A hybrid method for heartbeat classification via convolutional neural networks, multilayer perceptrons and focal loss

Tao Wang [Corresp., 1] , **Changhua Lu** [1] , **Mei Yang** [2] , **Feng Hong** [1] , **Chun Liu** [Corresp. 3]

[1] School of Computer and Information, Hefei University of Technology, Hefei, Anhui, China

[2] Beijing Huaru Technology Co., Ltd. Hefei Branch, Hefei, Anhui, China

[3] School of Electrical Engineering and Automation, Hefei University of Technology, Hefei, Anhui, China

Corresponding Authors: Tao Wang, Chun Liu
Email address: wtustc@mail.ustc.edu.cn, dqlch03@hfut.edu.cn

Background: Heart arrhythmia, as one of the most important cardiovascular diseases (CVDs), has gained wide attention in the past two decades. The paper proposes a hybrid method for heartbeat classification via convolutional neural networks, multilayer perceptrons and focal loss. Methods: In the method, a convolution neural network is used to extract the morphological features. The reason behind this is that the morphological characteristics of patients have inter-patient variations, which makes it difficult to accurately describe using traditional hand-craft ways. Then the extracted morphological features are combined with the RR intervals features and input into the multilayer perceptron for heartbeat classification. The RR intervals features contain the dynamic information of the heartbeat. Furthermore, considering that the heartbeat classes are imbalanced and would lead to the poor performance of minority classes, a focal loss is introduced to resolve the problem in the paper. Results: Tested using the MIT-BIH arrhythmia database, our method achieves an overall positive predictive value of 64.68%, sensitivity of 68.55%, f1-score of 66.09%, and accuracy of 96.27%. Compared with existing works, our method significantly improves the performance of heartbeat classification. Conclusions: Our method is simple yet effective, which is potentially used for personal automatic heartbeat classification in remote medical monitoring. The source code is provided on
https://github.com/JackAndCole/Deep-Neural-Network-For-Heartbeat-Classification .

1 **A hybrid method for heartbeat classification**

2 **via convolutional neural networks, multilayer**

3 **perceptrons and focal loss**

5 Tao Wang[1], Changhua Lu[1], Mei Yang[2], Feng Hong[1] and Chun Liu[3]

7 [1] School of Computer and Information, Hefei University of Technology, Hefei, Anhui, China

8 [2] Beijing Huaru Technology Co., Ltd. Hefei Branch, Hefei, Anhui, China

9 [3] School of Electrical Engineering and Automation, Hefei University of Technology, Hefei, Anhui,

10 China

12 Corresponding Author:

13 Chun Liu[3] and Tao Wang[1]

14 Fuicui Lake Campus of HeFei University of Technology, Hefei, Anhui, China

15 Email address: dqlch03@hfut.edu.cn (C. Liu) and wtustc@mail.ustc.edu.cn (T. Wang)

Abstract

Background: Heart arrhythmia, as one of the most important cardiovascular diseases (CVDs), has gained wide attention in the past two decades. The paper proposes a hybrid method for heartbeat classification via convolutional neural networks, multilayer perceptrons and focal loss.

Methods: In the method, a convolution neural network is used to extract the morphological features. The reason behind this is that the morphological characteristics of patients have inter-patient variations, which makes it difficult to accurately describe using traditional hand-craft ways. Then the extracted morphological features are combined with the RR intervals features and input into the multilayer perceptron for heartbeat classification. The RR intervals features contain the dynamic information of the heartbeat. Furthermore, considering that the heartbeat classes are imbalanced and would lead to the poor performance of minority classes, a focal loss is introduced to resolve the problem in the paper.

Results: Tested using the MIT-BIH arrhythmia database, our method achieves an overall positive predictive value of 64.68%, sensitivity of 68.55%, f1-score of 66.09%, and accuracy of 96.27%. Compared with existing works, our method significantly improves the performance of heartbeat classification.

Conclusions: Our method is simple yet effective, which is potentially used for personal automatic heartbeat classification in remote medical monitoring. The source code is provided on https://github.com/JackAndCole/Deep-Neural-Network-For-Heartbeat-Classification.

Introduction

Heart arrhythmia, one of the most important cardiovascular disease (CVD), refers to the irregular beating of the patient's heart. Most arrhythmias are asymptomatic and not severe, but some could cause heart disease symptoms such as passing out, lightheadedness, chest pain, shortness of breath, and even stroke and cardiac arrest such as ventricular fibrillation, ventricular escape and atrial fibrillation, which are extremely dangerous and need immediate treatment. According to statistics from the World Health Organization (WHO), the number of CVD deaths in 2015 is close to 17.7 million, accounting for about 31% of the total deaths (Shen et al. 2019).

Electrocardiogram (ECG), a device that records the electrical activity of the heart, is widely used to diagnose cardiac arrhythmias in clinical (Mondéjar-Guerra et al. 2019). An ECG signal consists of a series of periodically repeating heartbeats. Each heartbeat usually contains a QRS complex, a T wave, and a P wave, in a few cases there is a U wave (Vulaj et al. 2017). The most significant characteristic of an ECG signal is the QRS complex. By analyzing this complex, arrhythmia can be detected. However, the occurrence of arrhythmia is intermittent, especially in the early stages,

51 which makes it difficult to perform effective detection in a short time (Mondéjar-Guerra et al.
52 2019).

53 To solve this problem, a Holter monitor is often used to collect long-term heart electrical activity
54 recordings (Sannino & De Pietro 2018). In general, an ECG recording lasts several minutes or
55 even hours. Investigating a variety of abnormal arrhythmias beat-by-beat from long-term ECG
56 recordings is very exhausting, even for trained cardiologists. Therefore, there is an urgent need for
57 a computer-aided method to automatically detect abnormal heartbeats from long-term ECG data.
58 Over the past two decades, a lot of research works (de Albuquerque et al. 2018; De Chazal et al.
59 2004; Mondéjar-Guerra et al. 2019) have been spent on classifying heartbeats automatically. Most
60 of these methods are based on morphological characteristics of heartbeats and traditional signal
61 processing techniques. However, the ECG waveform and its morphological characteristics (e.g.
62 the shape of the QRS waves and P wave) of different patients are significantly different, and for
63 the same patient, there are also differences in different circumstances (Mondéjar-Guerra et al.
64 2019), so the fixed features used in these methods are not sufficient to accurately distinguish
65 arrhythmias for all patients. In recent years, some deep neural networks have been proposed, such
66 as convolutional neural networks (CNN), which can automatically extract morphological features
67 and adapt to variations between patients.

68 Nevertheless, there is another challenge when processing medical data. Due to the limited number
69 of rare classes, the number of one class may greatly exceed that of other classes, that is, the
70 distribution of classes is imbalanced. However, most algorithms try to minimize the overall
71 classification loss during the training process, which implies that these classes are equally
72 important and the same misclassification cost is allocated to all types of errors. As a result, the
73 classifier will tend to correctly classify and favor more frequent classes.

