IEEE Journal of Translational Engineering in Health and Medicine

Received 19 September 2020; revised 24 January 2021; accepted 4 March 2021.
Date of publication 9 March 2021; date of current version 15 March 2021.

Digital Object Identifier 10.1109/JTEHM.2021.3064675

MLBF-Net: A Multi-Lead-Branch Fusion Network for Multi-Class Arrhythmia Classification Using 12-Lead ECG

JING ZHANG[1], DENG LIANG[1], AIPING LIU[1], MIN GAO[2], XIANG CHEN[1], XU ZHANG[1], (Member, IEEE), AND XUN CHEN[3], (Senior Member, IEEE)

[1]Department of Electronic Science and Technology, University of Science and Technology of China, Hefei 230027, China
[2]Department of Electrocardiogram, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei 230001, China
[3]Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, China

CORRESPONDING AUTHOR: A. LIU (aipingl@ustc.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61922075, in part by the National Key Research and Development Program of China under Grant 2018YFB1005001, and in part by the USTC Research Funds of the Double First-Class Initiative under Grant YD2100002004.

ABSTRACT Automatic arrhythmia detection using 12-lead electrocardiogram (ECG) signal plays a critical role in early prevention and diagnosis of cardiovascular diseases. In the previous studies on automatic arrhythmia detection, most methods concatenated 12 leads of ECG into a matrix, and then input the matrix to a variety of feature extractors or deep neural networks for extracting useful information. Under such frameworks, these methods had the ability to extract comprehensive features (known as integrity) of 12-lead ECG since the information of each lead interacts with each other during training. However, the diverse lead-specific features (known as diversity) among 12 leads were neglected, causing inadequate information learning for 12-lead ECG. To maximize the information learning of multi-lead ECG, the information fusion of comprehensive features with integrity and lead-specific features with diversity should be taken into account. In this paper, we propose a novel Multi-Lead-Branch Fusion Network (MLBF-Net) architecture for arrhythmia classification by integrating multi-loss optimization to jointly learning diversity and integrity of multi-lead ECG. MLBF-Net is composed of three components: 1) multiple lead-specific branches for learning the diversity of multi-lead ECG; 2) cross-lead features fusion by concatenating the output feature maps of all branches for learning the integrity of multi-lead ECG; 3) multi-loss co-optimization for all the individual branches and the concatenated network. We demonstrate our MLBF-Net on China Physiological Signal Challenge 2018 which is an open 12-lead ECG dataset. The experimental results show that MLBF-Net obtains an average F_1 score of 0.855, reaching the highest arrhythmia classification performance. The proposed method provides a promising solution for multi-lead ECG analysis from an information fusion perspective.

INDEX TERMS Arrhythmia classification, multi-lead ECG analysis, co-optimization, deep learning.

I. INTRODUCTION

Cardiovascular disease (CVD) is the leading cause of global mortality. It was announced by the World Health Organization (WHO) that an estimated 17.9 million people died from CVD in 2016, accounting for 31% of global deaths [1]. Cardiac arrhythmia is a very common type of CVD, which manifests as abnormal heart rhythms. According to statistics, about half of all cardiovascular deaths are sudden cardiac deaths and about 80% of these are caused by cardiac arrhythmia [2]. Electrocardiogram (ECG) is a widely accessed, non-invasive, and inexpensive tool for arrhythmia diagnosis in clinic. It records the heart's electrical activities over time through electrodes attached to the skin surface. Recently, intelligent healthcare has become increasingly prominent. Automatic arrhythmia detection based on ECG could assist doctors in clinical practice, and also provide ordinary people with daily monitoring using wearable devices. Therefore, how to promote the accuracy of automatic arrhythmia detection is a critical issue.

Over the last decades, a large number of traditional ECG classification methods have been developed. Traditional methods primarily comprise three procedures involving preprocessing (e.g. denoising and heartbeat segmentation), feature extraction, and classification. Of these, feature extraction

is the most crucial step, which relies on professional knowledge to construct a set of hand-craft features. In recent years, deep learning has made great success in the field of healthcare with its powerful capability to extract high-level abstract features automatically, avoiding laborious manual feature design. Many studies have designed deep learning-based approaches for arrhythmia detection using ECG signals.

A standard ECG record contains 12 leads (i.e. I, II, III, avR, avL, avF, V1, V2, V3, V4, V5, V6), which is widely used in clinical arrhythmia diagnosis. The 12-lead ECG has two inherent properties: integrity and diversity [3]. On the one hand, the 12-lead ECG signal contains comprehensive information by recording the electrical potential from different spatial angles of the heart. It gives an overall reflection on the heart's condition. Thus, the 12-lead ECG could be treated as an integrated one to make a diagnosis. On the other hand, different leads correspond to different anatomical areas of the heart, providing distinct perspectives. Thus, the ECG signal under each lead contains lead-specific features, and the 12-lead ECG signal has diverse information across leads. To maximize information learning for 12-lead ECG, integrity and diversity should be both taken into account. For multi-lead ECG analysis, most existing studies concatenated multi-lead ECG into a matrix, and then input the matrix to a variety of feature extractors or deep neural networks for extracting useful information. However, such approaches lacked explicit mechanisms to realize lead-specific features extraction and only considered comprehensive features extraction of multi-lead ECG. Furthermore, they were difficult to utilize the adequate information fusion of diversity and integrity to enhance the detecting performance.

To address the limitations mentioned above, how to fully utilize the diversity and integrity of multi-lead ECG and thereby maximize the information learning is investigated in this paper. The main contributions of this paper are concluded below:

1) We propose a novel Multi-Lead-Branch Fusion Network (MLBF-Net) architecture with multiple branches for arrhythmia classification using 12-lead ECG. For realizing diversity learning, each branch of MLBF-Net is designed to classify the ECG signal under a specific lead, which could learn lead-specific features. Specifically, each branch introduces a hierarchical network structure consisting of the convolutional layers, bidirectional gated recurrent unit (BiGRU), and an attention module to mine the discriminative information further. For realizing integrity learning, the output feature maps from all branches are concatenated to form the concatenated network of MLBF-Net. It is responsible for learning to classify the ECG signal based on all 12 leads, which could extract comprehensive features.

2) We design a collaborative optimization strategy with multiple losses, specialized for all the individual branches and the concatenated network. This strategy not only optimizes comprehensive features of multi-lead ECG, but also realizes lead-specific features learning simultaneously during the training process, which could achieve the information fusion of diversity and integrity.

The remainder of this paper is arranged as follows. Section II outlines the related works. Section III describes the architecture of the proposed Multi-Lead-Branch Network. Section IV presents the experimental result. Section V gives the discussions. Finally, Section VI summaries this paper.

II. RELATED WORKS

Traditional ECG classification methods designed a number of hand-craft features. Typical hand-craft features include statistical features [4]–[6], P-QRS-T features [7]–[9], morphological features [9]–[12], and wavelet features [13]–[16]. Also, mathematical transformations that transform the high-dimensional ECG signal into a lower-dimensional space can be used for extracting meaningful information, such as independent component analysis (ICA) [17]–[19], principal component analysis (PCA) [19]–[21], and linear discriminant analysis (LDA) [19], [21]. Following feature extraction, a variety of classifiers are carried out to classify the extracted features. This can be implemented by artificial neural network (ANN) [7], [13], support vector machine (SVM) [12], [14], [16], [22], k nearest neighbor (KNN) [10], [17], decision tree [15], [17] and bayesian classifier [12], [18].

