Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2021) 29: 2886 – 2900

doi:10.3906/elk-2105-92

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

Murat SURUCU[1], **Yalcin ISLER**[2,3,*], **Resul KARA**[1]
[1]Department of Computer Engineering, Faculty of Engineering, Düzce University, Düzce, Turkey
[2]Department of Biomedical Engineering, Faculty of Engineering, İzmir Katip Çelebi University, İzmir, Turkey
[3]İslerya Medical and Information Technologies Company, Izmir, Turkey

Received: 13.05.2021 • **Accepted/Published Online:** 12.07.2021 • **Final Version:** 04.10.2021

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, paraoxysmal 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

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/

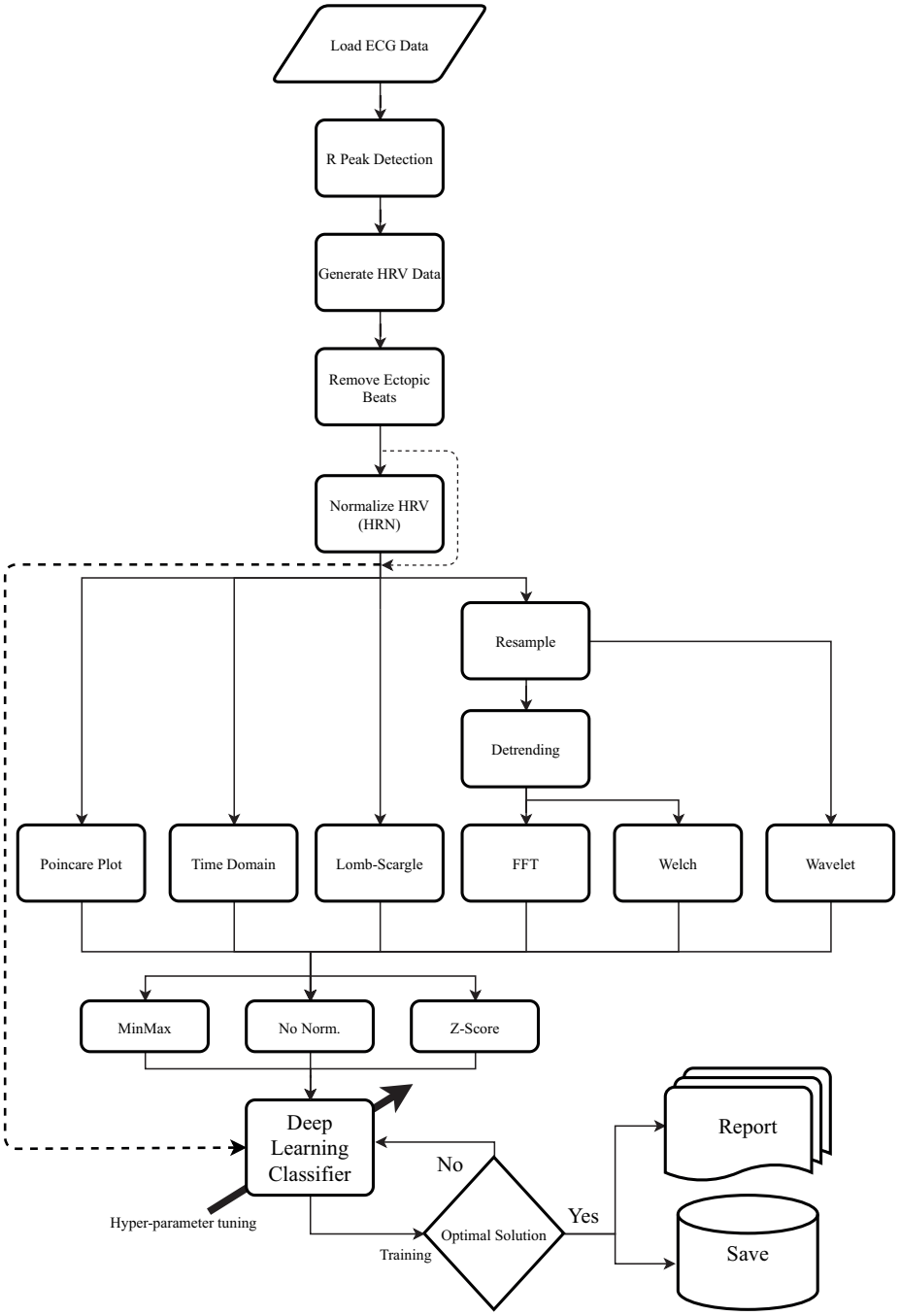

Figure 1. Block diagram of the proposed study.

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

$$data_{HRN} = \frac{60}{new\_bpm} \times \frac{data_{HRV}}{mean(data_{HRV})} \tag{1}$$

$$data_{HRN} = \frac{1000}{new\_mean\_period} \times \frac{data_{HRV}}{mean(data_{HRV})} \tag{2}$$

where $data_{HRN}$ is HRN data, $data_{HRV}$ is HRV data, new_bpm is the new beat per minutes, new_mean_period is the mean of new heartbeats, $mean(data\_HRV$ 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).

$$R_{detrended} = (I - (I + \lambda D_2^T D_2)^{-1})R \tag{3}$$

