Diagnosis of paroxysmal atrial fibrillation from thirty-minute heart rate variability data using convolutional neural networks

MURAT SÜRÜCÜ
YALÇIN İŞLER
RESUL KARA

Follow this and additional works at: https://journals.tubitak.gov.tr/elektrik

Part of the Computer Engineering Commons, Computer Sciences Commons, and the Electrical and Computer Engineering Commons

Recommended Citation
Available at: https://journals.tubitak.gov.tr/elektrik/vol29/iss8/20

This Article is brought to you for free and open access by TÜBİTAK Academic Journals. It has been accepted for inclusion in Turkish Journal of Electrical Engineering and Computer Sciences by an authorized editor of TÜBİTAK Academic Journals. For more information, please contact academic.publications@tubitak.gov.tr.
Diagnosis of paroxysmal atrial fibrillation from thirty-minute heart rate variability data using convolutional neural networks

Murat SURUCU¹, Yalcın ISLER²,³,∗, Resul KARA¹

¹Department of Computer Engineering, Faculty of Engineering, Düzce University, Düzce, Turkey
²Department of Biomedical Engineering, Faculty of Engineering, İzmir Katip Çelebi University, İzmir, Turkey
³İslerya Medical and Information Technologies Company, İzmir, Turkey

Abstract: Paroxysmal atrial fibrillation (PAF) is the initial stage of atrial fibrillation, one of the most common arrhythmia types. PAF worsens with time and affects the patient’s life quality negatively. In this study, we aimed to diagnose PAF early, so patients can start taking precautions before this disease gets worse. We used the atrial fibrillation prediction database, an open data from Physionet and constructed our approach using convolutional neural networks. Heart rate variability (HRV) features are calculated from time-domain measures, frequency-domain measures using power spectral density estimations (fast Fourier transform, Lomb–Scargle, and Welch periodogram), time-frequency-domain measures using wavelet transform, and nonlinear Poincare plot measures. We also normalized these features using min-max normalization and z-score normalization methods. In addition, we also applied alternatively the heart rate normalization (HRN), which gave promising results in a few HRV-based research, before calculating these features. Thus, HRV data, HRN data, and HRV features extracted from six different combinations of these normalizations, in addition to no normalization cases, were applied to the convolutional neural networks classifier. We tuned the classifiers using 90% of samples and tested the classifiers’ performances using 10% of data. The proposed approach resulted in 95.92% accuracy, 100% precision, 91.84% recall, and 95.74% f1-score in HRV with z-score feature normalization. When the heart rate normalization was also applied, the proposed approach reached 100% accuracy, 100% precision, 100% recall, and 100% f1-score in HRV with z-score feature normalization. The proposed method with heart rate normalization and z-score normalization methods resulted in better classification performance than similar studies in the literature. In addition, although deep learning models offer no use of separate feature extraction processes, this study reveals that using HRV-specific feature extraction techniques may improve the performance of deep learning algorithms in HRV-based studies. Comparing the existing studies, we concluded that our approach provides a much better tool to diagnose PAF patients.

Key words: Deep learning, convolutional neural networks, paroxysmal atrial fibrillation, heart rate variability, classification

1. Introduction

Atrial fibrillation (AF) is one of the most common heart rhythm disorder diseases that can cause blood clotting, heart failure, stroke, and even death [1]. AF usually begins in a paroxysmal (self-terminating) form. Paroxysmal AF (PAF) is a type of AF that lasts within 2 min to 7 days. It gradually worsens and transforms into more permanent forms [2]. As with all diseases, it is essential to diagnose AF at an early stage (PAF stage) and

*Correspondence: islerya@yahoo.com

This work is licensed under a Creative Commons Attribution 4.0 International License.
initiate treatment to increase patient comfort and reduce the risk of death [3]. However, PAF is usually not diagnosed earlier because it is asymptomatic in general [4]. When PAF cases reach the clinically detectable stage, symptoms that decrease the quality of life emerge suddenly [5]. This study aims to distinguish individuals with PAF disease from individuals without known heart disease.

