



# OPEN Assessment of the long RR intervals using convolutional neural networks in single-lead long-term Holter electrocardiogram recordings

Tan Lyu, Miao Ye, Minjie Yuan, Haina Chen, Sijie Han, Lifang Yu & Chen Li✉

Advancements in medical technology have extended long-term electrocardiogram (ECG) monitoring from the traditional 24 h to 7–14 days, significantly enriching ECG data. However, this poses unprecedented challenges for physicians in analyzing these extensive datasets. While artificial neural networks have shown great potential in the automatic recognition of arrhythmias and are gradually being adopted widely, their effectiveness still requires rigorous evaluation in clinical applications. Although there have been many research on the evaluation of atrial fibrillation recognition performance, systematic evaluation of automatic detection algorithms for prolonged RR interval (RRI) remains lacking. This study aims to comprehensively evaluate these algorithms based on convolutional neural networks (CNN). We collected 480 single-lead long-term dynamic ECG records from our hospital, all from patients with prolonged RRI. Both CNN algorithms and manual analysis methods were used to analyze the ECGs, with manual analysis serving as the gold standard for performance evaluation. Results indicated that the CNN algorithm achieved an average recall rate of 98.41%, an average precision of 98.68%, and an average F1 score of 98.54%. These metrics demonstrate that CNN algorithms can meet the clinical demands for recognizing prolonged RRI, thereby enhancing physicians' confidence, especially when dealing with large volumes of RRI data. AI provides invaluable support for quantitative analysis and significantly improves diagnostic efficiency. This evaluation offers a reliable basis for deploying single-lead intelligent monitoring devices in households, communities, nursing homes, and other settings.

The dynamic electrocardiogram (ECG) is long-term, noninvasive and has a high detection rate for arrhythmias<sup>1</sup>. There are no special contraindications to its operation, so it is widely used in clinical practice. The RRI of the ECG, which refers to the distance between two R peaks on the ECG, refers to the time interval between two ventricular muscle depolarizations<sup>2,3</sup>. Under normal circumstances, the R-R interval should be between 0.6 and 1 s, corresponding to a heart rate of 60 to 100 beats per min<sup>4–6</sup>. Compared to sinus rhythm, a significantly prolonged R-R interval is more likely to cause other arrhythmias, and the standard  $RRI \geq 2s$  is widely used in clinical practice<sup>7–9</sup>. When the RRI is too long, it may lead to a series of adverse effects and disadvantages<sup>10,11</sup>. A prolonged R-R interval reflects issues with cardiac rhythm and conduction system, such as sinus bradycardia, sinus arrest, atrioventricular block (AVB) (especially second-degree to third-degree AVB), atrial fibrillation, or atrial flutter<sup>12</sup>. The most direct manifestation of a prolonged RRI is bradycardia, which refers to a heart rate below the normal range<sup>13</sup>. This can lead to a decrease in heart pumping function, which in turn affects the blood supply of the whole body, and may lead to symptoms such as dizziness and maurosis<sup>14,15</sup>. Long-term ambulatory ECG monitoring is essential to record variations in cardiac electrical activity and is particularly significant for detecting prolonged RRI.

In evaluating patients with suspected bradycardia, dynamic ECG monitoring is recommended to correlate rhythm disturbances with symptoms. This monitoring enables clinicians to diagnose underlying cardiac conditions, assess the severity of the condition, and develop personalized treatment plans. Pacemaker implantation is indicated for symptomatic patients with the bradycardia-tachycardia form of sinus node dysfunction to correct bradyarrhythmias and allow pharmacological treatment, unless ablation of the tachyarrhythmia is

Electrocardiogram Department, Sir Run Run Shaw Hospital, Affiliated with the Zhejiang University School of Medicine, Hangzhou 310016, Zhejiang, China. ✉email: 1138800052@zju.edu.cn

preferred<sup>16</sup>. It is also indicated for atrial arrhythmias (mainly atrial fibrillation) and permanent or paroxysmal third- or high-degree AVB regardless of symptoms<sup>17</sup>. Furthermore, spontaneous documented symptomatic asystolic pause  $\geq 3$  sec or asymptomatic pause  $\geq 6$  sec due to sinus arrest or AVB are Class IA indications for the implantation of dual chamber pacemakers<sup>18</sup>. Early detection of prolonged RRI is crucial to determine the need for pacemaker implantation, as delayed treatment can lead to decreased cardiac function, syncope, or other severe cardiac problems. Through continuous monitoring, doctors can continuously assess the electrical activity and rhythm of the heart, allowing them to make timely decisions. The timely implantation of a pacemaker can not only effectively correct cardiac rhythm problems but also significantly improve the patient's quality of life and prognosis.

When the RRI is too long to some extent, such as exceeding 3 sec, it can lead to an insufficient blood supply to important organs such as the brain, which in turn can cause syncope or even a transient loss of consciousness<sup>19,20</sup>. This situation is particularly common in older people or patients with severe heart disease. In addition, prolonged RRI can be one of the manifestations of various arrhythmias, such as atrioventricular block, complete heart block, sinus arrhythmia, etc.<sup>21–23</sup>. These arrhythmias can further aggravate the burden on the heart, increase the risk of serious events such as cardiac arrest, and increase the risk of coronary heart disease, heart failure, and other cardiovascular diseases, thus affecting clinical treatment<sup>7</sup>. Artificial intelligence algorithms have been extensively researched in many fields<sup>24,25</sup>, and have also attracted widespread attention in the field of medical treatment and health care. There have been many artificial intelligence algorithms for the prediction of arrhythmias in academia and industry. Kwon et al.<sup>26</sup> used deep learning technology to predict cardiac arrest, and LEE et al.<sup>27</sup> used machine learning models to predict cardiac arrest in real time. Andrew et al.<sup>28</sup> used deep learning models to classify various arrhythmias. The application of artificial intelligence has improved the precision and efficacy of cardiac disease diagnosis through the development of advanced ECG recognition and prediction algorithms. Using deep learning techniques, artificial intelligence enables high-resolution diagnostics, real-time monitoring and alerting, and efficient processing of large-scale datasets.

