This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2910082, IEEE Journal of Biomedical and Health Informatics

MFB-CBRNN: a hybrid network for MI detection using 12-lead ECGs

Wenhan Liu, Fei Wang, Qijun Huang*, Sheng Chang, *Member, IEEE*, Hao Wang, *Member*, *IEEE* and Jin He, *Senior Member*, *IEEE*

***Abstract*—This paper proposes a novel hybrid network named MFB-CBRNN for Myocardial Infarction (MI) detection using 12-lead ECGs. The model efficiently combines CNN-based and RNN-based structures. Each feature branch consists of several 1-D convolutional and pooling layers, corresponding to a certain lead. All the feature branches are independent from each other, which are utilized to learn the diverse features from different leads. Moreover, a Bidirectional Long Short Term Memory (BLSTM) network is employed to summarize all the feature branches. Its good ability of feature aggregation has been proved by the experiments. Furthermore, the paper develops a novel optimization method, Lead Random Mask (LRM), to alleviate overfitting and implement an implicit ensemble like dropout. The model with LRM can achieve a more accurate MI detection. Class-based and subject-based 5-fold cross validations are both carried out using PTB diagnostic database. Totally, there are 148 MI and 52 Healthy Control (HC) subjects involved in the experiments. The MFB-CBRNN achieves an overall accuracy of 99.90% in class-based experiments, and an overall accuracy of 93.08% in subject-based experiments. Compared with other related studies, our algorithm achieves a comparable or even better result on MI detection. Therefore, the MFB-CBRNN has a good generalization capacity and is suitable for MI detection using 12-lead ECGs. It has a potential to assist the real-world MI diagnostics and reduce the burden of cardiologists.**

***Index Terms* — Convolutional Neural Network(CNN), Bidirectional Long Short Term Memory(BLSTM), electrocardiogram (ECG), Myocardial Infarction (MI), automated diagnosis, Lead Random Mask (LRM).**

I. Introduction

Myocardial Infarction (MI) is a typical kind of cardiovascular diseases (CVD) that is caused by blockages of coronary arteries. These blockages lead to a severe reduction of blood flow to the myocardium. As the oxygen-rich blood supply is insufficient, cardiac muscle damage occurs and may result in heart attacks. MI is a leading cause of death for human beings. According to the World Health Organization (WHO), more than 25% of all deaths around the world are due to MI-related diseases [1]. Therefore, early and accurate diagnosis of MI is crucial to increase the life expectancy and improve the life quality of the patients.

The electrocardiogram (ECG) is an efficient tool for MI detection. It is noninvasive and economical, commonly used by many medical institutes. Also, it can continuously monitor the heart status for a long time. In terms of the different positions of acquiring electrodes, the ECG is usually collected along 12 leads, including I, II, III, aVR, aVL, aVF, and V1~V6 [2]. A typical beat of ECG usually consists of P, QRS, and T waves. MI can be diagnosed by some significant morphological changes of ECG, such as ST-segment elevations and appearances of pathological Q-waves [3]. However, MI manual detection based on ECG is a tedious and time-consuming task for cardiologists, not possible in the regions lacking of medical support. Thus, research on computer-aided diagnosis systems has gained much concentration worldwide, and one of the key techniques of these systems is the design of intelligent algorithms.

To accomplish an automatic detection of MIs, many researchers adopted conventional machine learning (ML) algorithms including feature extraction, selection, and final classification. According to medical experience, significant time-domain features for MI detection are always extracted, such as QRS amplitude, QRS width, ST-segment level, and the appearance of inverted T-waves [4]-[7]. Also, mathematical transforms are applied to extract useful information that may indicates MIs. For instance, Wavelet Transform (WT) [8]-[10] and Discrete Cosine Transform (DCT) [11] [12] are often employed to decompose ECGs. Their coefficients are regarded as efficient features. After elaborative feature selection, conventional ML classifiers are used to discriminate MIs from healthy cases. Support Vector Machine (SVM) [9]-[11] [13], K-Nearest Neighbor (KNN) [4] [6] [9] [11] [13], and Multilayer Perceptron (MLP) [4] [5] [10] [11] are usually utilized because of their good performances in classification tasks. But the limitations of ML-based methods are also obvious. The aforementioned feature extraction and selection are highly dependent on human expertise, and useful features are often hard to obtain. In addition, some time-domain feature extraction (e.g. features related to ST-T changes) depends on the accuracy of fiducial points detection, but only QRS detection algorithms are stable enough to be used in real-world

This work was supported by the National Natural Science Foundation of China (61574102, 61774113 and 61331007), the Fundamental Research Fund for the Central Universities, Wuhan University (2042017gf0052 and 2042018gf0045) and the Natural Science Foundation of Hubei Province, China (2017CFB660).

The authors are with the School of Physics and Technology, Wuhan University, Wuhan 430072, China (*corresponding author: Qijun Huang*, e-mail: huangqj@whu.edu.cn).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2910082, IEEE Journal of Biomedical and Health Informatics

applications [14].

To overcome the limitations of the conventional ML-based schemes, Deep Learning (DL) algorithms are employed by other researchers to detect MIs based on ECG, because the DL models can learn useful features automatically from raw data [15]. The preferred model is Convolutional Neural Network (CNN). CNNs have made great achievements in bioinformatics because of its superior ability of feature learning [16]. In particular, there were some studies proposed CNN-based algorithms to diagnose MIs recently. Acharya et al. [17] designed an 11-layer 1-dimension (1-D) CNN for MI detection using single-lead ECG. They achieved an accuracy of 95.22% in their experiments, but the 1-D CNN can't utilize the information from more leads; Reasat et al. [18] proposed a shallow CNN to detect inferior MIs. They used 3-lead ECGs, including lead II, III, and aVF. Also, the effective Inception structure [19] was employed in their model. Accuracy of their model was 84.54%; Liu et al. [20] developed an ML-CNN using 4-lead ECGs (V2, V3, V5, aVL) for Generalized Anterior MI (GAMI) detection. Novel techniques like sub 2-D convolution and Lead Asymmetric Pooling (LAP) were adopted to optimize the conventional CNNs. Although these two multilead models can utilize more leads, they only concentrated on a certain type of MI. In our recent work [21], a Multiple-Feature-Branch CNN (MFB-CNN) using 12-lead ECG was proposed. Based on the comprehensive information from all the 12 leads, the MFB-CNN can be applied to classify multiclass MIs, including Anterior MI (AMI), Anterior-Septal MI (ASMI), Anterior-Lateral MI (ALMI), Inferior MI (IMI), and Inferior-Lateral MI (ILMI).

Besides, another preferred deep learning model named Recurrent Neural Network (RNN) is also employed to ECG-based detection. RNN is especially designed for sequential data processing [22]. It has been successfully applied to natural language processing, speech recognition, machine translation, and so on [39]-[41]. However, due to the vanishing and exploding gradients problems, standard RNNs are seldom used in real-world applications. In contrast, a variant of RNN, Long Short Term Memory (LSTM) has been widely adopted to overcome the limitations of standard RNNs [23]. As for the LSTM's applications in ECG processing, Zhang et al. [24] used a patient-specific LSTM to learn time correlation among ECG signal points. They achieved a state-of-the-art result in arrhythmia classification. Similarly, Chauhan et al. [25] implemented a deeper LSTM for accurate anomaly detection in ECG signals. To enhance the performance of LSTM, Yildirim [26] employed Bidirectional LSTM (BLSTM) [27] in ECG classification. According to their experiments, BLSTMs provided better performance than standard LSTMs. Furthermore, LSTMs can be combined with CNNs to generate hybrid models, which may lead to a significant performance improvement. For instance, a hybrid model named Convolutional Recurrent Neural Network (CRNN) has been used in ECG classification. It can be employed to detect Coronary Artery Disease (CAD) [28], Arial Fibrillation (AF) [29], and arrhythmia [30]. Impressive results were obtained in these studies. Considering the significance of MI diagnosis, application and optimization of hybrid models for MI detection deserve further explorations in future.

