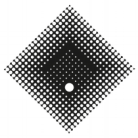


An Improved ConvNeXt with Multimodal Transformer for Physiological Signal Classification

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

An Improved ConvNeXt with Multimodal Transformer for Physiological Signal Classification

Jiajian Zhu[1], Yue Feng[1], Qichao Liu[1], Hong Xu[1,2], Yuan Miao[2], Zhuosheng Lin[1], Jia Li[1], Huilin Liu[3], Ying Xu[4], and Fufeng Li[4]

[1.]Faculty of Intelligent Manufacturing, Wuyi University, Jiangmen 529020, Guangdong, China;
[2.]Victoria University, Melbourne 8001, Australia;
[3.]Basic Medical College, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China
[4.]Laboratory of TCM Four Processing, Shanghai University of TCM, Shanghai 201203, China.

Corresponding author: Yue Feng (e-mail: J002443@wyu.edu.cn) and Fufeng Li (e-mail: li_fufeng@aliyun.com)

This work was supported by the Basic Research and Applied Basic Research Key Project in General Colleges and Universities of Guangdong Province, China (2021ZDZX1032); the Special Project of Guangdong Province, P. R. China (2020A1313030021); and the Scientific Research Project of Wuyi University, China (2018TP023, 2018GR003).

ABSTRACT With escalating mortality rates associated with cardiovascular disease, the early detection of arrhythmias assumes ever-increasing significance. This study introduces a novel multimodal network that concurrently classifies electrocardiogram (ECG) and wrist pulse signal (WPS). Both ECG and WPS, as human physiological signals, share closely related distributions and characteristics, holding potential to accurately reflect underlying cardiovascular conditions. The proposed ICMT-Net utilizes continuous wavelet transform to partition 5-second ECG and WPS segments into spectrograms. It incorporates an improved ConvNeXt, a multimodal transformer layer, and a fused multi-layer perceptron to extract and fuse multimodal features for ECG classification. Subsequently, the network is adapted to WPS and coronary heart disease classification tasks through transfer learning techniques. In comparison to existing methods, our approach achieves heightened sensitivity in detecting supraventricular and ventricular ectopic segments, while also outperforming established WPS classification methodologies. Importantly, the proposed network adeptly handles multimodal signals and excels in classification accuracy, particularly within the realm of physiological signals.

INDEX TERMS Electrocardiogram, wrist pulse signal, transfer learning, coronary heart disease classification.

I. INTRODUCTION

The electrocardiogram (ECG) serves as a crucial diagnostic tool for monitoring the heart health of patients. Its ability to detect abnormal states in real time allows for early intervention and treatment. Nevertheless, the scarcity of medical resources poses a significant challenge to healthcare accessibility [1]. Automating ECG analysis and disease diagnosis not only saves medical staff time but also enhances the availability, accuracy, and consistency of diagnosis.

In the past decade, rapid advancements in ECG analysis techniques have led to the categorization of methodologies into four key domains. The first group predominantly encompasses methods that establish a comprehensive pre-processing process to convert the ECG signal into a clean single-cycle signal. Subsequently, feature extraction methods are employed to extract key features. Finally, machine learning-related methods, such as Support vector machine (SVM) [2], K-nearest neighbor (KNN), weighted XGBoost, and Linear discrimination (LD) algorithm, are applied to accomplish ECG classification. However, in this approach, there is interdependence among the steps, leading to redundancies in the extracted features. Consequently, these methodologies necessitate feature selection [3], introducing increased computational complexity and time demand in practical applications. The key drawback of these methods lies in their low efficiency for fast ECG classification. Compared to deep learning, traditional machine learning exhibits suboptimal performance in the representation learning of large, complex samples of data [4].

The second group of models aims to address the aforementioned problems by devising techniques for automatic feature extraction and classification. Representative models in this group include Convolutional neural network (CNN) [5], Tuned dedicated convolutional neural network (TDCNN) [6], Deep bidirectional LSTM network-based wavelet sequences (DBLSTM-WS) [7], DNN [8,9], a CNN, Transformer and Long Short-Term Memory (LSTM) based assemble neural network framework named CLSTM-Transformer [10], dual attention hybrid network (DA-Net) [11]. All these models demonstrate high classification performance. However, they all use segmented single ECG signals, which requirement cannot be satisfied in a real-time ECG signal classification.

The third group of models takes a distinct approach by avoiding the extraction of the R peak and focusing solely on the single beat of ECG signals. Many methods within this category advocate segmenting the ECG signal into 2, 5 or 10-second heartbeats [12], which are subsequently classified using CNN, Densely Connected CNN (DenseNet) [13], or an improved deep residual network [14]. While this strategy yields a classification performance comparable to the most current state-of-the-art heartbeat classification methods, it exhibits a limitation in sensitivity to abnormal heartbeats, such as type S and type V heartbeats.

The fourth group of methods, a recent emergence in the field, aims to enhance capability in handling diverse data formats and pushing the limits of accuracy. Various techniques based on 2D ECG images have been developed for this purpose. Corresponding 2D CNN has been designed to classify the ECG grayscale image[15], 2D matrix[16], and 2D spectrograms[17]. The

classification performance of these methods has seen improvement through advanced data enhancement strategies. However, a notable limitation is the underutilization of time-domain features inherent in the 1D signals, leaving room for the improvement in classification.

After reviewing the literature on ECG signals, it is noteworthy that few studies have harnessed information from both 1D and 2D signals. Many research gaps exist in the multimodal fusion of 1D signals and their corresponding spectrograms, despite the demonstrated effectiveness of multimodal approaches. Recent studies have illustrated the superiority of multi-module recurrent CNN with the transformer that utilize the 1D signal segments, 2D spectrograms, and metadata. These approaches have achieved better classification results than the 1D and 2D networks in a number of applications [18].

On the other hand, the wrist pulse signal (WPS) holds a wealth of information not only related to cardiac health disorders but also providing crucial insights into blood viscosity, blood vessel wall elasticity, blood flow velocity, and various other physiological and pathological data on diverse human organs [19]. WPS and ECG signals, both driven by cardiac activity, share many similarities in waveform. However, in contrast to the ECG signal, the WPS travels through various parts of the body, including nerves, muscles, skin, and arterial walls. Consequently, the WPS contains more physiological information than the ECG signals[20]. Notably, WPS has been employed as one of the four diagnostic methods in traditional Chinese medicine (TCM). The study of pulse signals bears substantial theoretical and practical significance. Currently, most of studies on WPS are confined to feature engineering, with pattern recognition relying on classical machine learning methods. Several classification methods such as SVM and Gradient boosting decision tree (GBDT) [21] have been utilized to classify the denoised and feature extracted WPS.

Similar to the development process of ECG signal analysis methods, WPS was first investigated in feature extraction, and then deep learning techniques were applied to pulse signal analysis and recognition. However, compared to the progress of deep learning research on ECG signal analysis, few studies have concentrated on the applications of deep learning for WPS. Noteworthy methods in this domain include GA-BP neural network [22], 1D CNN [23], and TCN [24], proposed for the classification of WPS. Most classification methods for both ECG and WPS have been adapted from architectures like ResNet or DenseNet. Additionally, CNN structure have evolved to ConvNeXt [25]. These networks draw inspiration from the visual transformer (ViT) and incorporate various modules, such as inverted bottlenecks and separate downsampling layers, to enhance classification performance. The integration of these modules into a network architecture has proven effective in improving signal classification performance.

To summarize, our rationale for aiming to address the problems of existing models can be outlined as follows:

1) Network Structure Enhancement: The current network structure of physiological signal classification needs an update with the latest CNN architecture to enhance overall classification performance.

2) Segmentation Sensitivity: Although a heartbeat overlapping segmentation method has been proposed to enhance ECG segmentation, it exhibits poor sensitivity to type S and type V segments, highlighting the need for improved methodologies.

3) Multimodal Fusion: The field of multimodal fusion for physiological signals has received limited attention. Therefore, there is necessity to explore and leverage the multimodal properties of signals to enhance the model's performance.

4) Data-Driven WPS Classification: Most of classification methods for WPS rely on traditional techniques. To improve WPS classification, there is a need for the development of more sophisticated, data-driven classification methods.