Electrocardiography (ECG) is the most commonly used functional test to diagnose and monitor the treatment progress of several heart diseases [6]. The doctors examine ECG graphs directly, or a computer program analyzes numerical ECG data. Many studies in the literature present methods to detect heart diseases using ECG numerical data [7, 8]. Heart rate variability (HRV), an analysis of variations between the peaks of ECG signals, has become widely used [9].

In HRV studies, there are various features extracted from HRV data. These features include time-domain statistical calculations, frequency-domain measures from Fourier transform or Lomb–Scargle periodogram, time-frequency-domain features from Wavelet transform, and nonlinear measurements from Poincare plot [10–12]. Although HRV studies use different data lengths (mostly 5 min) in the literature, Seker et al. [13] concluded that nonlinear HRV features provide reproducible results after 10,000 heartbeat samples at least. For this reason, since this study covers nonlinear HRV measurements, 30-min HRV data were used in this study. We extracted all these commonly used features from the AF prediction database (atrial fibrillation prediction database, AFPDB, distributed for ”The Computers in Cardiology Challenge 2001”) in this study. It is freely available and frequently used in PAF studies [14].

In addition, Hallstrom et al. [15] suggested fixing the average heartbeat to 75 bpm (or 800 ms) to minimize individuals’ average heart rates on some HRV features. HRV features, calculated from HRV data after this heart rate normalization, are called normalized HRV (HRN) features. HRN analysis increased classifier performances in CHF diagnosis Isler2010, systolic dysfunction diagnosis [16] and PAF diagnosis [17]. Surucu and colleagues also observed that feature normalization affects the classifier accuracies in PAF diagnosis [17]. Hence, three feature normalization (direct use, min-max normalization, and z-score normalization) methods were tested on both HRV and HRN features, resulting in six different feature combinations addition to HRV and HRN data in this study.

A general pattern recognition study reveals various classification methods after the features were calculated. Many studies, based on disease diagnosis using HRV, have employed different well-known classifier algorithms like k-nearest neighbors (kNN), linear discriminant analysis (LDA), decision tree (DT), fuzzy logic (FL), multilayer perceptron (MLP), support vector machines (SVM), stochastic gradient descent (SGD), and radial basis function (RBF), etc. [3, 5, 16, 18]. Similarly, various studies evaluated different classifiers to discriminate PAF patients from normal subjects. For example, Maier and colleagues [19] achieved 80.00% accuracy using LDA and polynomial LDA. Another study resulted in 82.00% using FL with morphological ECG features [20]. Boon and colleagues evaluated different periods of 5 [21], 10, 15, and 30 [22] min to diagnose PAF patients. They achieved 83.90% accuracy using an SVM classifier with 30-min HRV data. They also reached 87.70% accuracy with 5-min HRV data when they practiced genetic algorithms (GAs) as a feature selection. Surucu and colleagues obtained 81.00% accuracy from min-max normalized HRV features and 86.00% accuracy from z-score normalized HRN features using kNN [17]. Thong et al. [23] achieved 90.00% accuracy using DT on premature atrial contraction timings. Ros and colleagues achieved 92.00% accuracy using the kNN classifier [24]. Using the same classifier, Ozcan and Kuntalp obtained 92.20% accuracy with the help of feature selection based on a genetic algorithm [25]. Chesnokov et al. reached 95.50% accuracy using MLP classifier [26]. Martinez et
al. [27] achieved the highest accuracy of 96.05% using SGD classifier on morphological ECG measurements. Among these studies, Surucu et al. [17] emphasized the importance of both feature normalization and heart rate normalization in the diagnosis of PAF patients.