Thus, in this work, a novel hybrid model for MI detection using 12-lead ECG is proposed. It combines CNN and BLSTM into a single model. Since this model derives from our previous study about MFB-CNN [21], it is named Multiple-Feature-Branch Convolutional Bidirectional Recurrent Neural Network (MFB-CBRNN). Compared with conventional ML-based schemes, our MFB-CBRNN requires no explicit feature extraction and selection as it is a DL-based model. It automatically learns how to diagnose MI by analyzing a large number of heart beats. Moreover, the structure of MFB-CBRNN is specially designed for 12-lead ECG processing, which is more suitable for an accurate MI detection than the previous DL models. In detail, the contributions of this work are listed as follows:

1) A novel combination of CNN and BLSTM is designed to exploit the information from all the 12 leads. Each lead is assigned to an independent feature branch consisting of convolutional and pooling layers [21], so diverse features from different leads can be learned; the 12 leads constitute an interdependent entity, a BLSTM is employed to summarize all the 12 feature branches as a whole since it is proven to be efficient in logical dependencies [29]. A fully-connected layer is used to make a prediction. Therefore, compared with the existing DL models, our model is more suitable for MI detection using 12-lead ECG.

2) An original optimization method, Lead Random Mask (LRM), is proposed to enhance the performance of the MFB-CBRNN. It is a dropout-like method. In detail, a certain number of feature branches are randomly "dropped" in an iteration of training phase, whereas all the feature branches are used in testing phase. According to our experiments, models with LRM can achieve a more accurate detection of MIs.

3) Class-based and subject-based experiments are both performed in this paper. In particular, the data are divided into training and testing set in terms of the subjects under subject-based evaluation. That is, there are no patient overlaps between the training and testing sets. Thus, it is a more practical approach for performance evaluations. Also, 5-fold cross validation is adopted to obtain more reliable results. Compared with other studies, the MFB-CBRNN shows a state-of-the-art performance on MI detection.

The paper is organized as follows: Section II introduces the database used in this paper and preprocessing of the ECG signals; the proposed algorithm is illustrated in Section III; Section IV shows the results of class-based and subject-based experiments, and a detailed discussion is presented in Section V. Section VI concludes the paper.

II. Data Used

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2910082, IEEE Journal of Biomedical and Health Informatics

TABLE I
STATISTICS OF THE ECG SIGNALS USED

Class	No. of subjects	No. of records	No. of 12-lead heartbeats
MI	148	368	53712
HC	52	80	10638
Total	200	448	64350

In this work, the ECG data for MI detection are obtained from Physikalisch-Technische Bundesanstalt (PTB) diagnostic ECG database [31]. All the signals are sampled at 1000 Hz with a 16-bit resolution over a range of ± 16.384 mV. To alleviate the computation burden, all the signals from the database are downsampled to 250 Hz in this work, which is enough to reveal the important information of ECGs [42]. The signals last at least 30 seconds, whereas most of the signals are typically of ~2 minute duration. Each ECG record in this database provides all the 12 leads with a diagnostic result given by the cardiologists. In total, there are 549 12-lead records from 290 subjects in the PTB database. In particular, 368 records from 148 MI and 52 Healthy Control (HC) subjects are selected for the evaluation of our algorithm. Obviously, a subject in the database has one or more ECG records, including a certain number of beats. The statistics of the used ECGs are shown in Table I.

As many previous studies, our algorithm is also developed and evaluated on the basis of beats. Therefore, the Pan-Tompkins algorithm [32] is used to detect the QRS waves in the ECGs at first. Using efficient filters and adaptive thresholds, the Pan-Tompkins algorithm is robust to noise and beat deformations, like premature ventricular contractions (PVCs). Then the beats are segmented according to the QRS positions, including 49 samples before the QRS peaks and 100 samples after the QRS peaks. In other words, each beat contains 150 samples, lasting for 0.6 seconds at a sampling rate of 250 Hz. Furthermore, a 12-lead beat consists of the single-lead beats that simultaneously appeared in all the 12 leads. Z-score is used to normalize each beat, which can be calculated as:

$$Z = \frac{X - \mu}{\delta} \tag{1}$$

where X denotes the beat, μ and δ are the mean value and standard deviation of the beat, respectively. Z-score can alleviate the effects of baseline offset and amplitude scaling [17], which can benefit the model training.

III. METHOD

A. Feature branches for a lead

The proposed MFB-CBRNN is developed and improved based on our previous study about MFB-CNN [21]. The MFB-CNN is a modified version of conventional CNNs, which is designed for MI detection using 12-lead ECG. Exploiting the integrity and diversity of 12-lead ECGs (discussed in [21]), it achieves good performances in class-based and patient-specific experiments. Considering its excellent feature learning ability, feature branches of MFB-CNN are reused here to process the 12-lead ECGs at first. Each feature branch corresponds to a certain lead, consisting of 1-D convolutional and pooling layers:

1) 1-D convolutional layers: 1-D convolutions between kernels and the input signals are performed in convolutional layers. The kernels can be treated as a group of 1-D filters, which are essentially trainable weights in neural networks. The number of filters is a hyperparameter that needs to be determined manually, and the actual value of the filters are obtained by training on a huge number of ECG data. Signals filtered by different kernels represent various characteristics of the raw input. Thus, useful features can be learned and extracted by the convolutions. In addition, bias and activation function are employed to introduce shifting and nonlinearity, enhancing the capacity of the convolutional layers. The operations of a 1-D convolutional layer can be formulated as:

$$\boldsymbol{h}_j^l = f\left(b_j^l + \sum_{i \in M_j} \boldsymbol{h}_i^{l-1} * \boldsymbol{w}_{ij}^l\right) \tag{2}$$

where $\boldsymbol{h}_j^l$ and $\boldsymbol{h}_i^{l-1}$ are the jth output and the ith 1-D input feature map of the layer, and $\boldsymbol{w}_{ij}^l$ is the kernel that connecting them; b_j^l is the corresponding bias; * refers to the 1-D convolution; M_j denotes the connectivity between the input and output in this layer; Rectified Linear Unit (ReLU) activation function is used here, which is shown as $f(\bullet)$.

2) 1-D pooling layers: Pooling layers can be regarded as specific convolutional layers. It is always adopted to reduce the dimension of feature maps. Also, it can enhance the translational invariance of the data internal representation. Thus, pooling layers are always employed to alleviate computational burden and overfitting. Max and mean pooling are the commonly used pooling methods [15]. Mean pooling layers are deployed in a feature branch as [21]. A 1-D mean pooling operates as:

$$\boldsymbol{h}_j^l(n) = \underset{r \in R}{\text{mean}}[\boldsymbol{h}_j^{l-1}(n \times S + r)] \tag{3}$$

where mean $(\bullet)$ is the function that computes the mean values of the pooling windows; n denotes the index of sampling point of the feature maps; R and S denote the pooling region and stride respectively, and r is the index in the pooling region R. Here the size of a pooling region is equal to the pooling stride in all pooling layers, which means that the pooling regions are not overlapped with each other.

Several convolutional and pooling layers are stacked to generate a feature branch. Each feature branch is deployed for a certain lead. In other words, 12 feature branches should be set up for MI detection using 12-lead ECGs. To strengthen the performance of the feature branches, Batch Normalization (BN) layers are deployed after the 1-D convolutions. BN can reduce the internal covariate shift of the data, speeding up the convergence in a training procedure [33]. Also, the model with BN can produce better results than the one without BN, which

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2910082, IEEE Journal of Biomedical and Health Informatics

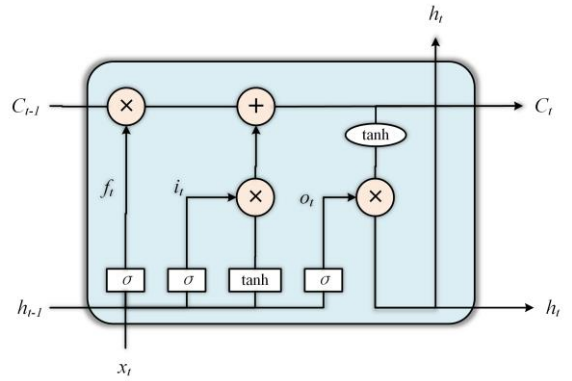

Fig. 1. A typical LSTM block. The operation is illustrated as equation (4)-(8).