where R is HRV data, $R_{detrended}$ is the non-stationary HRV data, I is the unity matrix, λ 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 1. Commonly used frequency-domain heart rate variability measures.

Features	Description	Frequency Range
P_{ULF}	Spectral power in the ultra-low frequency (ULF) band	0.000 – 0.003 Hz
P_{VLF}	Spectral power in the very-low frequency (VLF) band	0.003 – 0.040 Hz
P_{LF}	Spectral power in the low frequency (LF) band	0.040 – 0.150 Hz
P_{HF}	Spectral power in the high frequency (HF) band	0.150 – 0.400 Hz
P_{Total}	Total spectral power	0.000 - 0.400 Hz
$\frac{LF}{HF}$	Spectral power ratio of P_{LF} to P_{HF}	0.040 – 0.400 Hz

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:

$$SD_1 = \sqrt{\frac{1}{2}(SDSD)^2} \tag{4}$$

$$SD_2 = \sqrt{2(SDNN)^2 - \frac{1}{2}(SDSD)^2} \tag{5}$$

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

$$SD_1SD_2 = SD_1 \times SD_2 \tag{6}$$

$$RATIO = \frac{SD_1}{SD_2} \tag{7}$$

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 coeeficinet was calculated using the following equation:

$$E_j = C_j^2 \tag{8}$$

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

$$E_f = \sum_{j \in f} E_j \tag{9}$$

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

$$ENT_f = -\sum_{j \in f}(p_j log_2(p_j)) \tag{10}$$

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).