Aligned with the aforementioned challenges, we propose an improved ConvNeXt network with a multimodal transformer layer and a multi-layer perceptron (MLP) fused layer, which is called ICMT-Net to classify the ECG. In our approach, ECG signals are denoised using a discrete wavelet transform (DWT), followed by normalization. Subsequently, the ECG signals are segmented into 5-second intervals using the overlapping segmentation method. Within these segments, 2D spectrogram information is extracted using CWT. Following feature extraction by ConvNeXt with CBAM module [26], 1D and 2D ECG features are obtained and concatenated. These concatenated features serve as the input for the multimodal transformer layer and the MLP fusion layer. The constructed network demonstrates heightened sensitivity in detecting abnormal ECG signals, specifically type V and S segments. Subsequently, the network is transferred for WPS classification task. To optimize pulse classification and coronary heart disease (CHD) classification, the WPS is denoised and normalized. Leveraging the similarities and differences between ECG and WPS, the network pretrained on ECG signals is fine-tuned and transferred to the WPS. Finally, the transferred network fuses the features of 1D WPS and the corresponding spectrograms, achieving superior classification performance compared to existing WPS classification methods.

The main contributions of this paper can be summarized as follows:

- We introduce a groundbreaking network architecture that integrates two types of feature extraction networks, namely 1D CBAM-ConvNeXt and 2D CBAM-ConvNeXt. Local feature maps for each heartbeat, along with their resulting transformed 2D time-frequency maps, are extracted for both networks.
- A Transformer layer is incorporated to extract corresponding global feature maps from the aforementioned local feature maps. These global features are then fused using the MLP Fused module to generate final feature vectors. These vectors are subsequently input into the classification layer, completing the classification process.
- We present a meticulously designed multimodal network and a corresponding transfer learning architecture. The transfer learning framework enables a multimodal network fully trained on ECG signals to undergo full parametric tuning on WPS, yielding promising classification results. This anticipates a significant reduction in workload for various physiological signal classification tasks.

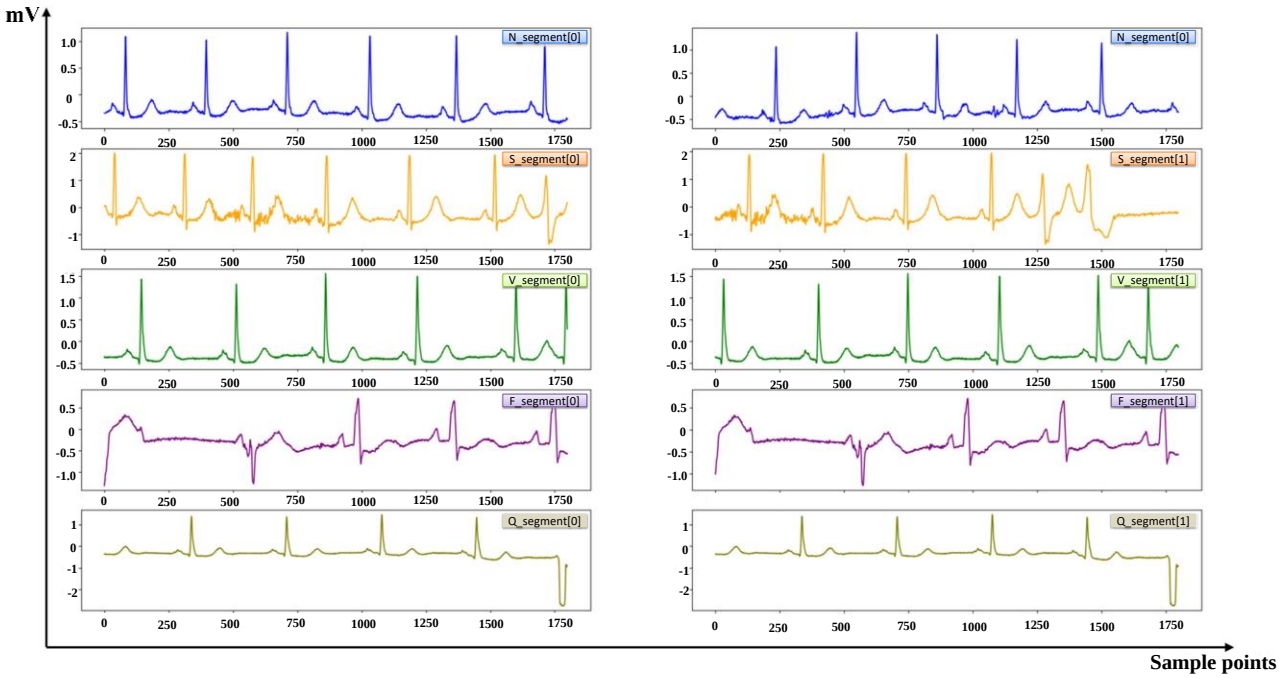


FIGURE 1. Five types of overlapping ECG data. The ECG signals in the left column are the first group of ECG signals of types N, S, V, F, and Q from top to bottom, while the second group of five types of ECG signals are in the right column.

II. Data Preprocessing

A. *ECG dataset*

The ECG dataset used in this study was sourced from the MIT-BIH arrhythmia database (MITDB), which is publicly accessible on PhysioNet [27]. The MITDB comprises 48 ECG recodes, each with a sampling rate of 360 Hz and a duration of 30 minutes. In most recordings, one channel of Modified limb leads II (MLII) is obtained by placing electrodes on the chest, with MLII being the standard practice for dynamic ECG recording. This paper specifically selects lead II as the primary lead. Based on the literature [28], The ECG dataset is further categorized into DS1 and DS2, with the training set divided into DS11 and DS12, as detailed in Table I based on literature guidelines[28].

In according with the Association for the Advancement of Medical Instrumentation (AAMI), there are 15 arrhythmia categories that can be classified into five super-classes: non-ectopic (N), supraventricular ectopic beat (SVEB), ventricular ectopic beat (VEB), fusion beat (F), and unknown beat (Q). The specific 15 arrhythmias within these categories are detailed in Table II.

TABLE I
DIVISION OF TRAINING SET (DS11), VALIDATION SET (DS12), AND TEST SET (DS2)

Dataset	Sub-dataset	Recording NO.
DS1	DS11	101 106 108 109 114 115 116 119 122 209 223
	DS12	112 118 124 201 203 205 207 208 215 220 230

TABLE II
SUPER CLASSES OF HEARTBEAT PRESENT IN THE MIT-BIH DATABASE

AAMI classes	Symbol	Heartbeat types
N(Non-ectopic)	N	Normal beat
	L	Left bundle branch block beat
	R	Right bundle branch block beat
	e	Atrial escape beat
	j	Nodal (junctional) escape beat
S (Supraventricular ectopic beat)	A	Atrial premature beat
	a	Aberrated atrial premature beat
	J	Nodal (junctional) premature beat
	S	Supraventricular premature beat
V (Ventricular ectopic beat)	V	Premature ventricular contraction
	E	Ventricular escape beat
F (Fusion beat)	F	Fusion of ventricular and normal beat
Q (Unknown beat)	P	Paced beat
	f	Fusion of paced and normal beat
	U	Unclassifiable beat

B. *Preprocessing of ECG*

In this study, the ECG signals are partitioned into 5-second segments using re-labeling rules and an overlapping segmentation approach. The overlapping segmentation method is specifically implemented in the training set (DS11), as outlined in Table III. It's important to note that the validation set (DS12) and test set (DS2) do not undergo overlapping segmentation. Table III provides the counts of different types of heartbeats in the dataset, highlighting the use of overlapping segmentation in DS11.

The visual representation of different types of overlapping ECG data (type N, S, V, F, and Q) is illustrated in Fig. 1. The subfigures on the left showcase the first ECG data of each category, while those on the right depict the second ECG data of each category. From Fig. 1, It is evident that compared to the first ECG data of type N, S, V, F, and Q, the second ECG data is shifted by 1800, 312, 446, 6, and 2, respectively. This strategy allows a smaller number of heartbeat classes to sample more ECG data.