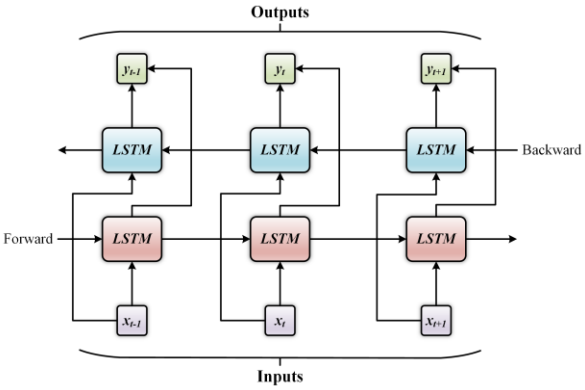

Fig. 2. A typical structure of BLSTM.

is illustrated in Section V.

B. BLSTM for leads summary

LSTM is a typical variant of standard RNN. It has been widely used since it reduces the negative effects of gradient vanishing and exploding, which are the main limitations of standard RNNs. With excellent capacity of data abstraction, LSTMs have been widely applied in sequential data processing [29] [39]-[41]. Moreover, according to the MI diagnostic rules in [3] and leads interpretation in [43], the relationship between leads is not a simply superposition but an abstract integration, so a BLSTM is employed to manage this relationship. The basic unit of a LSTM, namely LSTM memory block, contains three gate units: forget, input, and output (gate). These gates are the key approaches to control the data flow through the LSTM cell, making LSTM network a better choice for sequential learning than the standard RNN. The structure of a LSTM cell is shown as Fig. 1.

1) Forget gate: It is used to determine how much information from the previous memory block can be retained or forgotten. The operation in a forget gate can be formulated as:

$$f_t = \sigma(w_f[x_t, h_{t-1}] + b_f) \tag{4}$$

where w_f and b_f are the weight and bias of the forget gate, respectively; x_t is the block input; h_{t-1} denotes the output of the previous step. The sigmoid function $\sigma(\bullet)$ is utilized as activation function, calculating the final output of the forget gate. The output value, ranging from 0 to 1, represents how much previous information to be retained.

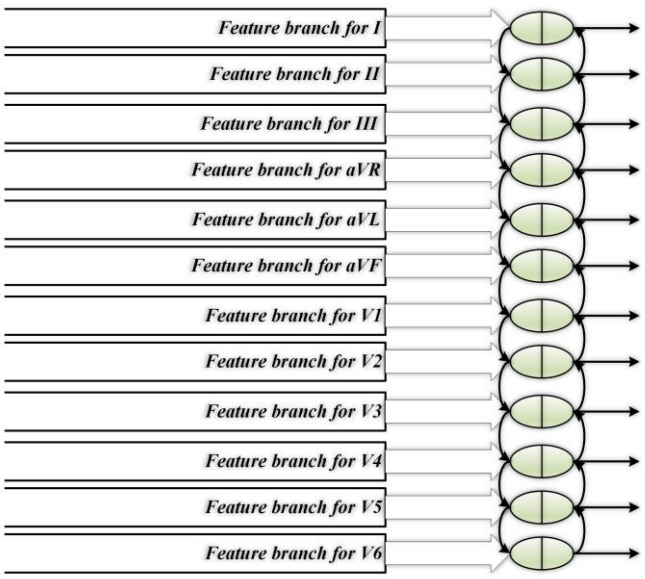

Fig. 3. Connection between feature branches and BLSTM.

2) Input gate: It learns to store useful information in the block. Also, the new status of the memory block is obtained using the input and forget gate:

$$i_t = \sigma(w_i[x_t, h_{t-1}] + b_i) \tag{5}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tanh(w_c[x_t, h_{t-1}] + b_c) \tag{6}$$

where i_t is the output of input gate, (w_i, w_c) and (b_i , b_c) are the corresponding weights and biases, respectively. The dot in (6) denotes element-wise multiplication. Here tanh activation function is used for calculating the current status C_t, which is controlled by the outputs of forget (f_t) and input gate (i_t).

3) Output gate: The final output of the LSTM block is generated by the output gate. The operations can be shown as (7) and (8):

$$o_t = \sigma(w_o[x_t, h_{t-1}] + b_o) \tag{7}$$

$$h_t = \tanh(C_t) \cdot o_t \tag{8}$$

where o_t and h_t are the outputs of the output gate and the whole LSTM block, respectively. The form and parameters of (7) are similar to those of (5). To obtain the final output, another tanh function is employed to calculate the activation of current status C_t, multiplying by o_t.

Considering all the 12 leads simultaneously acquire signals from the same heart, they may have implicit logical correlations, even though signals from different leads may manifest as very different morphologies. As the aforementioned analysis, it is an abstract integration, since all the 12 leads jointly determine the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2910082, IEEE Journal of Biomedical and Health Informatics

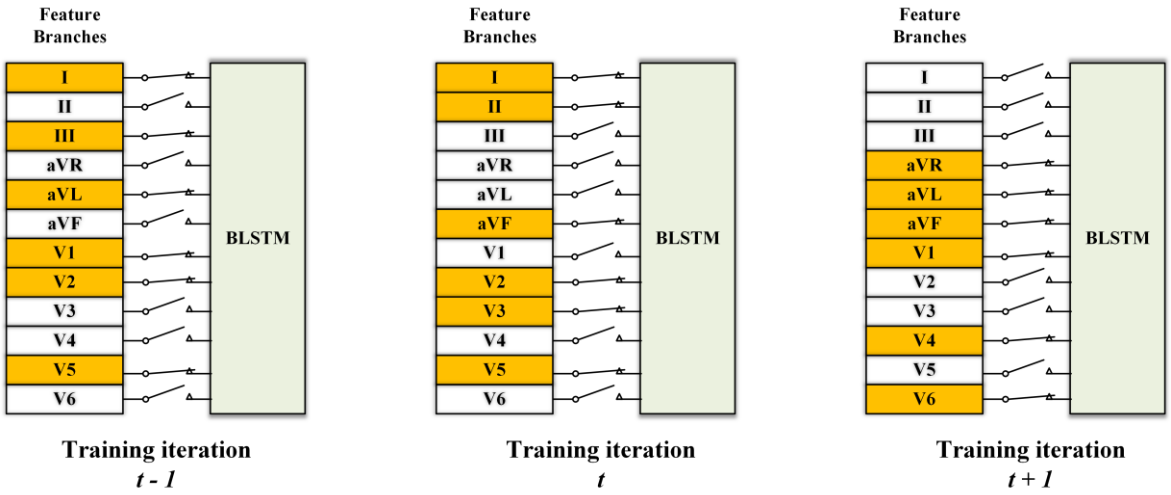

Fig. 4. An example of LRM with an inclusion probability of 0.5. For each iteration, 6 (12 ×0.5) feature branches are randomly removed.

final result. Here BLSTM is adopted to further summarize and explore the features from the 12 leads. The typical structure of a BLSTM is depicted as Fig. 2.

Each step of the BLSTM corresponds to a certain lead's feature branch, as shown in Fig. 3. Compared with the standard unidirectional LSTM, BLSTM can simultaneously process all available information from "forward" and "backward" sides [27]. Thus, more comprehensive information from 12 leads can be exploited. Finally, a fully-connected layer is proposed to utilize all the steps of the BLSTM, predicting the type of the input signal (MI or HC).

C. LRM for optimization

Inter-patient variation is one of the main challenges for computer-aided diagnostic algorithms, including MI detection. For the unique features of each one's heart and ECG, it is difficult for the algorithm trained on a certain group of patients to generalize well on other patients. From a perspective of ML theory, it can be regarded as a phenomenon of overfitting, because the models often focus on the particular features of the training patients and fail to get more general rules. To overcome the influence of inter-patient variation, Lead Random Mask (LRM) is proposed to improve the generalization of our model. It is applied to the connection between the 12 feature branches and the BLSTM. The operation and the significance behind the LRM are described as follows.

