Dear Author,

Here are the proofs of your article.

- You can submit your corrections **online**, via **e-mail** or by **fax**.
- For **online** submission please insert your corrections in the online correction form. Always indicate the line number to which the correction refers.
- You can also insert your corrections in the proof PDF and **email** the annotated PDF.
- For fax submission, please ensure that your corrections are clearly legible. Use a fine black pen and write the correction in the margin, not too close to the edge of the page.
- Remember to note the **journal title**, **article number**, and **your name** when sending your response via e-mail or fax.
- **Check** the metadata sheet to make sure that the header information, especially author names and the corresponding affiliations are correctly shown.
- **Check** the questions that may have arisen during copy editing and insert your answers/ corrections.
- **Check** that the text is complete and that all figures, tables and their legends are included. Also check the accuracy of special characters, equations, and electronic supplementary material if applicable. If necessary refer to the *Edited manuscript.*
- The publication of inaccurate data such as dosages and units can have serious consequences. Please take particular care that all such details are correct.
- Please **do not** make changes that involve only matters of style. We have generally introduced forms that follow the journal's style.
  Substantial changes in content, e.g., new results, corrected values, title and authorship are not allowed without the approval of the responsible editor. In such a case, please contact the Editorial Office and return his/her consent together with the proof.
- If we do not receive your corrections **within 48 hours**, we will send you a reminder.
- Your article will be published **Online First** approximately one week after receipt of your corrected proofs. This is the **official first publication** citable with the DOI. **Further changes are, therefore, not possible.**
- The **printed version** will follow in a forthcoming issue.

Please note

After online publication, subscribers (personal/institutional) to this journal will have access to the complete article via the DOI using the URL: http://dx.doi.org/[DOI].
If you would like to know when your article has been published online, take advantage of our free alert service. For registration and further information go to: http://www.link.springer.com.

Due to the electronic nature of the procedure, the manuscript and the original figures will only be returned to you on special request. When you return your corrections, please inform us if you would like to have these documents returned.

Metadata of the article that will be visualized in OnlineFirst

ArticleTitle	SAPTSTA-AnoECG: a PatchTST-based ECG anomaly detection method with subtractive attention and data augmentation	
Article Sub-Title		
Article CopyRight	The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature (This will be the copyright line in the final PDF)	
Journal Name	Applied Intelligence	
Corresponding Author	FamilyName	**Chen**
	Particle	
	Given Name	**Tieming**
	Suffix	
	Division	College of Geoinformatics
	Organization	Zhejiang University of Technology
	Address	Changhong East Road, Huzhou, 313299, Zhejiang, China
	Phone	
	Fax	
	Email	tmchen@zjut.edu.cn
	URL	
	ORCID	
Author	FamilyName	**Li**
	Particle	
	Given Name	**Yifan**
	Suffix	
	Division	College of Computer Science
	Organization	Zhejiang University of Technology
	Address	Liuhe Road, Hangzhou, 310023, Zhejiang, China
	Phone	
	Fax	
	Email	yifanli@zjut.edu.cn
	URL	
	ORCID	
Author	FamilyName	**Wang**
	Particle	
	Given Name	**Mengjue**
	Suffix	
	Division	
	Organization	Xixi Hospital
	Address	Liuhe Road, Hangzhou, 310023, Zhejiang, China
	Phone	
	Fax	
	Email	495185849@qq.com
	URL	
	ORCID	
Author	FamilyName	**Guan**
	Particle	
	Given Name	**Mingxiang**
	Suffix	

	Division	School of Electronic Communication Technology
	Organization	Shenzhen Institute of Information Technology
	Address	Longxiang Street, Shenzhen, 518172, Guangdong, China
	Phone	
	Fax	
	Email	gmx2020@126.com
	URL	
	ORCID	
Author	FamilyName	**Lu**
	Particle	
	Given Name	**Chen**
	Suffix	
	Division	School of Electronic Communication Technology
	Organization	Shenzhen Institute of Information Technology
	Address	Longxiang Street, Shenzhen, 518172, Guangdong, China
	Phone	
	Fax	
	Email	luchen_0826@qq.com
	URL	
	ORCID	
Author	FamilyName	**Li**
	Particle	
	Given Name	**Zhiyong**
	Suffix	
	Division	College of Information Science and Engineering
	Organization	Hunan University
	Address	South Lushan Road, Changsha, 310023, Hunan, China
	Phone	
	Fax	
	Email	zhiyong.li@hnu.edu.cn
	URL	
	ORCID	

Schedule	Received	
	Revised	
	Accepted	31 Oct 2024

Abstract

An electrocardiogram (ECG) is a crucial noninvasive medical diagnostic method that enables real-time monitoring of the electrical activity of the heart. ECGs hold a significant position in the rapid diagnosis and routine monitoring of cardiac diseases due to their user-friendly operation, prompt detection, broad range of diagnosable problems, and cost-effectiveness. However, thorough comprehension of ECG readings requires a high level of medical expertise due to the complex variations in ECG patterns, substantial interindividual differences, and numerous interfering factors. Consequently, current ECG machines and ECG Holters typically provide simplistic indications of ECG anomalies. Nonetheless, current ECG anomaly detection (EAD) algorithms lack precision; therefore, these medical devices cannot accurately report the specific types of diseases reflected in ECG results. In response to these challenges, this paper proposes enhancing the accuracy of electrocardiogram detection by improving algorithms. Therefore, we propose SAPTSTA-AnoECG, a PatchTST-based ECG anomaly detection method with subtractive attention and data augmentation. This method introduces a subtractive attention mechanism to make the Transformer architecture more suitable for time series data. We also use data augmentation to increase the robustness of the model. In addition, a patch-based approach is employed to reduce the algorithm's computational complexity of the model. Furthermore, we introduce a new publicly available ECG dataset named HCE in this paper and conduct comparative experiments using this dataset along with the PTB-XL and CPSC 2018 datasets. The experimental results demonstrate the effectiveness of this method.

Keywords (separated by '-') ECG anomaly detection - ECG classification - Patch time series transformer - Data augmentation - Subtractive attention mechanism

Footnote Information

Applied Intelligence _####################_
https://doi.org/10.1007/s10489-024-05881-5

SAPTSTA-AnoECG: a PatchTST-based ECG anomaly detection method with subtractive attention and data augmentation

Yifan Li[1] · Mengjue Wang[2] · Mingxiang Guan[3] · Chen Lu[3] · Zhiyong Li[4] · Tieming Chen[5]

Accepted: 31 October 2024

Abstract

An electrocardiogram (ECG) is a crucial noninvasive medical diagnostic method that enables real-time monitoring of the electrical activity of the heart. ECGs hold a significant position in the rapid diagnosis and routine monitoring of cardiac diseases due to their user-friendly operation, prompt detection, broad range of diagnosable problems, and cost-effectiveness. However, thorough comprehension of ECG readings requires a high level of medical expertise due to the complex variations in ECG patterns, substantial interindividual differences, and numerous interfering factors. Consequently, current ECG machines and ECG Holters typically provide simplistic indications of ECG anomalies. Nonetheless, current ECG anomaly detection (EAD) algorithms lack precision; therefore, these medical devices cannot accurately report the specific types of diseases reflected in ECG results. In response to these challenges, this paper proposes enhancing the accuracy of electrocardiogram detection by improving algorithms. Therefore, we propose SAPTSTA-AnoECG, a PatchTST-based ECG anomaly detection method with subtractive attention and data augmentation. This method introduces a subtractive attention mechanism to make the Transformer architecture more suitable for time series data. We also use data augmentation to increase the robustness of the model. In addition, a patch-based approach is employed to reduce the algorithm's computational complexity of the model. Furthermore, we introduce a new publicly available ECG dataset named HCE in this paper and conduct comparative experiments using this dataset along with the PTB-XL and CPSC 2018 datasets. The experimental results demonstrate the effectiveness of this method.

Keywords ECG anomaly detection · ECG classification · Patch time series transformer · Data augmentation · Subtractive attention mechanism

1 Introduction

The heart is one of the most important organs in the human body, and the number of deaths and disabilities due to cardiovascular disease remains high every year [1]. The ECG signal is valued by medical professionals as a type of cardiac information that can be obtained noninvasively, quickly and accurately. ECG has since become one of the most commonly used diagnostic tools for cardiovascular disease. However, the heavy workload of ECG reading increases the workload of medical personnel and makes it difficult for patients to understand the type of disease their ECG reflects in the first place. Therefore, many scholars have begun to focus on using computers to automatically analyze the meaning of ECGs.

First, we briefly introduce the basics related to electrocardiography. When conducting an ECG test, it is necessary to place several electrodes at fixed positions on the subject's body. Typically, the more electrodes placed on the patient, the more leads of ECG data are obtained. In hospitals, doctors typically utilize a 12-lead ECG machine or ECG Holter to examine patients. This is because the medical community considers the 12-lead configuration capable of providing more detailed and comprehensive cardiac electrical activity information, thereby enabling doctors to draw more accurate conclusions. The 12-lead ECG uses 6 limb leads (I, II, III, aVR, aVL, aVF) and 6 chest leads (V1, V2, V3, V4, V5, and V6). The limb leads were placed on the arms and legs, whereas the chest leads were placed on the chest wall. A 12-dimensional time series (i.e., a 12-lead ECG) is obtained by measuring over a period of time. The ECG signal obtained

Extended author information available on the last page of the article

Journal: 10489 MS: 5881 ☐ TYPESET ☐ DISK ☐ LE ☐ CP Disp.:2024/12/3 Pages: 15 Layout: Large

from the normal leads has a certain periodicity, with each cycle representing one heartbeat. Therefore, the change in cycle length can reflect the change in heart rate. The changes in the graph of each cycle are shown in Fig. 1. One cycle of the ECG signal contains P waves, QRS wave groups, T waves and U waves. P waves, QRS wave groups and T waves correspond to atrial contraction, ventricular contraction and ventricular diastole, respectively, whereas the mechanism of U wave generation is not fully understood.