Although the application of artificial intelligence in the field of ECG analysis is gradually emerging as a cutting-edge technology, its popularization in actual clinical settings and medical institutions still faces challenges, mainly due to the so-called “black box” problem<sup>29</sup>. This dilemma limits the transparency and interpretability of algorithms, which in turn hinders doctors' full trust in artificial intelligence (AI) models. To break this deadlock, the academic community has begun to explore the fit and consistency between the automatic analysis results of convolutional neural networks (CNN) algorithms and the traditional manual analysis results of doctors, in an effort to verify the effectiveness and reliability of the former. Hennings et al.<sup>30</sup> took an important step in this direction. They compared the estimation results of the atrial fibrillation burden in ambulatory ECG recordings between artificial intelligence algorithms and manual evaluation, and found that the two were highly consistent, thus strongly demonstrating the accuracy and efficiency of artificial intelligence-based tools in assessing atrial fibrillation burden. They proposed a novel and powerful candidate for this assessment method. Building on this foundation, this article delves deeper into the accuracy of artificial intelligence algorithms in predicting long RRI, and thoroughly examines the consistency between the algorithm's predictions and the longest RRI obtained through clinical standard methods. Furthermore, we aim to uncover the multiple causes of long RRI diseases and systematically evaluate the classification performance of artificial intelligence algorithms to distinguish these different causes, in order to provide a more comprehensive theoretical foundation and practical guidance for the technical application in this field.

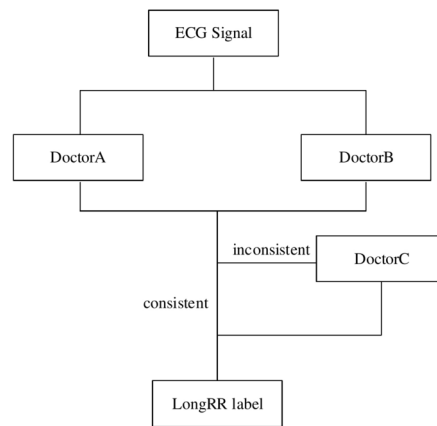
## Subjects and methods

### Introduction to data and assessment methods

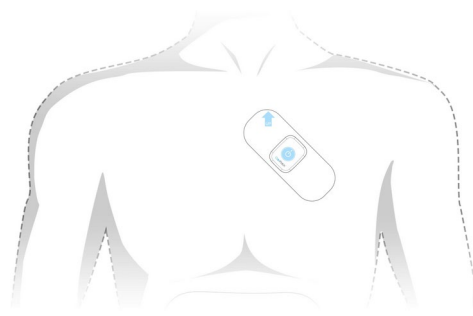
The diagnostic criterion for long RRI on an ECG is defined as the presence of at least one significantly prolonged RRI that exceeds the 2 sec threshold<sup>7</sup>. According to statistics, from August 12, 2020, to March 27, 2024, our hospital recorded a total of 3148 Holter ECG cases of 7 days (ECG-P01 Holter ECG recorder, Proton Technology, Hangzhou), among which 480 cases had  $\text{RRI} \geq 2.0$  sec, representing 15% of all Holter ECGs of 7 days. This proportion underscores the importance of paying attention to the persistent effects of long RRI.

During the data annotation process, we used a multistage verification strategy for the data preliminarily analyzed to ensure the quality of the data annotation, as shown in Fig. 1. Initially, Doctors A and B, each with over five years of experience, independently conducted preliminary annotations, focusing particularly on the identification and classification of long RRI (ie, prolonged RRI of the ECG). If their annotations for long RRI were consistent, the result was considered a high quality output and directly included in the ground truth data set. This step aimed to ensure the high accuracy and reliability of the ground truth dataset. Conversely, if Doctors A and B disagreed on the annotation of long RRI, the data were forwarded to the third stage for review and final determination by Doctor C, who has over ten years of professional experience. Doctor C's involvement aimed to resolve discrepancies between the first two doctors, providing authoritative and accurate annotations for controversial data through their profound professional knowledge and extensive experience, thereby further enriching and improving the content and quality of the ground-truth dataset. This process not only reflects the rigorous pursuit of annotation accuracy, but also underscores the importance of data rigor and reliability in scientific research.

This analysis relied on the industry-leading CarePatch dual-electrode single-channel monitoring device, which has undergone rigorous validation and obtained Class II medical device certification from the National Medical Products Administration (CFDA) (Registration Certificate No.: Zhe Med Device Reg No. 20202070050). The device model is ECG-P01, which features a sampling rate of 256 Hz, 12-bit ADC resolution, and the ability to operate continuously for 168 h, guaranteeing precision and dependability of data acquisition. Except for special circumstances, the device is worn with a 45-degree tilt, as shown in Fig. 2. During wear, patients are allowed to briefly remove the ECG recorder only when absolutely necessary. Our hospital encourages patients