1) Operation: The LRM is performed in the training procedure. Like the dropout [34], a random-selected sub-set of all the branches are involved at each training iteration, according to a hyperparameter named inclusion probability ranging from 0 to 1. For instance, when the inclusion probability is 0.5, 6 feature branches, corresponding to 6 leads, are randomly selected and retained. In the testing phase, all the feature branches are used to predict the final results. Fig. 4 shows an example of LRM with the inclusion probability of 0.5.

2) Significance: The reasons why LRM can enhance the classification performance, or the significance behind the LRM, may lie in the following two aspects:

a) *Regularization*: The random removal of feature branches in LRM reduces the effective capacity of the model. Thus, LRM can be treated as a regularization method. It is commonly known that appropriate regularizations can alleviate the effect of overfitting [22]. The regularization from LRM may lead to a more general model.

b) *Implicit ensemble*: Due to the random choice of the feature branches, LRM can generate many sub-structures of the raw network to train. In the testing phase, the whole network is used to predict, which can be regarded as an implicit ensemble of the aforementioned sub-networks. Inspired by researches on dropout [34], the implicit ensemble of sub-structures (with weights sharing) provides additional improvements for model generalization.

In summary, LRM is an effective optimization method that is specially designed for models using multiple leads. Notably, the aim of LRM is not to choose the "best" leads for detection, but regularize the network. Its performances and rules are discussed in the following sections about experiments and analysis.

IV. Experiments and Results

A. Architecture and configuration

The proposed architecture is described as Table II and Fig. 5. The structure of each feature branch is the same as those of the model in [21], including 3 convolutional layers and 3 mean pooling layers. However, the hyperparameters in a feature branch, such as kernel size and pooling stride, are reconsidered and optimized, since the data used has been changed. Furthermore, a global average pooling layer is appended to each feature branch, reducing the overfitting [35]. All the feature branches are summarized by a BLSTM with only 1 hidden layer, and then a fully-connected layer predicts the final result. According to our statistics, there are 2778 parameters in this model, including 2442 trainable and 336 un-trainable parameters, whereas there are 64350 samples in our dataset. This makes the model learn the general MI diagnostic rules rather than memorize the dataset, since the number of parameters is far less than the number of samples [22]. To avoid

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2910082, IEEE Journal of Biomedical and Health Informatics

TABLE II
DETAILED STRUCTURAL INFORMATION OF MODEL

	Layer	Kernel size	Pooling window size	Output map size	Number of output maps
Each feature branch	**I-0**	/	/	150	1
	C-1 + BN	7	/	150	2
	P-1	/	5	30	2
	C-2 + BN	5	/	30	4
	P-2	/	5	6	4
	C-3 + BN	3	/	6	8
	P-3	/	3	2	8
	GAP	/	/	1	8
	Input size		No. of steps		No. of hidden layers
BLSTM	8		12		1
Fully-Connected layer	Input size			Output size	
	24			2	

I: input layer **C**: 1-D convolutional layer **BN**: Batch Normalization **P**: 1-D mean-pooling layer
GAP: Global Average Pooling layer /: Not Available.
The kernel size and output map size refer to the number of sampling points included in the convolution kernel and the feature map, respectively.

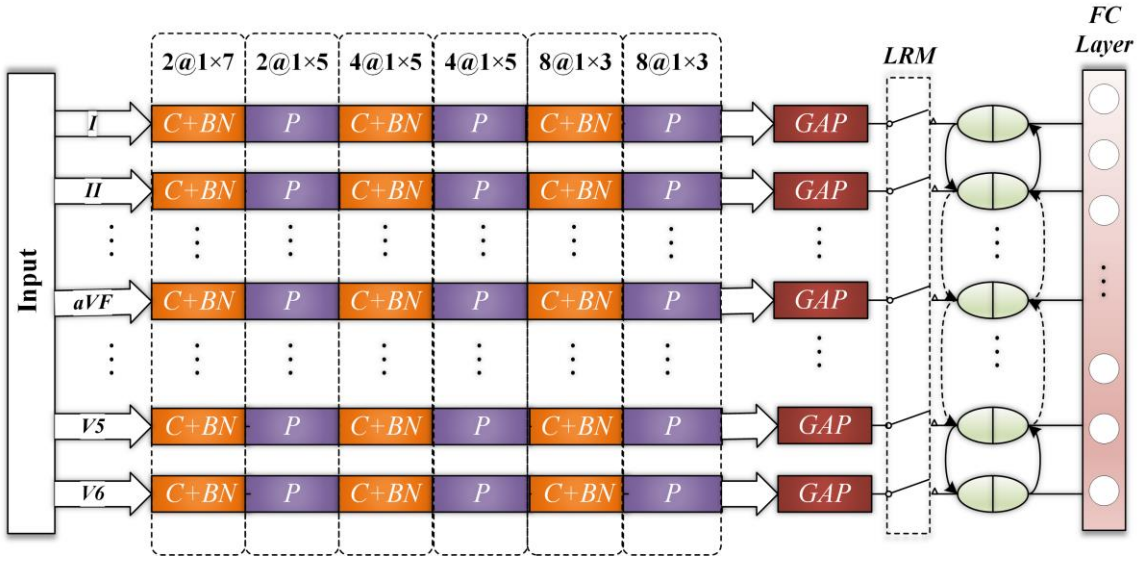

Fig. 5. An overview of the whole MFB-CBRNN, including 12 feature branches, a BLSTM and a fully-connected layer. **C**: 1-D convolutional layer **BN**: Batch Normalization **P**: 1-D mean-pooling layer **GAP**: Global Average Pooling layer m@1×n means the layer has m output maps, and 1×n kernels for convolutional layers or 1×n stride for pooling layers.

overfitting for a specific dataset, 5-fold cross validation is used to maintain a balance "bias" and "variance" [44].

The whole network is developed and trained using TensorFlow (Python version) [36], a prevalent DL framework. The loss is calculated by the categorical cross-entropy function. The Adam algorithm is employed to train the model, with an initial learning rate of 0.001. Batch size and the number of training iterations are set to 32 and 5000, respectively. To alleviate the impact of data imbalanced distribution, the HC beats in the training set are simply oversampled to match the number of MI beats. All the following experiments were performed on a computer with Intel Core i7-3770@3.40 GHz CPU and 12 GB RAM.

B. Performance evaluation

Here several commonly used metrics are used to evaluate the model performances, including accuracy (*Acc*), sensitivity (*Se*), specificity (*Sp*) and positive predicivity (*Pp*). *Acc*, *Se*, *Sp*, and *Pp* are defined as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

$$Se = \frac{TP}{TP + FN} \tag{10}$$

$$Sp = \frac{TN}{TN + FP} \tag{11}$$

$$Pp = \frac{TP}{TP + FP} \tag{12}$$

where *TP* and *FN* denote the number of MI beats predicted as MI and HC respectively, *TN* and *FP* represent the number of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2910082, IEEE Journal of Biomedical and Health Informatics

TABLE III
5-FOLD CROSS-VALIDATION RESULTS OF MI DETECTION OF CLASS-BASED EXPERIMENTS

No.of retained leads	Inclusion Probability (LRM)	Se (%)	Sp (%)	Pp (%)	Acc (%)
1	0.08	90.54	59.46	91.85	85.40
2	0.17	97.16	93.18	98.63	96.30
3	0.25	98.24	76.76	95.52	94.69
4	0.33	99.20	98.23	99.65	99.04
5	0.42	98.76	99.14	99.83	98.82
6	0.50	98.10	99.44	99.89	98.32
7	0.58	99.35	99.76	99.95	99.41
8	0.67	99.81	99.54	99.91	99.76
9	0.75	99.77	99.76	99.95	99.77
10	**0.83**	**99.97**	**99.54**	**99.91**	**99.90**
11	0.92	98.94	99.94	99.99	99.10
12	1.00	99.77	99.68	99.94	99.76

TABLE IV
CONFUSION MATRIX OF CLASS-BASED MI DETECTION ACROSS 5 FOLDS (INCLUSION PROBABILITY = 0.83)

Actual/Predicted	MI	HC
MI	53695	49
HC	17	10289

HC beats predicted as HC and MI respectively.