Deep learning is increasingly predominant in recent studies on ECG classification. Convolutional neural networks (CNNs) are a commonly adopted type of deep neural network due to its effectiveness in extracting features. Kiranyaz *et al.* [23] designed an adaptive CNN for patient-specific ECG heartbeat classification, which incorporates traditional feature extraction and classification into a single learning structure. Rahhal *et al.* [24] transformed ECG signals to image-like representations using continuous wavelet transform, and then fed these representations into a deep CNN pretrained on ImageNet with a large number of annotated images, which achieved a good detection performance for supraventricular ectopic beats and ventricular ectopic beats. Hannun *et al.* [25] presented a 34-layer residual CNN with a cardiologist-level accuracy in detecting twelve cardiac arrhythmias. In other studies, the ECG signal was viewed as a time-series and recurrent neural network (RNN) specialized for tackling sequential data was adopted. The representative variants of RNN include long short-term memory (LSTM) and gated recurrent unit (GRU). Saadatnejad *et al.* [26] developed a real-time heartbeat classification algorithm for personal wearable devices based on multiple LSTMs and wavelet transform. Lynn *et al.* [27] proposed a deep bidirectional GRU network for classifying biometric ECG signals. Further, many research works have designed hierarchical networks by stacking CNN and RNN. He *et al.* [28] stacked a deep residual CNN and a bidirectional LSTM layer for arrhythmia classification, and obtained a good performance. Yao *et al.* [29] classified multi-class arrhythmias using an integrated model consisting

IEEE Journal of Translational Engineering in Health and Medicine

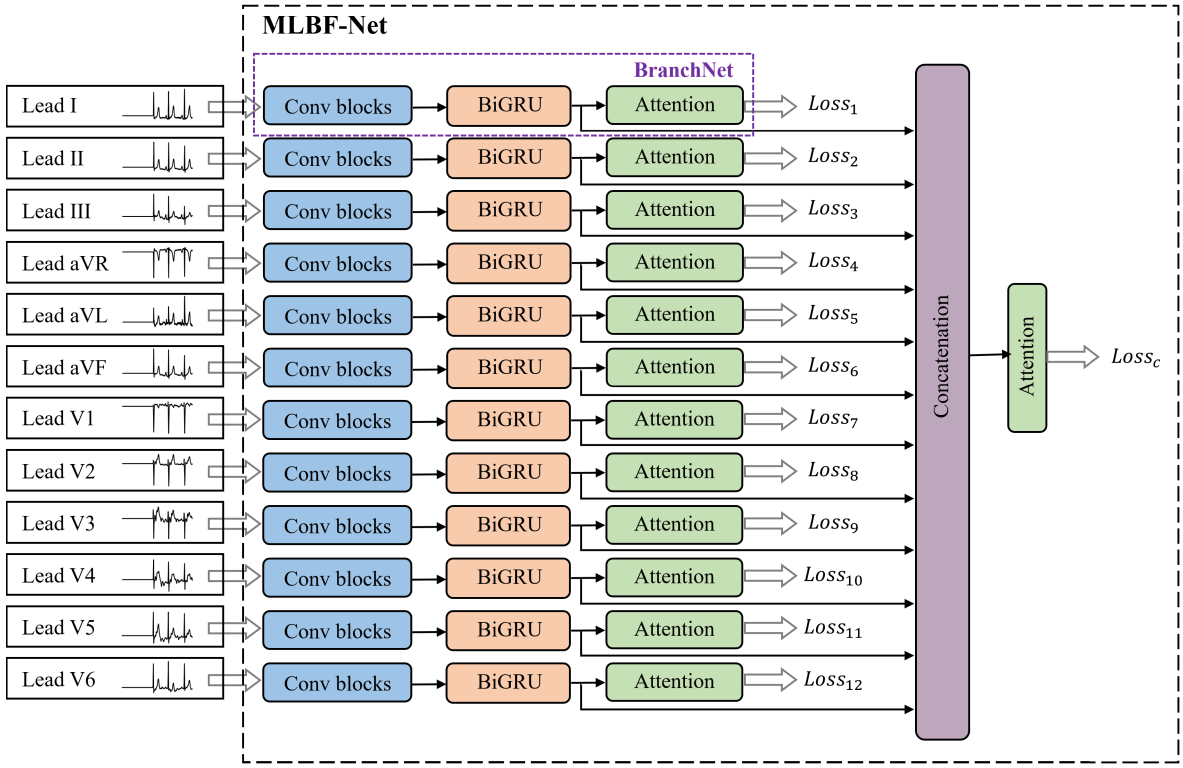


FIGURE 1. The proposed network architecture.

of VGGNet-based CNN and varied-length LSTMs, which allows for varied-length signal input and is effective in detecting paroxysmal arrhythmias.

Recently, Liu *et al.* [3] proposed a multiple-feature-branch convolutional neural network (MFB-CNN) for myocardial infarction detection using 12-lead ECG, aiming at exploiting both integrity and diversity. Particularly, 12 ECG leads were input into different branches of MFB-CNN for learning lead-specific features; a fully connected layer was used to summarize the output feature map of 12 feature branches, utilizing the integrity. However, the parameters of MFB-CNN were optimized by only a single loss function in an end-to-end way. This single-loss training strategy is unable to fully exploit the specific information of each individual branch in an isolated way, since the classification loss of training samples is backwardly propagated to all 12 branches [30]. Therefore, diversity learning is weakened significantly. In [31], Liu *et al.* refined the network architecture of MFB-CNN, in which the fully connected layer was replaced by LSTM for summarizing all the branches. In addition, several branches were randomly deactivated at each training iteration for improving the generalization of the model. Nevertheless, as with [3], the deficiency in diversity learning still exists. Currently, the existing works have not achieved joint learning for the diversity and integrity of multi-lead ECG. In other words, the fusion of diversity and integrity to maximize the information learning of multi-lead ECG is worthy of further investigation.

III. METHODS

A. MODEL OVERVIEW

The proposed Multi-Lead-Branch Fusion Network (MLBF-Net) is illustrated in Fig. 1. It is mainly composed of three components: 1) multiple lead-specific branches for learning the diversity of multi-lead ECG; 2) cross-lead features fusion by concatenating the output feature maps of all branches for learning the integrity of multi-lead ECG; 3) multi-loss co-optimization for all the individual branches and the concatenated network. These three components are described in detail below.

B. SINGLE LEAD-BRANCH FEATURE LEARNING

The preprocessed 12-lead ECG signals X are split into 12 single leads, denoted as $X_j \in \mathbb{R}^{L\times 1}, j \in \{1, 2, \ldots, 12\}$ where L is the length of the preprocessed signal. X_j is then fed into the jth branch. The configuration of all branches is the same. For convenience, we name the single branch network "BranchNet", of which the configuration is shown in Table 1. BranchNet combines 15 convolutional layers, a bidirectional gated recurrent unit (BiGRU) layer, and an attention module. Its internal behavior can be formulated as:

$$f_{cnn_j} = CNN(X_j) \tag{1}$$

$$f_{BiGRU_j} = BiGRU(f_{cnn_j}) \tag{2}$$

$$f_{att_j} = Attention(f_{BiGRU_j}) \tag{3}$$

TABLE 1. The configuration of each branch.