**Fig. 1.** Flow chart of data labeling.

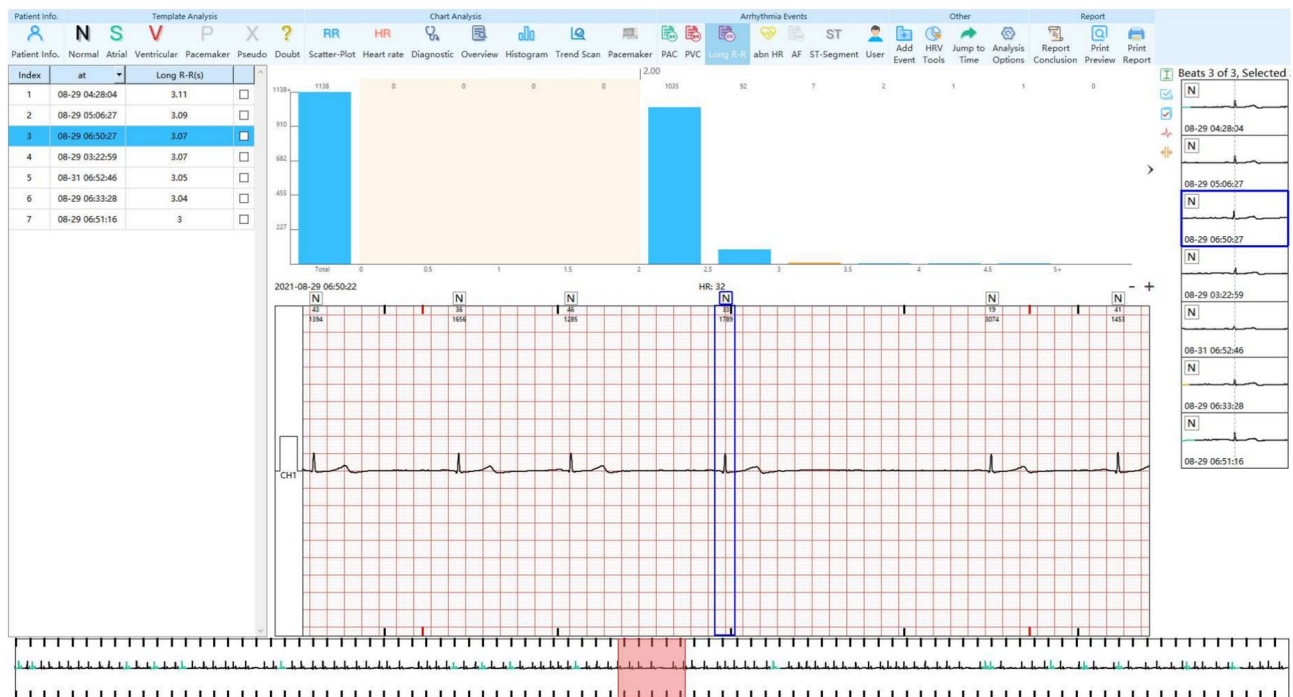


**Fig. 2.** Chest wearing diagram.

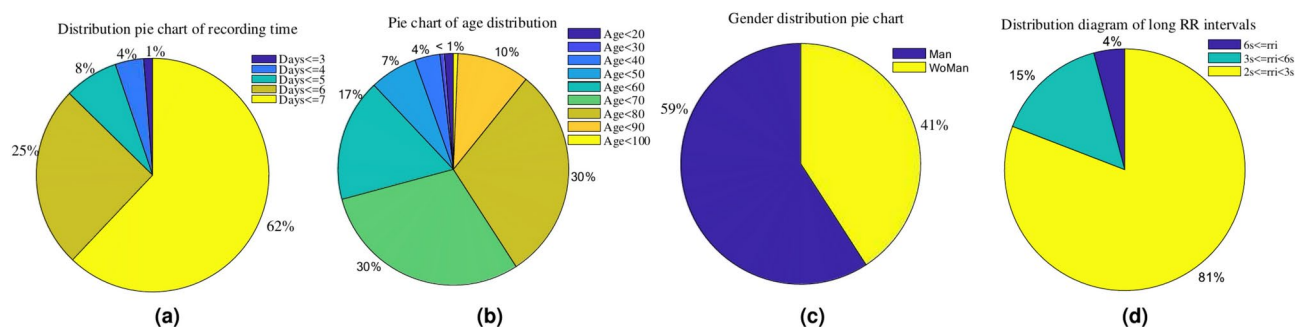
to maintain their daily routines while wearing the device to obtain ECG data that are closest to real-world conditions. To efficiently and accurately identify these long RR events, we introduced cloud-based AI\_ECG software as an analytical tool. The artificial intelligence algorithm utilized by this analysis tool is based on the CNN approach, with a model structure similar to that referenced in<sup>31</sup>, distinguished by the alteration of the data input layer from multilead to single-lead. Subsequent to the input layer of the model are three parallel convolution blocks, each comprising three layers of neural networks. Each layer encompasses a one-dimensional CNN layer, a Batch Normal layer, and a max-pooling layer. The three blocks employ different expansion rates. This software also has Class II medical device certification from the CFDA (Registration Certificate No.: Zhe Med Device Reg No. 20232211205), specializing in automatically screening for critical ECG abnormalities such as long RRI from massive Holter ECG data. Figure 3 presents the interface of this analysis software, whose core algorithm is deeply developed and trained based on the vast data accumulated by the CarePatch dual-electrode single-channel monitoring device independently developed by Hangzhou Proton Technology Co., Ltd., ensuring the precision and cutting-edge nature of the analysis results. This study was approved by the Ethics Committee of Sir Run Run Shaw Hospital, Zhejiang University School of Medicine (Approval No. 20241051NI). Informed consent was waived due to the retrospective nature of the study. The study was conducted in accordance with the Declaration of Helsinki.

### Data statistics

The total duration of 480 records reached 93,165.17 h (approximately 3882 days), with an average monitoring duration of 147.88 h (6.16 days) per patient and a standard deviation of 22.42 h. As shown in Fig. 4a, the percentage of different recording durations in the total ECG data recording is shown in days. Data with a recording duration of 6–7 days account for 62%, 5–6 days account for 25%, and data with a recording duration of less than or equal to 3 days only account for 1%. 87% of the patients wear the device for more than 5 days. The age range for patients with long RR is 19–93 years old ( $64 \pm 13$ ), with a significant age range, reflecting the broad representation of the study population. The age distribution is shown in Fig. 4b. Patients with long RR before the age of 60–70 and 70–80 account for 60%. In terms of gender composition, there are 284 males, representing 59.17%, while there are 196 females, representing 40.83%. As shown in Fig. 4c, this ratio distribution reflects the relative balance of gender in the research sample. It is particularly noteworthy that most patients with long RR have a maximum RRI of no more than 3 sec, as shown in Fig. 4d, the distribution of the long RR duration. 81% of the patients have a long RR of less than 3 sec, while 15% of the 62 records show that the RRI exceeds 3 sec but does not exceed 6 sec. There are 4% of the 30 RRI records that exceed 6 sec.