C. Class-based experiments

For class-based experiments, beats are randomly divided into training and testing sets regardless of the inter-patient variations. In other words, beats from a certain patient may exist in training and testing sets at the same time. To obtain more credible results, 5-fold cross validation is used to evaluate the model. Moreover, the MFB-CBRNN is tested on different values of inclusion probability in LRM to maintain certain number of leads, from 1 to 12. The *Acc*, *Se*, *Sp*, and *Pp* are summarized as Table III and Fig. 6.

From Table III, it can be observed that the proposed model performs best when 10 leads are randomly retained in the model. The MFB-CBRNN achieves averaged *Acc* of 99.90%, *Se* of 99.97%, *Sp* of 99.54%, *Pp* of 99.91%, respectively. Moreover, the confusion matrix of is shown in Table. IV. There are only 0.03% of the MI instances wrongly categorized as HC, while 0.46% of the HC instances are misclassified into MI class. It is known that different types of MI may affect ECG waveforms from different leads. In terms of MI diagnostic rules in [2] and [45], the relationship between MI location and ECG leads can be summarized as Table. V. The listed locations relate to all the MI types in the database. Regardless of inter-patient variations (class-based experiments), information access is the main factor that determines the model performance. When LRM randomly removes 2 leads, it is hard to completely screen information from a certain location. At the same time, the regularization of LRM still has an effect on the model (keep_prob < 1) to enhance the generalization, though it may reduce the model capacity. Therefore, according to these analyses and our experiments, it is reasonable to infer that keeping 10 leads in the network is a critical point that maintain

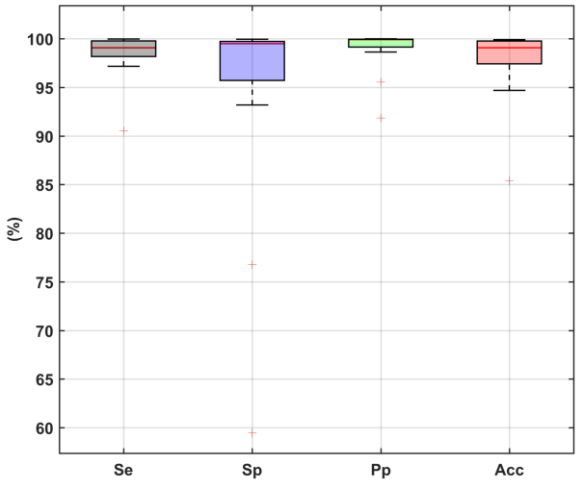

Fig. 6. The box plot of the class-based MI detection.

TABLE V
RELATIONSHIP BETWEEN HEART LOCATION AND LEADS

location	leads
Anterior/Antero-Septal	V1,V2,V3,V4
Lateral	V5,V6, I, aVL
Inferior	II, III, aVF
Posterior	V1,V2,V3

a balance between information access and regularization. Overall, impressive results are obtained in our class-based experiments, manifesting as high accuracies, sensitivities, specificities and positive predicivity. Although the class-based experiment may not reveal the actual generalization of the model due to the irrespective of the patients, it proves that the MFB-CBRNN has sufficient fitting capacity for MI detection using 12-lead ECG. Furthermore, subject-based evaluation that considers the inter-patient variations is performed in the following section.

D. Subject-based experiments

Compared with the aforementioned class-based experiments, subject-based evaluation is a more practical approach to demonstrate the generalization of the model. It takes the inter-patient variations into account. For a subject-based experiment, the model are trained on the data from certain patients, and tested on the data from new patients. There are no patient overlaps between training and testing set. Therefore, subject-based experiments are more similar to the real-world diagnostic applications. Also, subject-based 5-fold cross validation is adopted to assess the model performance. The accuracies, sensitivities, specificities and positive predicivity are collected and shown in Fig.7 and Table. VI, corresponding to different values of inclusion probability in LRM.

As summarized in Table. VI and Fig.7, the best inclusion probability of LRM is 0.5 for the subject-based experiments. In other words, the MFB-CBRNN demonstrates the highest averaged accuracy in the 5-fold cross validation with 6 leads randomly retained in the model. The corresponding averaged *Acc*, *Se*, *Sp* and *Pp* are 93.08%, 94.42%, 86.29%, and 97.21%,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2910082, IEEE Journal of Biomedical and Health Informatics

TABLE VI
5-FOLD CROSS-VALIDATION RESULTS OF MI DETECTION OF SUBJECT-BASED EXPERIMENTS

No.of retained leads	Inclusion Probability (LRM)	Se (%)	Sp (%)	Pp (%)	Acc (%)
1	0.08	87.58	77.66	95.19	85.94
2	0.17	91.64	80.64	95.98	89.82
3	0.25	94.03	75.61	95.11	90.98
4	0.33	93.57	81.71	96.27	91.61
5	0.42	93.30	87.11	97.34	92.27
6	**0.50**	**94.42**	**86.29**	**97.21**	**93.08**
7	0.58	92.62	85.30	96.95	91.41
8	0.67	95.77	76.87	95.43	92.65
9	0.75	94.15	80.16	95.99	91.84
10	0.83	92.71	83.70	96.63	91.22
11	0.92	94.18	76.55	95.30	91.26
12	1.00	93.37	80.94	96.11	91.32

TABLE VII
CONFUSION MATRIX OF SUBJECT-BASED MI DETECTION ACROSS 5 FOLDS (INCLUSION PROBABILITY = 0.5)

Actual/Predicted	MI	HC
MI	50716	2996
HC	1458	9180

TABLE VIII
IMPROVEMENTS STEMMING FROM EACH OPTIMIZATION STEP

Method	Se (%)	Sp (%)	Pp (%)	Acc (%)
1	92.99	55.29	91.31	86.76
1+2	93.65	61.49	92.47	88.33
1+2+3	93.37	80.94	96.11	91.32
1+2+3+4	**94.42**	**86.29**	**97.21**	**93.08**

*1: MFB-CNN. 2: BN. 3. BLSTM. 4. LRM
"1+2+3+4" denotes the proposed MFB-CBRNN.

respectively. According to the confusion matrix shown in Table. VII, 13.71% of HC beats are wrongly classified as MI, whereas the percentage of the misclassified MI beats is only 5.58%. Also, the *Sp* is always lower than *Se* regardless of the number of retained leads, according to Fig.7. This may due to the intrinsic imbalanced distribution that the database has more MI subjects than the HC ones, although oversampling is employed to augment the HC instances. In general, the model achieves an acceptable performance in the subject-based MI detection, which indicates that it has a good generalization on the unseen patients. In the next section, more elaborate discussion based on subject-based evaluation is explained.

V. DISCUSSION

In this section, detailed discussion and exploration on the advantages of the MFB-CBRNN are illustrated. First, the comparison between the raw MFB-CNN in [21] and MFB-CBRNN is illustrated. According to the experiments in Section 4 and [21], both of the model can achieve an excellent performance on the class-based MI detection: the MFB-CNN get an overall accuracy of 99.95% in class-based experiments, which is roughly consistent with the performance of the

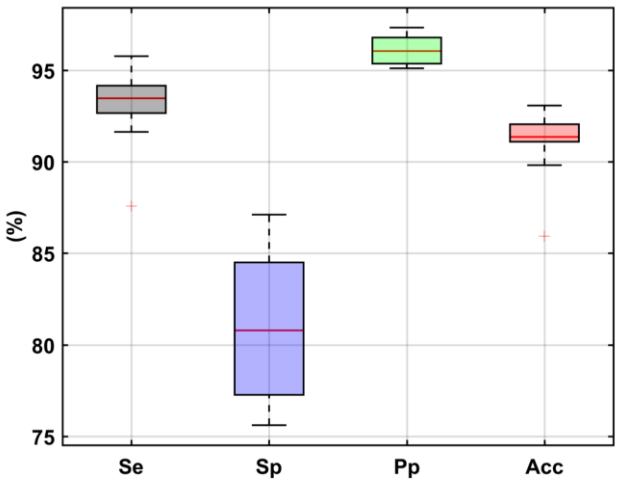

Fig. 7. The box plot of the class-based MI detection.