TABLE III
NUMBER OF SEGMENTS OF VARIOUS CATEGORIES OBTAINED BY USING THE ORDINARY AND THE OVERLAPPING SEGMENTATION METHOD

Dataset	Sub-dataset	N	SVEB	VEB	F	Q
DS1	DS11	3023	188	748	10	2
	DS12	2884	156	885	44	2
	DS11 (overlapping)	3023	3006	3010	2880	1800
DS2		5874	477	1468	118	5

Subsequently, the denoising process is implemented by applying the DWT to the segmented heartbeats. The wavelet basis function is from Daubechies D6 [29]. To ensure that the network can downsample, each ECG segment is resampled to consist of 1280 sample points. Finally, each ECG segment is normalized using Z-score standardization. The normalization formula is shown in the following equation:

$$\mu_\beta = \frac{1}{m}\sum_{i=1}^{m} x_i \quad (1)$$

$$\sigma_\beta^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_\beta) \quad (2)$$

$$x_{norm} = \frac{x_i - \mu_\beta}{\sqrt{\sigma_\beta^2}} \quad (3)$$

From (1) to (3), x_i represents the ECG segments, μ_β and σ_β represent the mean and standard deviation, respectively, and x_{norm} represents the normalized ECG segments.

For the adaptation of multimodal input, the ECG segments should be transformed into 2D spectrograms using CWT.

C. *Preprocessing of WPS*

The WPS as a distinct physiological signal, shares morphological similarities with the ECG signal. However, due to the absence of public WPS dataset, this study relies on a private pulse signal dataset collected at Wuyi University. Four types of wrist pulse signals were gathered for analysis, namely sunken pulse, moderate pulse, normal pulse, and skipping pulse signals. The entire data collection process adhered to the requirement of Human Research Ethics (approval number: [2019]18).

During the WPS collection, interference from human breathing, small fibrillation, and power frequency introduces high-frequency and low-frequency noise into pulse signals, necessitating their removed (below 3 Hz and above 20 Hz). A band-pass filter is employed for this process, effectively eliminating low-frequency and high-frequency noise and retaining the clean WPS between 3 Hz and 20 Hz range. Additionally, to address baseline drift issues caused by the acquisition instrument, baseline drift removal is performed to obtain a stable WPS signal. Following noise reduction and baseline drift removal, additional steps are taken to enhance the quality of the WPS. Median filtering and wavelet transform denoising using the sym8 wavelet are applied to further address baseline drift and overcome the artifacts in the signal. Cubic spline interpolation is employed to fit the WPS and identify the minimum value within a specific frequency window (a single pulse cycle). This process is repeated by applying cubic spline interpolation is to fit

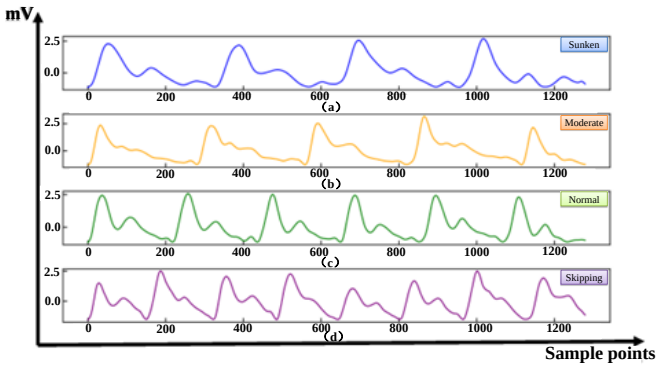


FIGURE 2. Four types of WPS and the corresponding spectrograms. (a) Sunken pulse. (b) Moderate pulse. (c) Normal pulse. (d) Skipping pulse.

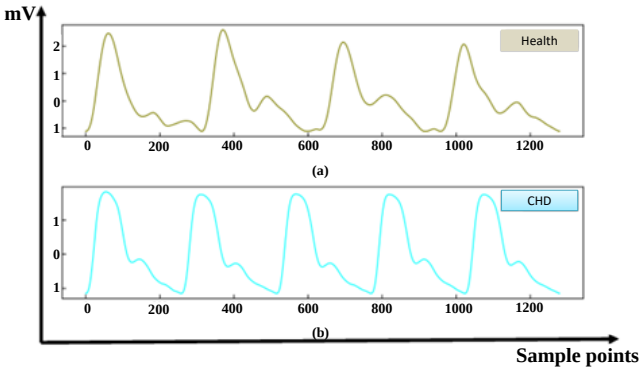


FIGURE 3. Two types of WPS and the corresponding spectrograms. (a) Healthy pulse signal. (b) CHD pulse signal.

the minimum value once again. Ultimately, subtracting the minimum curve from the original signal curve effectively eliminates the baseline drift in the WPS.

The WPS undergoes further processing by being divided into 5-second segments and subsequently re-labeled into four types of pulse after denoising. Fig. 2 displays the WPS along with their corresponding spectrograms, depicting sunken pulse, moderate pulse, normal pulse, and skipping pulse. To facilitate transfer learning, it is imperative that the input shapes of both the ECG and pulse signals are identical. Hence, the four types of WPS segments are resampled into 1280 sample points, mirroring the ECG segments. Finally, these WPS segments are normalized using Z-score standardization. Moreover, there are 186 cases of healthy pulse signals and 210 cases of CHD pulse signals, sourced from the Shanghai University of Traditional Chinese Medicine. This dataset also undergoes denoising and baseline drift removal, following the same preprocessing steps as the previously described pulse dataset. Fig. 3 illustrates the two types of pulse data are obtained after preprocessing. After denoising and segmenting the data into 5-second signal segments, the distribution of pulse sample for the two tasks is presented in Table IV and Table V, respectively.

D. *Correlation analysis and the feasibility of transfer learning*

The ECG signal and the pulse signal share numerous similarities, as both are generated by the systolic-diastolic movement

TABLE IV
FOUR CLASSES OF PULSES PRESENT IN THE TASK OF PULSE CLASSIFICATION

Dataset	Sunken	Moderate	Normal	Skipping
Pulse	5892	888	5304	3462

TABLE V
TWO CLASSES OF PULSES PRESENT IN THE TASK OF CHD CLASSIFICATION

Dataset	Health	CHD
CHD	2428	932

of the heart, establishing a morphological relationship between the two types of signals. After calculating the KL divergence [30] of the two types of datasets, the KL scatter values are distributed from 0.03 to 0.06, indicating that the distributions of the two signals are extremely similar. The detailed KL divergence formula is shown as follows:

$$KL(p||q) = \sum p(x) log \frac{p(x)}{q(x)} \quad (4)$$

In (4), $p(x)$ and $q(x)$ denote the distribution of two datasets.

Due to the similar distribution of the two datasets and the fact that both are 1D signal data, they are quasi-periodic physiological signals with multiple peaks and valleys. Thus, the classification network architecture trained in the ECG signal can be effectively transferred to the WPS. This transfer includes all layers of the network except for the final classification layer. The network parameters can then undergo globally fine-tuned to adapt the architecture for pulse signal classification, resulting in optimized network parameters tailored to the characteristics of the pulse signal.

III. Modeling Approach

A. *ConvNeXt-Trans*

The proposed network is built upon ConvNeXt, a convolutional neural network architecture traditionally utilized in image classification, where it has demonstrated superior performance. However, in the domain of signal classification, certain modifications are necessary, particularly at the normalization layer. In our proposed method for ECG signal classification, we replace the 2D convolution with 1D convolution in ConvNeXt. Additionally, during the training phase, a method of gradually changing the learning rate is employed. The learning rate initiates from a predetermined value and is systematically reduced by a factor of 10 times every 10 epochs, spanning a total of 50 epochs. This approach aims to ensure more consistent model convergence during the later stages of training. It is observed that the layer normalization (LN) cannot process 1D feature maps, which leads to convergence failure. To overcome this, the LN in the network is replaced with batch normalization (BN). Additionally, LN module in the 2D ConvNeXt network is also substituted with a BN, maintaining a bimodal network capable of smooth fusion in subsequent experiments. To improve the robustness of the network and reduce the network channel redundancy, the CBAM attention module is added to the modified ConvNeXt. Furthermore, to facilitate the successful fusion of two distinct modal data types, the transformer encoder module is introduced to the network. This layer aids the model in extracting global features from both the 1D signal and 2D spectrogram. Most importantly, it transforms the 2D spectrogram into the 1D sequence, enabling seamless fusion of the two modalities. The resulting network, termed ConvNeXt-Trans, is illustrated in Fig. 4. The architecture of ConvNeXt-Trans contains a pre-convolution module and four stage layers. The essential module includes four mechanisms detailed as follows:

