LOW-DIMENSIONAL DENOISING EMBEDDING TRANSFORMER FOR ECG CLASSIFICATION

Jian Guan[1], Wenbo Wang[1], Pengming Feng[2], Xinxin Wang[3], and Wenwu Wang[4]

[1]Group of Intelligent Signal Processing, Harbin Engineering University, Harbin, 150001, China
[2]State Key Laboratory of Space-Ground Integrated Information Technology, Beijing, 100095, China
[3]Alibaba Group, Hangzhou, 310051, China
[4]Centre for Vision Speech and Signal Processing, University of Surrey, Guildford, GU2 7XH, UK

ABSTRACT

The transformer based model (e.g., FusingTF) has been employed recently for Electrocardiogram (ECG) signal classification. However, the high-dimensional embedding obtained via 1-D convolution and positional encoding can lead to the loss of the signal's own temporal information and a large amount of training parameters. In this paper, we propose a new method for ECG classification, called low-dimensional denoising embedding transformer (LDTF), which contains two components, i.e., low-dimensional denoising embedding (LDE) and transformer learning. In the LDE component, a low-dimensional representation of the signal is obtained in the time-frequency domain while preserving its own temporal information. And with the low-dimensional embedding, the transformer learning is then used to obtain a deeper and narrower structure with fewer training parameters than that of the FusingTF. Experiments conducted on the MIT-BIH dataset demonstrates the effectiveness and the superior performance of our proposed method, as compared with state-of-the-art methods.

Index Terms— ECG classification, Feature integration, Low-dimensional embedding, Transformer

1. INTRODUCTION

Electrocardiogram (ECG) is widely used in clinical diagnosis to provide information about cardiac disease according to its three different components: P wave, QRS complex, and T wave in different conditions [1]. The characteristics of these waveforms are of great importance for the classification of the ECG signals and the diagnosis of cardiovascular disease.

Deep learning based models [2, 3, 4, 5, 6, 7] have been used for ECG classification, giving state-of-the-art performance. According to how the information is exploited, these approaches can be divided into two categories: methods based on convolutional neural network (CNN) and methods based on sequence module. The CNN-based methods aim to exploit morphological features of ECG signals either in one dimensional time series [2, 3] or in two dimensional image [4], and are shown to outperform the traditional machine learning methods such as support vector machines (SVM) [8, 9, 10, 11]. However, these methods focus on the morphological features in the area centred around a certain peak data point or in the wavelets decomposed from original signal via DWT [3], and they are ineffective in exploiting the latent feature of the temporal signal in either time or frequency domain. In contrast, the methods with sequence module structure often employ recurrent neural network (RNN) based models, such as long-short term memory (LSTM) [5, 6] or gated recurrent unit (GRU) [7], to capture the latent temporal information of the ECG signals for classification.

Recently, the transformer model has been proposed and applied to many applications including ECG classification [12], giving state-of-the-art performance [13]. In [12], a transformer based method, i.e., FusingTF, is presented for ECG classification, where the ECG signal is represented in a high-dimensional space (i.e., 64-D) via a simple 1-D convolutional layer to expand the original dimension (i.e., 1-D) for the transformer. After this, a positional encoder is applied to incorporate additional temporal information into the 64-D result, however, this method is limited in compensating for the loss of temporal information within the signal caused by the 1-D CNN. In addition, the high-dimensional embedding will result in a model with increased number of parameters. Although the original ECG signal could be used directly as the transformer input, the presence of noise in the signal can degrade the classification performance. Representing the ECG signal in the embedding space can help denoise the signal and reduce the number of parameters in the model.