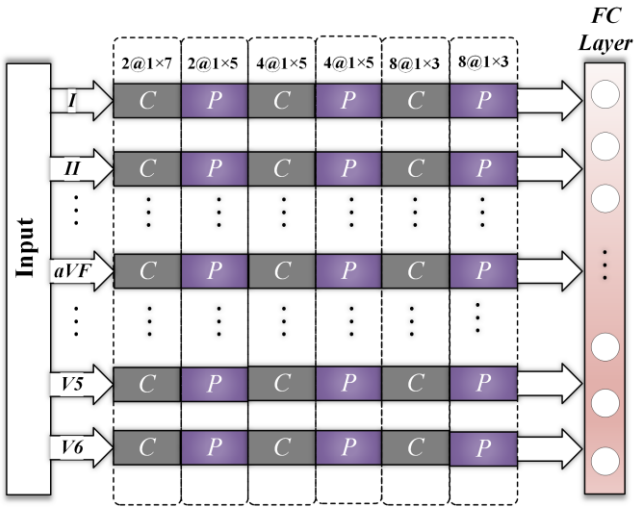

Fig. 8. The architecture of the MFB-CNN [21] used for comparison.
C: 1-D convolutional layer **P**: 1-D mean-pooling layer.
m@1×n means the layer has m output maps, and 1×n kernels for convolutional layers or 1×n stride for pooling layers.

MFB-CBRNN (99.90%). Thus, only subject-based evaluation is proposed to compare the generalization of these two algorithms. Notably, hyperparameters in MFB-CNN are re-optimized to adapt to the samples in this study. An overview of MFB-CNN used in this paper is shown as Fig.8. The same subject-based experiments as in Section IV are performed to compare these models. Furthermore, the improvements stemming from each optimization step, including the combination of MFB-CNN and BLSTM, the utilization of BN and LRM, are summarized in Table. VIII.

From Table. VIII, it can be noted that the MFB-CBRNN has the best generalization capacity due to several particular optimization approaches. BN is a commonly used method to enhance the performance of CNN-based algorithms [33], it is feasible in our model as well. The combination of MFB-CNN and BLSTM provides the most obvious improvement of all the optimization methods. Hence, the design of independent feature branches and BLSTM for leads summary may fit in the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2910082, IEEE Journal of Biomedical and Health Informatics

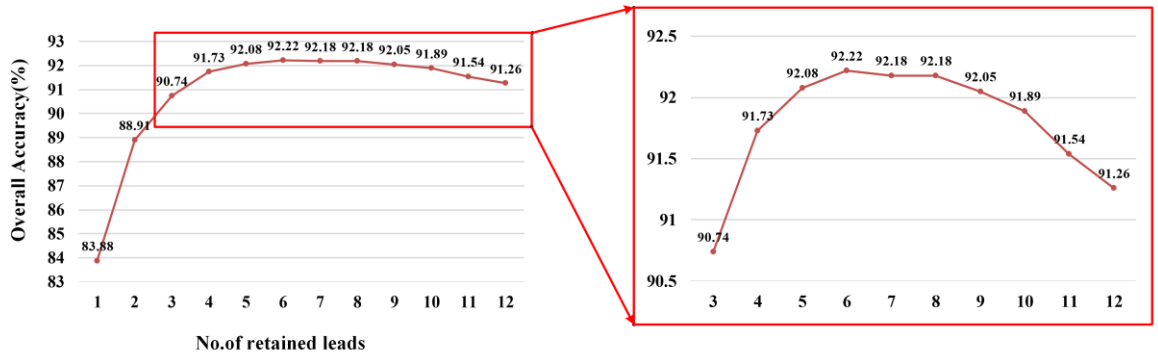

Fig. 9. The variation of overall accuracy along with the inclusion probability of LRM.

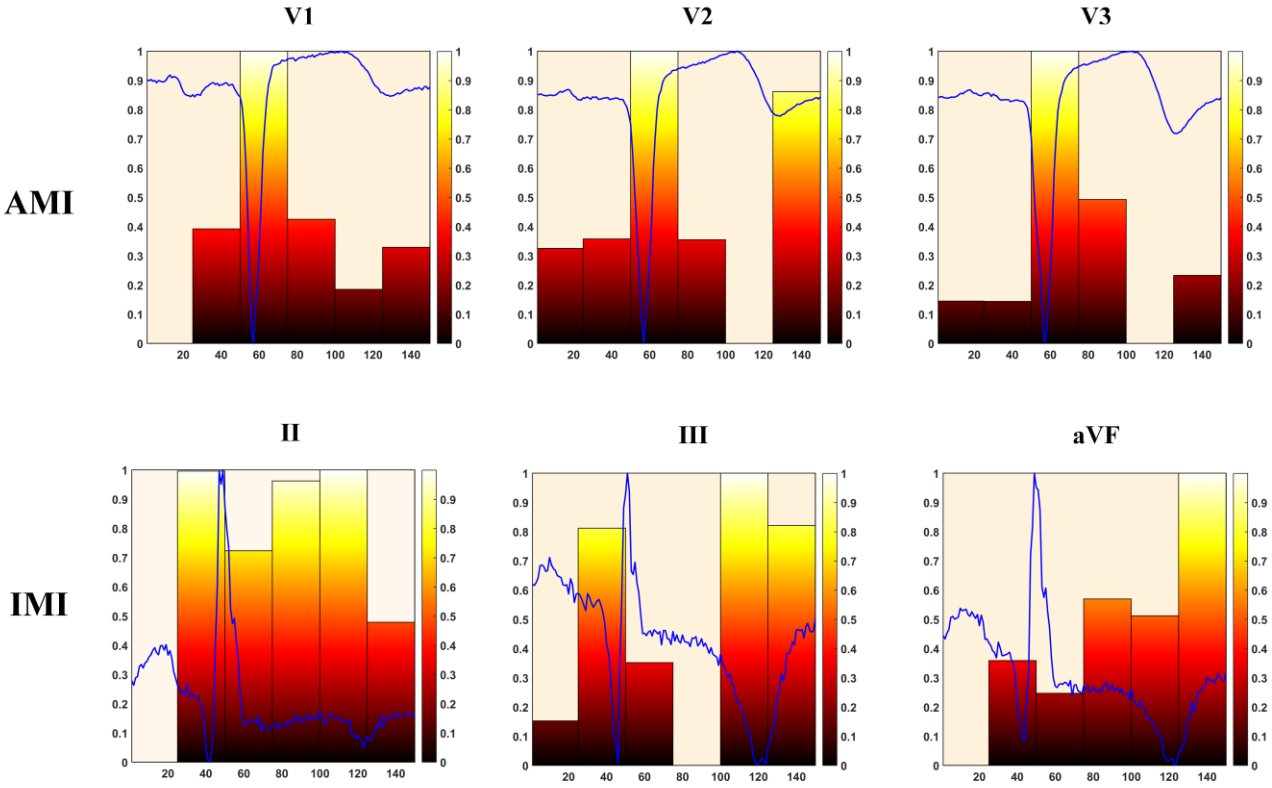

Fig.10. Examples of heatmaps for typical AMI and IMI in MFB-CBRNN .

specific structures and features of 12-lead ECGs efficiently. It is worth noticing that the LRM, an original optimization method proposed in the paper, can also boost the model performance. To further explore the underlying rules of LRM, more experiments corresponding to different values of inclusion probability are performed. Here, 50 trials of subject-based 5-fold cross validation are accomplished for each value of inclusion probability that represent the number of maintained leads ranging from 1 to 12. The averaged accuracies of MI detection are described in Fig. 9.