Although ECGs involve a great deal of specialized knowledge, their anomaly detection falls within the realm of time series anomaly detection [2], similar to industrial fault diagnosis [3], network traffic anomaly detection [4], traffic accident detection [5], etc, this implies that ECG anomaly detection can also be achieved via certain time series analysis methods [6]. Many researchers currently use machine learning or deep learning time series classification methods for the EAD problem. Typically, machine learning methods for ECG anomaly detection can be divided into three steps: preprocessing, feature extraction and signal classification. Preprocessing removes some of the noise and reduces the interference of artificial artifacts; feature extraction extracts various useful waveform components of the signal; and signal classification uses a suitable machine learning model to perform the final signal classification. However, deep learning methods tend to yield better results, so this paper focuses on deep learning-based EAD methods. The feature extraction step is sometimes downplayed because it tends to have some feature extraction capability itself.

Currently, there are still considerable challenges in using deep learning for solving EAD problems. First, normal ECG signals are temporal and periodic, there are certain patterns in the morphology of each cycle, and how to effectively use these characteristics of ECGs when building models is important. In addition, challenges common to the anomaly detection problem must be addressed, such as how to distinguish heterogeneous abnormality classes and how to distinguish abnormal data from noise; this corresponds to the presence of multiple diseases in the ECG and the condition of artificial artifacts in the ECG. To solve the problems above, this paper proposes the SAPTSTA-AnoECG method, which uses a Transformer-based model with data augmentation and improves the attention mechanism to handle time series effectively. The patch technique is subsequently used to reduce the time complexity of the algorithm while improving the accuracy. Specifically, our main contributions are as follows:

- We apply the PatchTST [7] model to our SAPTST model by reconstructing samples and using the reconstruction error to detect anomalies. This approach enables the model to retain the temporal data processing capabilities of Transformer while reducing time complexity.
- We propose a new attention mechanism called the subtractive attention mechanism based on R peaks in our SAPTSTA model, which makes the model more suitable for periodic time series data such as ECG signals.
- We perform data augmentation on samples through small-scale longitudinal compression and stretching to generate pseudo-ECG samples that closely resemble real ECG samples, thereby expanding the dataset. This approach enhances the model's robustness.
- We propose a new ECG dataset and validate the effectiveness of our method on this dataset.

Section 2 of this paper introduces related work, Section 3 outlines several foundational methods used in this study, Section 4 focuses on the proposed method, Section 5 presents experimental validation, and Section 6 provides a summary of the entire document.

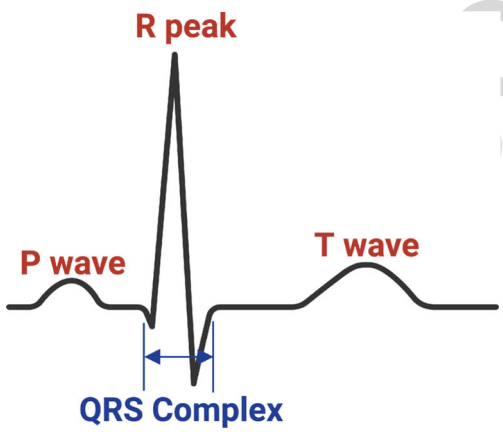

Fig. 1 Introduction of waves in ECG signal. A 12-lead ECG signal is a 12-dimensional time series data, with each dimension (corresponding to one ECG lead) being a periodic time series data. The morphology of the ECG signal for each lead is shown in the figure

2 Related work

In this section, we present the developmental history of the ECG anomaly detection and classification models. We provide an overview of both traditional methods and deep learning methods in this introduction.

2.1 Traditional methods for solving the EAD problem

Before computers are widely used, doctors must read ECGs manually, which requires considerable time. Then, rule-based methods for solving EAD problems are developed [8]. These methods utilize specific rules, such as heart rate and

Journal: 10489 MS: 5881 TYPESET DISK LE CP Disp.:2024/12/3 Pages: 15 Layout: Large

voltage, to determine whether an ECG is normal or not, which is similar to the reasoning process of a doctor. However, due to the diversity of ECG signals and the presence of multiple anomalies occurring simultaneously, rule-based methods have shown limited accuracy. Later, more advanced methods for solving EAD problems, such as nonlinear transformation methods, PCA-based methods and NN-based methods, were widely used and studied [9]. However, compared with deep learning, these algorithms still exhibit lower accuracy.

2.2 Deep learning methods for solving the EAD problem

Since 2019, researchers have focused on deep learning methods for solving EAD problems. In 2019, Strodthoff et al. [10] compared the performance of all kinds of deep learning algorithms (i.e., Inception1d [11], XResnet1d [12], Resnet-Wang, and FCN-Wang [13]) and further proposed an ensemble learning-based model to fuse these methods to increase performance. In 2021, Liu et al. [14] summarized deep learning methods based on stacked autoencoders (SAEs) or autoencoders (AEs) [15, 16], deep belief networks (DBNs) [17, 18], convolutional neural networks (CNNs) [19, 20], and long short-term memory networks (LSTMs) [21, 22] to solve the EAD problem. In 2022, Zhao et al. proposed the ECGNN [23], which uses the GNN to learn the relationships among the twelve leads of the ECG signals.

3 Preliminaries

This section briefly introduces the PatchTST method and the R peak detection method in Neurokit2, which are used in our methodology.

3.1 PatchTST

The patch time series Transformer (PatchTST) [7] is a time series prediction model that is based on the Transformer architecture; it is inspired by the well-known model ViT [24] in the field of computer vision and uses patches to divide the samples. By using a patch structure, the number of tokens T is reduced from the length of the time series L to the number of patches N. Since the complexity of attention is linearly related to T^2, using patches can significantly reduce the complexity of the algorithm. However, unlike ViT, PatchTST uses a method called channel-independence, which treats each channel (dimension) as an independent input, allowing each Transformer backbone to handle a 1-dimensional sequence. The framework of PatchTST is depicted in Fig. 2. The entire framework of PatchTST can be divided into three parts: data segmentation, model prediction,

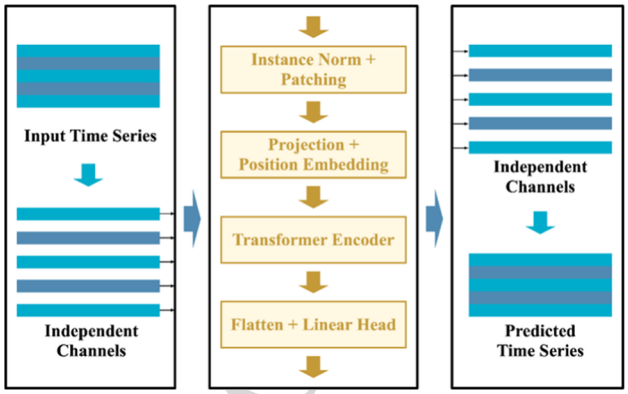

Fig. 2 Framework of PatchTST

and prediction value concatenation. First, in the data segmentation stage, PatchTST divides each k-dimensional time series sample into k 1-dimensional time series. After normalization, each 1-dimensional time series is further divided into patches of length P. When the patches are nonoverlapping, the number of patches is $\lceil L/P \rceil$. Next, in the model prediction stage, PatchTST inputs each 1-dimensional time series into a Transformer structure, treating each patch as a token for prediction. Finally, in the prediction value concatenation stage, PatchTST combines the output results from all channels to obtain the final prediction result. PatchTST demonstrates good performance in multidimensional time series prediction, indicating its ability to effectively learn features from each dimension of the time series. In this work, we employ a framework based on PatchTST, which uses well-reconstructed ECG signals as normal samples and poorly-reconstructed ECG signals as abnormal samples to achieve the diagnosis of various diseases via ECGs. According to Reference [7], PatchTST has a lower time complexity than the Transformer does, which allows it to detect ECG anomalies more quickly.

3.2 R peak detection method in neurokit2

A normal electrocardiogram (ECG) is a highly periodic time series where each cycle represents a heartbeat. Since many diseases can be diagnosed based on characteristics such as the length of the ECG cycle, regularity of the ECG cycle, and morphology of each cycle in the ECG, there has been significant interest among researchers in accurately and quickly obtaining the ECG cycle. Since the R peak is typically located at the extremum in the entire electrocardiogram, it is easier to detect than other waveforms. Therefore, the R–R interval is commonly used to calculate the cardiac cycle. In the Neurokit2 package of Python, a method for locating R peaks is provided [25], which first calculates the steepness of the absolute gradient of the ECG signals and infers the location

Journal: 10489 MS: 5881 TYPESET DISK LE CP Disp.:2024/12/3 Pages: 15 Layout: Large

of the QRS complexes from it. The position of the R peak is subsequently determined based on the local extremes in the inferred QRS complexes. In this work, we use the R peak obtained via this method to determine the range of each heartbeat.

4 Methodology

4.1 Problem statement

Given a set of n 12-lead ECG signal samples $\mathbf{X} := (\mathbf{x}_1, \mathbf{x}_2..., \mathbf{x}_n)$, the length of each sample can be denoted as $L = \{l_1, l_2..., l_n\}$, respectively (i.e. $\mathbf{x}_t \in \mathbb{R}^{l_t \times 12}$). The whole set can be divided into a training set $\mathbf{X}_{tr}$ and a testing set $\mathbf{X}_{te}$. Suppose there are a total of k classes of labels (including $(k-1)$ abnormal ECG statements and a normal ECG statement) for all the ECG signals; then, the label of each ECG signal can be represented by a k-dimensional vector. Therefore, the labels of X can be represented by $\mathbf{C} = \{c_1, c_2..., c_n\}$, where $|c_t| = k$. After training on $\mathbf{X}_{tr}$, the goal of EAD is to provide the anomaly score of the testing samples $\mathbf{S_{te}} := (\mathbf{s}_1, \mathbf{s}_2..., \mathbf{s}_{n_{te}}) \in \mathbb{R}^{n_{te} \times k}$. Using a threshold τ, the final label(s) of $\mathbf{X}_{te}$ denoted as $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2..., \mathbf{z}_{n_{te}}\}$ can be further obtained.