74 The paper presents a hybrid method for heartbeat classification via CNN, multilayer perceptrons
75 (MLP) and focal loss. An overall structure of the method is displayed in Fig. 1. The morphological
76 features are extracted by one-dimensional (1D) CNN and combined with the RR intervals features
77 as the input of MLP. The RR intervals features contain the dynamic information of the heartbeat,
78 which could help better capture the pattern of the ECG waveform. Furthermore, considering that
79 the heartbeat classes are imbalanced and would lead to the poor performance of minority classes,
80 a focal loss is introduced to solve the problem. It shows superior performance in various
81 application environments (Howland et al. 2002; Lin et al. 2017; Zhou et al. 2002). By testing in
82 the well-known MIT-BIH arrhythmia database (Moody et al. 2001), our method achieves superior
83 classification performance than existing heartbeat classification methods. Note that the accuracy
84 of the ECG classification method has been standardized according to AAMI's recommendations

85 (the Association for the Advancement of Medical Instrumentation). The proposed method obtains

86 an overall PPV of 64.68%, SE of 68.55%, F1 of 66.09%, and accuracy of 96.27%.

87 The paper is organized as follows: Section II presents the related works of heartbeat classification.

88 The proposed method and loss function are introduced in Section III and IV. The dataset and the

89 performance of our method against existing works are described in Section V. Section VI discusses

90 the conclusions.

Related works

92 The existing automatic heartbeat classification works can be divided into two paradigms: intra-

93 patient paradigm and inter-patient paradigm (De Chazal et al. 2004; Sannino & De Pietro 2018).

94 In the intra-patient paradigm, the dataset is based on the heartbeat label split into training and test

95 subsets, so an ECG recording will appear in two subsets (Sannino & De Pietro 2018). According

96 to de Chazal et al. (De Chazal et al. 2004), the results of this paradigm are biased, resulting in an

97 accuracy of about 100% in the test phase, because the patient's characteristics are learned during

98 the training phase (Sellami & Hwang 2019). However, in actual scenarios, the trained model must

99 be able to handle inter-patient variations during the training phase.

100 In the inter-patient paradigm, the training set and test set are from different patients (Sannino &

101 De Pietro 2018), so the differences between patients will be considered during the training process.

102 The classifier will show a better generalization capability. For instance, de Chazal et al. (De Chazal

103 et al. 2004) propose a linear discriminant heartbeat classification method based on heartbeat

104 morphological and dynamic features. Their method achieves a PPV of 38.5%, SE of 75.9% in the

105 SVEB class, and a PPV of 81.6%, SE of 80.3% in the VEB class. Ye et al. (Ye et al. 2012) apply

106 wavelet transform and independent component analysis (ICA) to extract morphological features

107 from heartbeats, and combined with dynamic RR interval features develop an SVM method to

108 classify heartbeat. A PPV of 52.3%, SE of 60.8% in the SVEB class, and a PPV of 63.1%, SE of

109 81.5% in the VEB class are obtained by their method. However, the classification accuracies of

110 these methods are significantly lower than the intra-patient paradigm-based methods. This is due

111 to variations of ECG characteristics between patients.

112 Recently, with the rapid development in deep learning, deep neural networks-based, especially

113 CNN-based, heartbeat classification methods have received a lot of attention. For example,

114 Yıldırım et al. (Yıldırım et al. 2018) develop a 1D-CNN for arrhythmia detection based on long-

115 term ECG signal. Their method achieves 91.33% overall accuracy in 17 cardiac arrhythmias.

116 Similarly, Sellami et al. (Sellami & Hwang 2019) develop a CNN with a batch-weighted loss

117 function for heartbeat classification. Hannun et al. (Hannun et al. 2019) present a deep neural

118 network with residual block to classify 12 rhythm classes. Romdhane et al. (Romdhane et al. 2020)

119 based on CNN and focal loss propose an ECG heartbeat classification method. Although the
120 performance of heartbeat classification is improved, these works mainly focus on using CNN to
121 extract the heartbeat morphological features, while ignoring the influence of RR intervals on
122 heartbeat classification. Research shows that by integrating RR interval features, the performance
123 of heartbeat classification can be significantly improved (De Chazal et al. 2004; Mondéjar-Guerra
124 et al. 2019; Sannino & De Pietro 2018). Romdhane et al. (Romdhane et al. 2020) try to use an
125 improved heartbeat segmentation method to make CNN capture RR interval information, but in
126 their work, CNN can only extract the previous RR interval information at most. This is due to the
127 incomplete division of the right interval. Different from existing works, we pre-extract the RR
128 interval information in advance, and then combine it with CNN-based morphological features as
129 the input of the classifier.

130 In addition to the above two classification paradigms, a hybrid paradigm has also been studied by
131 some scholars, namely patient-specific paradigm. In the patient-specific paradigm, a global model
132 is first built and then use part of patient data to tune the model to form a local model. de Chazal et
133 al. shows that this paradigm is superior to a pure inter-patient model. However, this paradigm
134 requires a professional doctor to label part of the ECG data, and an engineer to fine-tuning the
135 model in clinical. Meanwhile, the patient's ECG signal may change significantly over time, that is,
136 the current ECG signal may undergo large variations at some time in the future, and the use of a
137 previously fine-tuned local classifier may lead to larger misclassification. We focus on the
138 performance of our method in the inter-patient paradigm in the paper.

Methods

140 Fig. 1 shows the overall structure of the proposed method. The proposed method includes three
141 steps: ECG denoising, feature extraction, and classification. The feature extraction step contains
142 RR intervals features extraction and morphological features extraction via CNN architecture.

ECG denoising

144 The ECG signal is usually disturbed by various noises such as electromyography noise, power line
145 interference and baseline wandering (Chen et al. 2017), which makes useful features to be
146 difficultly extracted. In this step, most previous works typically perform a baseline wandering
147 removal and then high-frequency noise filtering (Mondéjar-Guerra et al. 2019). However,
148 excessive filtering will lead to the loss of some helpful information in the ECG signal. Since CNN
149 has better noise immunity (Huang et al. 2018), we only perform baseline wandering removal and
150 preserve as much information as possible from the raw ECG signal.

