MS-MLP: Multi-scale Sampling MLP for ECG Classification

Wenbo Wang[1], Jian Guan[1*], Xinyi Che[1], and Wenwu Wang[2]
[1]Group of Intelligent Signal Processing, Harbin Engineering University, Harbin, 150001, China
[2]Centre for Vision Speech and Signal Processing, University of Surrey, Guildford, GU2 7XH, UK

***Abstract*—Transformer-based models (i.e., Fusing-TF and LDTF) have achieved state-of-the-art performance for electrocardiogram (ECG) classification. However, these models may suffer from low training efficiency due to the high model complexity associated with the attention mechanism. In this paper, we present a multi-layer perceptron (MLP) model for ECG classification by incorporating a multi-scale sampling strategy for signal embedding, namely, MS-MLP. In this method, a novel multi-scale sampling strategy is first proposed to exploit the multi-scale characteristics while maintaining the temporal information in the corresponding dimensions. Then, an MLP-Mixer structure with token-mixer and channel-mixer is employed to capture the multi-scale feature and temporal feature from the multi-scale embedding result, respectively. Because of the mixing operation and attention-free MLP structure, our proposed MS-MLP method not only provides better classification performance, but also has a lower model complexity, as compared with transformer-based methods, in terms of experiments performed on the MIT-BIH dataset.**

***Index Terms*—ECG classification, Multi-scale embedding, Transformer, MLP-Mixer**

I. Introduction

The aim of electrocardiogram (ECG) signal classification is to classify the ECG signals according to their components (e.g., P wave, QRS complex and T wave) [1], and to provide assistance for the diagnosis of cardiovascular disease. Recently, deep learning technique has been successfully applied for ECG classification, achieving state-of-the-art performance [2]–[8]. ECG classification models based on deep learning mainly include two stages, namely, signal embedding and model learning, where the signal embedding stage aims to embed the original ECG signal into a latent space with a certain dimension that can facilitate the learning of the deep model, and the model learning stage is to train the deep model to classify the signal based on the latent feature. According to the methods used for signal embedding, existing deep learning methods for ECG classification can be divided into two categories: models using original ECG signal as the embedding feature [9]–[11], and models using preliminary feature of the original ECG signal as the embedding feature [3], [7], [8]. For the first type, the original ECG signal is converted to an image as the embedding, and then a deep model, such as convolutional neural network (CNN), is used to capture morphological features from the area centered at a certain peak data point in the model learning stage [9]–[11]. However, without feature extraction, the noise contained in the embedding often limits the learning ability of the deep learning model, and degrades its classification performance.

As for the second category, deep learning methods (e.g., 1-D CNN) [7] or traditional signal processing methods (e.g., discrete wavelet transform, DWT) [3], [8] are often employed to design the preliminary features as the embedding result, which helps the deep models learn more effective features and achieve better classification performance. For example, in [7], a transformer-based method, i.e., Fusing-TF, adopts a 1-D CNN layer and a positional encoding function to obtain a 64-D embedding result with additional temporal information. However, the use of 1-D CNN for signal embedding may lead to the loss of temporal information from the original signal. To address this limitation, another transformer-based method LDTF [8] is proposed, in which a low-dimensional denoising embedding (LDE) method is introduced to embed the signal into a low-dimensional space by using DWT and fast Fourier transform (FFT) to extract the temporal-spectral feature simultaneously. Therefore, the LDTF achieves better performance with fewer parameters than Fusing-TF. Although these transformer-based methods have achieved state-of-the-art performance for ECG classification, both Fusing-TF and LDTF suffer from low training efficiency due to the complex transformer structure and the use of self-attention mechanism [12].

In this paper, we propose a novel multi-scale sampling MLP architecture, namely MS-MLP, for ECG classification, which can improve the classification performance with high training efficiency and low model complexity. The proposed method includes two stages: multi-scale sampling based embedding (MSE) and MLP learning. First, a novel multi-scale sampling strategy is proposed to map the original ECG signal to a low-dimensional latent space. Here, the multi-scale sampling strategy can not only exploit the multi-scale characteristics of the original ECG signal, but also maintain the temporal information with respect to these multi-scale characteristics in the corresponding embedded dimensions.Then, in the MLP learning stage, an MLP-Mixer structure [13] with token-mixer and channel-mixer is employed to capture the multi-scale feature and temporal feature from the multi-scale embedding result, respectively. Because of the mixing operation, we can achieve effective feature learning, thus improve the classification performance. In addition, as the mixers in MLP-Mixer are mainly realized by fully connected neural network