Conventional machine learning methods require the use of suitable feature extraction techniques before classifying the data. Deep learning (DL) is a popular machine learning method to obtain features and classes together from the raw data [55]. Recently, deep learning algorithms have become very popular, especially in image classification studies [28, 29]. DL classifiers are more complex than conventional artificial neural network structures. They seem to have more layers to extract features themselves and classify the inputs as one of two or more decisions [52–54]. Convolutional neural network (CNN) is the most frequently preferred DL structures in the literature. CNN shows excellent success in classifying two-dimensional (2D) images and three-dimensional (3D) video data; recently, it also becomes popular in one-dimensional (1D) time-series data classification [30, 31].

For example, Pourbabaee and colleagues evaluated a deep learning model with the last stage as the kNN, and they achieved 91.00% accuracy [32] in the diagnosis of PAF patients. Due to its promising superior success in many classification problems, we decided to evaluate the deep learning algorithm with the convolutional neural network (CNN) structure to classify PAF patients. On the other hand, although the feature extraction is not necessary in DL studies, we experimented with both data itself and handy-crafted features for the classification performances to test whether the use of feature extraction gives another approach in classifying data using CNN classifiers in HRV-based studies.

In summary, this study investigates the effects of both heart rate normalization and feature normalization methods in the PAF diagnosis using CNN from 30-min HRV data. For this purpose, eight different input combinations (two data and six feature sets) were given to the inputs of the CNN-structure deep learning classifier. Consequently, these classifiers were trained and tested.

2. Materials and methods

We summarized our study in Figure 1 by determining blocks from the HRV-based literature and our previous studies. The dashed line indicates the alternative heart rate normalization method. The long dashed line shows the path when HRV (or HRN) data is used with skipping the feature extraction processes. Feature extraction methods were time-domain parameters, frequency-domain measures (using FFT, Lomb-Scargle, and Welch), time-frequency-domain measures (using Wavelet transform), and Poincare plot’s measures. Then, one of the feature normalization methods (no normalization, min-max normalization, and Z-score normalization) was applied to the extracted HRV (or HRN) features. Finally, a deep learning classifier decides whether the ECG data belongs to a PAF patient or a normal subject. The following subsections cover brief definitions of these blocks.

2.1. ECG data

The AFPDB database, which is freely available\(^1\), comprises two channels of 30-minute ECG data with a resolution of 16 bits at a sampling frequency of 128 Hz [14]. ECG data are acquired from 50 healthy subjects (named Group A) and 50 patients with PAF. Twenty-five patients (Group C) faced PAF attacks just after their ECG recordings, while other patients (Group B) didn’t have PAF attacks following ECG recordings. Two ECG data, obtained from n27 and p37 subjects, were excluded from the study because these involve many noises. As

---

\(^1\)Pyshio.net website https://www.physionet.org/content/afpdb/1.0.0/
a result, the 49 ECG records from PAF patients (Group B + Group C) and 49 ECG data from healthy subjects (Group A) were objected to the PAF diagnosis in this study.
2.2. Beat detection and generating HRV data

Ventricular depolarization, characterized by QRS peaks in ECG, resulted in blood pumping to the body. It is easy to detect these peaks due to their higher amplitudes. Therefore, the time differences between these peaks constitute the heartbeat periods.

First, the ECG signal is passed through a band-pass filter. Then, the signal is subjected to differentiation, squaring, time averaging, and finally peaks are detected by applying threshold logic \[33\]. HRV data is the time differences between these peaks versus their occurring times points. Since the peak detection time points are not equal, the HRV data becomes irregularly-sampled signal, naturally.

2.3. Removing ectopic beats

Heartbeats that are not originating from the sinoatrial node of the heart are called ectopic beats. The workgroup has offered to remove these irregular beats before the HRV analysis \[34\]. Langley and colleagues \[35\] presented an easy method to detect these premature beats. In this method, a heartbeat and its following heartbeat shape an ectopic beat together if the occurring time of the heartbeat is 20% less than the average occurring time. Simply eliminating these beats from the data is called the ectopic removal process.

