] W. Zhao, J. Hu, D. Jia, H. Wang, Z. Li, C. Yan, and T. You. Deep learning based patient-specific classification of arrhythmia on ecg signal. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1500–1503, July 2019.

[10] G. B. Moody and R. G. Mark. The impact of the mit-bih arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3):45–50, May 2001.

[11] Stuart Andrews, Ioannis Tsochantaridis, and Thomas Hofmann. Support vector machines for multiple-instance learning. In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems 15*, pages 577–584. MIT Press, 2003.

[12] Gary Doran and Soumya Ray. A theoretical and empirical analysis of support vector machine methods for multiple-instance classification. *Machine Learning*, 97(1):79–102, Oct 2014.