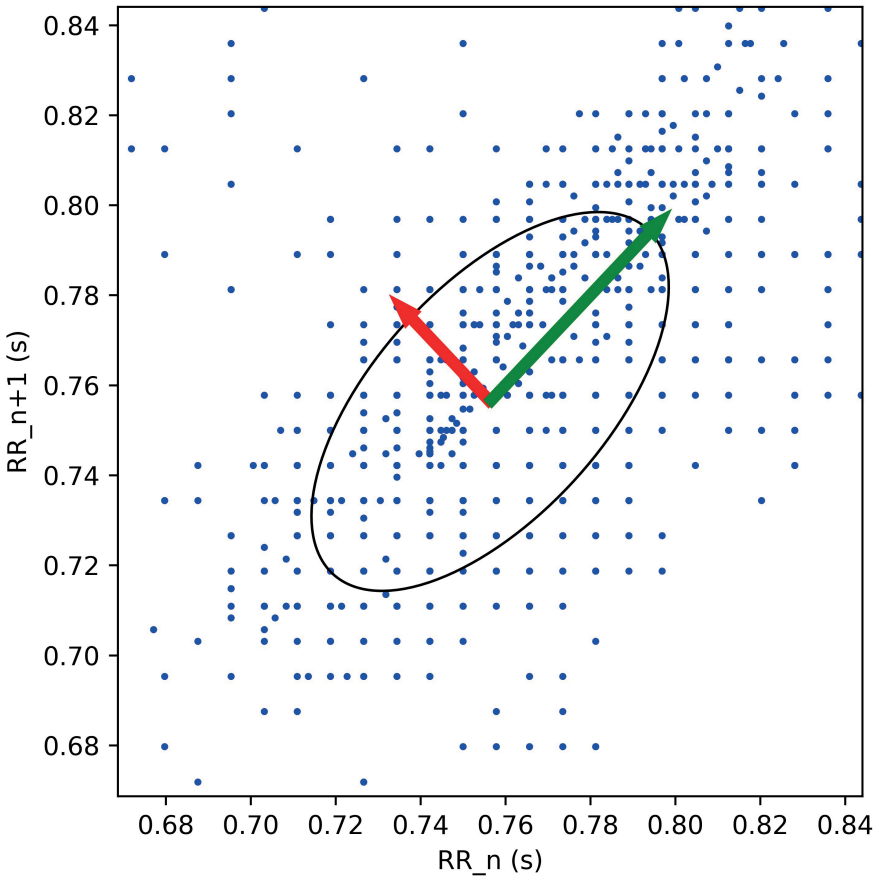

Figure 2. An example of Poincare plot. The red arrow shows SD1 feature, and the green arrow shows SD2 feature from the fitted ellipse.

Table 2. Time-frequency-domain heart rate variability measures.

Features	Description
ENT_{ULF}	Wavelet entropy of the ultra-low frequency (ULF) band
ENT_{VLF}	Wavelet entropy of the very-low frequency (VLF) band
ENT_{LF}	Wavelet entropy of the low frequency (LF) band
ENT_{HF}	Wavelet entropy of the high frequency (HF) band

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 ($f_i^{min-max}$) and z-score ($f_i^{z-score}$) normalized

samples can be calculated using

$$f_i^{min-max} = \frac{f_i - \min(f)}{\max(f) - \min(f)} \tag{11}$$

$$f_i^{z-score} = \frac{f_i - \mu_f}{\sigma_f} \tag{12}$$

where f_i is the i-th sample, $\min(f)$ is the minimum value, $\max(f)$ is the maximum value, μ_f is the average value, and σ_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.

2.8. Classifier: convolutional neural networks

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.