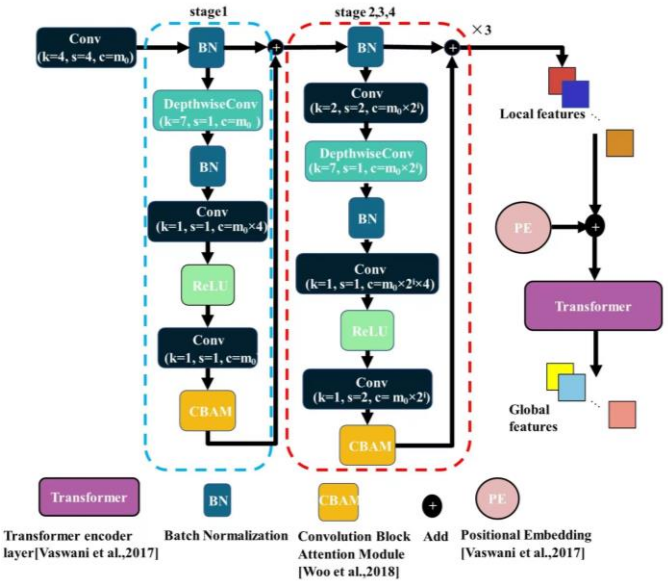


FIGURE 4. ConvNeXt-Trans network, which has three main modules: (I) The improved ConvNeXt extracts the local information of signal data and the corresponding spectrograms; (II) The CBAM calculates the weight of feature map in channel and space, and makes the model more goal focused; (III) The transformer layer further extracts the global feature maps that were weighted by multi-head self-attention (MHSA) module.

1) SEPARATE DOWNSAMPLING MODULE

In ConvNeXt architecture, the separate downsampling module (SD) is used for downsampling feature maps with a stride of 2. The module is constructed using a convolutional layer and BN, and the formula for the separate downsampling module is expressed as follows:

$$SD(x) = BN(Conv_{k=4,c=c_1*2^L}(x)) \quad (5)$$

In equation (5), x represents the input feature map, BN represents batch normalization. $Conv_{k=4,c=c_1*2^L}$ represents the convolution operation with a kernel size of 4 and c_1 denoting the original number of convolution kernels. In stage 2, L starts from 1 and increases by 1 per stage, resulting in the doubling of the number of convolution kernels with each down-sampling module.

2) INVERTED BOTTLENECK MODULE

The inverted bottleneck is inspired by MobileNetV2 [31], and it serves to reduces the network FLOPs (floating-point operations per second). The formula for the inverted bottleneck is expressed as follows:

$$IB(x) = Conv_{k=1,c=c_1*2^L}(\sigma(Conv_{k=1,c=4c_1*2^L}(BN(DWC_{k=7,c=c_1*2^L}(x))))) \quad (6)$$

In equation (6), DWC represents the depthwise convolution, σ represents the ReLU activation function, and $Conv_{k=1,c=c_1*2^L}$ denotes a 1x1 convolution with $c_1 * 2^L$channels. The inverted bottleneck structure is designed to increases the dimensionality slightly to compensate for information loss during the depthwise convolution.

3) CBAM MODULE

To enhance the feature extraction capability of the network, the CBAM is employed, which applies attention mechanisms to the feature map on both channel and spatial dimensions. After applying CBAM, the feature map receives the attention weights on both channel and spatial dimensions. This process significantly enhances the interdependence of individual features within the channel and space, making it more effective in extracting relevant and meaningful features from the current feature map. The attention mechanism allows the network to focus

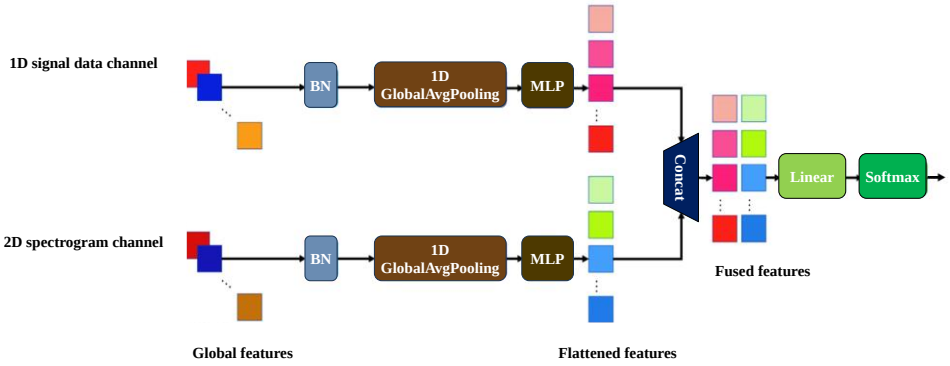


FIGURE 5. MLP Fused module. Firstly, two modal feature maps extracted from ConvNeXt-Trans are normalized by the BN layer, and the normalized feature maps need to be flattened by global average pooling. Then, the MLP module establishes weights for all features to carry out global awareness and obtains two sets of sequences. Finally, the two modal sequences are smoothly concatenated and fed to the classification layer to complete the ECG signal classification.

on the most informative parts of the feature map, contributing to improved feature extraction and, consequently, enhanced classification performance. The detailed expressions for CBAM are as follows:

$$AM_C(x) = \delta(W_1\left(W_0\left(GAP(x)\right)\right) + W_1\left(W_0\left(GMP(x)\right)\right)) \quad (7)$$

$$AM_s(x) = \delta(Conv_{k=7,c=1}(Concat([AM_C(x), MP_C(x)]))) \quad (8)$$

In equation (7) and (8), GAP, GMP, AM, and MP represent global average pooling, global max pooling, average pooling, and max pooling, respectively, and δ denotes the sigmoid function. The weights for channel attention and spatial attention are calculated using (5) and (6), respectively.

4) TRANSFORMER ENCODER MODULE

The transformer encoder in the proposed architecture involves a multi-head self-attention (MHSA) module and a feed-forward network (FFN) with a single hidden layer. To enhance the feature extraction network, both sub-layers are followed by a BN, and there is a residual connection around every two sub-layers. This design is beneficial for improving the training stability and the flow of gradient information through the network. The overall structure ensures effective learning and feature extraction capabilities.

Firstly, the key insight in the transformer encoder is to extract N non-overlapping patches from the 1D signal and 2D spectrogram image, denoted as $x_i^{1D} \in \mathbb{R}^t$ and $x_i^{2D} \in \mathbb{R}^{h\times w}$, respectively. These patches are then transformed into a series of 1D tokens, $z_i \in \mathbb{R}^d$, expressed as follows:

$$\mathrm{z} = d(\mathrm{x}; l) = [lx_1, lx_2, \ldots, lx_N] + PE \quad (9)$$

Here, l represents the linear projection that maps each token to $\mathbb{R}^d$, and $PE \in \mathbb{R}^{(N+1)\times d}$ is the positional embedding added to the tokens to provide positional information.

Secondly, the encoded tokens are input into the transformer encoder module that contains MHSA, BN and FFN. The corresponding expressions are as follows:

$$\mathrm{y} = MHSA\left(BN(\mathrm{z}^l)\right) + \mathrm{z}^l \quad (10)$$

$$\mathrm{z}^l = MLP\left(BN(\mathrm{y}^l)\right) + \mathrm{y}^l \quad (11)$$

$$MHSA(X) = DPA(W^Q X, W^K X, W^V X) \quad (12)$$

where $MHSA$ is the Dot-product attention (DPA) between the queries, keys and values are linear projections of the same tensor.

Finally, the weighted sequence obtained from the attention mechanism is projected linearly using the FFN. This process allows the network to capture complex relationships and dependencies in the input sequences.