4.2 Key ideas and the overall framework

To solve the problem of ECG signal anomaly detection, this section proposes the SAPTST-AnoECG method, which is based on subtractive attention, and further combines data enhancement to propose the SAPTSTA-AnoECG method. The overall frameworks of these two methods are shown in Fig. 3 (the submodules with "*" are specific to the SAPTSTA). The raw ECG signals are split into training and testing sets and then preprocessed to obtain the preprocessed data and the corresponding constructed series. Then, the preprocessed data and constructed series of the training set are input into the SAPTSTA model (or the SAPTST model) for model training. Finally, the preprocessed test set and the test set constructed series are input into the SAPTSTA model (or the SAPTST model) for anomaly detection.

The key idea of our methodology is as follows. First, we use the PatchTST model, which not only retains the characteristics of the Transformer model, such as parallel computation and the number of operations required for correlation between two positions not increasing with distance but also reduces the time complexity of the algorithm. Second, stretching and compressing methods are used for data augmentation, which enhances the robustness of the model. Third, it is assumed that the morphology of each R–R interval is similar, ensuring equal weights for the corresponding positions of each R–R interval. The subtractive attention mechanism is an optimization of traditional scaled dot-product attention, making it better suited for handling time series data.

4.3 Anomaly detection method based on subtractive attention and data augmentation

–ibes the proposed SAPTSTA-AnoECG and SAPTST-AnoECG ECG signal anomaly detection methods in detail, and explains the flow of the methods from three aspects: ECG

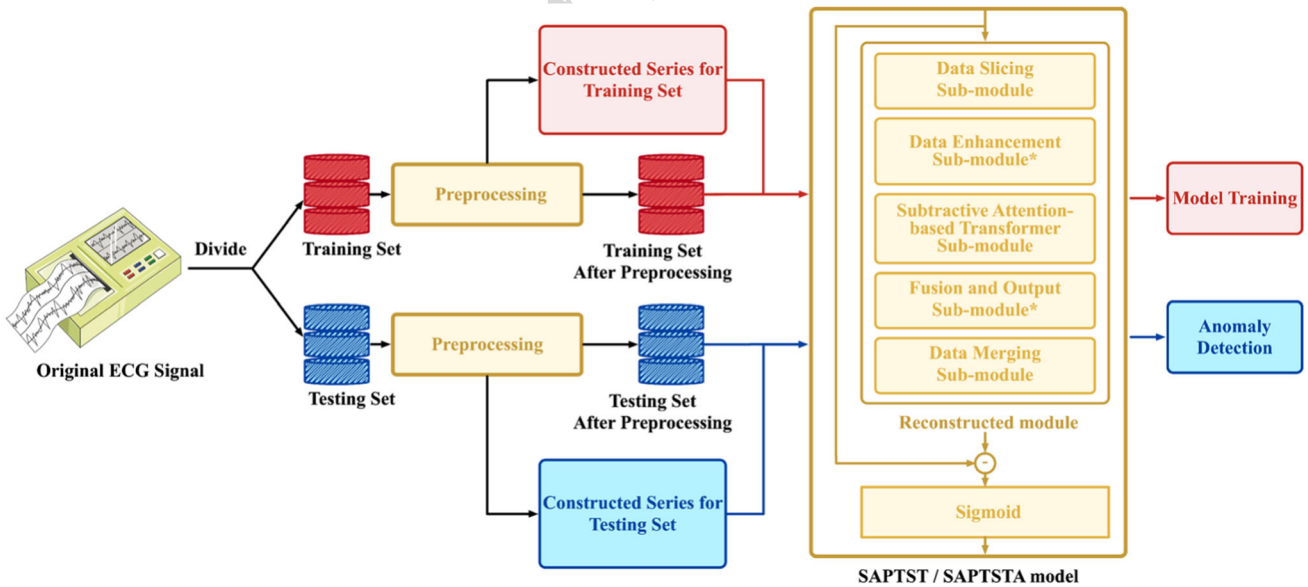

Fig. 3 The overall framework of our method

Journal: 10489 MS: 5881 TYPESET DISK LE CP Disp.:2024/12/3 Pages: 15 Layout: Large

signal preprocessing, model network structure, and model training & anomaly detection.

4.3.1 ECG signal preprocessing

Figure 4 shows the details of the signal preprocessing stage. Before training, we need to preprocess the ECG signals obtained from the original dataset (i.e., $\mathbf{X}$ in the Problem Statement section). The signal preprocessing stage mainly consists of data cleaning, attention method preparation and data segmentation.

In the data cleaning stage, the baseline drift of the ECG signal is removed and then normalized via min-max normalization and z-score standardization with the following formulas:

$$\begin{aligned} \mathbf{X}_{det} &= \mathbf{X} - \text{trend}(\mathbf{X}) \\ \mathbf{X}_{clean} &= \text{z-score}(\text{min-max}(\mathbf{X}_{det})) \end{aligned} \tag{1}$$

where trend$(\cdot)$ uses polynomial fitting to capture the overall trend of $\mathbf{X}$, min-max$(\cdot)$ can be formulated as $\frac{x-x_{min}}{x_{max}-x_{min}}$, z-score$(\cdot)$ can be formulated as $\frac{x-\mu_x}{\sigma_x}$.

In the attention method preparation stage, we use the attention method preparation module to generate similarity sequence segmentation in preparation for the improvements to the Transformer attention computation method in the following sections.

In the slicing stage, since the lengths of all samples are not exactly the same, and long samples are often difficult to learn, similar to most ECG signal anomaly detection methods, sliding window technology needs to be used for data segmentation. We use a sliding window of size w and stride p to divide each sample into several segments, and then assign the labels of the corresponding sample to all these segments. We use the same sliding window to process the similarity sequence, ensuring that the similarity sequence segments generated after the slicing stage correspond one-to-one with the data segments.

4.3.2 SAPTST model

In this section, a PatchTST-based ECG anomaly detection model with subtractive attention (SAPTST) is proposed, and its network structure is shown in Fig. 5. (The reconstruction module does not activate the optional submodules of "data enhancement" and "fusion and output" but only contains the submodules of "data slicing", "subtractive attention-based Transformer", and "data merging".)

PatchTST [7] performs well in time series prediction and is able to effectively reduce the time complexity of the model. In fact, the prediction effectiveness of PatchTST is attributed to the strong representational capacity of the Transformer model. Therefore, we believe that PatchTST can also be applied to the reconstruction of time series data. Building upon this idea, the SAPTST model uses PatchTST as the basis for its network structure. While the reconstruction of timing data can be achieved via only the PatchTST module, experiments have shown that simple use of the base PatchTST

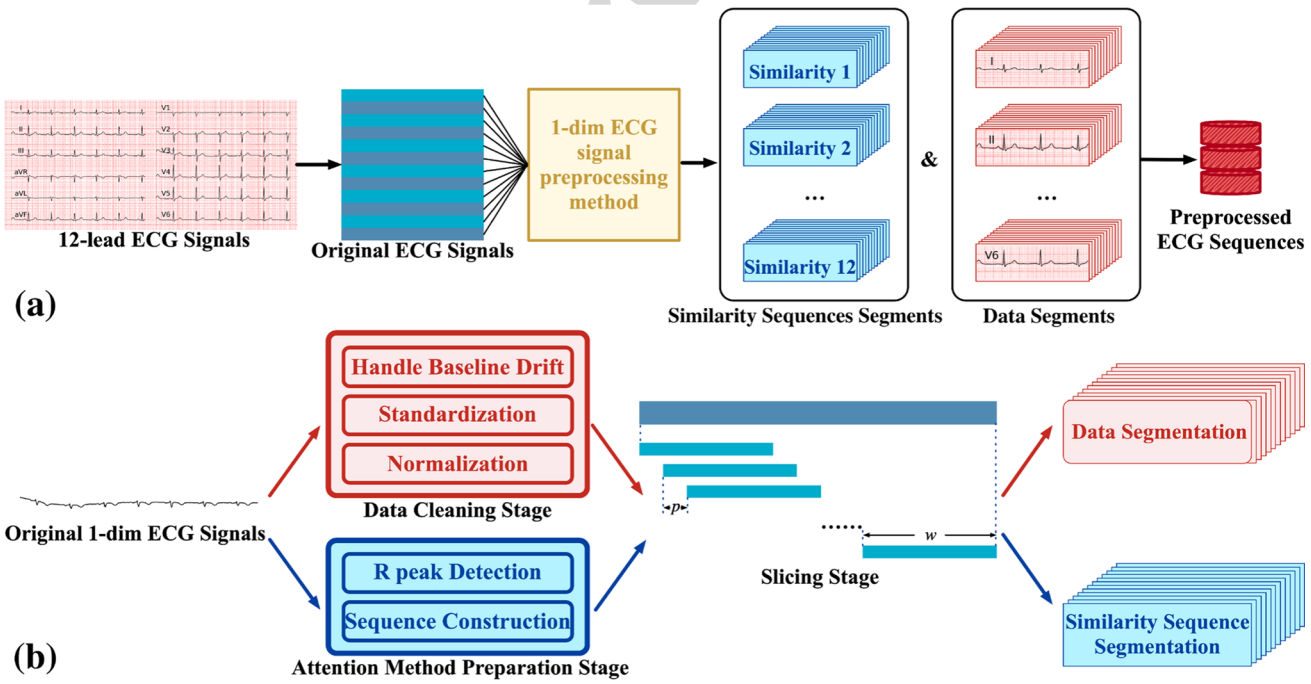

Fig. 4 Fig. (a) shows the whole preprocessing stage, Fig. (b) reveals the details of 1-dim ECG signal preprocessing block in Fig. (a)

Journal: 10489 MS: 5881 TYPESET DISK LE CP Disp.:2024/12/3 Pages: 15 Layout: Large

