

Detecting Atrial Fibrillation and Atrial Flutter in Daily Life Using Photoplethysmography Data

Linda M. Eerikäinen , Graduate Student Member, IEEE, Alberto G. Bonomi, Fons Schipper, Lukas R. C. Dekker, Helma M. de Morree , Rik Vullings , and Ronald M. Aarts , Fellow, IEEE

I. INTRODUCTION

Abstract—Objective: Photoplethysmography (PPG) enables unobtrusive heart rate monitoring, which can be used in wrist-worn applications. Its potential for detecting atrial fibrillation (AF) has been recently presented. Besides AF, another cardiac arrhythmia increasing stroke risk and requiring treatment is atrial flutter (AFL). Currently, the knowledge about AFL detection with PPG is limited. The objective of our study was to develop a model that classifies AF, AFL, and sinus rhythm with or without premature beats from PPG and acceleration data measured at the wrist in daily life. **Methods:** A dataset of 40 patients was collected by measuring PPG and accelerometer data, as well as electrocardiogram as a reference, during 24-hour monitoring. The dataset was split into 75%–25% for training and testing a Random Forest (RF) model, which combines features from PPG, inter-pulse intervals (IPI), and accelerometer data, to classify AF, AFL, and other rhythms. The performance was compared to an AF detection algorithm combining traditional IPI features for determining the robustness of the accuracy in presence of AFL. **Results:** The RF model classified AF/AFL/other with sensitivity and specificity of 97.6/84.5/98.1% and 98.2/99.7/92.8%, respectively. The results with the IPI-based AF classifier showed that the majority of false detections were caused by AFL. **Conclusion:** The PPG signal contains information to classify AFL in the presence of AF, sinus rhythm, or sinus rhythm with premature contractions. **Significance:** PPG could indicate presence of AFL, not only AF.

Index Terms—Photoplethysmography, Accelerometer, Atrial Fibrillation (AF), Atrial Flutter, Random Forest.

Manuscript received June 26, 2019; revised September 2, 2019 and September 27, 2019; accepted October 24, 2019. Date of publication November 4, 2019; date of current version June 5, 2020. This work was performed within the framework of the Eindhoven MedTech Innovation Center (e/MTIC) as a collaboration between the Eindhoven University of Technology, Catharina Hospital Eindhoven, and Philips Research, and supported in part by ITEA3 project, 15032, eWatch. (Corresponding author: Linda M. Eerikäinen.)

L. M. Eerikäinen and R. M. Aarts are with the Department of Electrical Engineering, Eindhoven University of Technology, 5612 AE Eindhoven, The Netherlands, and also with the Philips Research, 5656 AE Eindhoven, The Netherlands (e-mail: L.M.Eerikainen@tue.nl; r.m.aarts@tue.nl).

A. G. Bonomi, F. Schipper, and H. M. de Morree are with Philips Research, 5656 AE Eindhoven, The Netherlands (e-mail: alberto.bonomi@philips.com; fons.schipper@philips.com; helma.de.morree@philips.com).

L. R. C. Dekker is with Catharina Hospital, 5623 EJ Eindhoven, The Netherlands, and also with the Department of Electrical Engineering, Eindhoven University of Technology, 5612 AE Eindhoven, The Netherlands (e-mail: lukas.dekker@catharinaziekenhuis.nl).

R. Vullings is with the Department of Electrical Engineering, Eindhoven University of Technology, 5612 AE Eindhoven, The Netherlands (e-mail: r.vullings@tue.nl).

Digital Object Identifier 10.1109/JBHI.2019.2950574

ATRIAL fibrillation (AF) is a cardiac arrhythmia that has been estimated to affect approximately 3% of the adult population, the prevalence increasing at older age [1], [2]. The arrhythmia is associated with increased morbidity, such as stroke and heart failure [3], [4]. Therefore, a timely diagnosis and start of the treatment of AF is essential, and new solutions for unobtrusive, low-cost, and possibly prolonged monitoring are increasingly studied. Especially solutions for long-term monitoring that work in daily life are needed for detecting intermittent episodes of AF that may be missed if monitoring period is short.

Another cardiac arrhythmia causing a similar stroke risk as AF, but is less common, is atrial flutter (AFL) [5], [6]. In AFL the atrial rhythm is regular and the ventricular rate is dependent on atrioventricular conduction and on whether the flutter is typical or atypical. The guidelines for anticoagulation and aims for AFL management are similar as for AF [4]. In addition, many patients with AFL develop later AF [7] or both arrhythmias may coexist [6]. Although the aims in management are similar, the treatment strategies for the two arrhythmias differ. AF is more often treated with medication whereas cardiac ablation is more common in treating AFL [4], [5], the success rate of ablations for a specific type of AFL being 90–95% [4]. Rate control in AFL is often more difficult to achieve than in AF [4] and antiarrhythmic therapy of AF may also cause AFL [5]. Knowing the type of arrhythmia causing the stroke risk is therefore important as antiarrhythmic therapies differ.