B. MLP Fused Module

In order to fuse the two kinds of feature maps extracted from the 1D signal and 2D spectrogram, the BN module is applied to

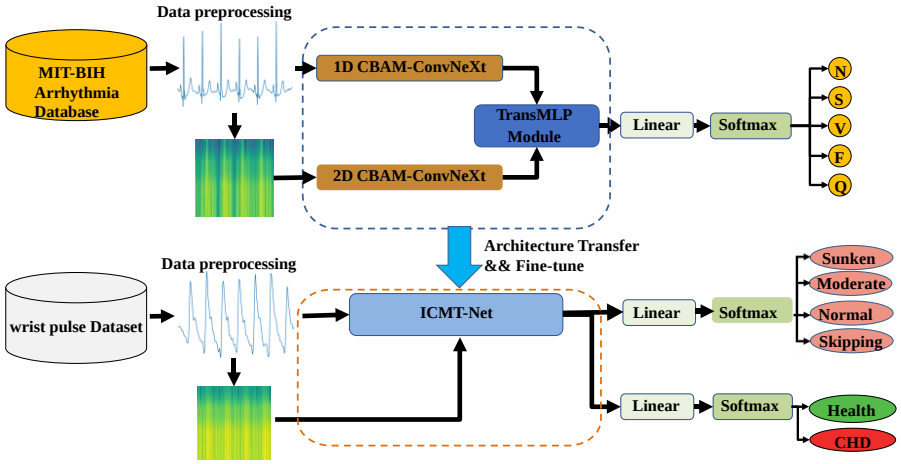


FIGURE 6. The overall architecture of our proposed method, which consists of four modules: (I) The preprocessing module denoises and normalizes the signal data, and the processed data are transformed into spectrograms by using the CWT. (II) ConvNeXt is the feature extractor of the 1D and 2D data, which can obtain the local feature of 1D signal and 2D spectrogram, and the global features of those feature maps can be extracted by the transformer encode layer. (III) Furthermore, for fusing two modal data, the 1D features and the 2D features are pooled by the global average pooling to obtain the flattened data, and two modal data are fused through the MLP layer. (IV) The fused network is fine-tuned and transferred to wrist pulse classification.

normalize these feature blocks. This normalization is crucial for converge, especially when dealing with large difference between the two sets of modal data. Subsequently, the 1D global average pool is used to transform the two feature maps into flattened feature sets, and these sets are concatenated into a fusion feature set. The detailed MLP module is depicted in Fig. 5.

Finally, a linear classification layer with the SoftMax function is used to classify the fused feature sets into five ECG types. It's noteworthy that due to the similarity in distributions and most features of the two datasets, and considering that the pulse dataset is large enough, the original structure and initial weights of the model are kept unchanged. Only the final classification layer is replaced with a fully connected layer capable of classifying into 4 classes for the pulse dataset. This final configuration is then re-trained on the WPS dataset using a learning rate of 0.001[32]. The final ICMT-Net structure is shown in Fig. 6.

IV. Results

A. Performance metrics

Four classification indicators, accuracy (Acc), sensitivity (Se), positive productivity (Pp), and F1 score (F1), are adopted according to the following formulas:

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \tag{13}$$

$$Se = \frac{TP}{TP+FN} \tag{14}$$

$$Pp = \frac{TP}{TP+FP} \tag{15}$$

$$F_1 = 2 \times \frac{Se \times Pp}{Se+Pp} \tag{16}$$

where *TP*, *TN*, *FP*, and *FN* are true positive, true negative, false positive, and false negative, respectively.

B. Performance evaluation

The ablation test is conducted through separate experiments, each focusing on specific aspects such as the choice of convolution kernel number, the inclusion or exclusion of attention mechanisms, and the impact of multimodal fusion and transfer learning on the pulse classification task. These experiments utilize the inter-patient paradigm for evaluation. In the training phase, the stochastic gradient descent with momentum (SGDM) is used as an optimizer to update the model parameters, with a base learning rate of L_{base}. A ten-epoch gradual warm-up training strategy[33] is utilized to approach L_{base} starting from a small value in the first 10 epochs. The learning rate decays based on the cosine annealing schedule:

$$L_{current} = \frac{L_{base}}{2}(1 + \cos(\frac{E_{current}}{E_{set}}\pi)) \tag{17}$$

where $E_{current}$ is the current epoch, and E_{set} is the total number of epochs. For fair comparisons, E_{set} and L_{base} are uniformly set to 50 and 0.01, respectively.

The experiments are implemented in Python using TensorFlow (version 2.5.0) as the programming language. The study is conducted on hardware featuring an NVIDIA Quadro RTX 5000 GPU, 64GB RAM, and the Windows 11 operating system.

C. Convolution kernel number selection

In the ablation experiments, various configurations of the network are explored, and the number of convolution kernels in the pre-convolution layer and each stage is adjusted. The choices are 16, 16, 32, 64, 128, 32, 32, 64, 128, 256 and 64, 64, 128, 256, 512 for the pre-convolution layer and each stage, respectively. These three numbers of convolution kernels are used, whether the data overlapping segmentation method is performed or not. The results are summarized in Table VI.

Table VI shows that ConvNeXt with channels starting from 32 achieves the best classification performance for ECG signals and their corresponding 2D spectrograms. The overlapping segmentation method further improves the detection ability of the dataset for ECG signals. Under the condition of using the overlapping segmentation method, the sensitivity, positive productivity, and F1 score of the 1D-ConvNeXt with channels starting from 32 are increased by 7.12%, 0.96%, and 1.93%, respectively. For 2D-ConvNeXt, the sensitivity, positive productivity, and F1 score are increased by 2.79%, 2.61%, and 0.96%, respectively. The next number of channels in the model is selected only for 16 and 32.

TABLE VI
THE SELECTION OF THE NUMBER OF CONVOLUTION KERNELS

Model	Overlapping	*C*	Acc (%)	Se (%)	Pp (%)	F1 (%)
1D-ConvNeXt	False	16	78.29	40.02	36.24	37.32
	False	32	77.84	40.25	36.66	37.33
	False	64	77.70	37.13	33.34	34.74
	True	16	69.87	46.18	37.46	38.05
	True	32	74.98	47.37	37.52	39.26
	True	64	72.40	46.15	36.77	38.37
2D-ConvNeXt	False	16	71.68	38.00	34.44	34.76
	False	32	75.81	39.75	34.83	36.59
	False	64	75.16	38.68	33.57	35.01
	True	16	72.53	41.10	33.04	34.15
	True	32	74.83	42.54	37.44	37.55
	True	64	68.61	42.38	31.43	31.84

These results indicate the importance of channel selection in achieving optimal classification performance for ECG signals.

D. ConvNeXt with attention mechanism

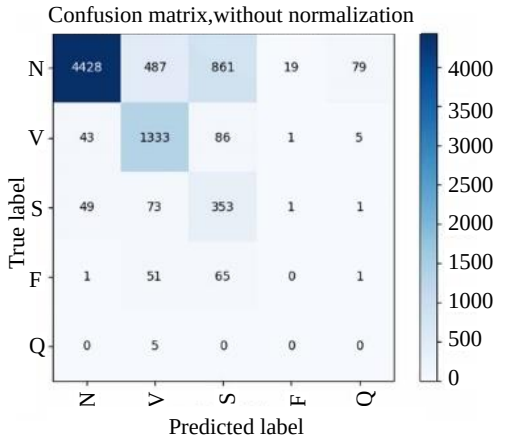
The effect of the CBAM module is investigated in the ablation experiments. The results are presented in Table VII. When using the overlapping segmentation method, the classification metrics (Acc, Se, Pp, and F1) of the 1D model increased by 1.64%, 0.35%, 1.39%, and 0.61%. However, for the 2D model, only the Se increased by 7.37%, while all other metrics are decreased. The confusion matrix in Fig. 7 illustrates that the 1D model emphasizes ECG signals of type V and S, while the 2D model focuses more on ECG signals of type Q. This leads to the different aspects of performance improvement of the model, with the 2D model contributing more to sensitivity improvement.