151 Two median filters are combined to remove the baseline wandering of the ECG signal in the paper.
152 First, the QRS complexes and P-waves are removed using a 200-ms width median filter, and then

153   a 600-ms width median filter is further adopted to remove T-waves. The output is the baseline
154   wandering of the ECG signal, and the baseline-corrected ECG signal can be achieved by
155   subtracting it from the original signal. An effect of baseline wandering removal is shown in Fig.
156   2.
157   After obtaining the baseline-corrected ECG signal, the ECG is further segmented into a series of
158   heartbeats based on the labeled R-peaks. In specific, for each heartbeat, we obtain 200 sampling
159   points of the ECG signal segment, 90 sampling points before and 110 sampling points after the
160   labeled R peak. The R-peak detection is not the focus of the paper and we directly use labeled R-
161   peaks in the dataset, as there are many high-precision (>99%) R-peak detection methods in the
162   literature (Adam 2012; Pan & Tompkins 1985).

163   **RR Interval features extraction**

164   The time interval between two consecutive R-peaks is normally called the RR interval
165   (Ruangsuwana et al. 2010), which contains the dynamic information of the heartbeat. To capture
166   this information for heartbeat classification, four features are extracted from the RR interval,
167   namely previous RR interval, post RR interval, ratio RR and local RR interval. The previous RR
168   interval refers to the distance between the current R-peak position and the previous R-peak, the
169   post RR interval is the distance between the current R-peak position and the following one. The
170   ratio RR represents the ratio of the previous RR interval and the post RR interval. These three
171   features reflect the instantaneous rhythm of a heartbeat. The average value of the 10 RR intervals
172   before the current heartbeat is taken as the local RR interval, which represents the overall rhythm
173   in the past. Due to the inter-patient variations in the ECG signal, the RR interval of different
174   patients cannot be directly compared, in this paper we use the entire patient's ECG signal to
175   calculate the average RR interval, and subtract it from all RR characteristics (expect the ratio RR)
176   to eliminate this effect.

177   **Morphological features extraction via CNN architecture**

178   CNN is a powerful deep neural network inspired by visual neuroscience (Chu et al. 2019b). It has
179   been successfully used in speech recognition, natural language processing, image classification,
180   and biomedical signal (Palaz & Collobert 2015; Pourbabaee et al. 2018; Yin et al. 2017). Given an
181   image, CNN can effectively learn high-level abstractions, which can then be input into the
182   classifier (e.g. fully connected neural network and support vector machine (SVM)) for
183   classification (Zhang et al. 2017). A CNN usually consists of convolutional layers, activation
184   functions, and pooling layers, and sometimes including batch normalization layers.

185   Convolutional Layer: It is the most important component in CNN and performs convolution
186   operation on the input data (Liu & Chen 2017). Let f_k and s be the filter and the 1D ECG signal,
187   respectively. The output of the convolution is calculated as follows:

188
$$C[i] = s(i) * f_k(i) = \sum_m s(m) f_k(i - m)$$

189   where m is the size of the filter and the filter f_k is realized by sharing the weights of adjacent
190   neurons.
191   Activation Function: The activation function is used to determine whether the neuron should be
192   activated. The purpose is to enable neurons to achieve nonlinear classification. Rectifier Linear
193   Unit (ReLU) is one of the most widely used activation function, which can be expressed as

194
$$f(x) = max(0, x)$$

195   where x is the output value of the neuron.
196   Pooling layer: The pooling layer, also known as the down-sampling layer, is an operation that
197   decreases the computational intensity by reducing the output neuron dimension of the
198   convolutional layer, and can handle some variations due to signal shift and distortion (Zhang et al.
199   2017). The most widely used pooling method is the max-pooling, which is to apply the maximum
200   function over input s. Let m be the filter size, and the output is:

201
$$M(x) = max \left\{ s(x + k) \| |k| \leq \frac{m - 1}{2} \right\}$$

202   Batch Normalization Layer: The batch normalization layer is a technology for standardizing
203   network input, applied to either the activations of a prior layer or inputs directly, which can
204   accelerate the training process, and provides some regularization, reducing generalization error.
205   Let $B = \{x_i, i = 1, \cdots, m\}$ be a mini-batch of the entire training set, the output of batch normalization
206   is as follows:

207
$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}$$

208   where σ_B and μ_B are the variance and the mean of training set B, respectively. ϵ is an arbitrarily
209   small constant to ensure the denominator is not zero.
210   A CNN is developed and utilized for heartbeat morphological feature extraction in this paper. The
211   CNN architecture is displayed in Fig. 1. It contains 3 convolutional blocks and 3 pooling layers.
212   Each convolutional block includes a convolution layer, a ReLU activation function and a batch
213   normalization layer. The kernel of the convolution is reduced as the network structure becomes
214   deeper. For instance, the first convolution kernel is 11, while the second is reduced to 5. A batch
215   normalization and ReLU activation are applied after each convolution operation, and a max-

216   pooling is used to reduce the spatial dimension. Note that the parameters of the convolutional

217   network are usually set based on the author's experience. The detailed parameters of CNN

218   architecture are listed in Table 1. The output of the last pooling layer is the morphological features

219   extracted by CNN from the heartbeat. An illustration of the morphological features extracted from

220   the heartbeat is shown in Fig. 1 (c).

221   **MLP classifier**

222   The CNN-based morphological features and RR interval features are combined as the input of the

223   classifier in the paper. In general, any classifier (i.e. support vector machine (SVM) and random

224   forest (RF)) can be used for heartbeat classification. Here, we adopt a multilayer perceptron (MLP,

225   also known as fully connected neural networks in deep learning) as the classifier. The reason

226   behind this is that CNN and MLP can be combined for parameter training (we call it one-step

227   training). Compared with other methods, this usually achieves better performance. Specifically,

228   our MLP classifier contains an input layer, a hidden layer and an output layer. The input layer

229   consists of two parts of information: CNN-based morphological features and RR interval features.

230   The hidden layer has 64 neurons, and each neuron is connected to the input features. The output

231   layer neurons are 4 in the paper, each representing a kind of arrhythmia or normal heartbeat. The

232   details of our method are shown in Fig. 1 and Table 1.

233   ## Loss function

234   Before training a deep neural network, a loss function is first needed. The cross-entropy loss is the

235   most widely used in deep neural network classification (Chu et al. 2019a). However, this loss

236   function does not address the class imbalance problem. A focal loss function is introduced in the

237   paper to deal with this problem.

238   **Cross-entropy loss**

239   The cross-entropy is a measure in information theory (Robinson et al. 2001). It is based on entropy

240   and calculates the difference between two probability distributions. Closely related to KL

241   divergence that computers the relative entropy between two probability distributions, but the cross-

242   entropy calculates the total entropy between the distributions. The cross-entropy is usually taken

243   as the loss function in deep neural network classification (Chu et al. 2019a).

244   Let t_i and p_i be the ground truth and the estimated probability of each category, the cross-entropy

245   loss is computed by:

246
$$CE = -\sum_{i}^{C} t_i \cdot \log(p_i)$$

247   where C refers to the category set of the heartbeat. In the cross-entropy loss, each category is treated

248   equally, which causes the majority category to overwhelm the loss and the model tends to classify

249   to the majority category in an imbalanced environment.