Configurations of BranchNet	
Layers	Output Size
Input (preprocessed single lead ECG signal)	15000 × 1
(Conv3_12, stride 1 Conv3_12, stride 1 Conv24_12, stride 2 Dropout, 0.2) × 4	938 × 12
(Conv3_12, stride 1 Conv3_12, stride 1 Conv48_12, stride 2 Dropout, 0.2) × 1	f_{CNN}: 469 × 12
BiGRU_12 Dropout, 0.2	f_{BiGRU}: 469 × 24
Attention Dropout, 0.2	f_{att}: 24
Fully-connected	f_{out}: 9

The convolutional layer is represented as Conv(kernel size)_(kernel number).

where *CNN*, *BiGRU*, and *Attention* denote the convolutional neural subnetwork, BiGRU layer, and attention module, respectively. f_{cnn_j}, f_{BiGRU_j}, and f_{att_j} denote the output feature map of these three modules of the jth branch.

The cross-entropy loss is employed for training single branch, calculated as:

$$L_j = -\frac{1}{N}\sum_{i=1}^{N} log(\frac{exp(p(X_j^{(i)}, y^{(i)}))}{\sum_c exp(p(X_j^{(i)}, c))}) \tag{4}$$

where N is the number of training samples, $y^{(i)}$ is the true label of ith sample, and $p(X_j^{(i)}, c)$ denotes the probability that the input $X_j^{(i)}$ is predicted as label c. By minimizing the cross-entropy loss function during training, lead-specific features are iteratively optimized for achieving diversity learning.

1) CONVOLUTIONAL NEURAL SUBNETWORK

As shown in Table 1, the convolutional neural subnetwork is composed of 5 convolutional blocks, with a total of 15 convolutional layers. Each convolution block includes three convolution layers, together with a dropout [32] layer. The output of each convolutional layer is nonlinear transformed by a leaky rectified linear unit (LeakyReLU) activation function, where the operation is omitted for brevity in Table 1. Although ReLU [33] is a more common choice, LeakyReLU [34] is applied due to the ability to avoid the dead neurons problem. To mitigate the neural network from overfitting, the dropout rate is set to 0.2.

2) BIDIRECTIONAL GRU

The output feature map f_{cnn} of convolutional neural subnetwork flows into a bidirectional GRU layer. GRU [35] and LSTM [36] are the evolutionary implementations of RNN. We select GRU since it has similar performance to LSTM but with less computational complexity. As illustrated in Fig. 2, the bidirectional GRU consists of a forward GRU and a backward GRU, which read the time-series feature map in the temporal and anti-temporal directions, respectively. At time t, the forward GRU aggregates the information of f_{cnn} from 1 to t and the backward GRU aggregates the information of f_{cnn} from T to t. In our model, T equals 469, which is the time length of f_{cnn}. The features from both opposite directions are incorporated by the bidirectional GRU to obtain the contextual information. Note that we achieve the incorporation by concatenation operation, summarizing the features centered around time t. In our experiments, the unit number of the bidirectional GRU layer is configured to 12, meaning that the dimension for each time step is 12.

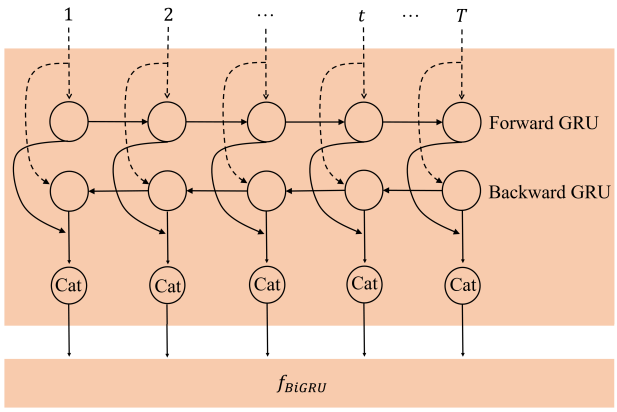


FIGURE 2. The illustration for a bidirectional GRU. "Cat" is an abbreviation for concatenation operation.

3) ATTENTION MODULE

Different ECG components such as P wave, QRS complex wave, and ST-segment contribute differentially to the pathological information of the whole ECG segment. Hence, attention mechanism [37] is introduced to extract the informative segments and merge the representation of these important segments. This process is formulated as:

$$u_{jt} = tanh(W_w f_{BiGRU_{jt}} + b_w) \tag{5}$$

$$\alpha_{jt} = \frac{exp(u_{jt}^T u_w)}{\sum_t exp(u_{jt}^T u_w)} \tag{6}$$

$$f_{att_j} = \sum_t \alpha_{jt} f_{BiGRU_{jt}} \tag{7}$$

That is, the encoded ECG signal $f_{BiGRU_{jt}}$ is fed into a one-layer multilayer perceptron to obtain u_{jt} as a hidden representation of $f_{BiGRU_{jt}}$. Then the similarity of u_{jt} and a trainable vector u_w is measured and a normalized importance weight vector α_{jt} is obtained through a softmax function. Afterwards, the weighted sum of the encoded ECG signal $f_{BiGRU_{jt}}$ and its corresponding weight vector α_{jt} is computed to get the weighted representation f_{att_j}. W_w, u_w, and b_w are randomly initialized trainable parameters.

IEEE Journal of Translational Engineering in Health and Medicine

C. MULTI-LEAD FEATURE FUSION LEARNING

For learning complementary cross-lead information and providing more robust diagnosis, we fuse the features extracted from each lead by concatenating the lead-specific feature maps f_{BiGRU_j} of all branches in channel axis:

$$F = Cat(f_{BiGRU_1}, f_{BiGRU_2}, \ldots, f_{BiGRU_{12}}) \tag{8}$$

Then the concatenated feature map F is fed through the same attention module used in BranchNet to build the concatenated network. The concatenated network shares the feature maps from the input X_j to the output f_{BiGRU_j} of BiGRU layer with each branch for obtaining the comprehensive prediction based on 12-lead ECG. Similar to single lead-branch training, a cross-entropy loss is also employed to train the concatenated network, calculated as:

$$L_c = -\frac{1}{N}\sum_{i=1}^{N} log(\frac{exp(p(X^{(i)}, y^{(i)}))}{\sum_c exp(p(X^{(i)}, c))}) \tag{9}$$

where $p(X^{(i)}, c)$ denotes the probability that the input $X^{(i)}$ is predicted as label c.

D. JOINT OPTIMIZATION WITH MULTIPLE LOSSES

It is noted that lead-specific branches are interrelated rather than independent. In order to jointly learning the diversity and integrity of multi-lead ECG, we design multiple dedicated losses for collaboratively optimizing multiple branches and the concatenated network. The loss for each branch is designed to optimize lead-specific features for maximizing the discriminative capability of single-lead ECG, and the loss for the concatenated network aims to optimize multi-lead comprehensive features simultaneously. For the training of the whole model, the final loss function is defined as:

$$L = L_c + \lambda(L_1 + \ldots + L_{12}) \tag{10}$$

where λ is a balance parameter used for determining the importance ratio between diversity and integrity, which is set to 1 in our experiments. L_c and L_j are the cross-entropy losses of the concatenated network and the jth branch, respectively.