E. Multimodal fusion experiment

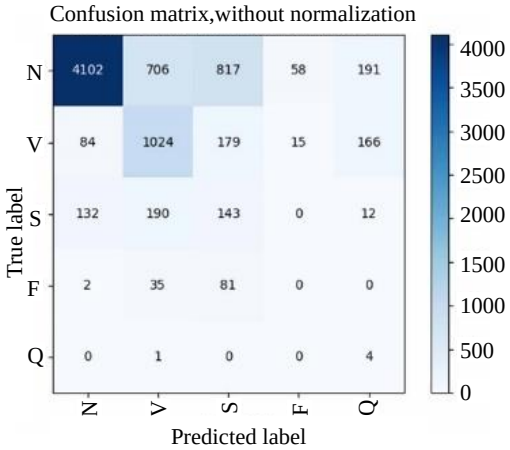
The fusion of 1D ConvNeXt with2D ConvNeXt using the transformer encoder layer and the MLP fusion layer in ICMT-Net is analyzed in Table VIII. Compared to the 1D CBAM-ConvNeXt, ICMT-Net shows improvements of 5%, 4.59%, 2.74%, and 4.08% in Acc, Se, Pp, and F1, respectively. Furthermore, compared to ICMT-Net and 2D CBAM-ConvNeXt, ICMT-Net achieves improvements of 15.23%, 2.4%, 9.63%, and 12.1% in Acc, Se, Pp, and F1, respectively. These results indicate that the transformer encoder layer and MLP fusion layer effectively fuse the data features from both modalities, enhancing the network's classification ability.

TABLE VII
CORRESPONDING EXPERIMENTAL INDICATORS OF THE IMPROVED CONVNEXT WITH ATTENTION MECHANISM

Model	C	Acc (%)	Se (%)	Pp (%)	F1 (%)
1D CBAM- ConvNeXt	16	74.41	47.17	38.80	39.82
	32	76.62	47.72	38.91	39.87
2D CBAM- ConvNeXt	16	64.47	47.48	31.67	31.65
	32	66.39	49.91	32.02	31.85



(a)



(b)

FIGURE 7. (a) The confusion matrix of 1D CBAM-ConvNeXt. (b) The confusion matrix of 2D CBAM-ConvNeXt.

Additionally, when comparing Fig. 7 and Fig. 8, it can be observed that the classification performance of the multimodal fusion network for each category is improved compared to that of the 1D and 2D network, except for the sensitivity and the positive productivity of the Q class. This suggests that the multimodal fusion network incorporates features from both 1D and 2D ECG data, leading to improved classification performance, especially for type V and S EGC segments.

F. Transfer learning applied to the pulse classification task

Finally, due to many similarities between WPS and ECG, ICMT-Net has been adapted for the pulse classification task and the CHD classification task by transferring the corresponding network structures and fine-tuning the corresponding parameters, excepting the final classification layer, which is pre-trained in the ECG classification task. It's important to note that the initial learning rate (L_{base}) is set to 0.0001.

TABLE VIII
CORRESPONDING EXPERIMENTAL INDICATORS OF THE ICMT-NET

Model	C	Acc (%)	Se (%)	Pp (%)	F1 (%)
ICMT-Net	16	81.62	52.31	41.65	43.95
	32	74.41	47.17	38.80	39.82

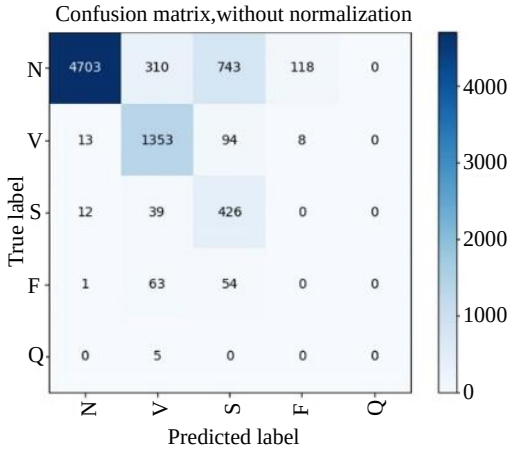


FIGURE 8. Confusion matrix of ICMT-Net when using the overlapping segmentation method.

In Table IX, a comparative analysis of the classification performance of 1D-ConvNeXt, 2D-ConvNeXt, and ICMT-Net reveals that three networks, pre-trained in the ECG classification task, demonstrate superior classification performance in both tasks. The variance in classification performance between the network without pre-training and those with pre-training fluctuates between 0.54% and 3.15%.

In the PWS classification task, among the models pre-trained in the task of ECG classification task, ICMT-Net shows improvements of 1.38%, 1.10%, 2.42%, and 1.77% in accuracy, sensitivity, positive productivity, and F1 scores, respectively, compared to 1D CBAM-ConvNeXt. Compared with 2D CBAM-ConvNeXt, ICMT-Net exhibits improvements of 1.93%, 2.96%, 2.92%, and 2.94%, respectively.

Conversely, in the CHD classification task, ICMT-Net demonstrates improvements of 1.19%, 2.14%, 0.82%, and 1.51% in accuracy, sensitivity, positive productivity, and F1 scores, respectively, compared to 1D CBAM-ConvNeXt. In compared with 2D CBAM-ConvNeXt, ICMT-Net shows improvements of 6.39%, 8.70%, 7.44%, and 8.11%, respectively.

TABLE IX
PERFORMANCE COMPARISON OF MODELS TRANSFERRED TO THE WPS AND CHD CLASSIFICATION TASK

Model	Task	Pretraining	Acc (%)	Se (%)	Pp (%)	F1 (%)
1D CBAM-ConvNeXt	WPS	False	95.98	96.23	95.44	95.79
		True	97.24	97.19	96.63	96.89
	CHD	False	97.03	94.98	97.62	96.19
		True	98.07	96.85	98.33	97.56
2D CBAM-ConvNeXt	WPS	False	95.56	92.18	95.54	93.68
		True	96.69	95.33	96.13	95.72
	CHD	False	91.98	89.67	90.21	89.94
		True	92.87	90.29	91.71	90.96
ICMT-Net (ours)	WPS	False	97.49	96.30	98.29	97.24
		True	98.62	98.29	99.05	98.66
	CHD	False	98.22	96.96	98.61	97.74
		True	99.26	98.99	99.15	99.07

These results highlight that ICMT-Net, pre-trained on the ECG classification task, effectively leverages the features of 1D signals and 2D spectrograms in the WPS and CHD classification, and achieving optimal classification results.

V. DISCUSSION

For evaluating the classification performance of our model, we conducted a comparative analysis against leading state-of-the-art ECG classification methods, and the results are presented in Table X.

All the methods listed in Table X are designed for classifying heartbeats across five different types within an inter-patient paradigm. In this paradigm, the available trainable ECG data, excluding overlapping segmentation techniques, is limited. Remarkably, when utilizing the same overlapping dataset, the ICMT-Net demonstrates significant performance improvements (as shown in Fig. 8). Specifically, it achieves a 6.12% increase in Pp for class N, a 3.82% enhancement in Se for class V, and an impressive 54.09% surge in Se for class S when compared to the improved deep ResNet. These findings underscore the heightened sensitivity of our model in accurately classifying type V and S ECG signals.

This paper also undertook a comparison of existing deep learning models for WPS classification, encompassing the GA-BP algorithm, 1D CNN, and TCN. These network architectures were employed for both WPS and coronary heart disease (CHD) classification tasks. The dataset was divided into training, validation, and testing sets using a 6:2:2 partition ratio. The comprehensive classification results are summarized in Table XI and XII. Across both tables, our proposed approach consistently demonstrates superior performance in terms of Se and Pp for both pulse and CHD classification tasks. These outcomes underscore the adaptability of our newly introduced pre-trained multimodal fusion network, reinforcing its superiority when compared to established methodologies for pulse classification.

In conclusion, our study sheds light on the ongoing complexity in the realm of cardiovascular disease diagnosis and model precision. The ICMT-Net, our innovative creation, introduces a new paradigm by harnessing the synergistic potential of multi-modal physiological signals. By seamlessly integrating inputs from both ECG and WPS sources, the ICMT-Net adapts adeptly to diverse application scenarios, elevating its accuracy and robustness. The architectural framework is meticulously detailed in Figure 6. This groundbreaking approach not only aligns with contemporary CNN architectures but also pioneers new dimensions through the incorporation of multi-modal learning modules. Leveraging techniques such as overlapping ECG segmentation and continuous wavelet transform (CWT) for 2D spectrogram extraction, our model transcends the limitations imposed by single-modal analyses. The ConvNeXt architecture, augmented with CBAM, plays a pivotal role in extracting pivotal features from both 1D and 2D ECG domains. The amalgamation of these features within the multi-modal transformer and subsequent MLP fusion layers generates a holistic representation that excels in classifying intricate ECG segments, including those of type V and S. Furthermore, our exploration ventures beyond the realms of ECG to embrace wrist pulse signals (WPS), uncovering the potential synergy between these two modalities. Despite the relative novelty of WPS diagnosis, our integrated approach holds promise in revealing distinct physiological dimensions. Through denoising, normalization, and transfer learning, the convergence of WPS and ECG signals harmoniously take place within the ICMT-Net. The fusion of 1D WPS signals with their corresponding spectrograms further accentuates the potential relationship inherent in these physiological signals, thereby achieving classification performance that surpasses traditional methodologies. Ultimately, this comprehensive approach effectively addresses the pressing need for accurate cardiovascular disease diagnosis, bridging the gap between demand and precision.