In this paper, we present a low-dimensional denoising embedding transformer model (LDTF) for ECG classification, which improves the performance with fewer training parameters by introducing a low-dimensional denoising embedding (LDE) approach for signal embedding and a deeper and narrower transformer model for signal classification. The pro-
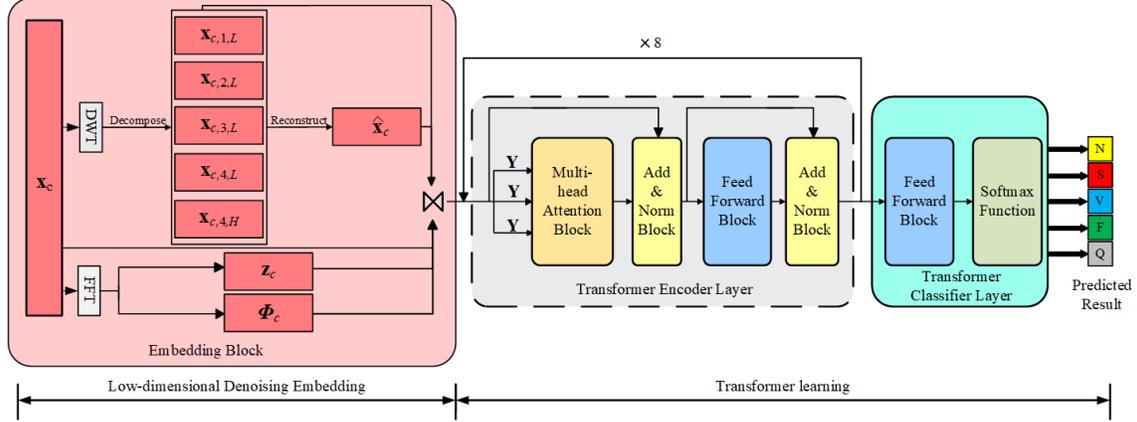
Fig. 1: Framework of the proposed LDTF model, which consists of two stages: low-dimensional denoising embedding and transformer learning. Here "⋈" denotes the concatenation operation, and c denotes the c-th channel.

posed method includes two stages, namely, low-dimensional denoising embedding and transformer learning. First, the original ECG signal is represented in a low-dimensional space by the LDE approach, which employs discrete wavelet transform (DWT) [6, 14, 15] and fast Fourier transform (FFT) to exploit the features from both time domain and frequency domain simultaneously.

The DWT of ECG signal can effectively expand the dimension of the original ECG signal and extract the primary features of the signal in time domain thereby preserving the temporal information of the signal, whereas FFT generates the features of the ECG signal in the frequency domain. Instead of mixing different kinds of features obtained above, we integrate them together by representing each feature with a specific dimension in the low-dimensional space. Such that, our LDE method can well preserve the characteristics of ECG signals. Then, with the low-dimensional embedding results for transformer learning, we can thereby use a narrower and deeper structure by stacking more transformer encoder layers to improve the classification performance with fewer training parameters in each layer. The experiments conducted on the MIT-BIH dataset [16] verify the effectiveness of our proposed model.

The remainder of the paper is organized as follows: Section 2 presents the proposed model in details; Section 3 shows the experimental results; and Section 4 summarizes the paper and gives the conclusion.

2. PROPOSED METHOD

In this section, we introduce the proposed LDTF model for ECG classification, which consists of a low-dimensional denioising embedding stage and a transformer learning stage. The framework of the proposed model is given in Figure 1. The details of the model are given next.

2.1. Low-Dimensional Denoising Embedding

In this stage, we propose an LDE method to well represent the original ECG signal in a low-dimensional space, where DWT and FFT are employed to extract the features from both time domain and frequency domain.

Here, DWT is used to decompose and reconstruct the original ECG signal to get its waveform representations in time domain, which aims to preserve the signal's own temporal information. For the purpose of a more comprehensive representation of the signal, fast Fourier transform is also adopted to extract features of the signal (i.e., magnitude and phase angle) in frequency domain. Then, we can obtain a low-dimensional embedding representation for transformer learning by concatenating the DWT decomposed wavelets, reconstructed signal, the original ECG signal and its magnitude and phase angle. As this representation already contains the information from both time and frequency domain, there is no need to use other function as in [12] to add additional temporal information to the embedding result. The details of the LDE method is given as follows.