The co-optimization strategy not only optimizes multi-lead comprehensive features, but also realizes lead-specific features learning simultaneously during the training process. Compared with the regular single-loss learning, this strategy achieves the information fusion of diversity and integrity, thereby promoting the maximum of multi-lead ECG information learning.

IV. EXPERIMENT

The proposed method is performed using Python language and Keras 2.2.4 framework. All experiments in this paper were run on a server with Xeon E5 2620 CPU, 128GB memory and four GeForce RTX cards.

TABLE 2. Distribution of ECG classes on the public dataset of CPSC 2018.

ECG class	#Record
Normal	918
Atrial fibrillation (AF)	1098
First-degree atrioventricular block (1-AVB)	704
Left bundle branch block (LBBB)	207
Right bundle branch block (RBBB)	1695
Premature atrial contraction (PAC)	556
Premature ventricular contraction (PVC)	672
ST-segment depression (STD)	825
ST-segment elevated (STE)	202
Total	**6877**

A. DATA DESCRIPTION

1) CPSC 2018 DATASET

China Physiological Signal Challenge 2018 (CPSC 2018) provides 12-lead ECG records, which is suitable for multi-lead ECG analysis. CPSC 2018 dataset includes publicly accessible 6877 12-lead ECG records (female: 3178; male: 3699) and the private test set consisting of 2954 12-lead ECG records. The private test set is inaccessible for researchers to ensure a fair comparison. These ECG records were acquired from 11 hospitals. The length of the records lengths varying from 6 to 60 seconds, and the sampling rate is 500Hz. The details of the public database can be found in Table 2, with a total of 9 ECG classes containing normal rhythm and 8 types of arrhythmias. Figure 3 gives an example of a 12-lead ECG record. More details about CPSC 2018 database can be seen in [38].

2) PTB-XL DATASET

PTB-XL dataset [39] contains 21837 clinical 12-lead ECG records of 10 seconds length from 18885 patients. The ECG statements used for annotation are conforming to the SCP-ECG standard, and each record may contain more than one statement. The dataset covers 71 different ECG statements, which can be further aggregated into 44 diagnostic, 19 form, and 12 rhythm statements. The three statement levels are non-mutually exclusive categories, and the diagnostic ECG statements include 4 form rhythm statements. For diagnostic statements, they can be organized into five coarse superclasses (*NORM*: normal ECG, *CD*: conduction disturbance, *MI*: myocardial infarction, and *STTC*: ST/T changes) and 23 subclasses, as described in Table 4. Table 3 shows the number of ECG statements per record for different annotation levels. The sampling rate of 100Hz was used in this paper.

B. EVALUATION METRIC

In this research, F_1 score is used to measure the model's classification performance for each class of CPSC 2018 dataset.

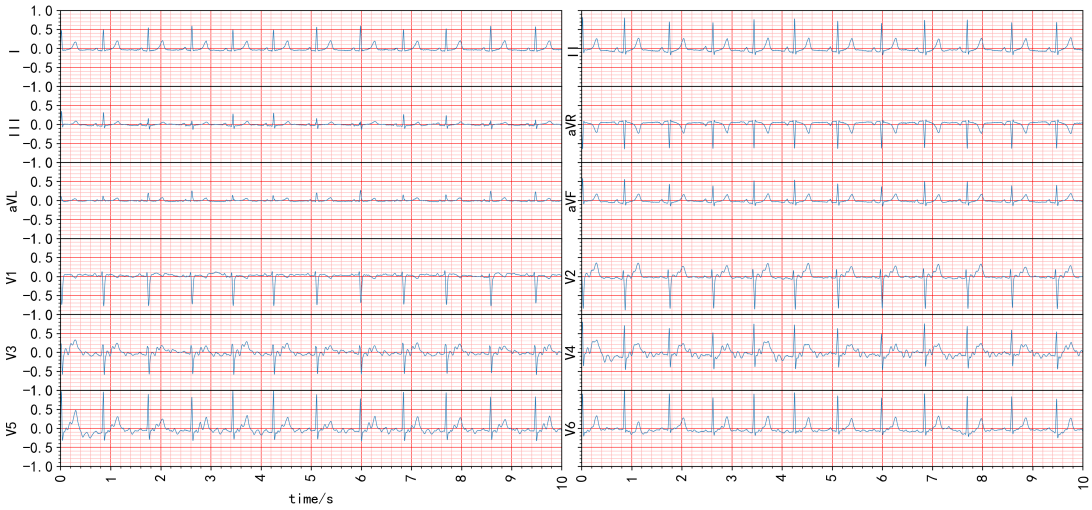


FIGURE 3. An example of normal 12-lead ECG record.

TABLE 3. Number of ECG statements per record for different annotation levels on PTB-XL dataset.

Level	# Classes	1	2	3	≥4	# Total
diag.	44	15019	4242	1515	654	21430
sub-diag.	23	16272	4079	920	159	21430
super-diag.	5	15239	4171	1439	581	21430
form	19	6693	1672	524	99	8988
rhythm	12	20923	142	1	0	21066
all	71	705	11247	5114	4771	21837

It is the harmonic mean of precision and recall, defined as:

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

$$F_1 = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \tag{13}$$

The terms TP, FP, and FN refer to the sample number of true positive, false positive, and false negative, respectively. The average of F_1 scores for all classes, namely macro-F_1 score, is computed to make a final evaluation.

Following the recommendation in [39], we use macro-averaged and threshold-free area under curve (AUC) metric to evaluate the classification performance of the proposed model on PTB-XL dataset.

C. IMPLEMENTATION DETAILS

1) PREPROCESSING

Similar to our previous work [44], the preprocessing for the original ECG signal provided by CPSC 2018 dataset includes two procedures: downsampling and cropping or padding. In the first step, the downsampling from 500Hz to 250Hz was performed to speed up the training. In the second step, the downsampled ECG signals were cropped or padded with zeros to the same length because convolutional neural networks do not accept varied-length input. In our setting, 60 seconds was the target length. It means that the signals longer than 60 seconds were cropped and those less than 60 seconds were padded with zero. For PTB-XL dataset, the original ECG signal was fed into the model without preprocessing.

TABLE 4. SCP-ECG acronym descriptions for superclasses and subclasses of PTB-XL dataset.

Superclasses	Subclasses	SCP statement Description
NORM	NORM	Normal ECG
CD	LAFB/LPFB	left anterior/left posterior fascicular block
	IRBBB	incomplete right bundle branch block
	ILBBB	incomplete left bundle branch block
	CLBBB	complete left bundle branch block
	CRBBB	complete right bundle branch block
	_AVB	AV block
	IVCB	non-specific intraventricular conduction disturbance (block)
	WPW	Wolff-Parkinson-White syndrome
HYP	LVH	left ventricular hypertrophy
	RHV	right ventricular hypertrophy
	LAO/LAE	left atrial overload/enlargement
	RAO/RAE	right atrial overload/enlargement
	SEHYP	septal hypertrophy
MI	AMI	anterior myocardial infarction
	IMI	inferior myocardial infarction
	LMI	lateral myocardial infarction
	PMI	posterior myocardial infarction
STTC	ISCA	ischemic in anterior leads
	ISCI	ischemic in inferior leads
	ISC_	non-specific ischemic
	STTC	ST-T changes
	NST_	non-specific ST changes

2) TRAINING SETTING

We trained the proposed model in an end-to-end way. The preprocessed (for CPSC 2018 dataset) or raw (for PTB-XL

IEEE Journal of Translational Engineering in Health and Medicine

TABLE 5. Comparison for classification performance of previous works and ours evaluated on the private test set of CPSC 2018.