**Fig. 3.** Interface of doctor analysis software.



**Fig. 4.** Data distribution diagram, (a) is the distribution diagram of recording time, (b) is the distribution diagram of age, (c) is the ratio of male to female, and (d) is the duration of long RRI.

### Analysis method

In this study, we compared the results of long RR occurrences on long-term dynamic ECG manually analyzed by doctors with those automatically analyzed by artificial intelligence algorithms. We evaluated the consistency of the longest RRI with the data obtained by the long RR detection tool of the artificial intelligence algorithm using the Pearson correlation coefficient and the Bland-Altman plot.

#### Analysis indicators for algorithm accuracy

We use the long RR results manually analyzed by physicians as the gold standard, and the long RR results automatically analyzed by algorithms as the control results. We measure the performance of long RR analysis results using commonly used algorithm analysis metrics in the field of artificial intelligence. The main indicators include True Positives (TP), False Positives (FP), False Negatives (FN) and three performance metrics derived from these basic parameters: Precision (Prec), Sensitivity (Se), and F1 score (F1). The specific formulas are shown in Eq. (1).

$$Prec = \frac{TP}{TP + FN} \quad Se = \frac{TP}{TP + FP} \quad F1 = \frac{2 * Prec * Se}{Prec + Se} \quad (1)$$

Indeed, TP represent the number of samples that are correctly predicted by the algorithm model as long RR when they actually are long RR. FP refers to the number of samples that are incorrectly predicted by the model as long RR when they are actually not long RR, that is, the number of false positives (misclassified long RR). FN

represents the number of samples that the model incorrectly predicted as not long RR when they are actually long RR, that is, the number of false negatives (missed long RR).

#### Consistency analysis indicator for the longest RRI time

We evaluated the agreement between the two techniques for the most extracted features using Pearson's correlation coefficients and Bland-Altman plots. Consistency assessment of long RRI. Pearson correlation coefficient is a statistic that measures the degree of linear correlation between two variables X and Y. The value of this coefficient is between  $-1$  and  $1$ , where  $1$  indicates a perfect positive correlation,  $-1$  indicates a perfect negative correlation, and  $0$  indicates that there is no linear correlation but that it may have other types of correlation, such as nonlinear correlation. The formula is shown in eq.2

$$r = \frac{\sum_{i=1}^n (Lrri_i - \overline{Lrri})(LrriAI_i - \overline{LrriAI})}{\sqrt{\sum_{i=1}^n (Lrri_i - \overline{Lrri})^2} \sqrt{\sum_{i=1}^n (LrriAI_i - \overline{LrriAI})^2}} \quad (2)$$

Where  $Lrri_i$  and  $LrriAI_i$  are the observed values of the annotated longest RRI ( $Lrri$ ) and the longest RRI analyzed by the artificial intelligence algorithm ( $LrriAI$ ), respectively.  $\overline{Lrri}$  and  $\overline{LrriAI}$  are the corresponding mean values and  $n$  is the number of observations, which is 480 in this study. The Bland-Altman plot<sup>32</sup> is a graphical tool used to evaluate the agreement between two methods of measuring continuous variables, also known as a difference mean plot or difference plot. This graphical tool has widespread applications in medical experiments, data analysis, biomedical research, and the evaluation of machine learning algorithms<sup>30,33</sup>. The Bland-Altman plot is presented in the form of a scatter plot, which visually displays the agreement between two measurement methods by plotting the difference between the measurement results of the two methods against their mean. Each point on the Bland-Altman plot represents a measurement object, with the horizontal axis typically representing the mean of the two measurement results (predicted and actual values), and the vertical axis representing the difference between these two measurement results.

The central solid line in the plot represents the mean difference of all sample points. The upper and lower dashed lines represent the 95% Limits of Agreement (95% LoA), which are calculated as the mean difference  $\pm 1.96$  times the standard deviation, as shown in Eqs.(3) and (4). Here,  $up95$  and  $down95$  represent the upper and lower limits of agreement, respectively, and the lines corresponding to these values constitute the primary reference range to evaluate the agreement between the two measurement methods.  $rriDiff_i$  represents the difference between the two sets of values, and  $\mu_{rriDiff}$  represents the mean of the differences between the two sets of values.

$$up95 = \frac{\sum_{i=1}^n rriDiff_i}{n} + 1.96 * \sqrt{\frac{1}{N} \sum_{i=1}^N (rriDiff_i - \mu_{rriDiff})^2} \quad (3)$$

$$down95 = \frac{\sum_{i=1}^n rriDiff_i}{n} - 1.96 * \sqrt{\frac{1}{N} \sum_{i=1}^N (rriDiff_i - \mu_{rriDiff})^2} \quad (4)$$

## Result

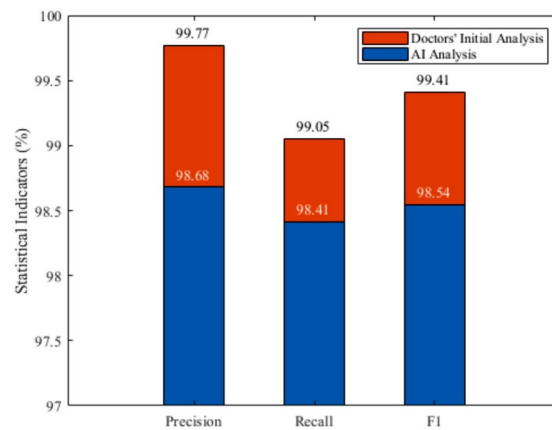
### Comparison between the results of automatic analysis by artificial intelligence algorithm and manual analysis by doctors