Let $\mathbf{X} \in \mathbb{R}^{2 \times \ell}$ be the two leads ECG signal and $\mathbf{x}_c \in \mathbb{R}^{1 \times \ell}$ be one lead of the ECG signal, where $c \in \{1, 2\}$ is the channel number and ℓ is the length of the signal. Then, it can be decomposed to wavelets via DWT, as follows

$$\mathbf{x}_{c,j,H}[n] = \sum_{k=0}^{K-1} \mathbf{x}_{c,j-1,H}[2n - k]\mathbf{h}(k) \qquad (1)$$

$$\mathbf{x}_{c,j,L}[n] = \sum_{k=0}^{K-1} \mathbf{x}_{c,j-1,L}[2n - k]\mathbf{g}(k) \qquad (2)$$

where n denotes the n-th element of the signal and K is the length of slide window, $k \in \{0, \cdots, K - 1\}$. Here, $\mathbf{h}(\cdot)$ and $\mathbf{g}(\cdot)$ denote the high pass filter and low pass filter, respectively. $\mathbf{x}_{c,j,H}$ and $\mathbf{x}_{c,j,L}$ are the decomposing results at j-th level of $\mathbf{h}(\cdot)$ and $\mathbf{g}(\cdot)$ respectively, where $j \in \{1, 2, 3, 4\}$.

Then, we can obtain the denoised reconstruction signal $\hat{\mathbf{x}}_c$ with an inverse decomposition operation as follows

$$\hat{\mathbf{x}}_{c,j-1,H} = \sum_{k=0}^{K-1} \hat{\mathbf{x}}_{c,j,H}[2n-k]\mathbf{h}(k) + \sum_{k=0}^{K-1} \mathbf{x}_{c,j,L}[2n-k]\mathbf{g}(k) \tag{3}$$

when $j = 1$, we will obtain the reconstructed signal $\hat{\mathbf{x}}_c = \hat{\mathbf{x}}_{c,0,H}$.

To obtain a better representation of the signal, we also extract the information from frequency domain, where the magnitude $\mathbf{z}_c \in \mathbb{R}^{1 \times \ell}$ and the phase angle $\boldsymbol{\phi}_c \in \mathbb{R}^{1 \times \ell}$ of the signal $\mathbf{x}_c$ are used as parts of the embedding result, which are obtained respectively, as follows

$$\mathbf{z}_c = |fft(\mathbf{x}_c)| \tag{4}$$

$$\boldsymbol{\phi}_c = angle(fft(\mathbf{x}_c)) \tag{5}$$

where $fft(\cdot)$ denotes the fast Fourier transform operation. Here, $|\cdot|$ and $angle(\cdot)$ denote the calculation of absolute value and phase angle respectively.

Finally, we can obtain a low-dimensional embedding result $\mathbf{X}_{lde} \in \mathbb{R}^{d \times \ell}$ by concatenating the two leads original ECG signals, decomposed wavelets, reconstructed signals, and magnitudes and phase angles, which is expressed as follows

$$\begin{aligned} \mathbf{X}_{lde} = [&\mathbf{x}_1; \mathbf{x}_{1,1,L}; \cdots; \mathbf{x}_{1,4,L}; \mathbf{x}_{1,4,H}; \hat{\mathbf{x}}_1; \mathbf{z}_1; \boldsymbol{\phi}_1; \\ &\mathbf{x}_2; \mathbf{x}_{2,1,L}; \cdots; \mathbf{x}_{2,4,L}; \mathbf{x}_{2,4,H}; \hat{\mathbf{x}}_2; \mathbf{z}_2; \boldsymbol{\phi}_2] \end{aligned} \tag{6}$$

Here the dimension d of $\mathbf{X}_{lde}$ is 18. Because of integrating different features in different dimensions separately, the latent features of $\mathbf{X}_{lde}$ can be more easily learned by the transformer model.