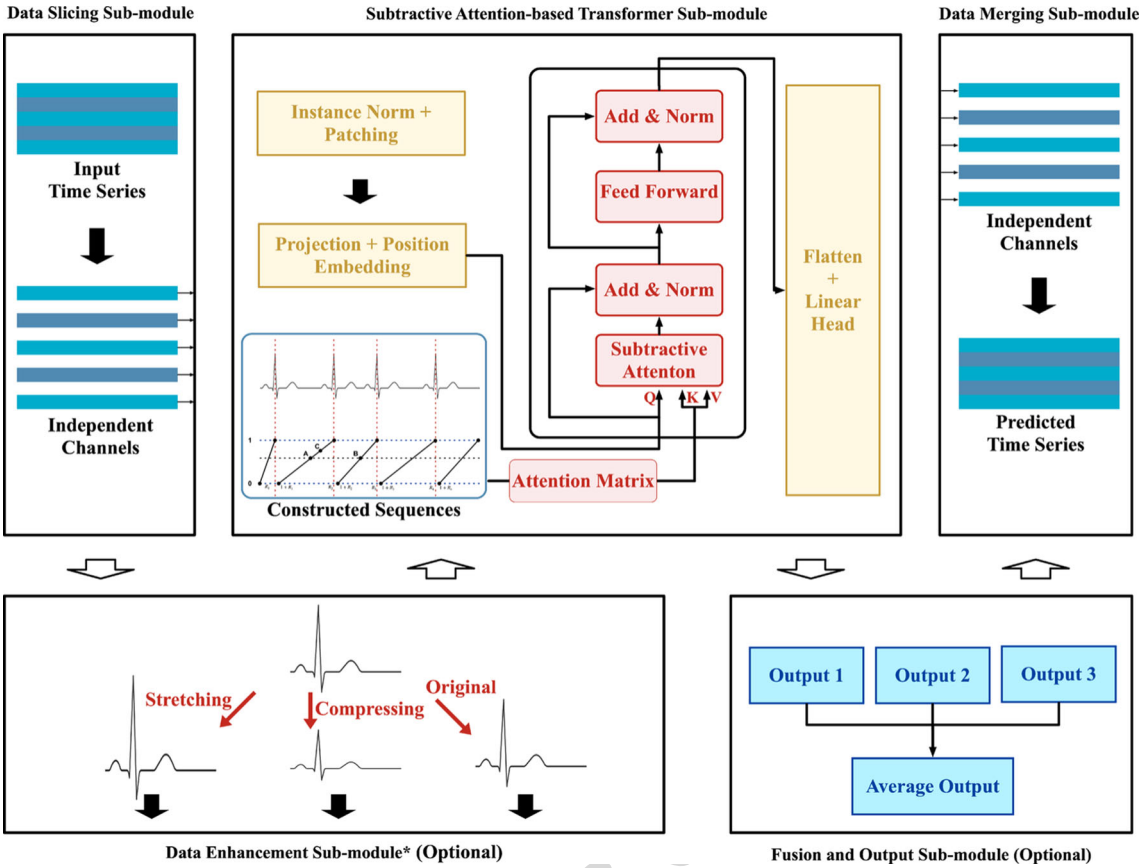

Fig. 5 The architecture of SAPTST/SAPTSTA block

model is ineffective in detecting anomalies, possibly because the original attention module in the Transformer is not suitable for ECG signals. For this reason, the SAPTST model changes the attention mechanism in the Vanilla Transformer. In the Vanilla Transformer, scaled dot-product attention is employed, and this attention mechanism is also inherited in PatchTST, which is formulated as follows:

$$A = \text{softmax}\left(\frac{QK^{\mathsf{T}}}{\sqrt{d_k}}\right)V \tag{2}$$

For the natural language processing task, Q represents the information of the word to be encoded, and K represents the information of the other words in the sentence, which are multiplied and scaled to obtain the weights of the rest of the words in the sentence with respect to the word to be encoded, while V represents the actual value. The Vanilla Transformer uses a self-attention mechanism (i.e., $Q = K = V$), where the weights are based solely on the data itself, and thus do not allow the periodicity of the ECG samples to be represented. Moreover, the positional encoding used in the Vanilla Transformer to reflect positional relationships also fails to take advantage of the periodicity of the data.

To solve the above problems, a subtractive attention mechanism is proposed in this section. This mechanism takes advantage of the similarity in the shape of each R–R interval of the ECG and assigns the same weight to the corresponding position in each R–R interval. First, we use the R peak detection method in the Neurokit2 package mentioned in Section III.B to find all the R-peaks of each lead from the 12-dimensional ECG signal. Then, we construct a line segment between each R–R interval, where the y-value at the left endpoint of the line segment is 0 and the y-value at the right endpoint is 1. Suppose that the R-peak series detected via the R-peak detection method is represented as $\mathcal{R} = \{R_1, R_2, ..., R_m\}$; then, the constructed function can be represented by the following formula:

$$f(t) = \begin{cases} \frac{t}{T_1} & 1 \le t \le T_1 \\ \frac{t-T_l-1}{T_r-T_l-1} & T_1 < t \le T_m \\ \frac{t-T_m-1}{n-T_m-1} & T_m < t \le n \end{cases} \tag{3}$$

where m is the length of $\mathcal{R}$, T_i is the timestamp of R_i, n is the total time duration of the ECG sample, $l = \arg\min_i(t - T_i)$ where $i \in \{x \mid T_x < t, x \in \mathbb{N}, x \in [1, m]\}$, $r = \arg\min_i(T_i - t)$ where $i \in \{x \mid T_x > t, x \in \mathbb{N}, x \in [1, m]\}$.

Journal: 10489 MS: 5881 TYPESET DISK LE CP Disp.:2024/12/3 Pages: 15 Layout: Large

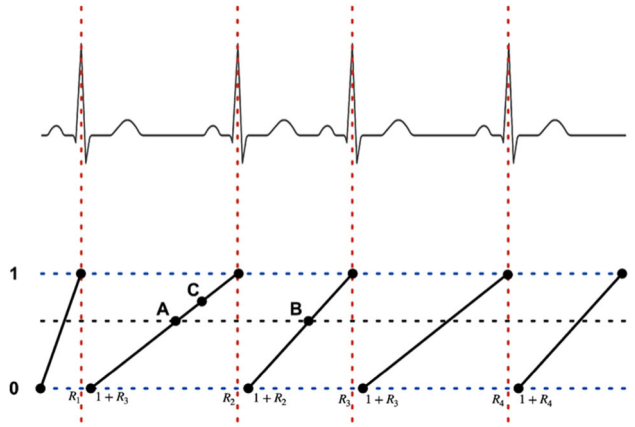

Fig. 6 Constructed series for attention strategy in SAPTST block

This method allows us to create the sequence shown in Fig. 6. The relative location of each point inside its corresponding R–R interval is represented by that point in the sequence. For example, even though there is a shorter time interval between point A and point C than between point A and point B, we still believe that the similarity between point A and point B is greater than the similarity between point A and point C because point A is in the same relative position as point B within the corresponding R–R interval. In the constructed series, the similarity increases with decreasing y-coordinate absolute differences, and vice versa.

Like self-attention in Vanilla Transformer, we also apply queries, keys and values in our framework. However, we use a subtractive attention mechanism based on R peaks. Suppose that the constructed series is $\mathcal{T} = \{f(1), f(2), f(3), ..., f(n)\}$; then, the newly constructed weight matrix of attention $\mathcal{W}_{\mathcal{T}}$ can be represented as:

$$\mathcal{W}_{\mathcal{T}} = \begin{pmatrix} |W_1 \cdot f(1) - W_2 \cdot f(1)| & ... & |W_1 \cdot f(1) - W_2 \cdot f(n)| \\ |W_1 \cdot f(2) - W_2 \cdot f(1)| & ... & |W_1 \cdot f(2) - W_2 \cdot f(n)| \\ ... & ... & ... \\ |W_1 \cdot f(n) - W_2 \cdot f(1)| & ... & |W_1 \cdot f(n) - W_2 \cdot f(n)| \end{pmatrix} \tag{4}$$

Then, the whole attention can be represented as:

$$\mathcal{A} = \text{softmax}\left(\frac{\mathcal{W}_{\mathcal{T}\,patch}}{\sqrt{d_k}}\right) * \tag{5}$$

where $\mathcal{W}_{\mathcal{T}\,patch}$ is the sub-matrix of $\mathcal{W}_{\mathcal{T}}$ which is corresponding to the certain patch of $\mathbf{X}$, i.e. $\mathbf{X}_{patch}$. W_1, W_2 and W_V refer to three different weighting matrices.

The use of this R-peak-based subtractive attention mechanism enables it to combine the weights with the characteristics of the actual ECG signals when processing the time series data and can lead to better integration of the ECG domain knowledge into the model.

4.3.3 SAPTSTA model

Using the SAPTST model alone can lead to problems of an insufficient amount of model input data and a lack of robustness. To increase the robustness of the model, we apply data augmentation in the model to the samples, and the samples processed via multiple data augmentation methods are input into our model separately. Finally, the average of the outputs is taken as the final output of the model. Specifically, since the difference between the ECGs obtained from the processing of the flip and rotation methods and the original ECG is too large and loses the meaning possessed by the ECG itself, we adopt the method of slight stretching and compression in the y-axis direction for data augmentation. This method can avoid certain small perturbations from having a large impact on the deep learning model. The formula of this method is as follows:

$$\begin{aligned} output_{ori} &= SAPTST(X, \mathcal{A}) \\ output_{str} &= SAPTST((1+\lambda)X, \mathcal{A}) \\ output_{com} &= SAPTST((1-\lambda)X, \mathcal{A}) \\ output &= \frac{output_{ori} + output_{str} + output_{com}}{3} \end{aligned} \tag{6}$$

where $SAPTST$ refers to the SAPTST algorithm, $\mathcal{A}$ is the attention matrix, λ is the extent of stretching and compression. Because excessive compression or stretching can cause changes in the diagnostic results of the ECG, we set $\lambda \ll 1$. The multidimensional data resulting from the enhancement of each lead are recombined into 1-dimensional data after passing through the subtractive attention-based Transformer sub-module.