250   **Focal loss**

251   A characteristic of cross-entropy loss is that even easy-to-classify examples can cause significant

252   losses, which will cause the loss of easy examples that constitute most of the dataset during the

253   training process to negatively affect rare classes (Lin et al. 2017). The focal loss is designed to

254   deal with this imbalanced problem by reshaping the cross-entropy loss function by reducing the

255   attention to easy examples and focusing on difficult ones. A general formula for focal loss is

256   expressed as:

257
$$FL = -\sum_{i}^{C} t_i \cdot (1 - p_i)^\gamma \log(p_i)$$

258   where γ acts as the modulating factor. As shown in Fig. 3, the higher the γ value, the lesser the

259   cost incurred by well-classified examples. In practice, the α-balanced variant of the focal loss is

260   usually used when one or more categories are highly imbalanced, which is defined as:

261
$$FL' = -\sum_{i}^{C} t_i \cdot \alpha_i (1 - p_i)^\gamma \log(p_i)$$

262   where α_i is the weighting factor of each category.

263   **Results**

264   **Data set**

265   The ECG dataset from the MIT-BIH arrhythmia database (Moody et al. 2001) is used to test our

266   proposed method. This dataset contains 48 30-minute ambulatory two leads ECG signal records

267   collected from 47 subjects. Each ECG signal is sampled at 360 Hz with an 11-bit resolution. The

268   first lead is the modified-lead II (ML II), and the second lead depends on the record, one of V1,

269   V2, V4 or V5. The heartbeat of these ECG signals is independently labeled by two or more doctors,

270   and there are about 110,000 heartbeats.

271   According to the recommendation of the Association for the Advancement of Medical

272   Instrumentation (AAMI), these heartbeats are further divided into five heartbeat classes. Table 2

273   shows the mapping of AAMI classes and MIT-BIH arrhythmia heartbeat types. Since Q is

274   practically non-existent, we ignore it like others (Mar et al. 2011; Zhang et al. 2014). Meanwhile,

275   four recordings with paced beats are removed in consistent with the AAMI recommended practice,

276   namely 102, 104, 107 and 217. Since all records have ML II ECG signals and they are widely used

277   in wireless body sensor network (WBSN) based ECG applications, this lead ECG signal is used

278   for heartbeat classification in the paper.

279   As we mentioned in related works, the paper focuses on the heartbeat classification under the inter-

280   patient paradigm. To facilitate comparison with existing works, we follow de Chazal et al. (De

281   Chazal et al. 2004) to split the dataset into two subsets. Each contains regular and complex

282   arrhythmia records and has roughly the same number of heartbeat types. Table 3 shows the details

283   of two subsets. The first (DS1) is used for training whereas the second (DS2) is used to test the

284   heartbeat classification performance (De Chazal et al. 2004). No patient appears in both subsets at

285   the same time.

286   **Model training and Performance metrics**

287   In the study, the general focal loss (non-α-balanced focal loss) is used as the loss function, and the

288   modulating factor γ is set to the default value ($\gamma=2$). Since Adam can accelerate the model training,

289   we use it as the optimizer. The batch size of the model is set to 512 and the maximum epoch is 50.

290   The initial learning rate is 0.001, and reduced by 0.1 times every 10 epochs. In addition, in order

291   to avoid overfitting, the l2 penalty is set to 1e-3 based on trial and error. The model is implemented

292   using Keras and trained on the NVIDIA GeForce RTX 2080Ti graphical processing unit.

293   To evaluate the performance of our proposed method, three widely used metrics are adopted,

294   namely positive predictive value (PPV), sensitivity (SE), and accuracy (ACC), which are defined

295   as:

296

$$PPV_i = \frac{TP_i}{TP_i + FP_i}$$

297

$$SE_i = \frac{TP_i}{TP_i + FN_i}$$

298

$$ACC_i = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}$$

299   where TP_i (true positive) refers to the number of the ith class is correctly classified, FP_i (false

300   positive) is equal to the number of heartbeats misclassified as the ith class, TN_i (true negative) is

301   the number of heartbeats that are not in the ith class and not classified into the ith class, and FN_i

302   (false negative) is equal to the number of heartbeats of the ith class classified as other classes. PPV_i

303   indicates the proportion of positive correct classification, and SE_i reflects the sensitivity of the

304   classifier in the ith class. ACC_i is the ratio of all correct classifications.

305   Since the heartbeat classes are imbalanced, *f1*-score (*F1*) is also selected as the performance
306   measure, defined as:

307
$$F1_i = \frac{2 \times PPV_i \times SE_i}{PPV_i + SE_i}$$

308   *f1*-score takes both the positive predictive value PPV_i and sensitivity SE_i into account, and is
309   generally useful than ACC_i in the imbalance class distribution (Chen 2009).

310   **Comparison with existing works**
311   Based on de Chazal et al. (De Chazal et al. 2004), the dataset is divided into DS1 and DS2 datasets.
312   DS1 is used for training and DS2 is used to test our proposed method. For fair evaluation, we
313   compare works (Chen et al. 2017; De Chazal et al. 2004; Garcia et al. 2017; Liu et al. 2019; Mar
314   et al. 2011; Zhang et al. 2014) that adopt the same strategy. As SVEB and VEB are more important
315   than other classes, we list the detailed information of these two classes in Table 4. The
316   experimental results show that the proposed method has better recognition in the inter-patient
317   paradigm, with F1s of SVEB and VEB of 74.29% and 92.40%, respectively. In particular, the PPV
318   of SVEB is 68.34%, indicating that the proposed method has better SVEB recognition ability. The
319   93.72% SE of VEB is superior to most reported works. The evaluation results of all four-classes
320   are listed in Table 5. The results related to PPV and SE are close to or surpass those obtained with
321   existing works except for F. For category F, it is mainly composed of the fusion of ventricular beat
322   and normal beat, which is very close to the normal heartbeat. Meanwhile, compared with other
323   categories, F has the least number and the most serious imbalance. As a result, the performance of
324   existing works is unstable in this category, usually a large number of N is predicted as F or F is
325   predicted as N. In the paper, although the focus loss is introduced, due to the high imbalance of F,
326   the proposed method cannot extract the discriminate features. Mar et al., although, obtains the best
327   PPV in F, a large number of N is incorrectly classified as F. We suggest that category F can be
328   included in other categories in future research.