2.2. Transformer Learning

Regarding transformer learning stage, thanks to the low dimensional embedding result as the input, we can obtain a narrower and deeper structure with less training parameters for ECG classification. This is because that here the space consuming mainly comes from the fully forward connected network structure (i.e., feed forward block) in each transformer encoder layer. Therefore, with the low-dimensional input, the training parameters of each transformer encoder layer will be reduced rapidly. Such that, we can stack more transformer encoder layers with fewer parameters (i.e., 9,258,742) in each layer, and fewer parameters (i.e., 74,087,228) for the whole model. Whereas the similar study in [12], i.e., FusingTF, has 32,029,382 parameters in each transformer encoder layer, and a total number of 192,237,988 parameters. In addition, this deeper and narrower structure can effectively improve the classification performance, which will be demonstrated in Section 3.2.

The transformer learning part of our model consists of 8 transformer encoder layers and one transformer classifier layer. Each transformer encoder layer includes one multi-head attention block (MAB) with 6 heads, one feed forward block (FFB) and two add & norm blocks (ANB) as shown in Figure 1.

Assuming $\mathbf{Y}$ is the input of MAB in each transformer encoder layer, which is initially set as $\mathbf{Y} = \mathbf{X}_{lde}$. Then, the linear mapping result $\mathbf{Y}\mathbf{W}_{i,q}$ of $\mathbf{Y}$ via transform matrix $\mathbf{W}_{i,q} \in \mathbb{R}^{\ell \times \ell}$ can be used to calculate the attention result $\boldsymbol{\mathcal{H}}_i \in \mathbb{R}^{d \times \ell}$ of the i-th head, where $i \in \{1, \cdots, 6\}$ and $q \in \{1, 2, 3\}$. Such that $\boldsymbol{\mathcal{H}}_i$ can be obtained as follows

$$\boldsymbol{\mathcal{H}}_i = \mathrm{softmax}\left(\frac{(\mathbf{Y}\mathbf{W}_{i,1})(\mathbf{Y}\mathbf{W}_{i,2})^\top}{\sqrt{d}}\right)(\mathbf{Y}\mathbf{W}_{i,3}) \tag{7}$$

where $\top$ denotes the transposition operation, and the softmax function is used to calculate the weights on the linear mapping result $\mathbf{Y}\mathbf{W}_{i,3}$.

To integrate the attention results of all heads, here we perform another linear mapping operation via transform matrix $\mathbf{W}_o \in \mathbb{R}^{6\ell \times \ell}$, then the output of MAB can be obtained as follows

$$\boldsymbol{\mathcal{O}}_{MAB} = [\boldsymbol{\mathcal{H}}_1, \cdots, \boldsymbol{\mathcal{H}}_i \cdots, \boldsymbol{\mathcal{H}}_6]\mathbf{W}_o \tag{8}$$

After that, a residual connection and layer norm operation are conducted in ANB to connect the FFB, here the FFB consists of two fully connected layers with a rectified linear unit (ReLU). Then, with an ANB again after the FFB, we can obtain the output of the transformer encoder layer, which will be used as the input of the next transformer encoder layer or the transformer classifier layer that contains one fully connected layer and a softmax function. Finally, we will get the ECG classification results after the transformer learning.

3. EXPERIMENTS AND RESULTS

3.1. Experimental Setup

Dataset We evaluate the proposed model on MIT-BIH, which are the two-channel recording ECG signals sampled from 48 different patients at a 360Hz sampling frequency for 30 minutes with 23 types. In our work, 5 essential arrhythmia groups from MIT-BIH are employed for evaluation, which are described in Table 1. These heartbeats classes are categorized by the American Association of Medical Instrumentation (AAMI). Here, 112,551 ECG segments are selected for experimental evaluation, where each segment is sampled with 241 sampling points centred at R peak point, which means the length of the signal is set as $\ell = 241$ in our experiments.