2.4. Heart rate normalization

Heart rate normalization, pioneered by Hallstrom and colleagues \[15\], is a process to remove the mean from HRV data by

\[
data_{HRN} = \frac{60 \times \text{data}_{HRV}}{\text{new}_\text{bpm} \times \text{mean(data}_{HRV})}
\]

(1)

\[
data_{HRN} = \frac{1000 \times \text{data}_{HRV}}{\text{new}_\text{mean}_\text{period} \times \text{mean(data}_{HRV})}
\]

(2)

where \(\text{data}_{HRN}\) is HRN data, \(\text{data}_{HRV}\) is HRV data, \(\text{new}_\text{bpm}\) is the new beat per minutes, \(\text{new}_\text{mean}_\text{period}\) is the mean of new heartbeats, \(\text{mean(data}_{HRV})\) is the mean of HRV data given.

We used mean values as 75 beats/min (or 800 ms in period) in this study as offered in the original article \[15\] and some similar research \[16, 18\].

2.5. Resampling and detrending

HRV data are unevenly sampled and contain non-stationary components. Some feature extraction methods (Fast Fourier Transform-FFT, Welch, and Wavelet) require the data sampled at equal time intervals \[36\]. Resampling (or interpolation) is the solution to cope with this issue. Although there are several interpolation methods in the literature, Clifford and Tarasenko \[37\] have offered a robust interpolation method, called Cubic-Spline, to resample HRV data. The various sampling frequencies (number of sample points in a second) of 1 to 10 Hz have been used \[38\]. We preferred 4 Hz, similar to our previous studies, in this study.

All linear feature extraction methods use linear data, naturally. The possible nonlinear components of the data disturb the results. Slowly changing polynomials or sinusoidal trends are common non-stationarity origins \[39\]. In recent years, Tarvainen et al. \[40\] pioneered the Smoothness Priors method to make the data non-stationary by the following equation (3).

\[
R_{\text{detrended}} = (I - (I + \lambda D_2^T D_2)^{-1}) R
\]

(3)
where $R$ is HRV data, $R_{\text{detrended}}$ is the non-stationary HRV data, $I$ is the unity matrix, $\lambda$ is the regularity parameter, $D_2$ is the second-order derivative operator, and $T$ is the transpose of a matrix. We used $\lambda = 1000$ as offered in [40].

2.6. Feature extraction

Feature extraction discovers the data to calculate features [41]. We decided to extract 16 time-domain features, 24 frequency-domain features from 4 different transforms, 4 time-frequency wavelet entropy measures, and 4 non-linear features, which equals to 48 features in total. Although these methods are outlined summarily here, a detailed information can be discovered in [3].

2.6.1. Time-domain features

The time-domain features are calculated from the raw time-series data, in general. These features hold statistical measures like mean, minimum, maximum, standard deviation (SDNN), root means square of successive differences (RMSSD), standard deviation of successive differences (SDSD), NN50 (the number of successive differences greater than 50 ms), NN20 (the number of successive differences greater than 20 ms), PNN50 (the ratio of NN50), PNN20 (the ratio of NN20), etc. as offered in [34].

2.6.2. Frequency-domain features

These features are estimated from the power spectral density (PSD) estimation. There are some methods to estimate PSD that require some preprocessing steps. Among them, the Lomb-Scargle algorithm doesn’t require resampling and detrending sub-steps, but it is a computationally expensive method by comparing to other methods [42, 43]. On the other hand, the Fast Fourier Transform (FFT) and Welch Periodogram algorithms require both resampling and detrending steps since these methods can work on evenly sampled stationary data only [37, 44].

The spectrum of the HRV analysis bases consists of components from four frequency bands: ultra-low-frequency (ULF), very-low-frequency (VLF), low-frequency (LF), and high-frequency (HF) bands [34]. Total power, powers of each frequency band, and the ratio of LF band to HF band powers are frequency-domain features (Table 1) [3, 34]. These features can be calculated from FFT-based periodogram, Welch periodogram, and Lomb–Scargle periodogram like similar studies [3, 5, 18]. There are several excellent signal processing textbooks, which describe these periodogram methods in detail.