Work	Year	Architecture	F_1 score									
			Normal (394)	AF (466)	I-AVB (295)	LBBB (97)	RBBB (756)	PAC (250)	PVC (276)	STD (340)	STE (80)	Average (2954)
[29]	2018	CNN+LSTM	0.753	0.900	0.809	0.874	0.922	0.638	0.832	0.762	0.462	0.772
[40]	2018	Expert features; CNN	0.82	0.91	0.87	0.87	0.91	0.63	0.82	0.81	0.60	0.81
[28]	2019	CNN+LSTM	-	-	-	-	-	-	-	-	-	0.806
[41]	2019	CNN+Attention	0.79	0.93	0.85	0.86	0.93	0.75	0.85	0.80	0.56	0.813
[42]	2020	CNN+LSTM+Attention	0.789	0.920	0.850	0.872	0.933	0.736	0.861	0.789	0.556	0.812
[43]	2020	Multi-Scaled CNN+Attention	0.82	0.90	0.86	0.87	0.93	0.78	**0.88**	0.80	0.62	0.828
[44]	2020	CNN+Attention+BiGRU; Ensemble model	0.819	**0.936**	0.866	0.862	0.926	0.789	0.865	0.812	0.640	0.835
[45]	2020	CNN+BiGRU+Attention; Ensemble model	0.801	0.933	0.875	0.884	0.910	**0.826**	0.869	0.811	0.624	0.837
Ours[1]		Multi-Lead-Branch; CNN+BiGRU+Attention	**0.850**	0.933	**0.885**	0.862	0.931	0.801	0.859	**0.820**	0.646	0.843
Ours[2]		Ensemble model based on Ours[1]	0.847	0.934	0.884	**0.896**	**0.939**	0.822	0.878	0.818	**0.677**	**0.855**

The "-" indicates the F_1 score was not reported. The number of ECG classes are in parentheses.
Ours[1] indicates the single model without ensemble. Ours[2] indicates ensemble model.
The highest score for each class is in bold.

dataset) training data was grouped into batches of 64 samples to fed into MLBF-Net. We set 64 to batch size by tuning this hyperparameter. The batched updating of network parameters takes less memory and is more computationally efficient. More importantly, a more robust convergence can be obtained by mini-batch, averting local minima. We applied adaptive moment estimation (Adam) optimizer [46] to update the weights of the whole model iteratively with a fixed learning rate of 0.001. Adam, which updates weights based on exponential decaying averages of past gradients and past squared gradients, is usually considered to converge to an excellent performance [47].

In order to alleviate the proposed model overfitting during the training process, another two strategies were adopted in addition to dropout layers of the network structure. The first strategy is earlystopping, which stops training if the classification performance of the model on validation data remains unimproved up to 50 epochs. The best-performing model was saved. The second strategy is that we set macro-F_1 score and macro-AUC of validation data as the stopping criterion for CPSC 2018 dataset and PTB-XL dataset, respectively. Different from the accuracy metric that is dominated by the classes with more samples, macro-F_1 score is an unbiased metric towards unbalanced ECG classes.

D. CLASSIFICATION PERFORMANCE

1) EVALUATION ON CPSC 2018 DATASET

Table 5 shows the classification performance of the proposed method, and makes a comparison with eight previous works in detail. To ensure a fair comparison, our models was trained using the public dataset and evaluated on the private test set of CPSC 2018 that is also used in the evaluation of these works. In order to provide a more robust prediction, We applied 10-fold cross validation to get an ensemble model, denoted as Ours[2]. In particular, the public dataset was randomly split into ten subsets, with each subset in turn as validation data and the rest ones as training data. Ten training data sets were separately preprocessed by the above preprocessing operations and then input into the proposed network architecture to obtain ten models. Finally, the prediction probabilities of these models were averaged as the final probabilities. In addition, a single model without ensemble was also evaluated, denoted as Ours[1].

Reference [45] was the first place among CPSC 2018 competitors, where their average F_1 score is 0.837. As shown in Table 5, the proposed method outperforms the existing methods in the average screening capability for 9 types of ECG. It even beats the previous first place by 1.8% average F_1 score, achieving the highest classification performance. It is observed that our model makes a superior diagnosis than other models for Normal, I-AVB, LBBB, RBBB, STD, and STE. Among them, the most significant superiority lies in the identification of Normal and STE, gaining 2.7% (0.82-0.847), 3.7% (0.64-0.677) F_1 score increase than the previous best-performing ones, respectively.

2) EVALUATION ON PTB-XL DATASET

Table 6 compares the macro-AUC score of seven reference models reported in [48] and ours under different annotation levels. Reference [39] split the whole PTB-XL dataset into ten folds, and recommended the tenth fold as the test set and the remaining nine folds as training and validation set. Following the recommendation, the tenth fold was only used when evaluating the model in the experiments of this paper and [48]. Similar to the ensemble method adopted for CPSC 2018 dataset, an ensemble model was obtained by nine-fold cross validation.

As shown in Table 6, the proposed MLBF-Net reaches the highest performance than other methods under "all", "diag.", "sub-diag.", "super-diag." and "rhythm" annotation levels, obtaining an overall classification AUC score of 0.934, 0.938, 0.943, 0.931 and 0.968, respectively. Compared to wavelet-based traditional algorithm, the macro-AUC score increases by about 8.5% in average(increases by 8.5%, 8.3%, 8.4%, 5.7%, 12.5% and 7.8% under the six annotation levels). It is noted that the performance of the feature-based classifiers is rather sensitive to the quality of extracted features. In comparison with the recurrent

TABLE 6. Comparison for classification performance of previous works and ours evaluated on the recommended test set of PTB-XL.

Model	macro-AUC score					
	all	diag.	sub-diag.	super-diag.	form	rhythm
*lstm	0.907	0.927	0.928	0.927	0.851	0.953
*inception1d	0.925	0.931	0.930	0.921	**0.899**	0.953
*lstm_bidir	0.914	0.932	0.923	0.921	0.876	0.949
*resnet1d_wang	0.919	0.936	0.928	0.930	0.880	0.946
*fcn_wang	0.918	0.926	0.927	0.925	0.869	0.931
*wavelet+NN	0.849	0.855	0.859	0.874	0.757	0.890
*xresnet1d101	0.925	0.937	0.929	0.928	0.896	0.957
Ours	**0.934**	**0.938**	**0.943**	**0.931**	0.882	**0.968**

The asterisk (*) indicates the models reported in [46].
The highest marco-AUC score for each annotation level is in bold.

TABLE 7. Classification performance comparison for single-branch and multi-branch frameworks with 12-lead ECG as input by 10-fold cross validation (mean±SD).