Longitudinal scaling-based data augmentation allows for slight changes in the inputs to the model so that the model will not be affected by slight changes in ECG morphology after sufficient training, which results in better robustness of the model.

4.4 Model training and anomaly detection

Figure 7 shows the details of the SAPTST / SAPTSTA model training and abnormality detection. In this section, the preprocessed ECG signal segment $\mathbf{X}_{tr}$ from the training set is input into the reconstruction module of the SAPTST / SAPTSTA model to obtain the reconstructed ECG signal segment, and the absolute value of the difference between the inputs and outputs of the SAPTST / SAPTSTA model is calculated by the sigmoid function as the model output. The absolute value of the difference between the input and output of the SAPTST/SAPTSTA model was calculated with the sigmoid function as the output of the model. If the similarity between the input ECG segments and the output ECG segments is

Journal: 10489 MS: 5881 TYPESET DISK LE CP Disp.:2024/12/3 Pages: 15 Layout: Large

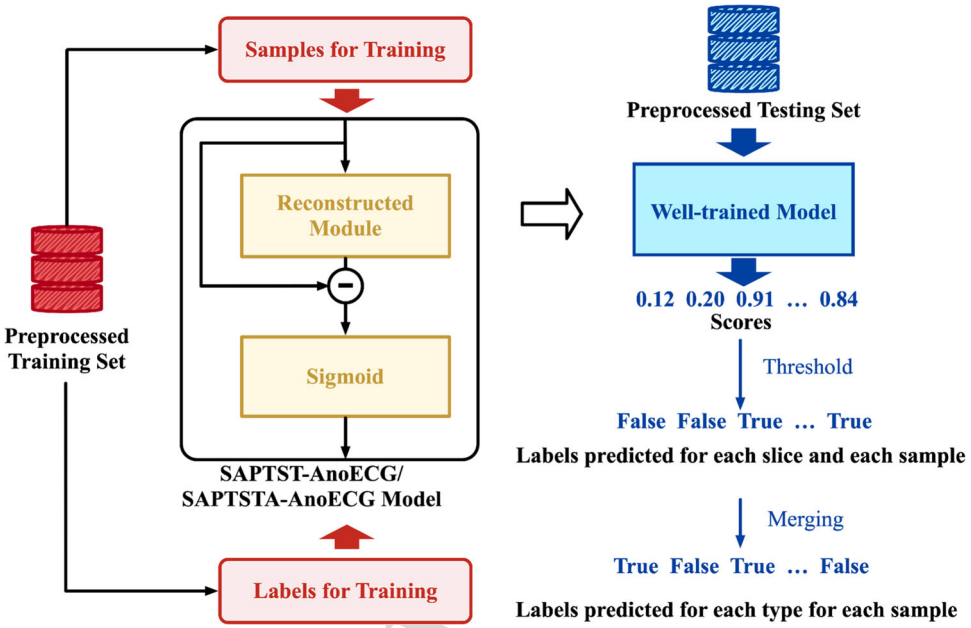

Fig. 7 Model training and classification for ECG data slices

greater, the model can reconstruct the sample better and the model output is closer to 0. If the similarity between the input ECG segments and the output ECG segments is lower, it means that the model has difficulty reconstructing the sample (i.e., the ECG produces an abnormality), and then the model output is closer to 1 under the effect of the sigmoid function. This logical relationship is consistent with the meaning of sample labeling. Moreover, since the differences between normal ECGs and the different types of abnormal ECGs are not the same, after sufficient training, each bit of the predicted value reflects the probability that the sample belongs to each category.

The training phase of the SAPTSTA-AnoECG method proposed in Fig. 7 is the same as that of most anomaly detection methods, iterating repeatedly until convergence. During training, BCE loss is used. After training, all outputs of the model should converge to 0 for normal samples, whereas for abnormal samples, the model outputs should converge to 1 for abnormal categories and 0 for other positions.

In the anomaly detection stage, we fuse all the test subsamples originating from the same test sample. Since for a certain ECG test sample $\mathbf{x}_u$, as long as a certain disease is detected in one of the subsamples $\mathbf{x}_{uv}$, it can be concluded that the patient has that disease, we use for each disease type the abnormal values of all the subsamples associated with $\mathbf{x}_u$ of the maximum value as the outlier of $\mathbf{x}_u$ for that disease type. In addition, we set separate thresholds for each disease. The formula for calculating the anomaly score is as follows:

$$\mathbf{s}_u = max(\mathbf{s}_{u1}, \mathbf{s}_{u2}, ..., \mathbf{s}_{ug}) \tag{7}$$

where g is the length of $\mathbf{s}_u$, and $\mathbf{s}_u$ is a k-dimensional vector. We further define another k-dimensional vector τ to represent the threshold of each class. Finally, for each $1 <= v <= k$ the predicted labels of the testing set should be as follows:

$$\mathbf{z}_{uv} = \begin{cases} 0 & \mathbf{s}_{uv} < \tau_v \\ 1 & \mathbf{s}_{uv} >= \tau_v \end{cases} \tag{8}$$

5 Experiments

In this section, we perform the experiments that are conducted to verify the effectiveness of the SAPTST-AnoECG. The dataset we use is introduced in Section 5.1, the results of the parameter tuning experiments are introduced in Section 5.2, the outcomes of the ablation experiments are described in Section 5.3, the results of the comparison experiments are reported in Section 5.4, and finally, the results of the statistical experiments are reported in Section 5.5.

5.1 Experimental setup

5.1.1 Datasets

In this paper, we use two public ECG datasets named PTB-XL and CPSC 2018 for all the experiments. In addition, we introduced a new publicly available ECG dataset named HCE in this paper and conducted comparative experiments using this dataset. The following is a brief introduction:

- PTB-XL Dataset [26]: PTB-XL is an open dataset used to diagnose ECG signals. It was first captured via equipment from Schiller AG between 1989 and 1996, and Physikalisch-Technische Bundesanstalt (PTB) later

Journal: 10489 MS: 5881 TYPESET DISK LE CP Disp.:2024/12/3 Pages: 15 Layout: Large

acquired it. PTB-XL comprises 21837 clinical 12-lead ECGs, each lasting 10 seconds, from 18885 patients. This dataset was labeled by two cardiologists, and 71 distinct ECG statements, including diagnostic, form, and rhythm statements, were discovered. The dataset is provided at sample frequencies of 100 Hz and 500 Hz, with 100 Hz being the sampling frequency utilized in this paper. Furthermore, the PTB-XL dataset can be divided into three categories, namely, diagnostic, rhythm, and form, according to [26]. The morphology of the ECG signal is described by a total of 44 diagnostic statements and 19 unique form statements. Four of these form statements are also diagnostic statements. In addition, the heart rhythm is described by 12 distinct rhythm assertions. Compared with the first fine-grained diagnostics, they can offer broader categories. The dataset can be downloaded from: https://physionet.org/content/ptb-xl/.

- CPSC 2018 Dataset [27]: The CPSC (China Physiological Signal Challenge) 2018 dataset was initially collected from 11 hospitals. It contains a total of 6877 12-lead samples, including 5501 for training, 687 for validation, and 689 for testing. These samples cover one normal ECG statement and 8 abnormal ECG statements (AF, I-AVB, LBBB, RBBB, PAC, PVC, STD, and STE). The sampling frequency of the CPSC is also 100 Hz. The dataset can be downloaded from: http://2018.icbeb.org/Challenge.html/.
- HCE Dataset: The HCE (Hangzhou–Changsha ECG) dataset is an ECG dataset collected by our team from Changsha, Hunan, and Hangzhou, Zhejiang, China. The data collection procedure required each participant to assume three different postures (i.e., standing, sitting, and lying), each for at least 60 seconds. A total of 76 volunteers, including 39 males and 37 females, participated in the ECG signal collection. The age range of the participants ranged from 19 to 94 years, encompassing 45 young adults (34 years old and younger), 26 middle-aged adults (35 to 59 years old), and 5 elderly individuals (60 years old and older). We recorded the ECG data generated by each volunteer while maintaining each posture into individual files, with a sampling rate of 100 Hz, resulting in a total of 228 ECG files. After excluding some invalid files, we obtained 215 valid ECG files (the reasons for invalid samples included circuit disconnection, lead misconnection, and program interruption). We divided the last 60 seconds of ECG data in each file into 6 segments, with each segment serving as a sample (lasting 10 seconds), yielding a total of 1290 ECG samples. These were annotated by cardiology experts, with 16 samples labeled as PVC and 12 samples labeled as RBBB. The HCE dataset and its more detailed description can be obtained from the following link: https://github.com/SLZWVICTOR/HCE_Dataset.

Table 1 Details about PTB-XL and CPSC 2018 datasets

Datasets	Task	n_{train}	n_{val}	n_{test}	n_{total}	k
PTB-XL	All	17411	2193	2203	21807	71
	Diagnostic	17111	2156	2163	21430	44
	Diag. Sub.	17111	2156	2163	21430	23
	Diag. Sup.	17111	2156	2163	21430	5
	Form	7202	904	882	8988	19
	Rhythm	16854	2109	2103	21066	12
CPSC 2018	All	5501	687	689	6877	9
HCE	All	–	–	–	2190	3

More information about the PTB-XL, CPSC 2018, and HCE datasets can be found in Table 1. For the PTB-XL and CPSC 2018 datasets, the table presents the sizes of the training set, validation set, and testing set for each task, along with the number of label classes for these tasks. In contrast, the HCE dataset does not have a designated training, validation, and testing split; instead, it is used solely as a testing set in this study. For specific usage details, please refer to Section 5.4.