<table>
<thead>
<tr>
<th>Features</th>
<th>Description</th>
<th>Frequency Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{ULF}$</td>
<td>Spectral power in the ultra-low frequency (ULF) band</td>
<td>0.000 – 0.003 Hz</td>
</tr>
<tr>
<td>$P_{VLF}$</td>
<td>Spectral power in the very-low frequency (VLF) band</td>
<td>0.003 – 0.040 Hz</td>
</tr>
<tr>
<td>$P_{LF}$</td>
<td>Spectral power in the low frequency (LF) band</td>
<td>0.040 – 0.150 Hz</td>
</tr>
<tr>
<td>$P_{HF}$</td>
<td>Spectral power in the high frequency (HF) band</td>
<td>0.150 – 0.400 Hz</td>
</tr>
<tr>
<td>$P_{Total}$</td>
<td>Total spectral power</td>
<td>0.000 - 0.400 Hz</td>
</tr>
<tr>
<td>$\frac{LF}{HF}$</td>
<td>Spectral power ratio of $P_{LF}$ to $P_{HF}$</td>
<td>0.040 – 0.400 Hz</td>
</tr>
</tbody>
</table>
2.6.3. Nonlinear features

Since heartbeats show nonlinear nature and the Poincare plot (Figure 2) reflects this nature, the Poincare plot becomes popular in HRV studies [45, 46]. There are two commonly used measures [47] derived from the fitted ellipse [48] on this plot:

\[
SD_1 = \sqrt{\frac{1}{2}(SDSD)^2}
\]

\[
SD_2 = \sqrt{2(SDNN)^2 - \frac{1}{2}(SDSD)^2}
\]

where \(SDSD\) and \(SDNN\) are standard time-domain features. In addition, there are two measures [3], derived from \(SD_1\) and \(SD_2\), also calculated as:

\[
SD_1SD_2 = SD_1 \times SD_2
\]

\[
RATIO = \frac{SD_1}{SD_2}
\]

2.6.4. Time-frequency-domain features

The wavelet transform can examine signals in both the time and frequency domains. It also eliminates polynomial non-stationarity [49]. Various features extracted from Wavelet transform have been used in HRV studies [3, 5].

Although choosing an appropriate mother wavelet is an important issue [50, 51], many HRV-related studies preferred Daubechies-4 as mother wavelet [5, 16, 18]. This study applied Daubechies-4 with the level of 7 to the resampled data as reported enough to group wavelet packets into specific HRV frequency bands [3].

The energy of each coefficient was calculated using the following equation:

\[
E_j = C_j^2
\]

where \(C_j\) is the wavelet coefficients. Total energy of an HRV band, \(E_f\), was calculated separately:

\[
E_f = \sum_{j \in f} E_j
\]

where, \(f\) represents the the HRV frequency band [3]. The wavelet entropy features (\(ENT_f\)) was calculated as follows:

\[
ENT_f = -\sum_{j \in f} (p_j \log_2(p_j))
\]

where the probability of energies of all frequency (\(f\)) values in the frequency band of interest is calculated as \(p_j\) [3, 49] and \(p_j\) is the value obtained by dividing the energy of the frequency of interest by the total band energy \(\frac{E_j}{E_f}\). The entropy features were calculated for the standard frequency bands (Table 2).
Figure 2. An example of Poincare plot. The red arrow shows SD1 feature, and the green arrow shows SD2 feature from the fitted ellipse.