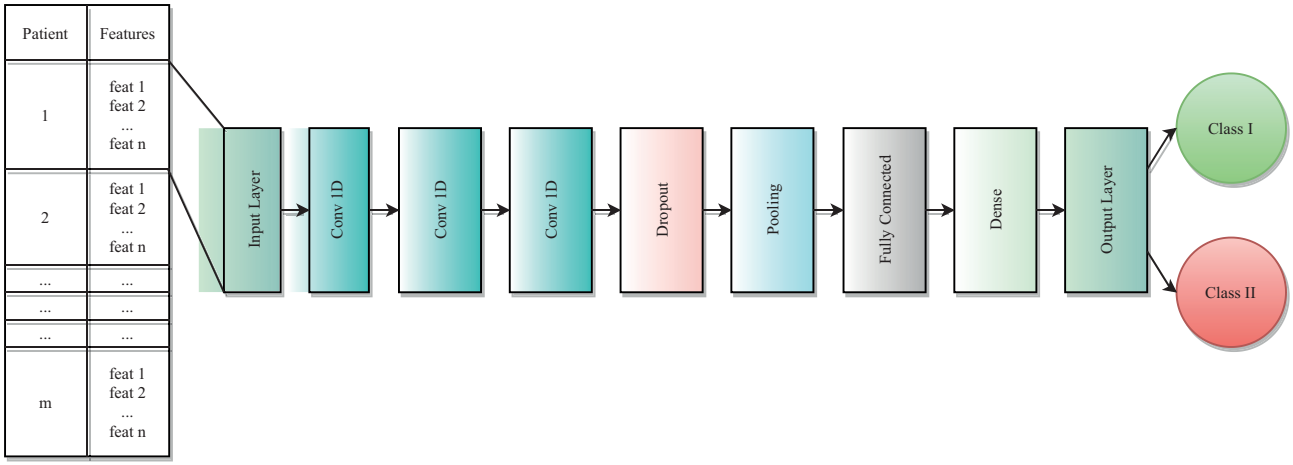

Figure 3. Block diagram of the proposed convolutional neaural network classifier.

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:

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

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

		Classifier output	
		Normal subject	PAF patient
Actual case	Normal subject	TN	FP
	PAF patient	FN	TP

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

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{14}$$

$$Recall = \frac{TP}{TP + FN} \tag{15}$$

$$Precision = \frac{TP}{TP + FP} \tag{16}$$

$$F1 - Score = \frac{Precision \times Recall}{Precision + Recall} \tag{17}$$

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.

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

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.

Table 5. PAF diagnosis studies using 30-min heart rate variability measures in the literature. The bold-indicated row emphasizes the highest classifier accuracy.

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

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

[1] Ryder KM, Benjamin EJ. Epidemiology and significance of atrial fibrillation. The American Journal of Cardiology 1999; 84 (9A): 131R-138R. doi: 10.1016/s0002-9149(99)00713-4

[2] Mohebbi M, Ghassemian H. Prediction of paroxysmal atrial fibrillation based on non-linear analysis and spectrum and bispectrum features of the heart rate variability signal. Computer Methods and Programs in Biomedicine 2012; 105 (1): 40-49. doi: 10.1016/j.cmpb.2010.07.011

[3] Isler Y, Kuntalp M. Combining classical hrv indices with wavelet entropy measures improves to performance in diagnosing congestive heart failure, Computers in Biology and Medicine 2007; 37 (10): 1502-1510. doi: 10.1016/j.compbiomed.2007.01.012

[4] Gladstone DJ, Spring M, Dorian P, Panzov V, Thorpe KE et al. Atrial fibrillation in patients with cryptogenic stroke. New England Journal of Medicine 2014; 370 (26): 2467-2477. doi: 10.1056/nejmoa1311376

[5] Narin A, Isler Y, Ozer M, Perc M. Early prediction of paroxysmal atrial fibrillation based on short-term heart rate variability. Physica A: Statistical Mechanics and its Applications 2018; 509: 56-65. doi: 10.1016/j.physa.2018.06.022

[6] Soudani A, Almusallam M. Atrial fibrillation detection based on ECG-Features extraction in WBSN. Procedia Computer Science 2018; 130: 472-479. doi: 10.1016/j.procs.2018.04.052

[7] Acharya UR, Fujita H, Lih OS, Adam M, Tan JH et al. Automated detection of coronary artery disease using different durations of ECG segments with convolutional neural network. Knowledge-Based Systems 2017; 132: 62-71. doi: 10.1016/j.knosys.2017.06.003

[8] Guruler H, Sahin M, Ferikoglu A. Feature selection on single-lead ECG for obstructive sleep apnea diagnosis. Turkish Journal of Electrical Engineering and Computer Science 2014; 22 (2): 465-478. doi: 10.3906/elk-1207-132

[9] Camm A, Malik JM, Bigger J, Brethardt TG, Cerutti S et al. Heart rate variability. Circulation 1996; 93 (5): 1043-1065. doi: 10.1161/01.CIR.93.5.1043

[10] Tsipouras MG, Fotiadis DI. Automatic arrhythmia detection based on time and time-frequency analysis of heart rate variability. Computer Methods and Programs in Biomedicine 2004; 74 (2): 95-108. doi: 10.1016/S0169-2607(03)00079-8