Discussion
330   **Focal loss vs. Cross-entropy loss**
331   Since the heartbeat has an imbalanced class distribution, the cross-entropy loss is replaced by the
332   focal loss as the loss function of the model in the paper. The performance comparison of the two
333   losses is listed in Table 6. Both losses have similar overall accuracy, but compared to cross-entropy
334   loss, the overall PPV, SE and F1 of the focal loss are significantly improved. An overall PPV of
335   64.68%, SE of 68.55%, and F1 of 66.09% are achieved by the focal loss, while the cross-entropy
336   loss obtains an overall PPV of 59.82%, SE of 66.77%, and F1 of 62.67%. The corresponding

337 metrics increased by 4.86%, 1.78%, and 3.42%, respectively. In addition, for each specific class,
338 the PPV, SE, and F1 of the focal loss also have achieved comparable or better performance than
339 the cross-entropy loss, especially in the F1.

340 The confusion matrix of the two losses is listed in Table 7. The focal loss achieves a total of 45,952
341 correct predictions, while the cross-entropy obtains 45,358 correct predictions. With the focal loss,
342 the total correct prediction has slightly increased, perhaps due to the suppression of easy-to-
343 classify samples by focal loss.

Conclusions

345 A hybrid method for heartbeat classification via CNN, MLP and focal loss is developed in the
346 paper. Among them, CNN is used to extract the morphological features of the heartbeat. Then the
347 morphological features are combined with the RR intervals features and input into the MLP to
348 perform heartbeat classification. Furthermore, in order to avoid the impact of heartbeat imbalance,
349 a focal loss function is introduced. Tested by using the MIT-BIH arrhythmia database, the
350 experimental results confirm that the method has good overall performance, with F1 of 66.09%
351 and accuracy of 96.27%. The superiority of the proposed method is due to multifactorial: (I)
352 Compared with traditional hand-craft features, CNN as an automatic extraction method can adapt
353 to small mutations in ECG signals to obtain powerful features; (II) Besides the CNN-based
354 morphological features, the pre-extracted RR interval features are also combined to build the
355 model, avoiding the loss of dynamic information due to heartbeat segmentation; (III) A focal loss
356 function is introduced to solve the class imbalance, preventing the model from biasing towards the
357 majority class; (IV) One-step training can improve the model to obtain better feature abstraction
358 capabilities. Due to the simple yet effective of the proposed inter-patient method, it has the
359 potential to be used for personal automatic heartbeat classification for surveillance in telemedicine.
360 The encouraging results have inspired continuous exploration. The future work will include (I)
361 testing the performance of the developed model with more ECG signals; (II) designing or
362 modifying CNN architecture to further improve the performance of our method; (III) trying to use
363 additional techniques such as wavelet transform to convert time-domain information to frequency-
364 domain information to reduce the difficulty of CNN feature extraction.

References

366 Adam GJBE. 2012. ECG Signal Processing, Classification and Interpretation.278.
367 Chen S, Hua W, Li Z, Li J, Gao XJBSP, and Control. 2017. Heartbeat classification using projected and dynamic
368   features of ECG signal. 31:165-173.
369 Chen YJDoCS, Iowa State University. 2009. Learning classifiers from imbalanced, only positive and unlabeled data
370   sets.

371 Chu J, Wang H, Lu WJJoMiM, and Biology. 2019a. A novel two-lead arrhythmia classification system based on CNN
372   and LSTM.   19:1950004.

373 Chu Y, Shen H, and Huang KJIA. 2019b. ECG authentication method based on parallel multi-scale one-dimensional
374   residual network with center and margin loss.   7:51598-51607.

375 de Albuquerque VHC, Nunes TM, Pereira DR, Luz EJdS, Menotti D, Papa JP, Tavares JMRJNC, and Applications.
376   2018. Robust automated cardiac arrhythmia detection in ECG beat signals.   29:679-693.

377 De Chazal P, O'Dwyer M, and Reilly RBJItobe. 2004. Automatic classification of heartbeats using ECG morphology
378   and heartbeat interval features.   51:1196-1206.

379 Garcia G, Moreira G, Menotti D, and Luz E. 2017. Inter-Patient ECG Heartbeat Classification with Temporal VCG
380   Optimized by PSO. *Scientific Reports* 7:10543. 10.1038/s41598-017-09837-3

381 Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, and Ng AY. 2019. Cardiologist-level
382   arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network.
383   *Nature Medicine* 25:65-69. 10.1038/s41591-018-0268-3

384 Howland DS, Liu J, She Y, Goad B, Maragakis NJ, Kim B, Erickson J, Kulik J, DeVito L, and Psaltis GJPotNAoS.
385   2002. Focal loss of the glutamate transporter EAAT2 in a transgenic rat model of SOD1 mutant-mediated
386   amyotrophic lateral sclerosis (ALS).   99:1604-1609.

387 Huang J, Liu H, Dai J, and Cai W. 2018. Reconstruction for limited-data nonlinear tomographic absorption
388   spectroscopy via deep learning.   218:187. 10.1016/j.jqsrt.2018.07.011

389 Lin T-Y, Goyal P, Girshick R, He K, and Dollár P. 2017. Focal loss for dense object detection. Proceedings of the
390   IEEE international conference on computer vision. p 2980-2988.

391 Liu J, Song S, Sun G, and Fu Y. 2019. Classification of ECG Arrhythmia Using CNN, SVM and LDA. Cham: Springer
392   International Publishing. p 191-201.

393 Liu Y, and Chen Y. 2017. Recognition of facial expression based on CNN-CBP features. 2017 29th Chinese Control
394   And Decision Conference (CCDC): IEEE. p 2139-2145.

395 Mar T, Zaunseder S, Martínez JP, Llamedo M, and Poll RJItoBE. 2011. Optimization of ECG classification by means
396   of feature selection.   58:2168-2177.

397 Mondéjar-Guerra V, Novo J, Rouco J, Penedo MG, Ortega MJBSP, and Control. 2019. Heartbeat classification fusing
398   temporal and morphological information of ECGs via ensemble of classifiers.   47:41-48.

399 Moody GB, Mark RGJIEiM, and Magazine B. 2001. The impact of the MIT-BIH arrhythmia database.   20:45-50.

400 Palaz D, and Collobert R. 2015. Analysis of cnn-based speech recognition system using raw speech as input. Idiap.

401 Pan J, and Tompkins WJJItobe. 1985. A real-time QRS detection algorithm.230-236.

402 Pourbabaee B, Roshtkhari MJ, Khorasani KJIToS, Man,, and Systems C. 2018. Deep convolutional neural networks
403   and learning ECG features for screening paroxysmal atrial fibrillation patients.   48:2095-2104.

404 Robinson S, Cattaneo A, and El-Said MJESR. 2001. Updating and estimating a social accounting matrix using cross
405   entropy methods.   13:47-64.