<table>
<thead>
<tr>
<th>Features</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ENT_{ULF}$</td>
<td>Wavelet entropy of the ultra-low frequency (ULF) band</td>
</tr>
<tr>
<td>$ENT_{VLF}$</td>
<td>Wavelet entropy of the very-low frequency (VLF) band</td>
</tr>
<tr>
<td>$ENT_{LF}$</td>
<td>Wavelet entropy of the low frequency (LF) band</td>
</tr>
<tr>
<td>$ENT_{HF}$</td>
<td>Wavelet entropy of the high frequency (HF) band</td>
</tr>
</tbody>
</table>

2.7. Feature normalization

Since ranges of features are very different, bigger-value features affect the classifier performances more than small-value features [41]. To eliminate this negative effect becomes very important in many pattern recognition applications [17]. There are two commonly used feature normalization methods: min–max normalization (11) and Z-score normalization (12). The min-max normalized ($f_i^{\text{min-max}}$) and z-score ($f_i^{\text{z-score}}$) normalized
samples can be calculated using

\[ f_i^{\text{min-max}} = \frac{f_i - \min(f)}{\max(f) - \min(f)} \]  

(11)

\[ f_i^{\text{z-score}} = \frac{f_i - \mu_f}{\sigma_f} \]  

(12)

where \( f_i \) is the i-th sample, \( \min(f) \) is the minimum value, \( \max(f) \) is the maximum value, \( \mu_f \) is the average value, and \( \sigma_f \) is the standard deviation of the feature \( f \).

This study iterates HRV and HRN features using no normalization, min-max normalization, and z-score normalization and reports corresponding classifier performances to examine the effect of feature normalization methods.


Among deep learning algorithms, convolutional neural networks (CNNs) include at least one grid-like convolution layer. A general CNN classifier has an input layer, one or more convolution blocks, a dropout block, a pooling block, one (or more) fully connected classifier block(s), a dense layer, and an output layer. The Input layer accepts 1D data from the real world, and the output layer reflects the decision given by the classifier. Other blocks process some mathematical operations. For example, Conv1D layers compute convolutions with different sizes, behaving filters to extract hidden information from the data. The Dropout layer prevents overfitting by randomly disconnecting some connections from input to output, named regularization [56]. The pooling layer reduces the number of parameters by calculating averages of small-size input boxes. The fully connected layer works as a hidden layer of a conventional multi-layer perceptron (MLP), in general. In some applications, other classification algorithms like k-nearest neighbors can be preferred instead of MLP [32]. The dense layer applies activation functions (ReLU, tanh, or sigmoid) to calculate raw output values. Figure 3 visualizes the flowchart of the proposed CNN classifier for this study. Class I is the decision of normal subject (or negative), and Class II is the decision of PAF patient (or positive) in this study.

![Figure 3. Block diagram of the proposed convolutional neural network classifier.](image)

The Adam algorithm is used to optimize parameters in the training of the proposed CNN model. It has
found common use in DL studies since it is computationally efficient and requires a small memory size [57]. We used this algorithm with its default parameters as given in the Keras library of Python. Also, we preferred the ReLU (rectified linear unit) activation function, which has been the most commonly used one in DL models:

$$ReLU(x) = \begin{cases} 
0, & \text{if } x < 0 \\
 x, & \text{if } x \geq 0 
\end{cases} \quad (13)$$

All layers between the input and output layers should be determined by trial-and-error or by experience [58] since the performance of a CNN network is strictly dependent on its hyper-parameters [59]. We proposed this model by tuning hyper-parameters in 1000 iterations.

2.9. Evaluating classifier performances

In pattern recognition applications, the performance of a classifier is calculated by its responses against unseen inputs before [41]. All samples were divided into two parts by 90% for training and 10% for testing the classifier in this study. The performance of the classifier is calculated using test samples, while hyperparameters of the classifier is adjusted using the training data.