5.1.2 Metrics

In this work, we use Macro-AUC for the PTB-XL dataset and Macro-AUC, Macro-F_2 and Macro-G_2 for the CPSC 2018 dataset. The AUC is the area under the receiver operating characteristic curve. The AUC curve is a curve with the FPR as the horizontal axis and the TPR as the vertical axis. The area below the curve is the AUC value, which is usually in the range of 0 to 1. The formula of the Macro-AUC is as follows:

$$AUC^{Macro} = \frac{AUC_1 + AUC_2 + ... + AUC_k}{k} \tag{9}$$

F_2 and G_2 can be calculated by:

$$\begin{aligned} F_2 &= \frac{5 \cdot TP}{5 \cdot TP + FP + 4 \cdot FN} \\ G_2 &= \frac{TP}{TP + FP + 2 \cdot FN} \end{aligned} \tag{10}$$

According to the formulas above, we can further define Macro-F_2 and Macro-G_2 as:

$$\begin{aligned} {F_2}^{Macro} &= \frac{F_{21} + F_{22} + ... + F_{2k}}{k} \\ {G_2}^{Macro} &= \frac{G_{21} + G_{22} + ... + G_{2k}}{k} \end{aligned} \tag{11}$$

The method for determining thresholds for each disease is similar to [10], which aims to maximize the value of

Journal: 10489 MS: 5881 TYPESET DISK LE CP Disp.:2024/12/3 Pages: 15 Layout: Large

$TPR - FPR$, thereby seeking a balance where the classifier performs well in distinguishing between true positives and false positives.

5.1.3 Environment

All the experiments are conducted on a computer equipped with an RTX 2080Ti GPU and Intel Core i9–9900K 3.60 GHz * 16 CPU.

5.2 Parameter tuning experiments

In this section, we conduct only the parameter tuning experiments on the CPSC 2018 dataset and "all statements" of the PTB-XL dataset, and we do not use the data augmentation part of the model. We mainly adjust the learning rate η, the early stopping strategy and the number of Transformer layers h. When adjusting η, we set the early stopping strategy to be "valid loss" and $h = 2$. When adjusting the early stopping strategy, we set $\eta = 2 \times 10^{-3}$ and $h = 2$. When adjusting h, we set the early stopping strategy to be "valid loss" and $\eta = 2 \times 10^{-3}$. The results of the parameter tuning experiments are reported in Tables 2 and 3.

For the learning rate η, we provide three optional numbers $\eta = 2 \times 10^{-2}$, $\eta = 2 \times 10^{-3}$ and $\eta = 2 \times 10^{-4}$. For the early stopping strategy, we provide three modes, including "no early stopping", "valid loss" and "valid AUC". "Valid loss" means that the basis of early stopping is the loss of the validation set. Similarly, "valid AUC" means that the basis of early stopping is the real-time AUC of the validation set when training.

The experimental results clearly indicate that the best learning rate is $\eta = 2 \times 10^{-3}$. The best early stopping strategy is valid loss for the PTB-XL dataset, and the early stopping strategy does not affect the performance of our method on the CPSC 2018 dataset. The best number of layers of the Transformer is $h = 2$ for the PTB-XL dataset and $h = 3$ for the CPSC 2018 dataset. To reduce the complexity of the algorithm while ensuring accuracy, we set $h = 2, \eta = 2 \times 10^{-3}$ in the following experiments and use the loss of the validation set as the basis for the early stopping mechanism.

Table 2 Results of parameter tuning experiments on PTB-XL dataset

Tuned parameter	Value	Macro-AUC
Learning Rate η	$\eta = 2 \times 10^{-2}$	0.828
	$\eta = 2 \times 10^{-3}$	0.930
	$\eta = 2 \times 10^{-4}$	0.887
Early Stopping Strategy	No Early Stopping	0.927
	Valid Loss	0.930
	Valid AUC	0.926
v Layer of Transformer h	$h = 1$	0.918
	$h = 2$	0.930
	$h = 3$	0.929

Table 3 Results of parameter tuning experiments on CPSC 2018 dataset

Tuned parameter	Value	Macro-AUC
Learning Rate η	$\eta = 2 \times 10^{-2}$	0.913
	$\eta = 2 \times 10^{-3}$	0.960
	$\eta = 2 \times 10^{-4}$	0.949
Early Stopping Strategy	No Early Stopping	0.960
	Valid Loss	0.960
	Valid AUC	0.960
Layer of Transformer h	$h = 1$	0.957
	$h = 2$	0.960
	$h = 3$	0.961

5.3 Ablation experiments

We conduct the ablation experiments to show the necessity of each component of our SAPTST-AnoECG model. To achieve this goal, we design the following models:

- **Dense-AnoECG:** We use the dense block to replace the SAPTST block in Fig. 5.
- **Transformer-AnoECG:** We simply use the vanilla Transformer block to replace the SAPTST block in Fig. 5.
- **PatchTST-AnoECG:** We use the PatchTST block mentioned in Fig. 2 to replace the SAPTST block in Fig. 5.
- **SAPTST-AnoECG:** We use the SAPTST block mentioned in Fig. 6 (without data augmentation and output integration) to replace the SAPTST block in Fig. 5.
- **SAPTSTA-AnoECG:** We use the SAPTST block mentioned in Fig. 6 (which includes data augmentation and output integration) to replace the SAPTST block in Fig. 5.

In this section, we use only "Macro-AUC" for measurement. We apply all these methods to "all statements" of the PTB-XL dataset and the CPSC 2018 dataset. The results of the ablation experiments are depicted in Tables 4 and 5. The results indicate that the Dense-AnoECG is the

Table 4 Results of ablation experiments on PTB-XL dataset

Method	Macro-AUC	Time (s/epoch)
Dense-AnoECG	0.565	1
Transformer-AnoECG	0.874	16
PatchTST-AnoECG	0.902	14
SAPTST-AnoECG	0.930	19
SAPTSTA-AnoECG	0.931	55

Journal: 10489 MS: 5881 TYPESET DISK LE CP Disp.:2024/12/3 Pages: 15 Layout: Large

Table 5 Results of ablation experiments on CPSC 2018 dataset

Method	Macro-AUC	Time(s/epoch)
Dense-AnoECG	0.549	1
Transformer-AnoECG	0.877	6
PatchTST-AnoECG	0.951	4
SAPTST-AnoECG	0.960	8
SAPTSTA-AnoECG	0.965	21

fastest method, but its performance is poor. Transformer-AnoECG takes slightly longer than the SAPTST-AnoECG and is less effective than the SAPTST-AnoECG. SAPTST-AnoECG takes longer than PatchTST-AnoECG does, which is due to other factors such as calculating the attention list, and its effect is significantly stronger than that of PatchTST. SAPTSTA-AnoECG is slightly better than SAPTSTA-AnoECG, but takes approximately three times as long as SAPTST-AnoECG.

5.4 Comparison experiments

In Tables 6 and 7, we compare our model SAPTST-AnoECG and SAPTSTA-AnoECG with Inception1d [11], XResnet1d101 [12], Resnet1d-Wang [13], FCN-Wang [13], Wavelet+NN [28], LFBT-FT [29], SelfONN [30] and MassMIB [31] on the PTB-XL and CPSC 2018 datasets. The value in parentheses indicates the maximum absolute error. In Fig. 8, we conduct experiments on the HCE dataset via the same model as in Tables 6 and 7. We use only the HCE dataset as the test set and continue to use the CPSC 2018 dataset as the training set. This is because the HCE dataset is too small to be further divided into training and test sets without significantly impacting its training effectiveness, and the anomaly types in the CPSC 2018 dataset conveniently encompass those in the HCE dataset.

Owing to the memory limitation of the runtime environment, we cannot adjust the number of parameters of the model to be sufficiently large. These tables clearly show that our SAPTST-AnoECG and SAPTSTA-AnoECG models outperform other advanced methods on the PTB-XL dataset and perform well on the CPSC 2018 dataset. On our own collected

Table 7 Results of comparison experiments on CPSC 2018 dataset

Methods	Macro-AUC	Macro-F_2	Macro-G_2
Inception1d	0.963(0.009)	0.807(0.030)	0.594(0.041)
XResnet1d101	0.974(0.005)	0.819(0.030)	0.602(0.037)
Resnet1d-Wang	0.969(0.006)	0.803(0.031)	0.586(0.037)
BiLSTM	0.959(0.011)	0.796(0.031)	0.573(0.036)
LSTM	0.964(0.006)	0.790(0.031)	0.561(0.037)
FCN-Wang	0.957(0.008)	0.787(0.031)	0.563(0.037)
Wavelet+NN	0.905(0.014)	0.665(0.034)	0.405(0.036)
LFBT-FT*	0.954	–	–
SelfONN*	0.956	0.802	0.583
SAPTST(ours)	0.960(0.007)	0.783(0.028)	0.564(0.039)
SAPTSTA(ours)	0.965(0.006)	0.785(0.029)	0.553(0.039)

(SAPTST refers to SAPTST-AnoECG, SAPTSTA refers to SAPTSTA-AnoECG. "*" indicates that the experimental result is obtained directly from the relevant paper, while "–" indicates that the relevant paper did not conduct the experiment. In addition, MassMIB's paper does not address the experiments related to this table