*Corresponding Author

ISBN: 978-1-6654-6798-8

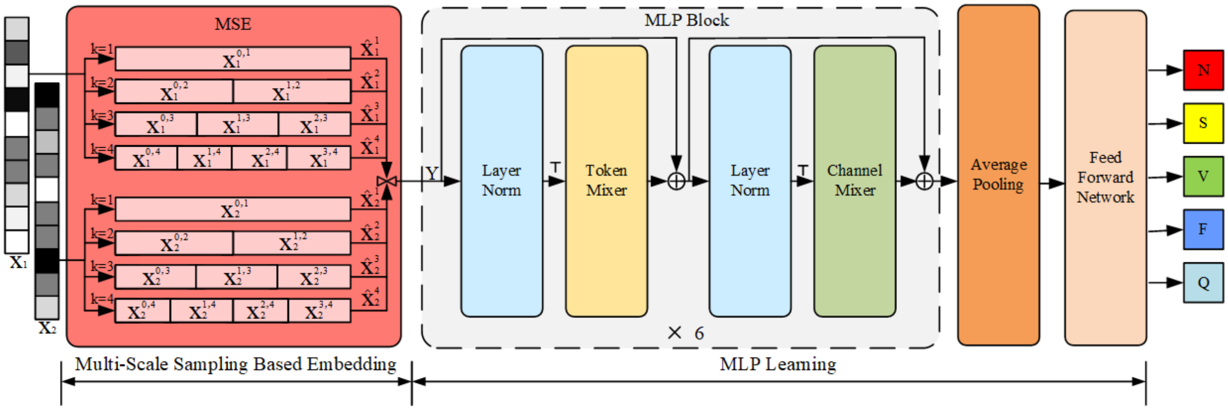

MSE
k=1
k=2
k=3
k=4
$X_1^{0,1}$
$X_1^{0,2}$ $X_1^{1,2}$
$X_1^{0,3}$ $X_1^{1,3}$ $X_1^{2,3}$
$X_1^{0,4}$ $X_1^{1,4}$ $X_1^{2,4}$ $X_1^{3,4}$
$\hat{X}_1^1$ $\hat{X}_1^2$ $\hat{X}_1^3$ $\hat{X}_1^4$
$X_2^{0,1}$
$X_2^{0,2}$ $X_2^{1,2}$
$X_2^{0,3}$ $X_2^{1,3}$ $X_2^{2,3}$
$X_2^{0,4}$ $X_2^{1,4}$ $X_2^{2,4}$ $X_2^{3,4}$
$\hat{X}_2^1$ $\hat{X}_2^2$ $\hat{X}_2^3$ $\hat{X}_2^4$
X_1
X_2
Y
MLP Block
Layer Norm
T
Token Mixer
Layer Norm
T
Channel Mixer
× 6
Average Pooling
Feed Forward Network
N
S
V
F
Q
Multi-Scale Sampling Based Embedding
MLP Learning

Fig. 1: Framework of the proposed MS-MLP model, which consists of two stages: multi-scale sampling based embedding and MLP learning. Here, "⊤" and "⋈" denote transposition operation and concatenation operation, respectively. "⊕" is the summation operation.

(FCN) without using the complex structure (e.g., encoder-decoder) and the attention mechanism as in transformer, our proposed MS-MLP can achieve better training efficiency with lower complexity. The experiments performed on the MIT-BIH dataset [14] demonstrate the effectiveness of our proposed method.

Our contributions can be summarized as follows:

- We propose a novel multi-scale sampling based embedding method to obtain a lower dimensional feature than LDTF and Fusing-TF, while preserving the temporal information in the corresponding dimension in the embedding, which is often lost in other embedding methods, e.g., 1-D CNN in Fusing-TF. The proposed embedding can be also applied in other signal processing tasks.
- We introduce the MLP-Mixer for classification, where token mixing is performed to exploit the multi-scale features with different frequency bands from inter-dimensions of the embedding, and channel-mixer is performed to exploit the temporal features from intra-dimension of the embedding.
- Thanks to the attention-free mechanism, our proposed MS-MLP can improve the training efficiency with 30.41 FLOPs and around 25 epochs, outperforming the-state-of-the-art methods, i.e., Fusing-TF (482.56 FLOPs) and LDTF (85.06 FLOPs) with more than 100 epochs.