406 Romdhane TF, Alhichri H, Ouni R, and Atri M. 2020. Electrocardiogram heartbeat classification based on a deep
407   convolutional neural network and focal loss. *Computers in Biology and Medicine* 123:103866.
408   https://doi.org/10.1016/j.compbiomed.2020.103866

409 Ruangsuwana R, Velikic G, and Bocko M. 2010. Methods to extract respiration information from ECG signals. 2010
410   IEEE International Conference on Acoustics, Speech and Signal Processing: IEEE. p 570-573.
411 Sannino G, and De Pietro GJFGCS. 2018. A deep learning approach for ECG-based heartbeat classification for
412   arrhythmia detection.   86:446-455.
413 Sellami A, and Hwang HJESwA. 2019. A robust deep convolutional neural network with batch-weighted loss for
414   heartbeat classification.   122:75-84.
415 Shen R, Yu Y, Lan R, Yu R, Yuan Z, and Xia ZJEP. 2019. The cardiovascular toxicity induced by high doses of
416   gatifloxacin and ciprofloxacin in zebrafish.   254:112861.
417 Vulaj Z, Draganić A, Brajović M, and Orović I. 2017. A tool for ECG signal analysis using standard and optimized
418   Hermite transform. 2017 6th Mediterranean Conference on Embedded Computing (MECO): IEEE. p 1-4.
419 Ye C, Kumar BV, and Coimbra MTJIToBE. 2012. Heartbeat classification using morphological and dynamic features
420   of ECG signals.   59:2930-2941.
421 Yıldırım Ö, Pławiak P, Tan R-S, and Acharya UR. 2018. Arrhythmia detection using deep convolutional neural
422   network with long duration ECG signals. *Computers in Biology and Medicine* 102:411-420.
423   https://doi.org/10.1016/j.compbiomed.2018.09.009
424 Yin W, Kann K, Yu M, and Schütze HJapa. 2017. Comparative study of cnn and rnn for natural language processing.
425 Zhang Q, Zhou D, and Zeng XJIA. 2017. HeartID: A multiresolution convolutional neural network for ECG-based
426   biometric human identification in smart health applications.   5:11805-11816.
427 Zhang Z, Dong J, Luo X, Choi K-S, Wu XJCib, and medicine. 2014. Heartbeat classification using disease-specific
428   feature selection.   46:79-89.
429 Zhou F-Q, Waterman-Storer CM, and Cohan CSJTJocb. 2002. Focal loss of actin bundles causes microtubule
430   redistribution and growth cone turning.   157:839-849.
431

Figure 1

A scheme of our proposed method.

a) Overview of our method. b) CNN block of our CNN architecture. c) An example of morphological features extracted by CNN, where the upper part heartbeat signal is the input of CNN, and the lower part is the features extracted by CNN, that is, the morphological features. These features will be flattened and combined with the RR interval features when used as the input of the MLP classifier.

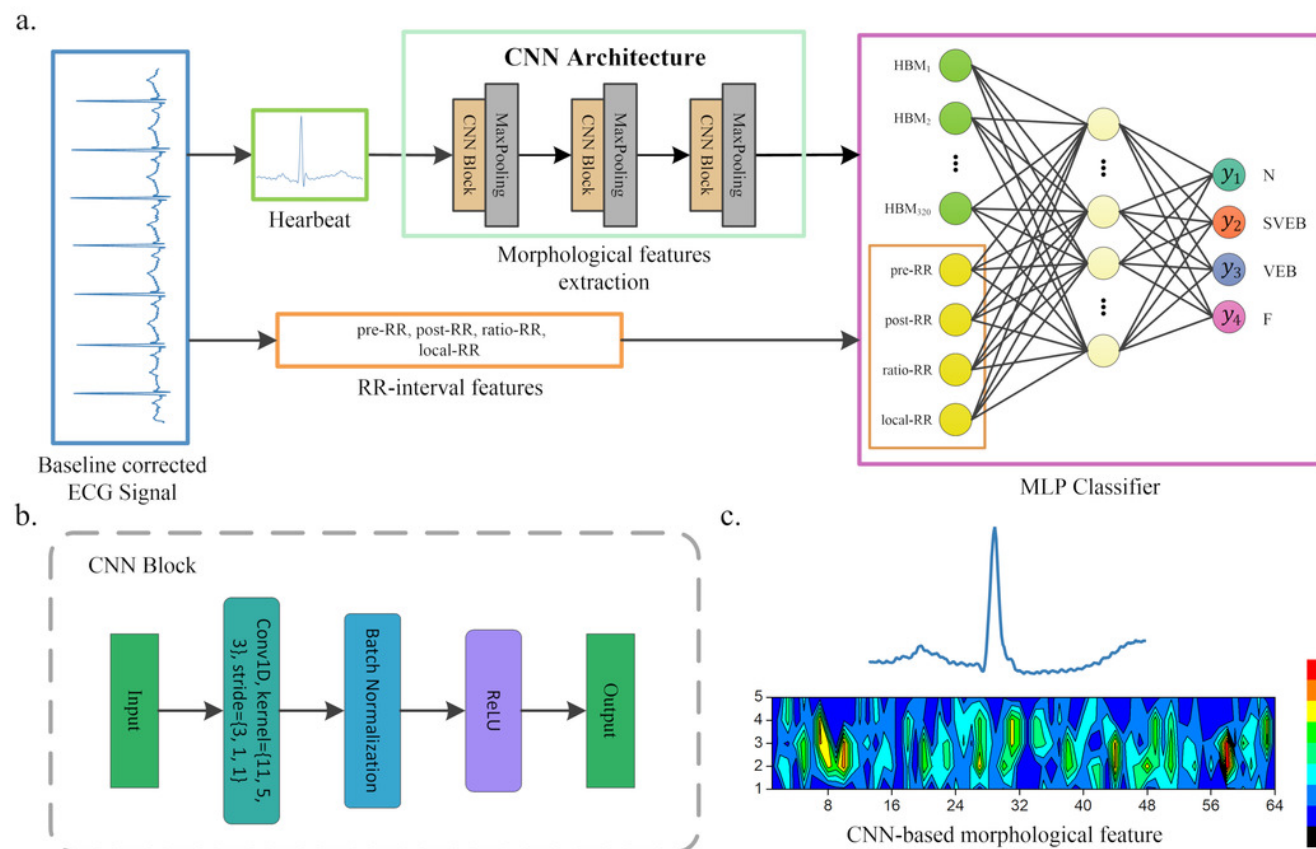



Figure 2

Example of baseline wandering removal.

a) Raw ECG signal. b) Raw heartbeat. c) Baseline-corrected ECG signal. d) Baseline-corrected heartbeat. It is easy to notice that after removing the baseline wandering, the heartbeat is shifted to zero.