Based on detailed algorithm evaluation criteria, we systematically summarized the recognition performance for long RRI covering  $RRI \geq 2s$ ,  $RRI \geq 3s$ , and  $RRI \geq 6s$ . The detailed results are shown in Table 1. From the data in the table, it can be clearly observed that for  $RRI \geq 2s$ , the precision and recall rate are as high as 98.6796 and 98.4106%, respectively, which can classify the vast majority of samples into their respective categories, highlighting the model's superior performance in complex classification tasks. At the same time, the F1 score reached 98.5448%, perfectly balancing precision and recall, further confirming the stability and reliability of the model. Further analysis of the recognition results of  $RRI \geq 3s$  shows that the precision rate of long RRI remains at a high level of 95.6352%, the recall rate is 97.7483%, and the F1 score also reaches 96.6802%. This indicates that the model also performs well when dealing with longer RRI, ensuring both comprehensiveness and precision in recognition. For the longer RRI with  $RRI \geq 6s$ , the model exhibits excellent zero-miss characteristics, with a recall rate of 100%. Although the precision slightly decreases to 92.5373%, the F1 score remains at a high level of 96.1240%, indicating that the model faces certain challenges in identifying longer RRI, but overall performs robustly, not missing any longer RRI, with only a few minor misjudgments. It is worth

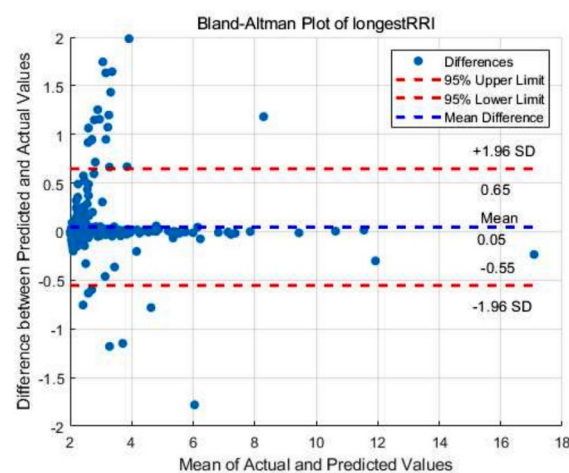
RRI	TP	FP	FN	Prec (%)	Recall (%)	F1 (%)
$\geq 2s$	248264	3322	4010	98.6796	98.4105	98.5448
$\geq 3s$	4645	212	107	95.6352	97.7483	96.6802
$\geq 6s$	62	5	0	92.5373	100	96.1240

**Table 1.** Indicators of long RRI identified by CNN algorithms under different standards.





**Fig. 5.** Comparison of the indicators of long RR between doctors' initial analysis and CNN algorithm analysis.



**Fig. 6.** The Bland-Altman plot for the longest RRI, where the blue line represents the mean of the differences and the red line represents the 95% consistency limit.

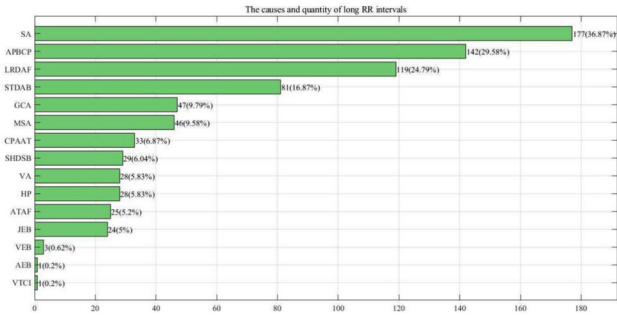
noting that compared with the cases of  $RRI \geq 2s$  and  $RRI \geq 3s$ , the cases of  $RRI \geq 6s$  was significantly reduced, with only five misclassified long RRI events and no missed identifications. This further highlights the stability and practicality of the model under extreme conditions. We also compared the long RR results of the initial analysis of the physicians with those of the AI analysis, as shown in Fig. 5. The result of the initial analysis of doctors is about 1% higher than that of the AI analysis, indicating that the results of the AI analysis are relatively close to the preliminary analysis results of doctors and can replace doctors in certain medical scenarios to complete the screening work for the long RR of patients, and the results are relatively reliable. Although the absolute number of FP and FN is relatively small compared to the number of TP, it is still necessary to pay attention to the possible impact of these erroneous predictions in practical applications, especially in scenarios where strict requirements are placed on the prediction results. In summary, this long-range RR classification model has shown good classification performance on the given dataset.

### Consistency analysis of the longest RRI

The longest RRI is a key indicator on the ECG, which measures the maximum time interval between two heartbeats. This indicator is of great importance for evaluating cardiac function and rhythm status. In this study, we first analyzed the linear relationship between the longest RRI (longestRR) and the corresponding CNN algorithm's longest RRI (longestRRAI) using the Pearson correlation coefficient. The results of the analysis showed that the Pearson correlation coefficient  $R$  between the two was 0.9412 ( $p < 0.001$ ), indicating a strong positive correlation between longestRRAI and longestRR, and this relationship was statistically highly significant. In addition, we further analyzed the consistency between the longest RRI automatically analyzed by the CNN algorithm and the longest RRI manually analyzed by the doctor through the Bland-Altman plot dimension. In Fig. 6, it can be seen that the distribution of the data points is horizontally banded, indicating that there is no obvious linear relationship between the measurement difference and the mean value. Most of the sample points

Abbreviation	Description
SA	Sinus arrest
APBCP	After premature beat compensatory pause
LRDAF	Long RRI during atrial fibrillation
STDAB	Second- to third-degree atrioventricular block
GCA	Global cardiac asystole
MSA	Marked sinus arrhythmia
CPAAT	Compensatory pause after atrial tachycardia
SHDSB	Second- to high-degree sinoatrial block
VA	Ventricular asystole
HP	Hypofunction of pacemaker
ATAF	After termination of atrial fibrillation
JEB	Junctional escape beat
VEB	Ventricular escape beat
AEB	Atrial escape beat
VTCT	After termination of ventricular tachycardia

**Table 2.** Abbreviations of the causes of prolonged RRI and their corresponding descriptions.



**Fig. 7.** The proportion of different causes of long RR. The horizontal axis in the figure is the number of cases, and the vertical axis is the cause of occurrence. The abbreviated description is shown in Table 2.

fall within the consistency limit 95%, while a small number of points are outside the consistency range. The bias of the Bland-Altman analysis is 0.05 (95% consistency limit is − 0.55 to 0.65).