Network	Normal	AF	I-AVB	LBBB	RBBB	PAC	PVC	STD	STE	average F_1
BranchNet	0.809±0.018	0.914±0.017	0.872±0.022	0.882±0.043	0.929±0.019	0.739±0.060	0.872±0.027	0.808±0.033	0.513±0.101	0.815±0.018
MLBF-Net	**0.832±0.033**	**0.932±0.017**	**0.902±0.025**	**0.911±0.032**	**0.944±0.013**	**0.814±0.051**	**0.889±0.018**	**0.834±0.038**	**0.608±0.126**	**0.852±0.021**

architectures (i.e. ''lstm'' and ''lstm_bidir'') and convolutional neural network models (i.e. ''resnet1d_wang'', ''fcn_wang'' and ''xresnet1d101''), the proposed model has a best-performing classification performance in all annotation levels. And compared with ''inception1d'', our model is superior in almost all annotation levels except for form statements where our model performs satisfactorily. More description about these reference models can be seen in [39].

V. DISCUSSION

A. ABLATION STUDIES

To verify the effectiveness of main components in MLBF-Net, we also evaluated several variants of the network. The following experiments involved in the Discussion section were evaluated through 10-fold cross validation, not the private test set. Unlike the experiment conducted in the Experiment section, the public dataset was randomly divided into ten subsets, with each subset taking turns as the test set. The remaining records were further divided into training data and validation data, of which 11.12% was as validation data. Finally, the ratio of training data, validation data, and test data in each fold is 8: 1: 1.

1) SINGLE-BRANCH FRAMEWORK VS. MULTI-BRANCH FRAMEWORK

When dealing with multi-lead ECG signals, the regular frameworks are that multi-lead ECG is concatenated into an integrated matrix and then input into a feature extractor or deep neural network. They can be regarded as single-branch. Despite the simplicity of such network frameworks, the diversity of multi-lead ECG signal is neglected. To analyze the effect of multi-lead diversity, we take the BranchNet to classify 12-lead ECG signals. Table 7 shows the performances of BranchNet with a single branch against MLBF-Net with multiple branches. In terms of network architecture, BranchNet is a branch of MLBF-Net, but it should be noted that in this experiment, the input to BranchNet is 12-lead ECG signals, not single-lead ones. We can observe that multi-branch model is far superior to single-branch model for all classes in the F_1 score, gaining 3.7% increase in the average F_1 score. It is proved that lead-specific features do considerably improve the classification performance of MLBF-Net.

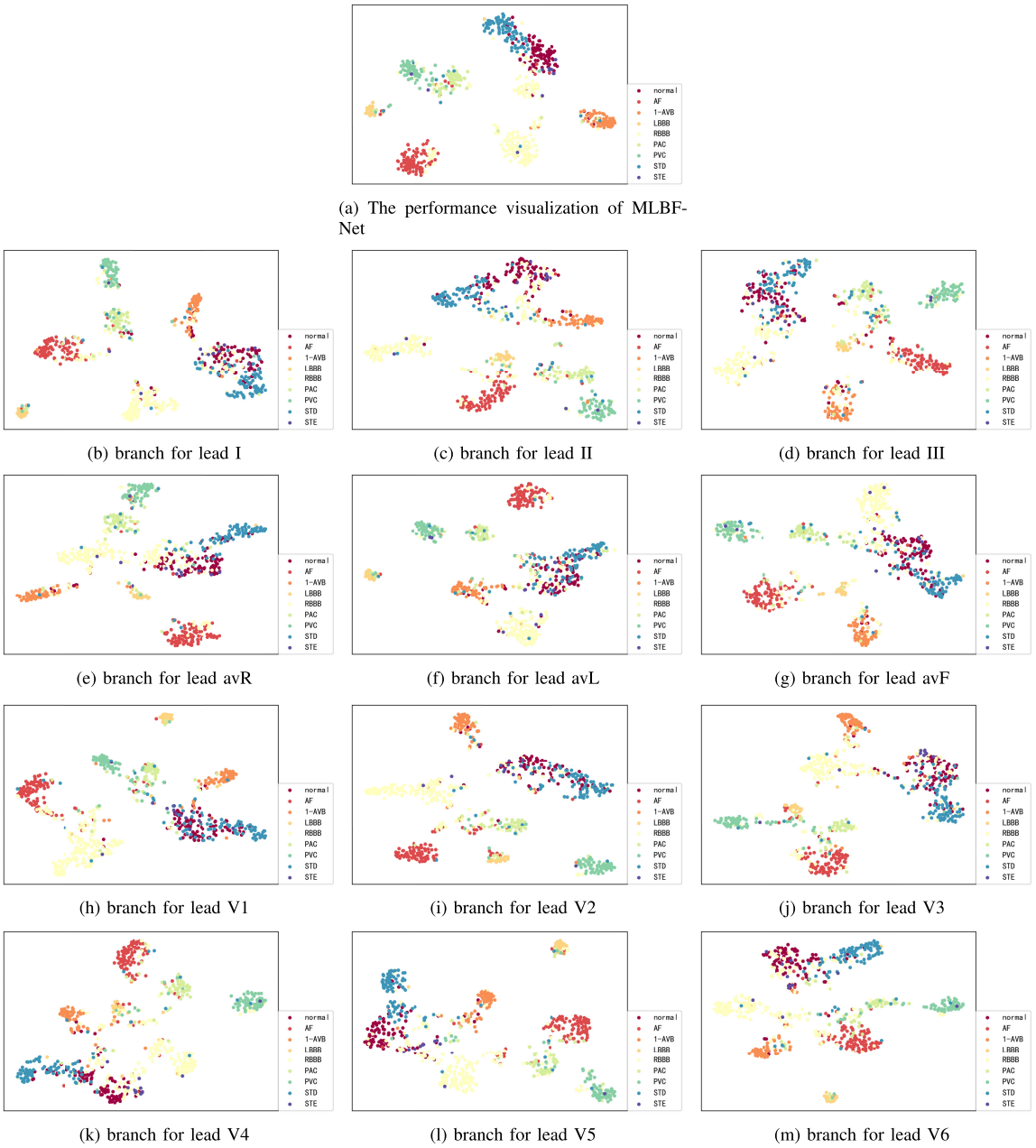
TABLE 8. Classification performance comparison for single-loss and multi-loss optimization by 10-fold cross validation (mean±SD).

Setting	average F_1
single loss	0.828±0.021
λ=0.1	0.837±0.020
λ=0.2	0.844±0.017
λ=0.5	0.849±0.017
λ=1	**0.852±0.021**
λ=2	0.843±0.014
λ=5	0.848±0.023
λ=10	0.847±0.023

2) SINGLE-LOSS OPTIMIZATION VS. MULTI-LOSS OPTIMIZATION

The hyperparameter λ governs the participation rate of lead-specific features in our model. Here, we conducted some experiments to explore the sensitiveness of this hyperparameter. The first experiment was implemented for evaluating the effectiveness of the multi-loss co-optimization strategy, in which the attention modules from all branches are removed. These attention modules are used to generate lead-specific features that correspond to all branch losses. Finally, we only keep the loss for the concatenated network. Table 8 shows the above experimental result. It is observed that multi-loss co-optimization improves the F_1 score by 2.4% (0.828-0.852). Furthermore, extra experiments were

IEEE Journal of Translational Engineering in Health and Medicine



(a) The performance visualization of MLBF-Net

(b) branch for lead I (c) branch for lead II (d) branch for lead III

(e) branch for lead avR (f) branch for lead avL (g) branch for lead avF

(h) branch for lead V1 (i) branch for lead V2 (j) branch for lead V3

(k) branch for lead V4 (l) branch for lead V5 (m) branch for lead V6

FIGURE 4. The t-SNE visualization for the performance of MLBF-Net (a) and each branch (b-m).

conducted to analyze the effect of different participate rates of lead-specific features by setting λ from 0 to 10 in our model. As we can see from Table 8, the classification performance is highest on the condition that lead-specific features with diversity and comprehensive features with integrity are equally important (λ=1). When either of lead-specific features and comprehensive features is more predominant, the classification performance is slightly worse than that of their equal participation rate. It is clear that the arrhythmia detection performance is significantly improved through multi-loss co-optimization to jointly learning diversity and integrity.