[11] Ming-Yuan L, Sung-Nien Y. Multiscale sample entropy based on discrete wavelet transform for clinical heart rate variability recognition. Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2012; 4299-4302. doi: 10.1109/EMBC.2012.6346917

[12] Park J, Lee S, Jeon M. Atrial fibrillation detection by heart rate variability in Poincare plot. BioMedical Engineering OnLine 2009; 8 (1): 38. doi: 10.1186/1475-925X-8-38

[13] Seker R, Saliu S, Birand A, Kudaiberdieva G. Validity test for a set of nonlinear measures for short data length with reference to short-term heart rate variability signal. Journal of Systems Integration 2000; 10: 41-53. doi: 10.1023/A:1026507317626

[14] Moody GB, Goldberger AL, McClennen S, Swiryn SP. Predicting the onset of paroxysmal atrial fibrillation: The computers in cardiology challenge. Computers in Cardiology 2001; 28: 113-116. doi: 10.1109/CIC.2001.977604

[15] Hallstrom AP, Stein PK, Schneider R, Hodges M, Schmidt G et al. Structural relationships between measures based on heart beat intervals: potential for improved risk assessment. IEEE Transactions on Biomedical Engineering 2004; 51 (8): 1414-1420. doi: 10.1109/TBME.2004.828049

[16] Isler Y. Discrimination of Systolic and Diastolic Dysfunctions using Multi-Layer Perceptron in Heart Rate Variability Analysis. Computers in Biology and Medicine 2016; 76: 113-119. doi: 10.1016/j.compbiomed.2016.06.029

[17] Surucu M, Isler Y, Kara R. Investigation of the effect of normalization techniques on discriminating patients with paroxysmal atrial fibrillation. In: 2nd International Conference of Applied Sciences, Engineering and Mathematics (IBU-ICASEM 2020), June 4-6, Skopje/North Macedonia, Book of Abstracts, 2020, ISBN: 978-608-4868-02-6, 12.

[18] Isler Y, Kuntalp M. Heart rate normalization in the analysis of heart rate variability in congestive heart failure. Proceedings of the IMechE Part H: Journal of Engineering in Medicine 2010; 224 (3): 453–463. doi: 10.1243/09544119JEIM642

[19] Maier C, Bauch M, Dickhaus H. Screening and prediction of paroxysmal atrial fibrillation by analysis of heart rate variability parameters. Computers in Cardiology 2001; 28: 129-132. doi: 10.1109/cic.2001.977608

[20] Schreier G, Kastner P, Marko W. An automatic ECG processing algorithm to identify patients prone to paroxysmal atrial fibrillation. Computers in Cardiology 2001; 28: 133-135. doi: 10.1109/cic.2001.977609

[21] Boon KH, Khalil-Hani M, Malarvili MB. Paroxysmal atrial fibrillation prediction based on HRV analysis and non-dominated sorting genetic algorithm. Computer Methods and Programs Biomedicine 2018; 153: 171-184. doi: 10.1016/j.cmpb.2017.10.012

[22] Boon KH, Khalil-Hani M, Malarvili MB, Sia CW. Paroxysmal atrial fibrillation prediction method with shorter HRV sequences. Computer Methods and Programs Biomedicine 2016; 134: 187-196. doi: 10.1016/j.cmpb.2016.07.016

[23] Thong T, McNames J, Aboy M, Goldstein B. Prediction of paroxysmal atrial fibrillation by analysis of atrial premature complexes. IEEE Transactions on Biomedical Engineering 2004; 51: 561-569. doi: 10.1109/TBME.2003.821030

[24] Ros E, Mota S, Fernandez FJ, Toro FJ, Bernier JL. ECG characterization of paroxysmal atrial fibrillation: parameter extraction and automatic diagnosis algorithm. Computers in Biology and Medicine 2004; 34 (8): 679-696. doi: 10.1016/j.compbiomed.2003.10.002

[25] Ozcan N, Kuntalp M. Determining best HRV indices for PAF screening using genetic algorithm. 10th International Conference on Electrical and Electronics Engineering (ELECO), 2017; 1385-1388.