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

II. PROPOSED METHOD

In this section, we introduce our proposed MS-MLP method, which consists of a multi-scale sampling based embedding stage and an MLP learning stage. The overall framework of the MS-MLP architecture is illustrated in Fig. 1, and the details are given next.

A. *Multi-Scale Sampling Based Embedding*

In [15], a multi-resolution CNN structure with different kernel sizes is presented to extract features corresponding to different frequency bands of the electroencephalogram (EEG) signal, and obtain effective multi-scale features for sleep stage classification. Inspired by this idea, we propose a multi-scale sampling strategy for ECG signal embedding, which can extract the characteristics from the original signal with different sampling intervals corresponding to different embedding dimensions, such that we can achieve multi-scale embedding while preserving the temporal information of the signal embedded in each dimension. This is because the sampling interval used provides an indication of the sampling rate used for data collection.

Let $\mathbf{X} \in \mathbb{R}^{2\times\ell}$ be the two-channel input ECG signal with the length ℓ, and $\mathbf{x}_i \in \mathbb{R}^{1\times\ell}$ be the i-th channel ECG signal, $i \in \{1, 2\}$. Then, $\mathbf{x}_i$ can be expressed as

$$\mathbf{x}_i = [x_{i,0}, x_{i,1}, \cdots, x_{i,j}, \cdots, x_{i,\ell-1}] \quad (1)$$

where $x_{i,j}$ represents the j-th data point of $\mathbf{x}_i$, and $j \in \{0, 1, \cdots, \ell-1\}$.

To achieve multi-scale embedding, we define sampling interval as k with different values that correspond to different embedded dimensions. This means, for each dimension, we use a different sampling frequency to obtain data points from the ECG signal.

Then, the sampling points with an interval of k are successively selected from the original ECG signal to form a series of new sub-sampled signals $\mathbf{x}_i^{m,k}$, $m \in \{0, 1, \cdots, k-1\}$. Here, $\mathbf{x}_i^{m,k}$ represents the signal sub-sampled from the m-th sampling point of $\mathbf{x}_i$ for sampling interval k, denoted as follows

$$\mathbf{x}_i^{m,k} = [x_{i,m}, x_{i,m+k}, \cdots, x_{i,m+(n-1)k}, x_{i,m+nk}] \quad (2)$$

where $m + nk \leq \ell - 1$, and n is a natural number.

As a result, we can obtain the embedding result $\hat{\mathbf{x}}_i^k \in \mathbb{R}^{1\times\ell}$ with sampling interval k for one specific embedding dimension, as follows

$$\hat{\mathbf{x}}_i^k = [\mathbf{x}_i^{0,k}, \mathbf{x}_i^{1,k}, \cdots, \mathbf{x}_i^{m,k}, \cdots, \mathbf{x}_i^{k-1,k}] \tag{3}$$

Note that, the sampling interval k is different for each embedding dimension. Here, k with four interval values (i.e., $k \in \{1, 2, 3, 4\}$) are used for sampling the embedding corresponding to each dimension, respectively.

Therefore, for the original two-lead ECG signal $\mathbf{X}$, we can obtain a more effective multi-scale representation $\mathbf{X}_{mse} \in \mathbb{R}^{8\times\ell}$ by concatenating the embedding vector of each dimension, as follows

$$\mathbf{X}_{mse} = [\hat{\mathbf{x}}_1^1; \hat{\mathbf{x}}_2^1; \hat{\mathbf{x}}_1^2; \hat{\mathbf{x}}_2^2; \hat{\mathbf{x}}_1^3; \hat{\mathbf{x}}_2^3, \hat{\mathbf{x}}_1^4; \hat{\mathbf{x}}_2^4] \tag{4}$$

B. MLP Learning

With the multi-scale embedding, we can use an MLP architecture [13] with token-mixer and channel-mixer which allow more effective latent feature to be exploited from inter-dimension (i.e., mixing multi-scale features of different dimensions) and intra-dimension (i.e., mixing temporal information within the same dimension) due to the mixing operation. In addition, with the simple attention-free MLP structure, our proposed MS-MLP method achieves efficient model training with low complexity, which will be illustrated in Section III.