B. VISUALIZATION OF LEARNED FEATURES

The t-distributed stochastic neighbor embedding (t-SNE) [49] visualizes high dimensional data in a two or three-dimensional map. Here, we introduced t-SNE algorithm to evaluate the proposed method visually. The attention modules in each branch and the concatenated network output the 24-dimensional features. These features are as the input of t-SNE to visualize the performance of learned the comprehensive features and lead-specific features, shown in 4a and Fig. 4b-4m. As we can see from Fig. 4b-4m, the dots of different colors represent the extracted features of different types of ECG signals with varying degrees of overlap, meaning

TABLE 9. Parameter amount for the proposed model, MLBF-Net.

Layers		Parameters	
BranchNet			
Conv3_12, stride 1	× 4	(1 × 12 × 3+12) + (12 × 12 × 3 + 12) × 3	× 12
Conv3_12, stride 1		(12 × 12 × 3 + 12) × 4	
Conv24_12, stride 2		(12 × 12 × 24 + 12) × 4	
Conv3_12, stride 1	× 1	12 × 12 × 3 + 12	
Conv3_12, stride 1		12 × 12 × 3 + 12	
Conv48_12, stride 2		12 × 12 × 48 + 12	
BiGRU_12		1872	
Attention		576	
BatchNormalization		96	
Fully-connected_9		(24 × 9) + 9	
BranchNet × 12			
Attention		82944	
BatchNormalization		1152	
Fully-connected_9		(24 × 12 × 9) + 9	
Total		418005	

that only lead-specific features of a single lead are incapable of distinguishing multi-class arrhythmias well. Due to the small number of STE samples, the purple dots representing STE are difficult to observe. Excluding STE, the heaviest overlapping is in distinguishing STD from Normal signals, probably because the tiny changes in ST-segment are prone to be contaminated by various noises. In Fig. 4a, the colorful dots are separated obviously. Thus, the features learned by multi-branch with multi-loss co-optimization are discriminative for classifying 9 ECG classes. It can be observed from this figure that the yellow dots representing RBBB have two clusters, possibly because the used dataset contains both complete RBBB and incomplete RBBB in which QRS complex durations are different.

C. MODEL PARAMETER

In this subsection, the parameter amount of the proposed MLBF-Net is analyzed, shown in Table. 9. The multi-branch architecture seems to introduce a large number of parameters exponentially. In fact, our model is still lightweight in comparison with many previous deep learning-based studies for arrhythmia detection. In [25], the comparable performance to cardiologists was reported for the identification of 12 types of arrhythmias by using a 34-layer ResNet. However, the training parameters were as high as about 10.47 million. Yao *et al.* [42] evaluated their proposed method on the same independent test set as ours, and obtained a F_1 score of 0.812. In their study, the parameters were about 4.98 million. As calculated in Table. 9, only 0.42 million parameters need to be trained in our model, realizing a tens of times parameter reduction than [25] and [42].

VI. CONCLUSION

In this paper, we propose a novel end-to-end Multi-Lead-Branch Fusion Network (MLBF-Net) for ECG classification using 12-lead ECG records. MLBF-Net fully utilizes the diversity and integrity of multi-lead ECG by integrating multiple losses to optimize lead-specific features and comprehensive the 12-lead features collaboratively. We demonstrate that our MLBF-Net reaches the highest arrhythmia classification performance on China Physiological Signal Challenge 2018 which is an open 12-lead ECG dataset. In addition, compared with many existing deep neural networks, MLBF-Net is a parameter-efficient model that is less prone to overfitting, despite its multiple branches architecture. The proposed model has the advantages of both high screening capability and light weight. Therefore, it has the potential to be applied in clinical applications and daily monitoring.

REFERENCES

[1] *Cardiovascular Diseases (CVDs)*. Accessed: May 17, 2017. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[2] R. Mehra, "Global public health problem of sudden cardiac death," *J. Electrocardiol.*, vol. 40, no. 6, pp. S118–S122, Nov. 2007.

[3] W. Liu, Q. Huang, S. Chang, H. Wang, and J. He, "Multiple-feature-branch convolutional neural network for myocardial infarction diagnosis using electrocardiogram," *Biomed. Signal Process. Control*, vol. 45, pp. 22–32, Aug. 2018.

[4] Q. Li, C. Rajagopalan, and G. D. Clifford, "Ventricular fibrillation and tachycardia classification using a machine learning approach," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 6, pp. 1607–1613, Jun. 2014.

[5] S. Dilmac and M. Korurek, "ECG heart beat classification method based on modified ABC algorithm," *Appl. Soft Comput.*, vol. 36, pp. 641–655, Nov. 2015.

[6] T. Adjei, W. Von Rosenberg, V. Goverdovsky, K. Powezka, U. Jaffer, and D. P. Mandic, "Pain prediction from ECG in vascular surgery," *IEEE J. Transl. Eng. Health Med.*, vol. 5, pp. 1–10, 2017.

[7] H. H. Haseena, A. T. Mathew, and J. K. Paul, "Fuzzy clustered probabilistic and multi layered feed forward neural networks for electrocardiogram arrhythmia classification," *J. Med. Syst.*, vol. 35, no. 2, pp. 179–188, Apr. 2011.

[8] G.-M. Lin and K. Liu, "An electrocardiographic system with anthropometrics via machine learning to screen left ventricular hypertrophy among young adults," *IEEE J. Transl. Eng. Health Med.*, vol. 8, pp. 1–11, 2020.

[9] P. deChazal, 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.

[10] Y. Kutlu and D. Kuntalp, "A multi-stage automatic arrhythmia recognition and classification system," *Comput. Biol. Med.*, vol. 41, no. 1, pp. 37–45, Jan. 2011.

[11] J. L. Rodríguez-Sotelo, D. Cuesta-Frau, and G. Castellanos-Dominguez, "Unsupervised classification of atrial heartbeats using a prematurity index and wave morphology features," *Med. Biol. Eng. Comput.*, vol. 47, no. 7, pp. 731–741, Jul. 2009.

[12] Z. Zhang, J. Dong, X. Luo, K.-S. Choi, and X. Wu, "Heartbeat classification using disease-specific feature selection," *Comput. Biol. Med.*, vol. 46, pp. 79–89, Mar. 2014.

[13] Y. Özbay, "A new approach to detection of ecg arrhythmias: Complex discrete wavelet transform based complex valued artificial neural network," *J. Med. Syst.*, vol. 33, no. 6, p. 435, 2009.