The confusion matrix is calculated for evaluating classifier performances. It is generated by comparing the responses of the classification algorithm to the test set with the actual values in the data set. In the case of two-class problems, it is a table consisting of four different situations (Table 3). True positive (TP) is the number of patients classified correctly, and true negative (TN) is the number of healthy subjects classified correctly. On the other hand, false negative (FN) is the number of patients misclassified as healthy ones, and false positive (FP) is the number of healthy subjects misclassified as patients [41]:

<table>
<thead>
<tr>
<th>Actual case</th>
<th>Classifier output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal subject</td>
<td>Normal subject</td>
</tr>
<tr>
<td>Normal subject</td>
<td>PAF patient</td>
</tr>
<tr>
<td>PAF patient</td>
<td>FN</td>
</tr>
</tbody>
</table>

Table 3. 2-by-2 confusion matrix for normal subjects versus PAF patients.

Four commonly used performance measures (Accuracy, Recall, Precision, and F1-Score) were used to evaluate classifiers in this study [41, 60]:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (14)$$

$$Recall = \frac{TP}{TP + FN} \quad (15)$$

$$Precision = \frac{TP}{TP + FP} \quad (16)$$

$$F1 - Score = \frac{Precision \times Recall}{Precision + Recall} \quad (17)$$
3. Results and discussion

This study investigated the effects of both heart rate normalization and feature normalization methods in the PAF diagnosis using CNN from 30-min HRV data. Eight different input combinations were given to the inputs of the CNN-structure deep learning classifier. Consequently, these classifiers were trained and tested using programming language of Python 3.8.

The first part of our study repeated the classifiers using the HRV data and the HRV features for no normalization, min-max normalization, and z-score normalization cases. The achieved classifier accuracies are 89.38%, 82.65%, 95.92%, and 95.92% (Table 4), respectively. Since both the min-max and z-score normalizations give higher accuracies, the CNN classifier using one of these two normalization methods seems enough in the PAF diagnosis based on HRV data.

Table 4. Classifier performances achieved in this study. The bold-indicated row emphasizes the highest classifier accuracy among them. HRV is heart rate variability and HRN is the heart rate normalized HRV where the 'data' stands the raw data and the 'features' shows the extracted features from the raw data.

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Feature Normalization</th>
<th>Accuracy (%)</th>
<th>Precision (%)</th>
<th>Recall (%)</th>
<th>F1-Score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRV data</td>
<td>-</td>
<td>89.38</td>
<td>89.70</td>
<td>89.38</td>
<td>89.54</td>
</tr>
<tr>
<td>HRV features</td>
<td>-</td>
<td>82.65</td>
<td>86.36</td>
<td>77.55</td>
<td>81.72</td>
</tr>
<tr>
<td>HRV features</td>
<td>Min-max</td>
<td>95.92</td>
<td>97.87</td>
<td>93.88</td>
<td>95.83</td>
</tr>
<tr>
<td>HRV features</td>
<td>Z-Score</td>
<td>95.92</td>
<td>100.0</td>
<td>91.84</td>
<td>95.74</td>
</tr>
<tr>
<td>HRN data</td>
<td>-</td>
<td>55.44</td>
<td>53.50</td>
<td>94.76</td>
<td>68.39</td>
</tr>
<tr>
<td>HRN features</td>
<td>-</td>
<td>78.57</td>
<td>75.00</td>
<td>85.71</td>
<td>80.00</td>
</tr>
<tr>
<td>HRN features</td>
<td>Min-Max</td>
<td>98.98</td>
<td>100.0</td>
<td>97.96</td>
<td>98.97</td>
</tr>
<tr>
<td>HRN features</td>
<td>Z-Score</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

The second part of our study repeated the classifiers using the HRN data and the HRN features for the same normalization cases. The achieved classifier accuracies become 55.44%, 78.57%, 98.98%, and 100.0% (Table 4), respectively. Although the min–max normalization gives satisfactorily noticeable accuracy again, the z-score normalization method results in excellent accuracy. As a result, the CNN classifier using both the heart rate normalization and z-score feature normalization methods together is enough to diagnose PAF patients from 30-min HRN data.