The MLP learning part of our model consists of 6 MLP blocks, an average pooling layer and a fully connected layer. Here, each MLP block contains two mixers, i.e., token mixer and channel mixer, as shown in Fig. 1.

Denote $\mathbf{Y}^{h-1}$ as the input of the h-th MLP block, which is initially set as $\mathbf{Y}^0 = \mathbf{X}_{mse}$, $h \in \{1, 2, \cdots, 6\}$, and $\mathbf{Y}^{h-1}_{*,s}$ as the s-th column of $\mathbf{Y}^{h-1}$, with $s \in \{1, 2, \cdots, \ell\}$, where "$*, s$" means the s-th column is selected. Then, to exploit the multi-scale features from inter-dimensions, a mixing operation via token-mixer is conducted on each column of $\mathbf{Y}^{h-1}$ with shared transform matrices $\mathbf{W}_1^{h-1}$, and $\mathbf{W}_2^{h-1}$ as follows

$$\mathbf{U}^h_{*,s} = \mathbf{Y}^{h-1}_{*,s} + \mathbf{W}_2^{h-1} \cdot \sigma(\mathbf{W}_1^{h-1} \cdot \mathcal{N}(\mathbf{Y}^{h-1}_{*,s})) \tag{5}$$

where $\mathbf{U}^h_{*,s}$ is the s-th column of $\mathbf{U}^h$, $\sigma(\cdot)$ and $\mathcal{N}(\cdot)$ are the GELU function and layer normalization function, respectively.

After that, for the purpose of learning temporal feature from intra-dimension, another mixing operation via channel-mixer is performed on each row of $\mathbf{U}^h$ with shared transform matrices $\mathbf{W}_3^{h-1}$ and $\mathbf{W}_4^{h-1}$ as follows

$$\mathbf{Y}^h_{q,*} = \mathbf{U}^h_{q,*} + \mathbf{W}_3^{h-1} \sigma(\mathbf{W}_4^{h-1} \cdot \mathcal{N}(\mathbf{U}^h_{q,*})) \tag{6}$$

where $\mathbf{U}^h_{q,*}$ and $\mathbf{Y}^h_{q,*}$ denote the q-th row of $\mathbf{U}^h$ and $\mathbf{Y}^h$, respectively, and "$q, *$" means the q-th row is selected, with $q \in \{1, 2, \cdots, 8\}$.

Finally, we can get the output of MLP blocks $\hat{\mathbf{Y}} = \mathbf{Y}^6$, and obtain the classification result via an average pooling layer and a fully connected layer, expressed as follows

$$\mathbf{z} = \mathrm{softmax}(\mathbf{W} \cdot \mathrm{AvgPooling}(\hat{\mathbf{Y}})) \tag{7}$$

where $\mathbf{W}$ denotes the weight matrix of the fully connected layer, $\mathrm{softmax}(\cdot)$ and $\mathrm{AvgPooling}(\cdot)$ represent the softmax function and average pooling operation, respectively.

III. EXPERIMENTS AND RESULTS

A. Experimental Setup

Dataset We evaluate our method on the MIT-BIH dataset [14], which consists of two-channel ECG recordings from 48 patients, sampled at 360 Hz. In our experiments, 5 essential arrhythmia groups (i.e., N, S, V, F, Q) from MIT-BIH are employed for evaluation, which are specified by the American Association of Medical Instrumentation (AAMI) standard as shown in Table I.

TABLE I: ECG signal classification standard specified by ANSI/AAMI EC57 and the number of samples for 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

In our study, 80% ECG segments from each class are used as the training set, and the remaining 20% are used as the test set, following [3], [16], [17]. For each ECG segment, 300 data points are sampled centred at the R peak point [18], i.e. the largest sampling point in one heart beat cycle, which indicates depolarization of the main mass of the ventricles.

Implementation Details For data pre-processing, the synthetic minority over-sampling technique (SMOTE) algorithm and Z-score normalization are adopted to solve the class imbalance and amplitude scaling problems, respectively [19]. The hyper-parameters settings are provided in Table II.