To alleviate the categories unbalanced problem and remove the amplitude scaling problem, synthetic minority oversampling technique (SMOTE) and Z-score normalization are adopted for data preprocessing respectively [17]. Then, we select 80% ECG segments from each class as training set, and the remaining 20% ECG segments from each class as test set.

Table 1: ECG signal classification standard specified by ANSI/AAMI EC57 and the number of each class in our dataset

Groups	ECG classes	Number
N	Normal(N)	75,017
	Left Bundle Brunch Block(L)	8,071
	Right Bundle Brunch Block(R)	7,255
	Atrial Escape (e)	16
	Nodal (junctional) escape(j)	229
S	Atrial Premature(A)	2,546
	Aberrant Atrial Premature(a)	150
	Nodal (Junctional) Premature(J)	83
	Supra-ventricular Premature(S)	2
V	Premature Ventricular Contraction(V)	7,129
	Ventricular escape(E)	106
F	Fusion of Ventricular and Normal(F)	802
Q	Paced(/)	7,023
	Fusion of Paced and Normal(f)	982
	Unclassifiable(Q)	33

Table 2: Hyperparameters settings for model training

Selected hyperparameters
Loss function = Cross-Entropy, Optimizer = SGD, Learning rate = 0.001, Batchsize = 64, Dropout ratio = 10% ECG segment length = 241, Head number = 6

The number of each category is also given in Table 1. The hyperparameters for our model training is given in Table 2.

Performance Metrics We use the performance metrics in terms of Recall (Rec) and Precision (Pre) to evaluate the ECG classification performance of the proposed model, as follows

$$\text{Rec} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad (9)$$

$$\text{Pre} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad (10)$$

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

3.2. Experimental Results

Ablation Study To show the effectiveness of both LDE and the deeper and narrower structure, an ablation study is conducted for a four categories ECG classification task, which is the same as FusingTF in [12] for the purpose of fair comparison. The baseline is FusingTF with 6 transformer encoder layers, and its embedding stage consists of a 1-D CNN layer to embed the original ECG signal into 64-D space and a positional encoding operation via a sinusoidal function.

Here, we simply modify FusingTF with 8 transformer encoder layers, and use the 1-D CNN layer to embed the signal

Table 3: Validation of the LDE method and the deeper and narrower structure of our model

Category	Rec(%)			Pre(%)		
	LDTF	FTFD	FusingTF	LDTF	FTFD	FusingTF
N	**95.77**	94.40	93.20	**96.81**	93.60	96.30
S	**97.65**	96.20	93.39	**97.58**	96.87	96.72
V	**97.72**	97.03	93.45	**97.38**	95.51	97.43
F	**99.01**	95.83	95.12	**98.38**	97.55	96.33
Average	**97.54**	95.87	93.79	**97.54**	95.88	96.70

into an 18-D space, such that we can obtain a deeper and narrower structure, i.e., FusingTF-Deep (FTFD). Compared to LDTF, the only difference is that FTFD used same embedding method (i.e., 1-D CNN and a positional encoder) as FusingTF for embedding. The experimental results are given in Table 3. From Table 3, we can see that FTFD outperforms FusingTF, which demonstrates the validity of the deeper and narrower structure. The proposed LDTF can achieve the best performance, as compared with FTFD and the baseline FusingTF, which also verifies the effectiveness of the LDE method.

Table 4: Performance comparison in terms of average Recall and average Precision

Method	Average Recall(%)	Average Precision(%)
LDTF	**98.39**	98.41
WCNN	93.55	96.72
MWFE	96.86	**98.92**
Ad-CNN	93.90	98.90

Performance Comparison We also compare our proposed method with other state-of-the-art methods (i.e., MWFE [15], Ad-CNN [18] and WCNN [3]) for five categories ECG classification task. All these three methods use DWT for data preprocessing. MWFE and Ad-CNN only apply DWT for denoising, whereas WCNN uses the decomposed wavelets obtained by DWT for feature learning. The results are given in Table 4. From Table 4, we can see our proposed LDTF can achieve the best performance in terms of average Recall, as compared with other state-of-the-art methods. Although, the average precision of LDTF is slightly lower than MWFE, LDTF improves the overall performance.