Table 6 Results of comparison experiments on PTB-XL dataset

Methods	Sub-dataset					
	All	Diag.	Diag. sub.	Diag. sup.	Form	Rhythm
Inception1d	0.925 (0.007)	0.933 (0.008)	0.926 (0.010)	0.918 (0.007)	0.880 (0.020)	0.937 (0.023)
XResnet1d101	0.924 (0.007)	0.930 (0.009)	0.926 (0.014)	0.932 (0.006)	0.899 (0.011)	0.955 (0.024)
Resnet1d-Wang	0.920 (0.006)	0.936 (0.007)	0.928 (0.014)	0.929 (0.007)	0.889 (0.013)	0.942 (0.013)
BiLSTM	0.912 (0.008)	0.928 (0.009)	0.922 (0.013)	0.922 (0.007)	0.868 (0.019)	0.951 (0.013)
LSTM	0.910 (0.007)	0.924 (0.008)	0.924 (0.011)	0.927 (0.006)	0.849 (0.017)	0.948 (0.015)
FCN-Wang	0.906 (0.009)	0.928 (0.010)	0.922 (0.014)	0.926 (0.006)	0.873 (0.013)	0.918 (0.015)
Wavelet+NN	0.836 (0.013)	0.835 (0.017)	0.846 (0.024)	0.871 (0.008)	0.777 (0.022)	0.882 (0.037)
LFBT-FT*	0.918	–	–	–	–	–
SelfONN*	0.930	0.923	0.925	0.923	0.886	0.952
MassMIB*	0.929	0.936	0.931	0.928	0.878	0.968
SAPTST(ours)	0.930 (0.008)	0.932 (0.005)	0.930 (0.008)	0.913 (0.007)	0.878 (0.010)	0.948 (0.012)
SAPTSTA(ours)	0.931 (0.016)	0.933 (0.007)	0.928 (0.013)	0.914 (0.007)	0.881 (0.010)	0.947 (0.012)

(SAPTST refers to SAPTST-AnoECG, SAPTSTA refers to SAPTSTA-AnoECG. All experimental results in the table are measured using Macro-AUC. "*" indicates that the experimental result is obtained directly from the relevant paper, while "–" indicates that the relevant paper did not conduct the experiment

Journal: 10489 MS: 5881 TYPESET DISK LE CP Disp.:2024/12/3 Pages: 15 Layout: Large

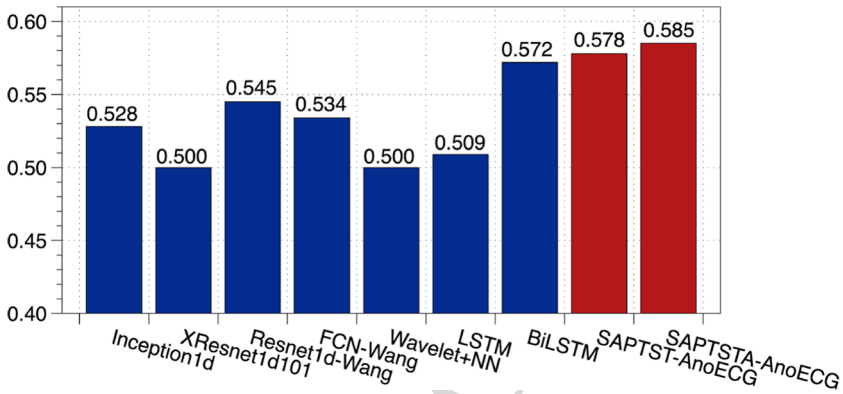

Fig. 8 Results of Comparison Experiments on HCE Dataset. (SAPTST refers to SAPTST-AnoECG, SAPTSTA refers to SAPTSTA-AnoECG. All experimental results in the table are measured using Macro-AUC.)

real-world dataset, the SAPTST-AnoECG and SAPTSTA-AnoECG methods also outperform other advanced methods.

5.5 Statistical tests

This section presents the statistical tests, including Friedman test and Nemenyi post-hoc test. Please note that we did not consider the methods LFBT-FT, SelfONN, and MassMIB in the statistical tests, as we were unable to validate their effectiveness using the HCE dataset. Furthermore, LFBT-FT and MassMIB do not yield complete experimental results even on the PTB-XL and CPSC 2018 datasets.

5.5.1 Friedman test

Friedman test is a nonparametric statistical method used to determine whether there are significant differences among three or more algorithms across multiple measurements. Assuming that k algorithms are applied to B datasets, the steps for performing Friedman test are as follows:

First, we systematically rank the performance of each model on each evaluation metric across all performance measurements in ascending order.

Second, we utilize the ranked performance data to compute the test statistic. When $B > 15$ or $k > 4$, the Friedman test employs the chi-squared distribution χ_F^2 as the test statistic, which can be calculated via the following formula:

$$\chi_F^2 = \frac{12B}{k(k+1)}\left[\sum_{j=1}^{k} R_j^2 - \frac{k(k+1)^2}{4}\right] \tag{12}$$

where R_j is the average rank for the j-th algorithm. Otherwise, Friedman test table is used.

Third, we compute the F statistic. The F statistic addresses the conservative tendencies of the original Friedman test, thereby enhancing its efficacy even with a small number of conditions ($k < 30$). The formula for calculating the F statistic is provided below:

$$F = \frac{(B-1)\chi_F^2}{B(k-1) - \chi_F^2} \tag{13}$$

The F statistic follows the F-distribution with $k-1$ and $(k-1)(B-1)$ degrees of freedom.

Finally, we compare the computed F statistic, F, with the critical value, denoted as F_c, which is determined based on the significance level α and the degrees of freedom d_1 and d_2. The critical value can be obtained from the relevant statistical tables or via statistical software. If $F > F_c$, we reject the null hypothesis, concluding that at least one model performs significantly differently than the other models do. Conversely, if $F \leq F_c$, we do not reject the null hypothesis, indicating that there is no significant difference in performance among the models.

In this work, we apply Friedman test on Macro-AUC, and set $\alpha = 0.05$, When $k = 9$ and $B = 8$, the degrees of freedom should be $d_1 = 8$ and $d_2 = 56$, and we can find $F_c = 2.109$ by searching the table. After calculation, $F = 56.149$, so $F > F_c$, which means that we can reject the null hypothesis and further conduct Nemenyi post-hoc test.

5.5.2 Nemenyi post-hoc test

The Nemenyi post-hoc test is a nonparametric multiple comparison procedure used to compute the critical difference (CD) between pairs of classifiers. We performed the Nemenyi post-hoc tests for Macro-AUC. Below is a detailed description of the specific steps involved in the Nemenyi post-hoc test.

First, the average rank of each algorithm across all datasets is computed. The formula for calculating the average rank of the i-th algorithm is:

$$\bar{r}_i = \frac{1}{B}\sum_{j=1}^{B} r_{ij} \tag{14}$$

Journal: 10489 MS: 5881 TYPESET DISK LE CP Disp.:2024/12/3 Pages: 15 Layout: Large

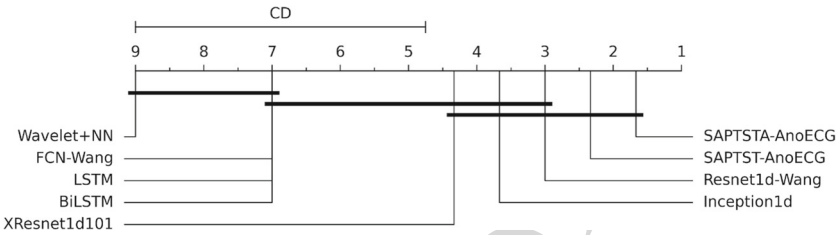

Fig. 9 Results of the Nemenyi post-hoc test

where r_{ij} refers to the rank of the i-th algorithm on the j-th dataset.

Second, calculate the critical difference (CD) value:

$$CD = q_\alpha \sqrt{\frac{k(k+1)}{6B}} \tag{15}$$

q_α can be obtained from a reference table.

Third, the difference between the average ranking values of the two algorithms is computed and compared against the critical difference (CD) value. If the difference exceeds the CD value, the two algorithms are considered significantly different; otherwise, they are deemed not significantly different.

After checking the table, when $k = 9$ and $\alpha = 0.05$, we can obtain $q_\alpha = 3.102$ and $CD = 4.247$, and Fig. 9 shows the CD plot We find that SAPTSTA-AnoECG and SAPTST-AnoECG significantly outperform Wavelet+NN, FCN-Wang, LSTM and BiLSTM, and are slightly better than Resnet1d-Wang and Inception1d.

6 Conclusion

In conclusion, we propose the SAPTST-AnoECG and SAPTSTA-AnoECG models, which are aimed at solving the ECG anomaly detection problem. We use PatchTST to combine the Transformer with the patch-based strategy, thus enhancing the accuracy and reducing the time complexity when handling time series. We also propose a subtractive attention mechanism based on R peaks, thereby learning the periodicity of ECG signals and further improving the accuracy of ECG anomaly detection. Experiments demonstrate the effectiveness of the method proposed in this paper.

However, the SAPTSTA-AnoECG model may require a long training time. Despite current limitations due to the low level of automation and accuracy in ECG anomaly detection, improving accuracy remains a major challenge in this field. However, as the automation level of models increases in the future, real-time performance is expected to become a bottleneck issue for ECG anomaly detection. Therefore, in future work, we can consider the use of methods such as model simplification and knowledge distillation to enhance real-time performance. Additionally, employing large models and distributed training can enable the handling of larger datasets, allowing the model to learn more about rare ECG anomalies and further improve accuracy and robustness.

Author Contributions Yifan Li: Conceptualization, Methodology, Writing-Original draft preparation; Mengjue Wang: Writing-Reviewing and Editing; Mingxiang Guan and Chen Lu: Data Curation, Validation; Zhiyong Li and Tieming Chen: Supervision, Writing-Reviewing and Editing.

Funding This work was partially supported by National Natural Science Foundation of China (No. 62403431, No. U21A20518, No. U23A20341, No. U22B2028, No. U1936215), the Fundamental Research Funds for the Provincial Universities of Zhejiang (No. RF-A2023009), Zhejiang Provincial Key Research and Development Projects (No. 2021C01117) and Zhejiang Provincial Natural Science Foundation of China (No. LD22F020002).

Data Availability and Access The PTB-XL dataset and CPSC 2018 dataset are both public datasets. PTB-XL dataset can be accessed at https://storage.googleapis.com/ptb-xl-1.0.1. physionet.org/ptb-xl-a-large-publicly-available-electrocardiography-dataset-1.0.1.zip, CPSC 2018 dataset can be accessed at http://2018.icbeb.org. We downloaded them in 2023. The HCE dataset is released in this paper, it can be accessed at https://github.com/SLZWVICTOR/ HCE_Dataset.