TABLE II: Setting of hyper-parameters for model training

Selected hyperparameters
Loss function = Cross-Entropy, Optimizer = SGD,
Learning rate = 0.001, Batch size = 64
ECG segment length = 300, MLP-Mixer block number = 6

Performance Metrics The performance metrics in terms of Recall (Rec), Precision (Pre) and Accuracy (Acc) are employed to evaluate the classification performance of our method, which are calculated as

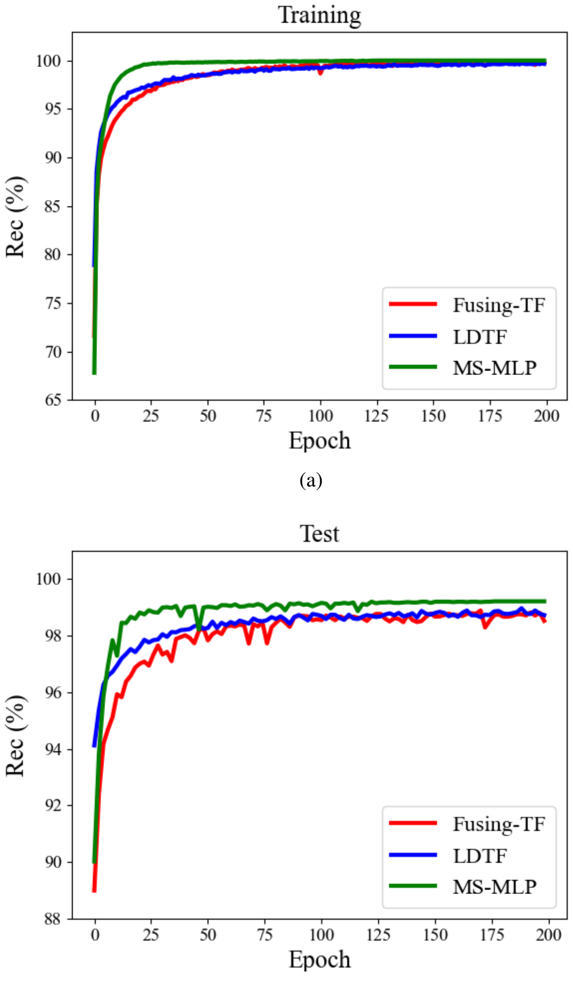
$$\mathrm{Rec} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{8}$$

$$\mathrm{Pre} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{9}$$

$$\mathrm{Acc} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN}} \tag{10}$$

TABLE III: Comparison of our proposed method with baseline methods in classification performance and model complexity.

Model	FLOPs (M)	Rec (%)					Pre (%)					Acc (%)				
		N	S	V	F	Q	N	S	V	F	Q	N	S	V	F	Q
WCNN	-	**99.66**	87.68	98.05	82.76	99.58	**99.44**	93.35	97.11	94.12	99.58	99.25	99.54	**99.68**	99.82	**99.94**
Fusing-TF	482.56	98.47	98.14	97.72	99.73	99.44	97.82	99.10	98.61	98.30	**99.75**	99.26	99.47	99.29	99.60	99.83
LDTF	85.06	98.00	98.49	97.86	99.93	99.50	98.40	99.10	98.40	98.24	99.69	99.28	99.53	99.28	99.63	99.83
MS-TF	85.06	92.95	93.19	95.44	95.22	98.45	93.26	94.82	92.37	95.72	99.06	97.25	97.70	97.61	98.19	99.47
MS-MLP	**30.41**	**98.60**	**99.17**	**98.55**	**100.00**	**99.63**	98.73	**99.17**	**99.10**	**99.27**	99.69	**99.47**	**99.68**	99.55	**99.85**	99.85



(a)

(b)

Fig. 2: Recall curves comparison between our proposed MS-MLP and the transformer-based models on the training and test set, respectively.

Here, TP and TN denote the true positive and the true negative, respectively. FP and FN denote the false positive and the false negative, respectively. In addition, we employ FLOPs as the metric for the evaluation of model complexity.

B. Experimental Result

Performance Comparison We compare our proposed MS-MLP with other state-of-the-art methods (i.e., WCNN [3], Fusing-TF [7] and LDTF [8]). Here, WCNN is the method exploiting multi-scale features, whereas Fusing-TF and LDTF are the two state-of-the-art transformer-based methods. In addition, to show the effectiveness of the proposed multi-scale embedding strategy, we introduce a multi-scale sampling based transformer method (i.e., MS-TF) for performance evaluation, by replacing the LDE of the LDTF model with our MSE for signal embedding. The results are given in Table III.