[26] Chesnokov YV, Holden AV, Zhang H. Screening patients with paroxysmal atrial fibrillation (PAF) from non-PAF heart rhythm using HRV data analysis. Computers in Cardiology 2007; 2007: 459–462. doi: 10.1109/cic.2007.4745521

[27] Martinez A, Alcaraz R, Rieta JJ. Study on the P-wave feature time course as early predictors of paroxysmal atrial fibrillation. Physiological Measurement 2012; 33 (12): 1959-1974. doi: 10.1088/0967-3334/33/12/1959

[28] Balli O, Kutlu Y. Effect of deep learning feature inference techniques on respiratory sounds. Journal of Intelligent Systems with Applications 2020; 3 (2): 134-137 (in Turkish with an abstract in English). http://asd.islerya.com/indir.php?id=135

[29] Camgozlu Y, Kutlu Y. Examining the difference between image size, background color, gray picture and color picture in leave classification with deep learning. Journal of Intelligent Systems with Applications 2020; 3 (2): 130-133 (in Turkish with an abstract in English). http://asd.islerya.com/indir.php?id=133

[30] Altan G, Kutlu, Y, Pekmezci AO, Nural S. Deep learning with 3D-second order difference plot on respiratory sound. Biomedical Signal Processing and Control 2018; 45: 58-69. doi: 10.1016/j.bspc.2018.05.014

[31] Altan G, Kutlu Y, Allahverdi N. Deep learning on computerized analysis of chronic obstructive pulmonary disease. IEEE Journal of Biomedical and Health Informatics 2020; 24 (5): 1344-1350. doi: 10.1109/JBHI.2019.2931395

[32] Pourbabaee B, Roshtkhari MJ, Khorasani K. Deep convolutional neural networks and learning ecg features for screening paroxysmal atrial fibrillation patients. IEEE Transactions on Systems, Man, and Cybernetics: Systems 2018; 48 (12): 2095-2104. doi: 10.1109/tsmc.2017.2705582

[33] Pan J, Tompkins WJ. A real–time QRS detection algorithm. IEEE Transactions on Biomedical Engineering 1985; 32 (3): 230-236. doi: 10.1109/TBME.1985.325532

[34] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation and clinical use. Circulation 1996; 93: 1043-1065. doi: 10.1161/01.CIR.93.5.1043

[35] Langley P, Di Bernardo D, Allen J, Bowers E, Smith F et al. Can paroxysmal atrial fibrillation be predicted? Computers in Cardiology 2001; 28: 121-124. doi: 10.1109/CIC.2001.977606

[36] Berntson G, Bigger J, Eckberg D, Grossman P, Kaufmann P et al. Heart rate variability: origins, methods, and interpretive caveats. Psychophysiology 1997; 34: 623-648. doi: 10.1111/j.1469-8986.1997.tb02140.x

[37] Clifford GD, Tarassenko L. Quantifying errors in spectral estimates of HRV due to beat replacement and resampling. IEEE Transactions on Biomedical Engineering 2005; 52 (4): 630-638. 10.1109/TBME.2005.844028

[38] Clifford GD, Azuaje F, McSharry PE. Advanced Methods and Tools for ECG Data Analysis. Norwood, MA, USA: Artech House, 2006.

[39] Weber EJM, Molenaar P, Molen MW. A nonstationarity test for the spectral analysis of physiological time series with an application to respiratory sinus arrhythmia. Psychophysiology 1992; 29 (1): 55-65. doi: 10.1111/j.1469-8986.1992.tb02011.x

[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

[41] Duda RO, Hart PE, Stork DG. Pattern Classification. 2nd edition, New York, NY, USA: Wiley-Interscience, 2000.