Declarations

Competing Interests All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Ethical and Informed Consent for Data Used This study utilizes the PTB-XL dataset and CPSC 2018 dataset. They are publicly available and follows the CC BY 4.0 license. Under this license, we are authorized to use this dataset for research purposes. All personal information contained in the dataset has been anonymized to protect the privacy of the data subjects. Our study further ensures that no re-identification of any personal information will occur. We commit to adhering to all relevant laws and ethical standards in the use of data, ensuring the legality and morality of data use. We will take full responsibility for any violations.

References

1. Roth GA, Mensah GA, Johnson CO (2020) Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the gbd 2019 study. J Am Coll Cardiol 76(25):2982–3021. https://doi.org/10.1016/j.jacc.2020.11.010
2. Li Y, Peng X, Zhang J et al (2021) Dct-gan: Dilated convolutional transformer-based gan for time series anomaly detection. IEEE

Journal: 10489 MS: 5881 TYPESET DISK LE CP Disp.:2024/12/3 Pages: 15 Layout: Large

Trans Knowl Data Eng 35(4):3632–3644. https://doi.org/10.1109/TKDE.2021.3130234
3. Algburi RNA, Gao H, Al-Huda Z (2022) Improvement of an industrial robotic flaw detection system. IEEE Trans Automat Sci Eng 19(4):3953–3967. https://doi.org/10.1109/tase.2022.3141248
4. Zhong Y, Wang Z, Shi X et al (2024) Rfg-helad: A robust fine-grained network traffic anomaly detection model based on heterogeneous ensemble learning. IEEE Trans Inf Forensic Secur 19:5895–5910. https://doi.org/10.1109/TIFS.2024.3402439
5. Yu H, Zhang X, Wang Y et al (2024) Fine-grained accident detection: database and algorithm. IEEE Trans Image Process 33:1059–1069. https://doi.org/10.1109/TIP.2024.3355812
6. Algburi RNA, Gao H, Al-Huda Z (2022) A new synergy of singular spectrum analysis with a conscious algorithm to detect faults in industrial robotics. Neural Comput Appl 34(10):7565–7580. https://doi.org/10.1007/s00521-021-06848-0
7. Nie Y, Nguyen NH, Sinthong P, et al (2023) A time series is worth 64 words: Long-term forcasting with transformers. In: International Conference on Learning Representations., pp 1–24, https://doi.org/10.48550/arXiv.2211.14730
8. Xu M, Wei S, Qin X, et al (2015) Rule-based method for morphological classification of st segment in ecg signals. In: J Med Biol Eng, pp 816–823, https://doi.org/10.1007/s40846-015-0092-x
9. Maglaveras N, Stamkopoulos T, Diamantaras K et al (1998) Ecg pattern recognition and classification using non-linear transformations and neural networks: A review. Int J Med Inform 52(1):191–208. https://doi.org/10.1016/S1386-5056(98)00138-5
10. Strodthoff N, Wagner P, Schaeffter T et al (2021) Deep learning for ecg analysis: Benchmarks and insights from ptb-xl. IEEE J Biomed Health 25(5):1519–1528. https://doi.org/10.1109/JBHI.2020.3022989
11. Ismail FH, Lucas B, Forestier G (2020) Inceptiontime: Finding alexnet for time series classification. Data Min Knowl Disc 34(6):1936–1962. https://doi.org/10.1007/s10618-020-00710-y
12. He T, Zhang Z, Zhang H, et al (2019) Bag of tricks for image classification with convolutional neural networks. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 558–567. https://doi.org/10.1109/CVPR.2019.00065
13. Wang Z, Yan W, Oates T (2017) Time series classification from scratch with deep neural networks: A strong baseline. In: International Joint Conference on Neural Networks., pp 1578–1585. https://doi.org/10.1109/IJCNN.2017.7966039
14. Liu X, Wang H, Li Z et al (2021) Deep learning in ecg diagnosis: A review. Knowl-Based Syst 2021:107187. https://doi.org/10.1016/j.knosys.2021.107187
15. Nurmaini S, Darmawahyuni A, Sakti MAN et al (2020) Deep learning-based stacked denoising and autoencoder for ecg heartbeat classification. Electronics 9(1):135. https://doi.org/10.3390/electronics9010135
16. Kazim H (2019) Deep neural network based approach for ecg classification using hybrid differential features and active learning. IET Signal Process 13(2):165–175. https://doi.org/10.1049/iet-spr.2018.5103
17. Lixin S, Dongzi S, Qian W et al (2019) Automatic classification method of arrhythmia based on discriminative deep belief networks. J Biomed Eng 36(3):444–452. https://doi.org/10.7507/1001-5515.201810053
18. Taji B, Chan AD, Shirmohammadi S (2018) False alarm reduction in atrial fibrillation detection using deep belief networks. IEEE T Instrum Meas 67(5):1124–1131. https://doi.org/10.1109/TIM.2017.2769198
19. Xu X, Liu H (2020) Ecg heartbeat classification using convolutional neural networks. IEEE Access 8:8614–8619. https://doi.org/10.1109/ACCESS.2020.2964749
20. Niu J, Tang Y, Sun Z et al (2020) Inter-patient ecg classification with symbolic representations and multi-perspective convolutional neural networks. IEEE J Biomed Health 24(5):1321–1332. https://doi.org/10.1109/JBHI.2019.2942938
21. Saadatnejad S, Oveisi M, Hashemi M (2020) Lstm-based ecg classification for continuous monitoring on personal wearable devices. IEEE J Biomed Health 24(2):515–523. https://doi.org/10.1109/JBHI.2019.2911367
22. Qiao F, Zhang X, Deng J (2020) Learning evolutionary stages with hidden semi-markov model for predicting social unrest events. In: Discrete Dyn Nat Soc, pp 1–16, https://doi.org/10.1155/2020/3915036
23. Zhao X, Liu Z, Han L, et al (2022) Ecgnn: Enhancing abnormal recognition in 12-lead ecg with graph neural network. In: IEEE International Conference on Bioinformatics and Biomedicine., pp 1411–1416. https://doi.org/10.1109/BIBM55620.2022.9995419
24. Dosovitskiy A, Beyer L, Kolesnikov A, et al (2020) An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations., pp 1–22. https://doi.org/10.48550/arXiv.2010.11929
25. Makowski D, Pham T, Lau ZJ et al (2021) NeuroKit2: A python toolbox for neurophysiological signal processing. Behav Res Methods 53(4):1689–1696. https://doi.org/10.3758/s13428-020-01516-y
26. Wagner P, Strodthoff N, Bousseljot RD et al (2020) Ptb-xl, a large publicly available electrocardiography dataset. Scientific Data 7(1):154. https://doi.org/10.1038/s41597-020-0495-6
27. Liu F, Liu C, Zhao L et al (2018) An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection. J Med Imag Health 8(7):1368–1373. https://doi.org/10.1166/jmihi.2018.2442
28. Lee G, Gommers R, Waselewski F, et al (2019) Pywavelets: A python package for wavelet analysis. J Open Source Softw 4(36):1–2. https://doi.org/10.21105/joss.01237
29. Liu W, Pan S, Li Z et al (2024) Lead-fusion barlow twins: A fused self-supervised learning method for multi-lead electrocardiograms. Inf Fusion 114:1–14. https://doi.org/10.1016/j.inffus.2024.102698
30. Qin K, Huang W, Zhang T et al (2024) A lightweight selfonn model for general ecg classification with pretraining. Biomed Signal Process Contr 89:1–12. https://doi.org/10.1016/j.bspc.2023.105780
31. Yang S, Lian C, Zeng Z et al (2024) Masked self-supervised ecg representation learning via multiview information bottleneck. Neural Comput Appl 36(14):7625–7637. https://doi.org/10.1007/s00521-024-09486-4

Authors and Affiliations

Yifan Li[1] · Mengjue Wang[2] · Mingxiang Guan[3] · Chen Lu[3] · Zhiyong Li[4] · Tieming Chen[5]

College of Computer Science and Technology

✉ Tieming Chen
tmchen@zjut.edu.cn

Yifan Li
yifanli@zjut.edu.cn

Mengjue Wang
495185849@qq.com

Mingxiang Guan
gmx2020@126.com

Chen Lu
luchen_0826@qq.com

Zhiyong Li
zhiyong.li@hnu.edu.cn

1 ~~College of Computer Science,~~ Zhejiang University of Technology, Liuhe Road, Hangzhou 310023, Zhejiang, China

2 Xixi Hospital, Liuhe Road, Hangzhou 310023, Zhejiang, China

3 School of Electronic Communication Technology, Shenzhen Institute of Information Technology, Longxiang Street, Shenzhen 518172, Guangdong, China

4 College of Information Science and Engineering, Hunan University, South Lushan Road, Changsha 310023, Hunan, China

5 College of Geoinformatics, Zhejiang University of Technology, Changhong East Road, Huzhou 313299, Zhejiang, China

Journal: **10489** MS: **5881** ☐ TYPESET ☐ DISK ☐ LE ☐ CP Disp.:**2024/12/3** Pages: **15** Layout: **Large**

Journal: 10489
Article: 5881

Author Query Form

Please ensure you fill out your response to the queries raised below and return this form along with your corrections

Dear Author

During the process of typesetting your article, the following queries have arisen. Please check your typeset proof carefully against the queries listed below and mark the necessary changes either directly on the proof/online grid or in the 'Author's response' area provided below

Query	Details required	Author's response
1.	Please check and confirm if the authors and their respective affiliations have been correctly identified. Amend if necessary.	The affiliation of Yifan Li is incorrect, it should be: "College of Computer Science and Technology". Please refer to Line 905.
2.	Photography and Biography are desired. Please provide. Otherwise, please confirm if unnecessary.	We have provided the Photography and Biography; please refer to the email attachment for further details.
3.	Please check figures if captured and presented correctly.	Yes
4.	Please check tables if captured and presented correctly.	Yes
5.	Please check Funding if captured and presented correctly.	Yes
6.	Please check references if captured and presented correctly.	Yes