As shown in Fig.9, the MFB-CBRNN performs best at the inclusion probability of 0.5. This may due to the influence of the implicit ensemble in LRM. From a perspective of combinatorics, the most substructures can be generated by the LRM when 6 leads are randomly retained in the model, because:

$$\binom{12}{6} > \left(\binom{12}{5} = \binom{12}{7}\right) > \ldots > \left(\binom{12}{i} = \binom{12}{12-i}\right) > \ldots > \left(\binom{12}{1} = \binom{12}{11}\right) > \binom{12}{12}, i = 1,2,\ldots,12 \tag{13}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2910082, IEEE Journal of Biomedical and Health Informatics

TABLE IX
COMPARISON OF THE PROPOSED METHOD AGAINST EXISTING METHODS ON MI DETECTION

Method	Number of hand-crafted features	Number of leads	Data-base	Dataset used	K-fold cross validation	Performance	Evaluation	
							Class-based	Subject-based
[6]	36	12	PTB	MI detection 16960 MI beats 3200 HC beats	×	Se=99.97% Sp=99.90% Acc=99.96%	√	×
[9]	72	12	PTB	MI detection 100 MI frames 100 HC frames	√	Se=93.00% Sp=99.00% Acc=96.00%	√	×
[10]	23	12	PTB	MI detection 369 MI records 82 HC records	×	Acc=91.07%	√	×
[13]	10	3	PTB	IMI detection 3240 IMI segments 3037 HC segments	√	Se=99.35% Sp=98.29% Pp=98.41% Acc=98.84%	√	×
[13]	10	3	PTB	IMI detection 3240 IMI segments 3037 HC segments	√	Se=79.01% Sp=79.26% Pp=80.25% Acc=81.71%	×	√
[37]	47	1	PTB	MI detection 485753 MI beats 125652 HC beats	√	Se=99.45% Sp=96.27% Acc=98.80%	√	×
[17]	0	1	PTB	MI detection 40182 MI beats 10546 HC beats	√	Se=95.49% Sp=95.19% Pp=98.43% Acc=95.22%	√	×
[18]	0	3	PTB	IMI detection 3222 IMI segments 3055 HC segments	√	Se=85.33% Sp=84.09% Acc=84.54%	×	√
[20]	0	4	PTB	PTB database GAMI detection 34769 beats	√	Se=95.40% Sp=97.37% Acc=96.00%	×	√
[38]	0	3	PTB	PTB database MI detection 6720 frames	×	Se=96.00% Sp=84.00% Acc=86.00%	×	√
Proposed (class-based)	**0**	**12**	**PTB**	**PTB database MI detection 53712 MI beats 10638 HC beats**	√	**Se=99.97% Sp=99.54% Pp=99.91% Acc=99.90%**	√	×
Proposed (subject-based)	**0**	**12**	**PTB**	**PTB database MI detection 53712 MI beats 10638 HC beats**	√	**Se=94.42% Sp=86.29% Pp=97.21% Acc=93.08%**	×	√

where $\binom{n}{k}$ denotes the number of k-element combination of a n-element set. More substructures may lead to a more efficient implicit ensemble, providing a more general model. However, the model performance deteriorates more obviously when less leads are included. It can be inferred that an accurate detection of MI also requires sufficient information from more leads.

To reveal the correlation of the features that system automatically learns and the features typically used by clinicians when diagnosing MIs, a visualizing method named "heat map" is adopted to interpret the network [46]. This method is useful for understanding which parts of a given sample led a network to its final classification decision. Given typical examples of AMI and IMI, the normalized heatmaps

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2910082, IEEE Journal of Biomedical and Health Informatics

and raw signals are shown as Fig.10. For AMIs, typical symptoms such as pathological Q-waves and ST-T changes are usually shown in lead V1~V3[45]. Thus, the raw signals and corresponding heatmaps of lead V1~V3 are depicted. It can be observed that the pathological Q-waves and ST-T changes are strongly activated, especially the former. That means these features make the network determine AMIs. As for IMIs, the aforementioned symptoms are usually shown in lead II, III, and aVF, the raw signals and heatmaps are depicted in the figure. Also, according to the heatmap, the network can capture the aforementioned pathological changes of IMIs in the ECG. In summary, there are significant correlations between the features that system automatically learns and the features typically used by clinicians when diagnosing a MI condition. Thus, the MFB-CBRNN can obtain useful information by training on large-scale data.

As shown in Table. IX, several existing studies on MI detection are listed to obtain a brief comparison with our MFB-CBRNN. The algorithms in [6] [9] [10] [13] [37] are all based on conventional ML scheme, which requires extra feature extraction and selection. For instance, entropy-based and WT-based features are generated for each beat in [37], and KNN is employed to predict the final results. The algorithm achieved good performance in their experiments. From a perspective of performance evaluation, The methods in [13], [18], [20], and [38] adopt subject-based experiments. In [13], Stationary Wavelet Transform (SWT) is used to decompose the ECGs, then entropy-based features are generated for IMI detection. Also, feature selection using gain ratio is performed to obtain the most useful features. Obviously, this scheme needs more expertise to get a good performance. It achieves an average accuracy of 81.71% in the subject-based 10-fold cross validation experiments, which is lower than the other literatures. Another IMI detection algorithm is proposed in [18]. Based on a CNN-based structure, the algorithm demonstrates an average accuracy of 84.54% without hand-crafted features. In [20], Generalized Anterior MI (GAMI) is chose as the positive class. A modified CNN-based model named ML-CNN is used to discriminate GAMI and HC samples. the ML-CNN achieves a high accuracy(96.00%) in subject-based GAMI detection since it is specially designed and optimized for multilead ECGs. However, these methods only focus on a certain category of MI, their generalization for other MI types is unclear, which may limit their application range. For the model in [38], it can detect more MI types, but it only uses 3 leads (V1,V2,V3), causing a latent information loss.

According to the comparison, it can be observed that our MFB-CBRNN has more advantages than the other listed algorithms. Firstly, the MFB-CBRNN is a DL model that can learn the efficient features automatically; secondly, our model is suitable for the unique structure of the multilead signals. Each lead has its unique characteristics. Also, different leads will exhibit different features for the same disease, which is an important basis in the diagnosis process [2][3][45]. Independent branches for each lead can emphasize this significant aspect, whereas directly combining features from the 12 leads at the CNN level can blur the difference between the leads. Thus, from the perspective of clinical experience, independent feature learning for each lead is benefit for accurate diagnosis. To get the final result, the model needs to further abstract the characteristics and relationships of the various leads, rather than simply stacking them. Here a BLSTM is used to combine and "abstract" the features from the 12 leads. According to our experiments in Section V, the model with BLSTM summary can achieve better performance than the one without BLSTM. Therefore, the MFB-CBRNN can not only capture unique characteristics from each lead but also efficiently summarize them using BLSTM, which is suitable for 12-lead ECG analysis. Furthermore, it has excellent generalization ability, and achieves state-of-the-art results on MI detection. Based on the 5-fold cross validation, the class-based and subject-based accuracy are 99.90% and 93.08%, respectively. In summary, the MFB-CBRNN has impressive advantages and achieves significant results, which indicates that it is a good choice for MI detection using 12-lead ECGs.

VI. Conclusion

In this paper, a hybrid MFB-CBRNN is proposed for MI detection using 12-lead ECGs. It has a novel structure and achieves state-of-the-art performance. Based on DL framework, CNN and RNN are fused into a single entity in the MFB-CBRNN. No hand-designed feature is required in our method. The development of independent feature branches and BLSTM for leads summary can exploit the unique properties of 12-lead ECGs. Also, a useful optimization method, LRM, is designed to enhance the model performance. Its efficiency has been proved by our detailed experiments. The MFB-CBRNN is evaluated under class-based and subject-based approaches. In particular, its impressive performance in subject-based experiments indicates that our model has good generalization capacity on unseen patients. In summary, the MFB-CBRNN is very suitable for MI detection using 12-lead ECGs, it has a potential to assist the cardiologists in real-world diagnostics. In the future, the authors will work with the hospital to get more clinical data and attempt to extend the MFB-CBRNN to the detection of other CVDs, making it a more universal algorithm for computer-aided diagnostic systems.

References

[1] WHOFactSheet.http://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds), 2017.

[2] B. Surawicz, T.K. Knilans, Chou's Electrocardiography in Clinical Practice, Saunders Elsevier, Philadelphia, PA, 2008.

[3] K. Thygesen et al., "Third universal definition of myocardial infarction," *Circulation*, vol. 126, no. 16, pp. 2020–2035, 2012.

[4] N. Safdarian, N. J. Dabanloo, and G. Attarodi, "A new pattern recognition method for detection and localization of myocardial infarction using T-wave integral and total integral as extracted features from one cycle of ECG signal," *J. Biomed. Sci. Eng.*, vol.7, no.10, pp. 818–824, 2014.