**Different causes of prolonged RRI**

Prolonged RRI can reflect abnormalities in cardiac rhythm, and a detailed analysis of the various causes of prolonged RRI is crucial for clinical diagnosis and treatment. Figure 7 presents 15 different causes of prolonged RRI and their respective proportions among 480 data records. It is important to note that if a patient exhibits multiple causes of prolonged RRI, each cause is counted separately. Furthermore, prolonged RRI following compensatory pauses after premature beats include those after atrial premature beats, ventricular premature beats, and non-conducted atrial premature beats.

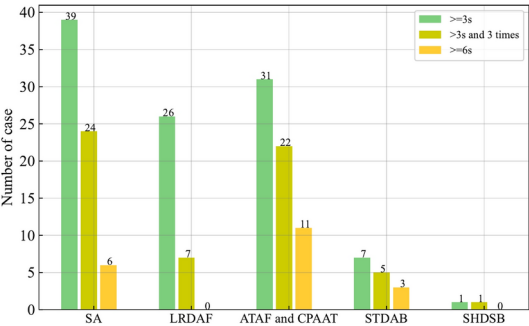
As evident in Fig. 7, sinus arrest and compensatory pauses after premature beats are the main causes of prolonged RRI, accounting for a significant proportion of all cases. This highlights the need to pay special attention to sinus arrest and compensatory pauses after premature beats in clinical practice. Subsequent causes, such as prolonged RRI in atrial fibrillation, second- to third-degree atrioventricular block, global cardiac arrest, and marked sinus arrhythmia, although less frequent in number, are still important factors contributing to prolonged RRI and should not be overlooked.

To investigate whether there is a difference in the indicators of artificial intelligence algorithms to analyze the extension of RRI caused by different reasons, we evaluated the classification performance of the algorithms for each reason. Table 3 shows the precision, recall and F1 classification score for each reason.

From the comprehensive index of the F1 score, the three main reasons for the highest classification performance are the moderate to high sinus node block F1 = 99.51%, severe sinus arrhythmia F1 = 98.95%, and the compensatory interval after atrial tachycardia F1 = 98.92%. In contrast, the result of the classification of prolongation of RRI caused by termination of atrial fibrillation is relatively low, F1=95.75%. This suggests that when long RRI occur due to the termination of atrial fibrillation, special attention should be paid to verify the indicators of the CNN algorithm classification. In general, this analysis emphasizes the importance of validating the performance of CNN algorithms in different causes of prolongation of the RRI to ensure precision of the clinical diagnosis.

Index (%)	SA	APBCP	LRDAF	STDAB	GCA	MSA	CPAAT	SHDSB	VA	HP	ATAF	JEB
precision	98.69	98	98.5	99.23	97.89	99.41	98.41	99.58	97.14	98.43	97.81	97.83
recall	98.87	95.93	97.53	98.51	98.37	98.5	99.44	99.44	98.17	97.6	93.78	98.63
F1	98.78	96.95	98.01	98.87	98.13	98.95	98.92	99.5	97.65	98.01	95.75	98.23

**Table 3.** Algorithm indicators for long RR recognition with different reasons. • The abbreviated description is shown in Table 2



**Fig. 8.** Classification of causes for RRI  $\geq 3$  sec, with the vertical axis representing the number of cases and the horizontal axis representing the abbreviated causes of occurrence. Descriptions of the abbreviations can be found in Table 2.

Analysis of RRI longer than 3 seconds

Among the 480 cases with long RRI in this study, there were 91 cases with long RRI of  $\geq 3.0$  sec, including 18 cases with long RRI of  $\geq 6.0$  sec, and the longest was 17.20 sec. As shown in Fig. 8, 52 cases with long RRI  $\geq 3.0$  sec and 24 h  $\geq 3.0$  times triggered dynamic ECG critical value reporting, the standard for critical values refers to the “Zhejiang Province Electrocardiogram Critical Value Standard Trial Version”<sup>34</sup>.

In ECG diagnosis, an R-R interval of  $\geq 3.0$  sec is defined as ventricular arrest. If it is accompanied by a P-P interval of  $\geq 3.0$  sec, the diagnosis is complete heart arrest. According to cause statistics, among the causes of the RRI  $\geq 3.0$  sec long, 39 cases were sinus arrest (42.85%); The critical value represents 46.15%; 31 cases of paroxysmal atrial fibrillation were terminated, accounting for 34. 07%, and the critical values represented 42. 31%. The number of cases with long RR caused by compensatory intermittence is relatively high at 2 sec, while the proportion of cases with long RR at 3 sec or more is almost nonexistent. For patients with frequent sinus arrests that last more than 3 sec, it is necessary to actively investigate the cause and provide timely treatment. If necessary, an artificial cardiac pacemaker can be implanted; Patients with paroxysmal atrial fibrillation should actively control the occurrence of atrial fibrillation through medication or radiofrequency ablation surgery to prevent complete cardiac arrest during the termination of atrial fibrillation. For cases that trigger the reporting of critical values, it is important to contact patients as soon as possible and remind them to seek medical attention promptly to prevent further adverse cardiac events. The value of F1 of the AI analysis for long RRI greater than 3 sec is 96.6802%, the statistical results are better than expected, and the AI analysis can provide effective and reliable help for the manual analysis of physicians, reducing the analysis time. It is believed that with further improvement of AI analysis accuracy, in future long-term dynamic reports provided by AI analysis will include quantitative data on RRI. The focus of manual analysis can be shifted towards qualitative cause analysis for long RRI. AI and manual analysis complement each other to improve efficiency, ensure quality and quantity, and better serve clinical needs.