From Table III, we can see that our proposed MS-MLP outperforms all the baseline methods in terms of the overall classification performance. In addition, the comparison between MS-MLP and MS-TF verifies the effectiveness of the proposed multi-scale embedding strategy for mixing operation of MLP, thus providing better classification performance. Regarding model complexity, we can see that the proposed MS-MLP has the lowest model complexity as compared with the transformer-based methods. Note that, the reason why LDTF and MS-TF have the same model FLOPs is that we just simply replace the LDE with MSE, and maintain the same embedding dimension and transformer structure.

Convergence Analysis To show the efficiency of our proposed method, we conduct experiments for convergence analysis, where the transformer-based methods LDTF and Fusing-TF are employed for comparison. The recall curves of these methods for model training and testing are given in Fig. 2.

As can be seen from Fig. 2, our proposed MS-MLP can converge rapidly and achieve stable Recall performance after 25 epochs without over-fitting problem, whereas the transformer-based methods (i.e., LDTF and FusingTF) both require more than 100 epochs. The results in Fig. 2 and Table III show that our proposed MS-MLP can achieve better ECG classification performance with higher training efficiency and lower model complexity as compared with the transformer-based methods, which verified the effectiveness of the proposed attention-free MLP model with the mixing operation for effective multi-scale feature learning.

IV. CONCLUSION

In this paper, we have presented a new method for ECG classification, where a novel multi-scale sampling strategy was proposed to exploit the multi-scale characteristics with corresponding temporal information for signal embedding, and an MLP-Mixer structure was employed to learn the latent feature. The proposed method can not only achieve better ECG classification performance, but also improve training efficiency with lower model complexity. Experimental results demonstrated the effectiveness and the superiority of our method, as compared with the state-of-the-art methods.

Acknowledgement

This work was partly supported by the Fundamental Research Funds for the Central Universities under Grant No. 3072022CFJ0601, and the Natural Science Foundation of Heilongjiang Province under Grant No. YQ2020F010. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

References

[1] G. D. Clifford, F. Azuaje, P. McSharry *et al.*, *Advanced Methods and Tools for ECG Data Analysis*. Boston: Artech house, 2006, vol. 10.

[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] 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.

[5] 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.

[6] 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.

[7] 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.

[8] J. Guan, W. Wang, P. Feng, X. Wang, and W. Wang, "Low-dimensional denoising embedding transformer for ECG classification," in *Proceddings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 1285–1289.

[9] A. Isin and S. Ozdalili, "Cardiac arrhythmia detection using deep learning," *Procedia Computer Science*, vol. 120, pp. 268–275, 2017.

[10] 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.

[11] L. Fu, B. Lu, B. Nie, Z. Peng, H. Liu, and X. Pi, "Hybrid network with attention mechanism for detection and location of myocardial infarction based on 12-lead electrocardiogram signals," *Sensors*, vol. 20, no. 4, p. 1020, 2020.

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

[13] I. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai, T. Unterthiner, J. Yung, D. Keysers, J. Uszkoreit, M. Lucic *et al.*, "MLP-Mixer: An all-MLP architecture for vision," *arXiv preprint arXiv:2105.01601*, 2021.

[14] 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.

[15] E. Eldele, Z. Chen, C. Liu, M. Wu, C.-K. Kwoh, X. Li, and C. Guan, "An attention-based deep learning approach for sleep stage classification with single-channel EEG," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 809–818, 2021.

[16] K. Jiang, S. Liang, L. Meng, Y. Zhang, P. Wang, and W. Wang, "A two-level attention-based sequence-to-sequence model for accurate inter-patient arrhythmia detection," in *Proceedings of IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2020, pp. 1029–1033.

[17] Y. Wang, L. Sun, and S. Subramani, "CAB: Classifying arrhythmias based on imbalanced sensor data," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 15, no. 7, pp. 2304–2320, 2021.

[18] S. Mousavi and F. Afghah, "Inter-and intra-patient ECG heartbeat classification for arrhythmia detection: a sequence to sequence deep learning approach," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 1308–1312.

[19] 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.