TABLE X
COMPARISONS BETWEEN OUR METHOD AND PREVIOUS METHODS FOR THE CLASSIFICATION PERFORMANCE OF TYPE N, S, AND V UNDER THE INTER-PATIENT PARADIGM.

Method	N Se/Pp (%)	V Se/Pp (%)	S Se /Pp (%)
Linear Discriminants [28]	**99.16**/86.86	81.59/77.74	38.53/**75.94**
PSO-SVM [2]	94.00/98.00	87.34/59.44	61.96/52.96
DNN [9]	91.89/97.00	89.23/50.85	62.49/55.86
CNN with Batch-Weighted Loss [8]	88.51/98.80	92.05/72.13	82.04/30.44
Improved Deep ResNet [14]	94.54/93.33	88.35/**79.86**	35.22/65.88
ICMT-Net (ours)	80.06/**99.45**	**92.17**/76.44	**89.31**/32.35

TABLE XI
COMPARISONS BETWEEN OUR METHOD AND PREVIOUS METHODS FOR THE CLASSIFICATION PERFORMANCE OF TYPE N, S, AND V UNDER THE INTER-PATIENT PARADIGM.

Method	Sunken Se/Pp(%)	Moderate Se/Pp(%)	Normal Se/Pp (%)	Skipping Se/Pp(%)
GA-BP algorithm [22]	98.47/95.24	**96.63/100**	95.10/98.82	98.27/97.42
1D CNN [23]	97.63/95.92	92.70/97.06	96.80/96.61	97.40/99.56
TCN [24]	95.76/97.24	**96.63/96.63**	96.80/9492	96.39/96.81

TABLE XII
COMPARISON OF THE CLASSIFICATION PERFORMANCE OF OUR METHOD WITH EXISTING METHODS FOR THE HEALTHY PULSE AND CHD IN THE PULSE CLASSIFICATION TASK.

Method	Health Se/Pp (%)	CHD Se/Pp (%)
GA-BP algorithm [22]	98.97/96.59	90.91/97.14
1D CNN [23]	99.18/96.21	89.84/97.67
TCN [24]	97.74/95.38	87.70/93.71
ICMT-Net (ours)	**99.59/99.38**	**98.40/98.92**

VI. Conclusion

This paper introduces an ICMT-Net network for the classification of arrhythmia, pulse patterns, and CHD. Notably, this network operates solely on raw ECG data and random 5-second ECG segments, bypassing the need for resource-intensive pre-

processing or complex feature extraction methods. The conducted experiments highlight the effectiveness of the CBAM attention mechanism, which refines the network's focus on crucial feature maps, thus elevating overall classification performance. By incorporating both the transformer layer and MLP fused layer, the network becomes adept at harnessing the attributes of both 1D signals and 2D spectrograms. This amplifies its feature extraction capabilities and subsequently improving classification performance. In a comparative analysis against existing ECG classification methods within the inter-patient paradigm, our network demonstrates exceptional sensitivity in type V and S heartbeat segments.

Furthermore, the ICMT-Net, initially pre-trained for ECG classification, showcases seamless transferability to WPS classification. When compared to established pulse classification techniques, our proposed method excels in overall classification performance for WPS and CHD classification tasks. Consequently, the model presented in this paper effectively addresses the persistent challenge of suboptimal classification performance observed in previous models. Additionally, our proposed methodology has versatility to extend its utility to the classification of various physiological signal types, significantly amplifying its pragmatic utility.

This study also fills signification research gap by addressing the dearth of methods that integrate both 1D and 2D approaches within physiological signals analysis. However, this approach belongs to the field of supervised learning, and the scarcity problem of physiological signals leads to the inability to achieve optimal network parameters and classification capability, regardless of how the network framework and hyperparameters are adjusted.

In the future, we plan to explore semi-supervised learning methods as a solution to the signal classification challenge. Semi-supervised learning enables the utilization of more unlabeled data, which can effectively mitigate sample scarcity and bring the solution closer to real-time applicability. Another avenue for potential enhancement involves establishing pre-trained networks that leverage transferred learning from the realm of ECG signal classification to other tasks in physiological signal classification.

Declaration of competing interest

The manuscript is confirmed to be exclusively submitted to this journal and is not under consideration or published elsewhere. The authors affirm that there are no known financial or personal relationships that could have affected the findings presented in this article. No potential conflicts of interest exist related to the research, authorship, or publication of this work. (Jiajian Zhu and Yue Feng contributed equally to this article.)

ACKNOWLEDGMENT

This work is supported by the Basic Research and Applied Basic Research Key Project in General Colleges and Universities of Guangdong Province, China (Grant No. 021ZDZX1032); the Special Project of Guangdong Province, China (Grant No. 2020A1313030021); and the Scientific Research Project of Wuyi University (Grants No. 2018TP023 and 2018GR003).

REFERENCES

[1] G. Zhang and N. J. Navimipour, "A comprehensive and systematic review of the IoT-based medical management systems: Applications, techniques, trends and open issues," Sustainable Cities and Society, vol. 82, p. 103914, 2022.

[2] G. De Lannoy, D. François, J. Delbeke, and M. Verleysen, "Weighted SVMs and feature relevance assessment in supervised heart beat classification," in Biomedical Engineering Systems and Technologies: Third International Joint Conference, BIOSTEC 2010, Valencia, Spain, January 20-23, 2010, Revised Selected Papers 3, 2011, pp. 212-223: Springer.

[3] K. K. Patro, A. Jaya Prakash, M. Jayamanmadha Rao, and P. Rajesh Kumar, "An efficient optimized feature selection with machine learning approach for ECG biometric recognition," IETE Journal of Research, vol. 68, no. 4, pp. 2743-2754, 2022.

[4] X. Liu, H. Wang, Z. Li, and L. Qin, "Deep learning in ECG diagnosis: A review," Knowledge-Based Systems, vol. 227, p. 107187, 2021.

[5] E. Kıymaç and Y. Kaya, "A novel automated cnn arrhythmia classifier with memory-enhanced artificial hummingbird algorithm," Expert Systems with Applications, vol. 213, p. 119162, 2023.

[6] Y. Li, Y. Pang, J. Wang, and X. Li, "Patient-specific ECG classification by deeper CNN from generic to dedicated," Neurocomputing, vol. 314, pp. 336-346, 2018.

[7] 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.

[8] A. Sellami and H. Hwang, "A robust deep convolutional neural network with batch-weighted loss for heartbeat classification," Expert Systems with Applications, vol. 122, pp. 75-84, 2019.

[9] J. Takalo-Mattila, J. Kiljander, and J.-P. Soininen, "Inter-patient ECG classification using deep convolutional neural networks," in 2018 21st Euromicro Conference on Digital System Design (DSD), 2018, pp. 421-425: IEEE.

[10] J. Wang, T. Liu, M. Zang, and Q. Wang, "CLSTM-Transformer: inter-patient ECG classification for ventricular arrhythmia," in International Conference on Optics and Machine Vision (ICOMV 2023), 2023, vol. 12634, pp. 108-113: SPIE.

[11] H. Lyu et al., "Automated inter-patient arrhythmia classification with dual attention neural network," Computer Methods and Programs in Biomedicine, vol. 236, p. 107560, 2023.

[12] S. Mousavi, F. Afghah, F. Khadem, and U. R. Acharya, "ECG Language processing (ELP): A new technique to analyze ECG signals," Computer methods and programs in biomedicine, vol. 202, p. 105959, 2021.

[13] L. Guo, G. Sim, and B. Matuszewski, "Inter-patient ECG classification with convolutional and recurrent neural networks," Biocybernetics and Biomedical Engineering, vol. 39, no. 3, pp. 868-879, 2019.