Table 5 summarizes PAF diagnosis studies using 30-min HRV data given in the literature. The achieved classifier accuracies varied from 80.00% to 95.96% in the table. Our proposed method, using CNN classifier with heart rate normalized and z-score normalized heart rate variability features, gives the highest classifier accuracy of 100%. By comparing the accuracy of the proposed classifier to other studies, our method has superior accuracy among them, to the best of our knowledge.

Consequently, our proposed method to diagnose PAF patients is to use CNN classifier with conventional HRV features derived from heart rate normalized HRV data. To our knowledge, this method resulted in better classification performance than similar studies in the literature. We may pretend that our approach provides a better tool to discriminate PAF patients from normal individuals.

In addition to the classifier performance, the other important outcome of this study is to use hand-crafted features in a CNN study. Since feature extraction is an embedded part of deep learning models, the researchers used raw data instead of extracted features as usual. In this study, we compared the CNN performances using raw HRV data with those using HRV features. In our opinion, it may indicate that the feature extraction part of deep learning models does not fit the HRV data well. 2896
Table 5. PAF diagnosis studies using 30-min heart rate variability measures in the literature. The bold-indicated row emphasizes the highest classifier accuracy.

<table>
<thead>
<tr>
<th>Study</th>
<th>Features</th>
<th>Classifier</th>
<th>Description</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>HRV</td>
<td>LDA, polynomial LDA</td>
<td>-</td>
<td>80.00</td>
</tr>
<tr>
<td>17</td>
<td>HRV</td>
<td>kNN</td>
<td>Min–max</td>
<td>81.00</td>
</tr>
<tr>
<td>20</td>
<td>Morphological ECG</td>
<td>FL</td>
<td>-</td>
<td>82.00</td>
</tr>
<tr>
<td>22</td>
<td>HRV</td>
<td>SVM</td>
<td>Genetic algorithms</td>
<td>83.90</td>
</tr>
<tr>
<td>17</td>
<td>HRN</td>
<td>kNN</td>
<td>Z-Score</td>
<td>86.00</td>
</tr>
<tr>
<td>23</td>
<td>Premature atrial contractions</td>
<td>DT</td>
<td>-</td>
<td>90.00</td>
</tr>
<tr>
<td>32</td>
<td>ECG</td>
<td>CNN with kNN</td>
<td>-</td>
<td>91.00</td>
</tr>
<tr>
<td>24</td>
<td>Morphological ECG</td>
<td>kNN</td>
<td>-</td>
<td>92.00</td>
</tr>
<tr>
<td>25</td>
<td>HRV</td>
<td>kNN</td>
<td>Genetic algorithms</td>
<td>92.20</td>
</tr>
<tr>
<td>26</td>
<td>HRV</td>
<td>MLP</td>
<td>-</td>
<td>95.50</td>
</tr>
<tr>
<td>This study</td>
<td>HRV</td>
<td>CNN</td>
<td>Z-Score</td>
<td>95.92</td>
</tr>
<tr>
<td>27</td>
<td>Morphological ECG</td>
<td>SGD</td>
<td>-</td>
<td>96.05</td>
</tr>
<tr>
<td>This study</td>
<td>HRN</td>
<td>CNN</td>
<td>Z-Score</td>
<td>100.0</td>
</tr>
</tbody>
</table>

On the other hand, CNN works well with big data. The database used in this study consists of ECG data obtained from only 49 patients with PAF and 49 healthy subjects, which is a weakness of this study. The findings obtained in our study, which is preliminary research in character, need to be supported by another study that will use a much larger data set.

Acknowledgment

There is no financial support for this study. The author contributions are as follows: The subject of this paper is the part of M. Surucu’s Ph.D. thesis. R. Kara is the supervisor and Y. Isler is the co-supervisor of the thesis. All authors equally contribute on writing this article.

Conflict of interest

The authors declare that there is no conflict of interest.

References


[40] Tarvainen MP, Ranta-Aho PO, Karjalainen PA. An advanced detrending method with application to HRV analysis. IEEE Transactions on Biomedical Engineering 2002; 49 (2): 172-175. doi: 10.1109/10.979357