[14] H. Khorrami and M. Moavenian, "A comparative study of DWT, CWT and DCT transformations in ECG arrhythmias classification," *Expert Syst. Appl.*, vol. 37, no. 8, pp. 5751–5757, Aug. 2010.

[15] M. Seera, C. P. Lim, W. S. Liew, E. Lim, and C. K. Loo, "Classification of electrocardiogram and auscultatory blood pressure signals using machine learning models," *Expert Syst. Appl.*, vol. 42, no. 7, pp. 3643–3652, May 2015.

[16] F. A. Elhaj, N. Salim, A. R. Harris, T. T. Swee, and T. Ahmed, "Arrhythmia recognition and classification using combined linear and nonlinear features of ECG signals," *Comput. Methods Programs Biomed.*, vol. 127, pp. 52–63, Apr. 2016.

[17] R. J. Martis, U. R. Acharya, H. Prasad, C. K. Chua, C. M. Lim, and J. S. Suri, "Application of higher order statistics for atrial arrhythmia classification," *Biomed. Signal Process. Control*, vol. 8, no. 6, pp. 888–900, Nov. 2013.

[18] R. J. Martis, U. R. Acharya, H. Prasad, C. K. Chua, and C. M. Lim, "Automated detection of atrial fibrillation using Bayesian paradigm," *Knowl.-Based Syst.*, vol. 54, pp. 269–275, Dec. 2013.

IEEE Journal of Translational Engineering in Health and Medicine

[19] R. J. Martis, U. R. Acharya, and L. C. Min, "ECG beat classification using PCA, LDA, ICA and discrete wavelet transform," *Biomed. Signal Process. Control*, vol. 8, no. 5, pp. 437–448, Sep. 2013.

[20] T. Ince, S. Kiranyaz, and M. Gabbouj, "A generic and robust system for automated patient-specific classification of ECG signals," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 5, pp. 1415–1426, May 2009.

[21] J.-S. Wang, W.-C. Chiang, Y.-L. Hsu, and Y.-T.-C. Yang, "ECG arrhythmia classification using a probabilistic neural network with a feature reduction method," *Neurocomputing*, vol. 116, pp. 38–45, Sep. 2013.

[22] G.-M. Lin and H. H.-S. Lu, "A 12-lead ECG-based system with physiological parameters and machine learning to identify right ventricular hypertrophy in young adults," *IEEE J. Transl. Eng. Health Med.*, vol. 8, pp. 1–10, 2020.

[23] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, Mar. 2016.

[24] M. M. Al Rahhal, Y. Bazi, M. Al Zuair, E. Othman, and B. BenJdira, "Convolutional neural networks for electrocardiogram classification," *J. Med. Biol. Eng.*, vol. 38, no. 6, pp. 1014–1025, 2018.

[25] A. Y. Hannun *et al.*, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Med.*, vol. 25, no. 1, p. 65, 2019.

[26] S. Saadatnejad, M. Oveisi, and M. Hashemi, "LSTM-based ECG classification for continuous monitoring on personal wearable devices," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 2, pp. 515–523, Feb. 2020.

[27] H. M. Lynn, S. B. Pan, and P. Kim, "A deep bidirectional GRU network model for biometric electrocardiogram classification based on recurrent neural networks," *IEEE Access*, vol. 7, pp. 145395–145405, 2019.

[28] R. He *et al.*, "Automatic cardiac arrhythmia classification using combination of deep residual network and bidirectional LSTM," *IEEE Access*, vol. 7, pp. 102119–102135, 2019.

[29] Q. Yao, X. Fan, Y. Cai, R. Wang, L. Yin, and Y. Li, "Time-incremental convolutional neural network for arrhythmia detection in varied-length electrocardiogram," in *Proc. IEEE 16th Intl Conf Dependable, Autonomic Secure Comput., 16th Int. Conf. Pervasive Intell. Comput., 4th Intl Conf Big Data Intell. Comput. Cyber Sci. Technol. Congr. (DASC/PiCom/DataCom/CyberSciTech)*, Aug. 2018, pp. 754–761.

[30] Y. Qi, S. Zhang, W. Zhang, L. Su, Q. Huang, and M. Yang, "Learning attribute-specific representations for visual tracking," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 1, 2019, pp. 8835–8842.

[31] W. Liu, F. Wang, Q. Huang, S. Chang, H. Wang, and J. He, "MFB-CBRNN: A hybrid network for MI detection using 12-lead ECGs," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 2, pp. 503–514, Feb. 2020.

[32] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[33] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, Jan. 2010, pp. 807–814.

[34] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. ICML*, vol. 30, no. 1, 2013, p. 3.

[35] K. Cho *et al.*, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1724–1734.

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

[37] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2016, pp. 1480–1489.

[38] F. Liu *et al.*, "An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection," *J. Med. Imag. Health Informat.*, vol. 8, no. 7, pp. 1368–1373, Sep. 2018.

[39] P. Wagner *et al.*, "PTB-XL, a large publicly available electrocardiography dataset," *Scientific Data*, vol. 7, no. 1, pp. 1–15, Dec. 2020.

[40] Z. Liu, X. Meng, J. Cui, Z. Huang, and J. Wu, "Automatic identification of abnormalities in 12-lead ECGs using expert features and convolutional neural networks," in *Proc. Int. Conf. Sensor Netw. Signal Process. (SNSP)*, Oct. 2018, pp. 163–167.

[41] R. Wang, Q. Yao, X. Fan, and Y. Li, "Multi-class arrhythmia detection based on neural network with multi-stage features fusion," in *Proc. IEEE Int. Conf. Syst., Man Cybern. (SMC)*, Oct. 2019, pp. 4082–4087.

[42] Q. Yao, R. Wang, X. Fan, J. Liu, and Y. Li, "Multi-class arrhythmia detection from 12-lead varied-length ECG using attention-based time-incremental convolutional neural network," *Inf. Fusion*, vol. 53, pp. 174–182, Jan. 2020.

[43] R. Wang, J. Fan, and Y. Li, "Deep multi-scale fusion neural network for multi-class arrhythmia detection," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 9, pp. 2461–2472, Sep. 2020.

[44] J. Zhang, A. Liu, M. Gao, X. Chen, X. Zhang, and X. Chen, "ECG-based multi-class arrhythmia detection using spatio-temporal attention-based convolutional recurrent neural network," *Artif. Intell. Med.*, vol. 106, Jun. 2020, Art. no. 101856. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0933365719312606

[45] T.-M. Chen, C.-H. Huang, E. S. C. Shih, Y.-F. Hu, and M.-J. Hwang, "Detection and classification of cardiac arrhythmias by a challenge-best deep learning neural network model," *iScience*, vol. 23, no. 3, Mar. 2020, Art. no. 100886.

[46] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–15.

[47] X. Fan, Q. Yao, Y. Cai, F. Miao, F. Sun, and Y. Li, "Multiscaled fusion of deep convolutional neural networks for screening atrial fibrillation from single lead short ECG recordings," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 6, pp. 1744–1753, Nov. 2018.

[48] N. Strodthoff, P. Wagner, T. Schaeffter, and W. Samek, "Deep learning for ECG analysis: Benchmarks and insights from PTB-XL," 2020, *arXiv:2004.13701*. [Online]. Available: http://arxiv.org/abs/2004.13701

[49] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.