Discussion

During the long-term acquisition process of the ECG signal, complex external environments often interfere with the signal, resulting in interference in the ECG signal, and there are many cases of missed or false judgments in the recognition of the R peak<sup>35,36</sup>. In clinical practice, the discovery of long RRI is made mainly by doctors manually calibrating heartbeats through dynamic ECG records, and supplementing and deleting heartbeats is very time-consuming, especially for patients with multiple long RR episodes. Long-term ECG data can help detect more arrhythmias, but long-term ECG data means more medical staff and medical resources. Therefore, a long RR assessment algorithm based entirely on artificial intelligence without excessive human reading is desirable, which will effectively aid in routine analysis of long RRI in clinical practice<sup>37,38</sup>. This article compares the accuracy of long RR automatically analyzed and manually analyzed by artificial intelligence algorithms in 480 long-term ECG data, obtaining a precision of 98.6796% and a recall rate of 98.4105%. The consistency of the longest RRI obtained by the two methods is compared and most of the sample points fall within the consistency limit 95%. Through our thorough and meticulous analysis, we have discovered an intriguing phenomenon in the delicate zone where the RRI approaches the critical threshold of 2 sec: a certain divergence emerges between





**Fig. 9.** Misjudgment of long RR caused by interference signal.



**Fig. 10.** Missed detection of long RR intervals caused by artifact heartbeats.

the professional annotations of doctors and the predictive outcomes of artificial intelligence algorithms. Upon further exploration of the causes of false positives and false negatives, we have summarized the following points.

- (1) Interference from signal disturbances: Minor disturbances in the signal constitute another significant factor that contributes to false positives. In such scenarios, while doctors can rely on experience to discern possible long RRI or interference-caused misjudgments, CNN algorithms occasionally fail in the automatic analysis process, unable to accurately differentiate between genuine and disruptive signals, as depicted in Fig. 9.
- (2) Challenges posed by artifact heartbeats: The unique physiological and pathological characteristics of certain diseases or the properties of certain devices can lead to artifact heartbeats, as shown in Fig. 10, which can lead to missed detections and false negatives in long RRI algorithms. Artifact heartbeats that occasionally appear in specific interference or disease states require enhancing the algorithm model through specific methods to address this phenomenon.

In summary, although CNN algorithms perform well in most cases, more optimization is needed when dealing with extreme or special situations to improve their completeness and accuracy. Of course, improving the anti-interference ability of recording equipment and improving the recognition rate of AI for interference can further improve the accuracy of AI recognition.

## Conclusion

This study provides a comprehensive evaluation of an automatic detection algorithm for prolonged RRI ( $\geq 2s$ ) in long-term single-lead ECG signals based on CNN algorithms. Initially, the efficiency and precision of the algorithm in identifying prolonged RRI caused by various factors were verified. This study not only confirms the tremendous potential of AI technology in automated ECG analysis, but also offers physicians a powerful quantitative analysis tool and significantly improves diagnostic efficiency and reduces the workload of medical personnel. Physicians can focus their limited time on analyzing the causes of prolonged RRI, as reflected in long-term dynamic reports. The combination of artificial intelligence and human expertise is expected to play an increasingly broad and in-depth role in future medical monitoring. In summary, this study establishes a solid foundation for the clinical application and promotion of automatic detection algorithms for prolonged RRI, highlighting its important practical significance and broad application prospects.

## Data availability

The data supporting this study's findings are available from the corresponding author upon request.

Received: 15 September 2024; Accepted: 31 March 2025

Published online: 07 April 2025

## References

1. Cai, Z. et al. An open-access long-term wearable ECG database for premature ventricular contractions and supraventricular premature beat detection. *J. Med. Imaging Health Inf.* **10**, 2663–2667 (2020).
2. McSharry, P. E. & Clifford, G. D. Models for ECG and RR interval processes. *Advanced Methods and Tools for ECG Data Analysis* 101–33 (2006).
3. Chen, W. Electrocardiogram. *Seamless Healthcare Monitoring: Advancements in Wearable, Attachable, and Invisible Devices* 3–44 (2018).
4. Peltola, M. *Analysis of heart rate variability from 24-hour ambulatory electrocardiographic recordings: significance of preprocessing of RR interval time series* University of Oulu (2011).
5. Viitasalo, M. & Karjalainen, J. QT intervals at heart rates from 50 to 120 beats per minute during 24 h electrocardiographic recordings in 100 healthy men. *Effects atenolol. Circulation* **86**, 1439–1442 (1992).
6. Davidovic, G., Iric-Cupic, V. & Milanov, S. Associated influence of hypertension and heart rate greater than 80 beats per minute on mortality rate in patients with anterior wall STEMI. *Int. J. Clin. Exp. Med.* **6**, 358 (2013).