4. CONCLUSION

In this paper, we presented a low-dimensional denoising embedding transformer model for ECG classification, where a low-dimensional embedding method is proposed for the well representation and embedding the signal into low-dimensional space for transformer learning. The proposed method can not only significantly reduce the training parameters, but also improve the overall performance for ECG classification. Experimental results demonstrated the effectiveness and the superiority of the our method, as compared with other state-of-the-art methods.

5. REFERENCES

[1] P. J. Zimetbaum and M. E. Josephson, "Use of the electrocardiogram in acute myocardial infarction," *New England Journal of Medicine*, vol. 348, no. 10, pp. 933–940, 2003.

[2] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych, and R. San Tan, "A deep convolutional neural network model to classify heartbeats," *Computers in Biology and Medicine*, vol. 89, pp. 389–396, 2017.

[3] L. El Bouny, M. Khalil, and A. Adib, "ECG heartbeat classification based on multi-scale wavelet convolutional neural networks," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 3212–3216.

[4] J. Li, Y. Si, T. Xu, and S. Jiang, "Deep convolutional neural network based ECG classification system using information fusion and one-hot encoding techniques," *Mathematical Problems in Engineering*, vol. 2018, 2018.

[5] Z. Al Nazi, A. Biswas, M. A. Rayhan, and T. A. Abir, "Classification of ECG signals by dot residual LSTM network with data augmentation for anomaly detection," in *Proceedings of the International Conference on Computer and Information Technology (ICCIT)*. IEEE, 2019, pp. 1–5.

[6] Ö. Yildirim, "A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification," *Computers in Biology and Medicine*, vol. 96, pp. 189–202, 2018.

[7] 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. 145 395–145 405, 2019.

[8] F. Y. Abdalla, L. Wu, H. Ullah, H. Mkindu, Y. Nie, and Y. Zhao, "ECG arrhythmia discrimination using SVM and nonlinear and non-stationary decomposition," in *Proceedings of the International Symposium on Signal Processing and Information Technology (ISSPIT)*. IEEE, 2019, pp. 1–6.

[9] S. El-Soudy, A. El-Sayed, A. Khalil, I. Khalil, T. E. Taha, A. El-Samie *et al.*, "An efficient method of ECG beats feature extraction/classification with multiclass SVM error correcting output codes," *Menoufia Journal of Electronic Engineering Research*, vol. 28, no. 2, pp. 65–78, 2019.

[10] Y. Zhang, S. Wei, L. Zhang, and C. Liu, "Comparing the performance of random forest, SVM and their variants for ECG quality assessment combined with nonlinear features," *Journal of Medical and Biological Engineering*, vol. 39, no. 3, pp. 381–392, 2019.

[11] Z. Ge, Z. Zhu, P. Feng, S. Zhang, J. Wang, and B. Zhou, "ECG-signal classification using SVM with multi-feature," in *Proceedings of the International Symposium on Next Generation Electronics (ISNE)*. IEEE, 2019, pp. 1–3.

[12] G. Yan, S. Liang, Y. Zhang, and F. Liu, "Fusing transformer model with temporal features for ECG heartbeat classification," in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2019, pp. 898–905.

[13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proceedings of Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.

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

[15] S. Sahoo, B. Kanungo, S. Behera, and S. Sabut, "Multiresolution wavelet transform based feature extraction and ECG classification to detect cardiac abnormalities," *Measurement*, vol. 108, pp. 55–66, 2017.

[16] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[17] S. K. Pandey and R. R. Janghel, "Automatic detection of arrhythmia from imbalanced ECG database using cnn model with smote," *Australasian Physical & Engineering Sciences in Medicine*, vol. 42, no. 4, pp. 1129–1139, 2019.

[18] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 664–675, 2015.