[5] N. Ouyang, M. Ikeda, K. Yamauchi, "Use of an artificial neural network to analyse an ECG with QS complex in V1-2 leads", *Med. Biol. Eng. Comput.*, vol.35, no.5, pp. 556–560, Sep.1997.

[6] M. Arif , I.A. Malagore , F.A. Afsar , "Detection and localization of myocardial infarction using K-nearest neighbor classifier", *J. Med. Syst.*, vol.36, no.1, pp. 279–289, Feb. 2012.

[7] R. Tafreshi, A. Jaleel, J. Lim, L. Tafreshi, "Automated analysis of ECG waveforms with atypical QRS complex morphologies", *Biomed. Signal Proces.* , vol.10 , no. 5, pp. 41–49. Mar. 2014.

[8] S. Banarjee, M. Mitra, "Cross wavelet transform based analysis of electrocardiogram signals", *Int. J. Electr. Electron. Comput. Eng.*, vol.1, no.2, pp. 88–92, Dec. 2012.

[9] L. N. Sharma, R. K. Tripathy and S. Dandapat, "Multiscale Energy and Eigenspace Approach to Detection and Localization of Myocardial Infarction," *IEEE Trans. Biomed. Eng.* , vol. 62, no. 7, pp. 1827-1837, Jul. 2015.

[10] N.A.Bhaskar, "Performance analysis of support vector machine and neural networks in detection of myocardial infarction," *Proc. Comput. Sci.*, vol.46, pp.20-30, 2015.

[11] L. Sun, Y.Lu, K.Yang et al., "ECG analysis using multiple instance learning for myocardial infarction detection," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 12, pp. 3348–3356, Dec. 2012.

[12] U.R. Acharya, H. Fujita, A. Muhammad, et al., "Automated characterization and classification of coronary artery disease and myocardial infarction by decomposition of ECG signals: a comparative study" , *Inf. Sci.*, vol. 377, pp. 17–29, Jan. 2017.

[13] L. D. Sharma, R. K. Sunkaria, "Inferior myocardial infarction detection using stationary wavelet transform and machine learning approach", *Signal. Image. Video. P.*, vol.12, no.2, pp. 199–206, Feb. 2018.

[14] S. Raj, K. C. Ray, and O. Shankar, "Development of robust, fast and efficient QRS complex detector: A methodological review," *Aust. Phys. Eng. Sci. Med.*, vol. 41, no. 3, pp. 581–600, Sep. 2018.

[15] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[16] D. Ravì, C.Wong, B.Lo, et al., "Deep Learning for Health Informatics," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 1, pp. 4–21,Jan. 2017.

[17] U.R. Acharya, H. Fujita, S.L. Oh, et al., "Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals", *Inf. Sci.*, vol. 415, pp. 190–198, Nov. 2017.

[18] T. Reasat, C. Shahnaz, "Detection of inferior myocardial infarction using shallow convolutional neural networks," *IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, 2017, pp. 718-721.

[19] C. Szegedy, et al., "Going deeper with convolutions," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.

[20] W. Liu, M. Zhang, Y. Zhang, et al., "Real-time multilead convolutional neural network for myocardial infarction detection," *IEEE J. Biomed. Health Inform.* , vol. 22, no. 5, pp. 1434-1444, Sep. 2018.

[21] W. Liu, Q. Huang, S. Chang, et al., "Multiple-feature-branch convolutional neural network for myocardial infarction diagnosis using electrocardiogram," *Biomed. Signal Proces.*, vol. 45, pp. 22-32, Aug. 2018.

[22] S. Haykin, Neural Networks and Learning Machines, 3rd ed. Cambridge, MA: Prentice-Hall, 2008.

[23] S. Hochreiter, and J. Schmidhuber, "Long short-term memory," *Neural Comput.* vol. 9, no. 8, pp. 1735-1780, Nov.1997.

[24] C. Zhang, G. Wang, J. Zhao, et al., "Patient-specific ECG classification based on recurrent neural networks and clustering technique," *Proc. of Int. Conf. Biomed. Eng. (BioMed)*, 2017, pp. 63–67.

[25] S. Chauhan, L. Vig, "Anomaly detection in ECG time signals via deep long short-term memory networks," *Proc. Of Int. Conf. Data Sci. Adv. Anal. (DSAA)*, 2015, pp. 1-7.

[26] Ozal Yildirim, "A novel wavelet sequences based on deep bidirectional LSTM network model for ECG signal classification," *Comput. Biol. Med.* vol. 96, pp. 189-202, May.2018.

[27] M. Schuster, K.K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.* , vol. 45, no. 11, pp. 2673-2681, Nov. 1997.

[28] J.H.Tan, Y. Hagiwara, W. Pang, et al., "Application of stacked convolutional and long short-term memory network for accurate identification of CAD ECG signals," *Comput. Biol. Med.*, vol. 94, pp.19-26, Mar. 2018.

[29] M. Zihlmann, D. Perekrestenko, and M. Tschannen, "Convolutional recurrent neural networks for electrocardiogram classification," arXiv preprint arXiv:1710.06122, 2017.

[30] P. Xie, G. Wang, C. Zhang et al., "Bidirectional Recurrent Neural Network And Convolutional Neural Network (BiRCNN) For ECG Beat Classification," *Proc. of Int. Conf. Eng. Med. Biol. Soc. (EMBC)*, 2018, pp. 2555-2558.

[31] A. L. Goldberger, L. A. N. Amaral, L. Glass, et al., "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[32] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, pp. 230–236, Mar. 1985.

[33] S. Ioffe and C. Szegedy., "Batch normalization: Accelerating deep network training by reducing internal covariate shift.," *Proc. Int. Conf on Mach. Learn. (ICML)*, 2015, pp.448-456.

[34] N. Srivastava, G. Hinton, A. Krizhevsky, et al. "Dropout: a simple way to prevent neural networks from overfitting." *J. Mach. Learn. Res.*, vol. 15, pp.1929-1958, 2014.

[35] M. Lin, Q.Chen, S. Yan, "Network in network," *Proc. Of Int. Conf. Learn. Repres. (ICLR)*, 2014.

[36] Abadi, Mart ń, et al. "Tensorflow: a system for large-scale machine learning," *Proc. of Oper. Syst. Des. Implementation. (OSDI)*, 2016, pp. 265-283.

[37] U.R. Acharya , H. Fujita , V.K. Sudarshan, et al. , "Automated detection and localization of myocardial infarction using electrocardiogram: a comparative study of different leads, " *Knowl.-Based Syst.*, vol.99, pp. 146–156, May. 2016.

[38] D. Rajan and J. Thiagarajan, "A Generative Modeling Approach to Limited Channel ECG Classification," ArXiv Preprint ArXiv:1802.06458, 2018.

[39] S. Bengio, O. Vinyals, N. Jaitly, and N. Shazeer, "Scheduled sampling for sequence prediction with recurrent neural networks," in *Proc. Advances Neural Info. Process. Syst.*, 2015, pp. 1171–1179.

[40] A. Graves, A. Mohamed, and G. E. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2013, pp. 6645–6649.

[41] M. Sundermeyer, T. Alkhouli, J. Wuebker, and H. Ney, "Translation modeling with bidirectional recurrent neural networks," in *Proc. Conf. Empir. Meth. Nat. Lang. Process.*, 2014, pp. 14–25.

[42] A. Hannun, P. Rajpurkar, M. Haghpanahi, et al. "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network", *Nat. Med.*, vol.25, no.1, pp.65-69, Jan. 2019.

[43] M.Sampson, A. McGrath. "Understanding the ECG Part 2: ECG basics". *Br. J. Card. Nurs.*, vol. 10, no.12, pp. 588-594, Dec. 2015.

[44] L. Breiman, "Bias, variance, and arcing classifiers," Univ. California, Berkeley, 1996. Tech. Rep.

[45] P. Zimetbaum, M.Josephson, "Use of the electrocardiogram in acute myocardial infarction", *N. Engl. J. Med.*, vol. 348, no. 10, pp. 933-940, Mar. 2003.

[46] R. Selvaraju, M. Cogswell, A. Das, et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization", *Proc. of Int. Conf. Comput. Vis.(ICCV)*, 2017, pp. 618-626.