[14] Y. Li, R. Qian, and K. Li, "Inter-patient arrhythmia classification with improved deep residual convolutional neural network," Computer Methods and Programs in Biomedicine, vol. 214, p. 106582, 2022.

[15] T. J. Jun, H. M. Nguyen, D. Kang, D. Kim, D. Kim, and Y.-H. Kim, "ECG arrhythmia classification using a 2-D convolutional neural network," arXiv preprint arXiv:1804.06812, 2018.

[16] Y. Xia, N. Wulan, K. Wang, and H. Zhang, "Detecting atrial fibrillation by deep convolutional neural networks," Computers in biology and medicine, vol. 93, pp. 84-92, 2018.

[17] J. Huang, B. Chen, B. Yao, and W. He, "ECG arrhythmia classification using STFT-based spectrogram and convolutional neural network," IEEE access, vol. 7, pp. 92871-92880, 2019.

[18] P. Lu et al., "Improving Classification of Tetanus Severity for Patients in Low-Middle Income Countries Wearing ECG Sensors by Using a CNN-Transformer Network," IEEE Transactions on Biomedical Engineering, 2022.

[19] Y. Chen, L. Zhang, D. Zhang, and D. Zhang, "Computerized wrist pulse signal diagnosis using modified auto-regressive models," Journal of Medical Systems, vol. 35, pp. 321-328, 2011.

[20] D. Wang, D. Zhang, and G. Lu, "A robust signal preprocessing framework for wrist pulse analysis," Biomedical Signal Processing and Control, vol. 23, pp. 62-75, 2016.

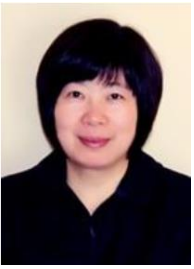
[21] X. Li et al., "Computerized wrist pulse signal diagnosis using gradient boosting decision tree," in 2018 IEEE international conference on bioinformatics and biomedicine (BIBM), 2018, pp. 1941-1947: IEEE.
[22] Z. Chen, A. Huang, and X. Qiang, "Improved neural networks based on genetic algorithm for pulse recognition," Computational Biology and Chemistry, vol. 88, p. 107315, 2020.
[23] S.-R. Zhang and Q.-F. Sun, "Human pulse recognition based on convolutional neural networks," in 2016 International Symposium on Computer, Consumer and Control (IS3C), 2016, pp. 366-369: IEEE.
[24] J. Yan, G. Zhu, R. Guo, Y. Wang, and H. Yan, "TCN-Based Diagnostic Model for the Severity of Coronary Atherosclerotic Heart Disease Using Wrist Pulse Wave Sequence," in The Fifth International Conference on Biological Information and Biomedical Engineering, 2021, pp. 1-6.
[25] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 11976-11986.
[26] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 3-19.
[27] 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.
[28] P. De Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," IEEE transactions on biomedical engineering, vol. 51, no. 7, pp. 1196-1206, 2004.
[29] 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.
[30] M. Thomas and A. T. Joy, Elements of information theory. Wiley-Interscience, 2006.
[31] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4510-4520.
[32] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in International conference on artificial neural networks, 2018, pp. 270-279: Springer.
[33] P. Goyal et al., "Accurate, large minibatch sgd: Training imagenet in 1 hour," arXiv preprint arXiv:1706.02677, 2017.

Jiajian Zhu was born in 1998 in Maoming, Guangdong Province, China. He is pursuing a master's degree in the Department of Intelligent Manufacturing at Wuyi University, Jiangmen City, Guangdong Province, China, in 2020, and will graduate in 2023. Research interests in advanced information processing and application technologies.

Yue Feng currently serves as a Graduate Advisor at Wuyi University. Born in 1959, he specializes in the academic domains of Computer Vision, Biometric Recognition, and Artificial Intelligence Algorithms.

Research Profile: Prof. Feng has accumulated over three decades of extensive experience in developing intelligent technologies. His career has encompassed work in both large-scale enterprises and collaborative research projects with universities, providing him with a strong foundation in applied research and team leadership. Presently, he holds the position of Distinguished Professor in the Faculty of Intelligent Manufacturing at Wuyi University, where he also mentors master's students and leads the Intelligent Medical Laboratory. Prof. Feng is recognized as a top-tier foreign expert (Category A) by the National Foreign Experts Bureau and holds the distinction of being a high-level talent in Jiangmen City.

Qichao Liu was born in 1996 and completed his undergraduate studies at Jianghan University in Wuhan, Hubei Province. In 2021, he embarked on his graduate studies at Wuyi University in Jiangmen, Guangdong Province. His research interests encompass deep learning algorithms, machine vision, as well as the analysis and processing of biomedical signals.



Dr. Hong Xu is a Ph.D. supervisor at the University of Victoria. Since 2013, she has been actively involved with the Information Technology Expert Team at the University of Victoria, focusing on the development of Traditional Chinese Medicine (TCM) and health information technology for the prevention and management of chronic diseases. Dr. Xu has also held a part-time position as a Distinguished Professor at Wuyi University since late 2018, where she collaborates with a medical intelligence research team to develop intelligent TCM diagnostic applications.

Yuan Miao earned his Ph.D. degree from the Department of Automation at Tsinghua University. He began his academic career as an Associate Professor in the Department of Computer Science and Technology at Tsinghua University and later extended his academic endeavors to the University of Melbourne in Australia, Nanyang Technological University in Singapore, and Victoria University in Australia. Currently, he serves as a Professor in the Faculty of Arts, Business, Law, Education, and Information Technology, where he also heads the Information Technology (IT) discipline at Victoria University.

Zhuosheng Lin received his B.Sc. degree in Electronic Information Science and Technology and his Ph.D. degree in Control Science and Engineering from Guangdong University of Technology, Guangzhou, China, in 2013 and 2018, respectively. He currently holds the position of Lecturer at Wuyi University in Jiangmen, China. Dr. Lin's research interests encompass chaos theory, secure communication, and multimedia privacy protection.

Jia Li possesses extensive research experience in the field of artificial intelligence applied to healthcare. Her primary research focus centers around the development of arrhythmia diagnosis methods and the exploration of key technologies based on deep neural network models featuring diverse structures. Additionally, Dr. Li has a solid research background in ECG signal monitoring technology, encompassing the application of intelligent wearable devices, critical medical devices, interactive technology, biosensors, and GPU parallel computing technology.

Huilin Liu Master's degree student of Pattern Recognition and Intelligent Systems, class of 2019. She has won several academic scholarships during her master's degree and is now pursuing her Ph.D. degree at Shanghai University of Traditional Chinese Medicine

Xu Ying, born in 1989 in Yinchuan, Ningxia Province, graduated from Shanghai University of Traditional Chinese Medicine (SUTCM) with a PhD majoring in TCM internal medicine, and is mainly engaged in the teaching of TCM four-diagnosis practical training and TCM four-diagnosis objectivity research, and is currently responsible for the Shanghai Shanghai Young Scientific and Technological Talents Sailing Program, the National Natural Fund for Young People, the China Postdoctoral Fund, and Shanghai University of Traditional Chinese Medicine (SUTCM) budgeted for a total of four scientific research projects, and has published 14 papers as the first author or corresponding author, and has participated in applying for 2 invention patents and 1 authorization.

Li Fufeng, Ph.D., professor and doctoral supervisor of Shanghai University of Traditional Chinese Medicine (SUTCM), now reserve expert of TCM diagnostics of key disciplines of the State Administration of Traditional Chinese Medicine (SATCM), academic backbone of TCM diagnostics of key disciplines of Shanghai Municipality, executive director of TCM Diagnostic Instrumentation Committee of the World Federation of Traditional Chinese Medicine (WFTCM), executive director of the Health Management Branch of the Chinese Society of Traditional Chinese Medicine (CSTCM), and registered expert of the international standardization of Chinese medicine in ISO/TC249. She is also the member and secretary-general of the Chinese Medicine Section Group of the Ministry of Education's Virtual Simulation Innovation Alliance in the field of medicine. In terms of scientific research, she has undertaken more than 30 projects such as the National Natural Science Foundation of China and the sub-themes of the National 13th Five-Year Plan.