[42] Lomb NR. Least-squares frequency analysis of unequally space data. Astrophysics and Space Science 1976; 39 (2): 447–462. doi: 10.1007/BF00648343

[43] Scargle JD. Studies in astronomical time series analysis (II): statistical aspects of spectral analysis of unevenly spaced data. Astrophysical Journal 1982; 263: 835-853. doi: 10.1086/160554

[44] Laguna P, Moody GB, Mark RG. Power spectral density of unevenly sampled data by least square analysis: performance and application to heart rate signals. IEEE Transactions on Biomedical Engineering 1998; 45 (6): 698-715. doi: 10.1109/10.678605

[45] Kamen PW, Tonkin AM. Application of the Poincare plot to heart rate variability: a new measure of functional status in heart failure. Australian and New Zealand Journal of Medicine 1995; 25 (1): 18–26. doi: 10.1111/j.1445-5994.1995.tb00573.x

[46] Kamen PW, Krum H, Tonkin AM. Poincare plot of heart rate variability allows quantitative display of parasympathetic nervous activity. Clinical Science (London, England) 1996; 92: 201–208. doi: 10.1042/cs0910201

[47] Brennan M, Palaniswami M, Kamen P. Do existing measures of Poincare plot geometry reflect nonlinear features of heart rate variability? IEEE Transactions on Biomedical Engineering 2001; 48 (11): 1342-1347. doi: 10.1109/10.959330

[48] Marciano ML, Migaux F, Acanfora D, Furgi G, Rengo F. Quantification of Poincaree maps for the evaluation of heart rate variability. Computers in Cardiology 1994; 1994: 557–580. doi: 10.1109/CIC.1994.470126

[49] Quiraga RQ, Rosso OA, Basar E, Schurmann M. Wavelet entropy in event-related potentials: a new method shows ordering of EEG oscillations. Biological Cybernetics 2001; 84 (4): 291-299. doi: 10.1007/s004220000212

[50] Sayilgan E, Yuce YK, Isler Y. Evaluation of wavelet features selected via statistical evidence from steady-state visually-evoked potentials to predict the stimulating frequency. Journal of the Faculty of Engineering and Architecture of Gazi University 2021; 36 (2): 593-605. doi: 10.17341/gazimmfd.664583

[51] Sayilgan E, Yuce YK, Isler Y. Evaluation of mother wavelets on steady-state visually-evoked potentials for triple-command brain-computer interfaces. Turkish Journal of Electrical Engineering & Computer Sciences 2021; in press. doi: 10.3906/elk-2010-26

[52] Schmidhuber J. Deep learning in neural networks: An overview. Neural Networks 2015; 61: 85-117. doi: 10.1016/j.neunet.2014.09.003

[53] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015; 521 (7553): 436-444. doi: 10.1038/nature14539

[54] Bengio Y, Courville A, Vincent P. Representation learning: A review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence 2013; 35 (8): 1798-1828. doi: 10.1109/tpami.2013.50

[55] Chollet F. Deep Learning with Python. 2nd edition, Shelter Island, NY, USA: Manning Publications, 2020.

[56] Srivastava N, Hinton G, Krizhevsky A, Sutskever A, Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research 2014; 15 (56): 1929-1958.https://jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf

[57] Kingma DP, Ba JL. Adam: a method for stochastic optimization. 3rd International Conference for Learning Representations, San Diego, 2015. https://arxiv.org/abs/1412.6980v9

[58] Bergstra J, Bengio Y. Random search for hyper-parameter optimization. Journal of Machine Learning Research 2012; 13: 281-305. https://www.jmlr.org/papers/volume13/bergstra12a/bergstra12a.pdf

[59] Abdi AH, Luong C, Tsang T, Allan G, Nouranian S et al. Automatic quality assessment of echocardiograms using convolutional neural networks: feasibility on the apical four-chamber view. IEEE Transactions on Medical Imaging 2017; 36 (6): 1221-1230. doi: doi: 10.1109/TMI.2017.2690836

[60] Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics 2020; 21: 1-6. doi: 10.1186/s12864-019-6413-7