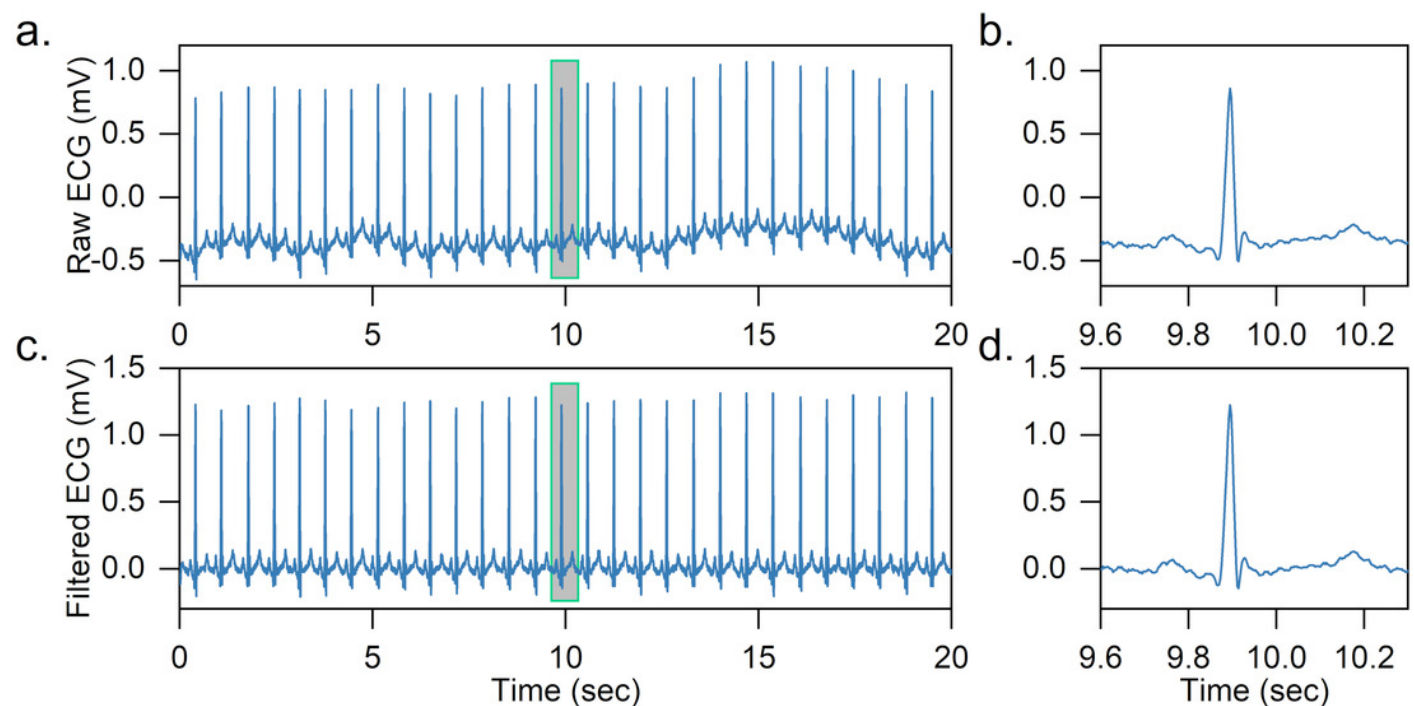
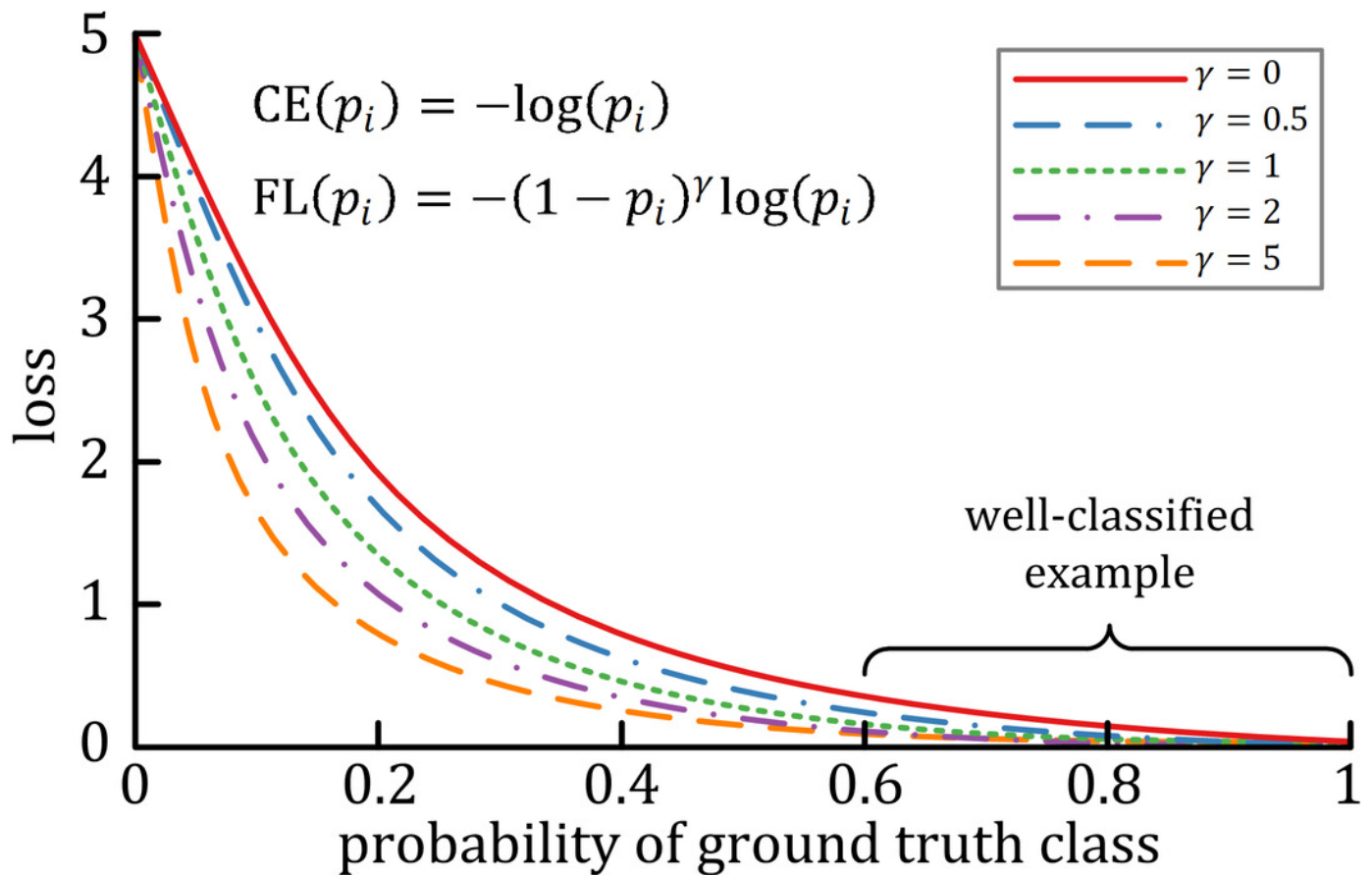



Figure 3

The relationship between the modulating factor γ and the cost of the well-classified examples (Lin et al. 2017) .