7. Xu, H. et al. Retracted characteristics of the dynamic electrocardiogram in the elderly with nonvalvular atrial fibrillation combined with long R–R intervals. *Evid. Based Complement. Alternative Med.* **2021**, 4485618 (2021).
8. Wan-tai, M. et al. SCD-HeFT: Use of RR interval statistics for long-term risk stratification for arrhythmic sudden cardiac death. *Heart Rhythm* **12**, 2058–2066 (2015).
9. Reimann, M. et al. RR interval variations influence the degree of mitral regurgitation in dogs with myxomatous mitral valve disease. *Vet. J.* **199**, 348–354 (2014).
10. Taylor, J. A., Carr, D. L., Myers, C. W. & Eckberg, D. L. Mechanisms underlying very-low-frequency RR-interval oscillations in humans. *Circulation* **98**, 547–555 (1998).
11. Achten, J. & Jeukendrup, A. E. Heart rate monitoring: Applications and limitations. *Sports Med.* **33**, 517–538 (2003).
12. Alboni, P., Holz, A. & Brignole, M. Vagally mediated atrioventricular block: Pathophysiology and diagnosis. *Heart* **99**, 904–908 (2013).
13. Sidhu, S. & Marine, J. E. Evaluating and managing bradycardia. *Trends Cardiovasc. Med.* **30**, 265–272 (2020).
14. Reeves, A. L., Nollet, K. E., Klass, D. W., Sharbrough, F. W. & So, E. L. The ictal bradycardia syndrome. *Epilepsia* **37**, 983–987 (1996).
15. Ye, F. et al. The clinical significance of relative bradycardia. *WMJ: Official publication of the State Medical Society of Wisconsin* **117**, 73–78 (2018).
16. Rucinski, P., Rubaj, A. & Kutarski, A. Pharmacotherapy changes following pacemaker implantation in patients with bradycardia-tachycardia syndrome. *Expert Opin. Pharmacother.* **7**, 2203–2213 (2006).
17. Glikson, M. et al. 2021 ESC Guidelines on cardiac pacing and cardiac resynchronization therapy: Developed by the Task Force on cardiac pacing and cardiac resynchronization therapy of the European Society of Cardiology (ESC) With the special contribution of the European Heart Rhythm Association (EHRA). *EP Eur.* **24**, 71–164 (2022).
18. Higuchi, S., Goldschlager, N. & Gerstenfeld, E. P. Atrioventricular block with narrow and wide qrs: The pause that refreshes. *Circulation* **144**, 1262–1264 (2021).
19. De Ponti, R., Marazzato, J., Bagliani, G., Leonelli, F. M. & Padeletti, L. Sick sinus syndrome. *Cardiac Electrophysiol. Clinics* **10**, 183–195 (2018).
20. Brignole, M. Sick sinus syndrome. *Clin. Geriatr. Med.* **18**, 211–227 (2002).
21. Higuchi, S., Goldschlager, N. & Gerstenfeld, E. P. Atrioventricular block with narrow and wide QRS: The pause that refreshes. *Circulation* **144**, 1262–1264 (2021).
22. Granfeldt, A. et al. Targeted temperature management in adult cardiac arrest: Systematic review and meta-analysis. *Resuscitation* **167**, 160–172 (2021).
23. Dhananjay, B. & Sivaraman, J. The role of heart rate variability in atrial ECG components of normal sinus rhythm and sinus tachycardia subjects. In *Intelligent System Design: Proceedings of Intelligent System Design: INDIA 2019*, 637–644 (Springer, 2021).
24. Zhu, H., Zhan, J., Lam, C.-T., Chen, B. & Ng, B. K. Machine Learning Based Blind Signal Detection for Ambient Backscatter Communication Systems. *IEEE Trans. Cognitive Commun. Netw.* (2024).
25. Yu, X. & Li, D. Attention mechanism aided signal detection in backscatter communications with insufficient training data. *IEEE Trans. Vehicular Technol.* (2023).
26. Kwon, J.-M. et al. Artificial intelligence algorithm for predicting cardiac arrest using electrocardiography. *Scand. J. Trauma Resuscitation Emerg. Med.* **28**, 1–10 (2020).
27. Lee, H. et al. Real-time machine learning model to predict in-hospital cardiac arrest using heart rate variability in ICU. *NPJ Digital Med.* **6**, 215 (2023).
28. Hannun, A. Y. et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat. Med.* **25**, 65–69 (2019).
29. Mittelstadt, B., Russell, C. & Wachter, S. Explaining explanations in AI. In *Proceedings of the conference on fairness, accountability, and transparency*, 279–288 (2019).
30. Hennings, E. et al. Assessment of the atrial fibrillation burden in Holter ECG recordings using artificial intelligence. *Europace* **25**, eua122–528 (2023).
31. Xie, H., Zheng, M., Zhu, H., Wu, F. & Pan, Y. Arrhythmia Detection Based on Semantic Segmentation for Multi-lead ECG. In *2022 Comput. Cardiol. (CinC)*, vol. 498, 1–4 (IEEE, 2022).
32. Myles, P. S. & Cui, J. I. Using the Bland–Altman method to measure agreement with repeated measures (2007).
33. Huttunen, R. et al. A comparison of signal combinations for deep learning-based simultaneous sleep staging and respiratory event detection. *IEEE Trans. Biomed. Eng.* **70**, 1704–1714 (2022).
34. Shenfeng, Y. et al. Zhejiang Province ECG Critical Value Standards (2021 Revision). *Electrocardiol. Circulation* (2021).
35. Pinto, J. R., Cardoso, J. S. & Lourenço, A. Evolution, current challenges, and future possibilities in ECG biometrics. *IEEE Access* **6**, 34746–34776 (2018).
36. Uwaechia, A. N. & Ramli, D. A. A comprehensive survey on ECG signals as new biometric modality for human authentication: Recent advances and future challenges. *IEEE Access* **9**, 97760–97802 (2021).
37. Gokana, V., Phua, C. T. & Lissorgues, G. Automatic detection of atrial fibrillation using RR interval from ECG signals. In *The 15th International Conference on Biomedical Engineering: ICBME 2013, 4th to 7th December 2013, Singapore*, 215–218 (Springer, 2014).
38. Faust, O. & Acharya, U. R. Automated classification of five arrhythmias and normal sinus rhythm based on RR interval signals. *Expert Syst. Appl.* **181**, 115031 (2021).

## Author contributions

The concept of this study was conceived by Tan Lyu, who also played a key role in the critical review of the article. Data analysis was jointly conducted by Tan Lyu, Miao Ye, Minjie Yuan, and Chen Li, while Haina Chen, Sijie Han, and Lifang Yu contributed to the data collection process. The writing of the article was performed skillfully by Tan Lyu, ensuring a cohesive and comprehensive presentation of the research findings.

## Declarations

### Competing interests

No competing interest is declared.

### Ethical approval

This study was approved by the Ethics Committee of Sir Run Run Shaw Hospital, Zhejiang University School of Medicine (Approval No. 20241051NI). The responsibilities, composition of personnel, operating procedures, and records of this ethics review committee follow the “Ethical Review Measures for Biomedical Research Involving Human Subjects”, “International Ethical Guidelines for Health-Related Research Involving Human Subjects”, “Declaration of Helsinki”, GCP and ICH-GCP international ethical guidelines, as well as relevant

domestic laws and regulations.

### Additional information

**Correspondence** and requests for materials should be addressed to C.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025