The detailed parameters of our proposed deep neural network.

Layers	Layer Name	Kernel Size	No. of Filters	Stride	Output Shape	No. of Trainable Parameters	No. of Non-trainable Parameters
0	Input1[a]	-	-	-	200 x 1	-	-
1	1D Convolution	11	16	3	64 x 16	192	-
2	Batch Normalization	-	-	-	64 x 16	32	32
3	ReLU	-	-	-	64 x 16	-	-
4	Max-Pooling	3	-	2	31 x 16	-	-
5	1D Convolution	5	32	1	27 x 32	2592	-
6	Batch Normalization	-	-	-	27 x 32	64	64
7	ReLU	-	-	-	27 x 32	-	-
8	Max-Pooling	3	-	2	13 x 32	-	-
9	1D Convolution	3	64	1	11 x 64	6208	-
10	Batch Normalization	-	-	-	11 x 64	128	128
11	ReLU	-	-	-	11 x 64	-	-
12	Max-Pooling	3	-	2	5 x 64	-	-
13	Flatten	-	-	-	320	-	-
14	Input2[b]	-	-	-	4	-	-
15	Concatenate	-	-	-	324	-	-
16	Dense	-	-	-	64	20800	-
17	Dense	-	-	-	4	260	-

[a] refers to the raw signal of the heartbeat. The morphological features of the heartbeat will be obtained through the CNN architecture.

[b] is the RR interval features of the heartbeat. It will be combined with the CNN-based morphological features to build the final classification model.

Mapping of AAMI classes and MIT-BIH arrhythmia heartbeat types.

AAMI Classes	MIT-BIH Types	MIT-BIH Annotate
Normal (N)	Normal beat (NOR)	N
	Nodal (junctional) escape beat (NE)	j
	Atrial escape beat (AE)	e
	Right bundle branch block beat (RBBB)	R
	Left bundle branch block beat (LBBB)	L
Supraventricular ectopic beat (SVEB)	Aberrated arial premature beat (aAP)	a
	Premature or ectopic supraventricular beat (SP)	S
	Nodal (junctional) premature beat (NP)	J
	Atrial premature beat (AP)	A
Ventricular ectopic beat (VEB)	Ventricular escape beat (VE)	E
	Premature ventricular contraction (PVC)	V
Fusion beat (F)	Fusion of ventricular and normal beat (fVN)	F
Unknown beat (Q)	Unclassifiable beat (U)	Q
	Fusion of paced and normal beat (fPN)	f
	Paced beat (P)	/

Detailed breakdown of the dataset.

Dataset	No. Of Samples Per AAMI Class				Total
	N	SVEB	VEB	F	
DS1	45,824	943	3,788	414	50,969
DS2	44,218	1,836	3,219	388	49,661
Total (DS1+DS2)	90,042	2,779	7,007	802	100,630

Performance comparison of our proposed method with existing works in SVEB and VEB classes.

Methods	SVEB					VEB			
	PPV (%)	SE (%)	F1 (%)	Accuracy (%)		PPV (%)	SE (%)	F1 (%)	Accuracy (%)
de Chazal et el.	38.53	75.98	51.13	94.61		81.67	80.31	80.98	97.62
Chen et al.	38.40	29.50	33.36	95.34		85.25	70.85	77.38	97.32
Zhang et al.	35.98	79.06	49.46	93.33		92.75	85.48	88.96	98.63
Mar et al.	33.53	83.22	47.80	93.28		75.89	86.75	80.96	97.35
Liu et al.	39.87	33.12	36.18	95.49		76.51	90.20	82.79	97.45
Garcia et al.	53.00	62.00	57.15	-		59.40	87.30	70.70	-
Our Proposed	68.34	81.37	74.29	97.92		91.12	93.72	92.40	99.00

Performance comparison of our proposed method with existing works in all four classes.

Methods	Accuracy[a]	Macro-F1[b]	N		SVEB		VEB		F	
			PPV (%)	SE (%)	PPV (%)	SE (%)	PPV (%)	SE (%)	PPV (%)	SE (%)
de Chazal et el.	86.24	60.12	99.17	87.06	38.53	75.98	81.67	80.31	8.57	89.43
Chen et al.	93.14	51.91	95.42	98.42	38.40	29.50	85.25	70.85	0.00	0.00
Zhang et al.	88.34	64.02	98.98	88.94	35.98	79.06	92.75	85.48	13.73	93.81
Mar et al.	88.99	62.24	99.12	89.64	33.53	83.22	75.89	86.75	16.57	61.08
Liu et al.	-	58.50	96.66	94.06	39.87	33.12	76.51	90.20	12.99	40.72
Garcia et al.	92.40	55.95	98.00	94.00	53.00	62.00	59.40	87.30	-	-
Our Proposed	92.53	66.09	98.20	93.67	68.34	81.37	91.12	93.72	1.06	5.41

[a] Accuray = $(TP_N + TP_{SVEB} + TP_{VEB} + TP_F)$ / number of testing heartbeats

[b] refers to the average f1-score of four AAMI classes.

Performance comparison of focal loss and cross-entropy loss.

Methods	AAMI Class	Performance Metrics			
		PPV (%)	SE (%)	F1 (%)	Accuracy (%)
	N	98.20	93.67	95.88	92.84
	SVEB	68.34	81.37	74.29	97.92
Focal Loss	VEB	91.12	93.72	92.40	99.00
	F	1.06	5.41	1.77	95.31
	Average[a]	64.68	68.55	66.09	96.27
	N	98.11	92.42	95.18	91.67
	SVEB	57.11	80.72	66.89	97.05
Cross-Entropy Loss	VEB	83.95	93.45	88.44	98.42
	F	0.11	0.52	0.18	95.54
	Average[a]	59.82	66.77	62.67	95.67

[a] refers to the average value of the corresponding metrics of four AAMI classes.

Confusion matrix of focal loss and cross-entropy loss.

	Focal Loss	Predicted Class				Total		Cross-Entropy Loss	Predicted Class				Total
		N	SVEB	VEB	F				N	SVEB	VEB	F	
Ground Truth	N	41420	671	206	1921	44218		N	40866	1082	462	1808	44218
	SVEB	282	1494	57	3	1836		SVEB	246	1482	95	13	1836
	VEB	141	21	3017	40	3219		VEB	173	31	3008	7	3219
	F	336	0	31	21	388		F	368	0	18	2	388
	Total	42179	2186	3311	1985	49661		Total	41653	2595	3583	1830	49661