Photoplethysmography (PPG) is an optical measurement modality that can be used in physiological measurements, such as heart rate monitoring [8], [9]. Reflective PPG is often used for wearable solutions, e.g. wristband devices. The potential of using PPG measured at the wrist to detect AF has been investigated in several studies with promising results [10]–[23]. Most of the approaches have focused on discriminating AF from normal sinus rhythm (NSR) [10]–[12] or, more in general, non-AF rhythms without further dividing the rhythms into different classes [13]–[21]. For classification of multiple rhythms, Corino *et al.* [22] proposed a method to classify the rhythms into AF, NSR, and other arrhythmias, whereas the approach of Fallet *et al.* [23] focused on classifying AF, NSR, and ventricular arrhythmias. Furthermore, the potential of using PPG signals to classify multiple cardiac rhythms has been studied with PPG measured with smartphones [24]–[26].

While AF detection with PPG has been widely studied, the literature about using PPG signals to classify AFL is limited. In

the study of Corino *et al.* [22], the class for other arrhythmias included 9 subjects having either AFL, ventricular premature beats (VPB), atrial tachycardia, and variable conduction. The sensitivity and specificity for this class were 75.8% and 76.8%, respectively. In other studies in which AFL is mentioned to appear in the dataset, the data has been often either excluded [16], [17], [21], [27] or it has been included in the group of non-AF rhythms [20]. Kashiwa *et al.* [21] excluded continuous AFL, but did not distinguish short AFL episodes among AF episodes. The approaches that included AFL in a different class than AF, were based on the information obtained from the inter-pulse intervals (IPIs) [20], [22]. From electrocardiography (ECG) studies we know that rhythm irregularity is not specific for AF, but occurs also during AFL [28]. Therefore, analysis on IPI irregularity patterns may not be sufficient for an accurate classification of these two arrhythmias. AFL can also manifest as a very regular rhythm, which can be a challenge when distinguishing from sinus rhythm. Adding additional information, e.g. from the PPG waveform, may be helpful in detecting different rhythms.

The objective of our study was to develop a classifier for classification of AF, AFL, and other rhythms, using PPG and acceleration data measured at the wrist in daily life. The category other rhythms included NSR and sinus rhythm accompanied with premature beats originating either from the atria or the ventricles. First, we developed an AF classifier based on rhythm irregularity, derived from the IPIs with commonly used features for AF detection, to benchmark the classification performance. This was done in three different ways: by considering AFL as a non-AF rhythm together with sinus rhythm and premature beats, considering AFL with AF because of the similar stroke risk, and by excluding AFL completely. Second, we added new features, such as features from the PPG waveform, to improve the classification performance and provide sufficient information for classifying multiple rhythms, i.e. AF, AFL, and other.

II. METHODS

A. Data

The dataset for this study consisted of simultaneous ECG, PPG, and accelerometry measurements in 40 patients undergoing a 24-hour Holter measurement as part of routine clinical care. The patients were contacted by a cardiologist and given at least one week to consider their participation in the study. Before the start of the measurements, the participants gave written informed consent. The study (NL53827.100.15) was approved by the medical ethical committee MEC-U (Medical Research Ethics Committees United) in the Netherlands, and the data was collected in the Catharina Hospital, Eindhoven, the Netherlands.

The ECG was measured with a 12-lead Holter monitor (H12+, Mortara, Milwaukee, WI, USA). The PPG and 3-axis accelerometer measurements at the non-dominant wrist were made with a data logging device equipped with the Philips Cardio and Motion Monitoring Module (CM3 Generation-3, Wearable Sensing Technologies, Philips, Eindhoven, the Netherlands). The PPG sensor was based on reflective mode using two green LEDs. The sampling frequency of both PPG and accelerometry was 128 Hz and the dynamic range of the accelerometer was ± 8 g.

TABLE I

RHYTHM TYPE DISTRIBUTIONS IN THE TRAINING AND TEST SETS

Rhythm type	Training set (N = 29)	Test set (N = 10)
NSR	48.5% (15)	45.7% (5)
SVPB	0.5% (15)	0.4% (5)
VPB	0.6% (12)	0.3% (5)
AFL	10.8% (3)	22.4% (2)
AF	39.6% (11)	31.0% (3)

NSR = normal sinus rhythm, SVPB = supraventricular premature beat, VPB = ventricular premature beat, AFL = atrial flutter, AF = atrial fibrillation.

The accuracy of heart rate measurement when using the same PPG-sensor has been previously reported in [29].

The ECG data were visually analyzed by a clinical expert using an automated rhythm detection software (Veritas, Mortara, Milwaukee, WI, USA). The software extracted beat times from the ECG and identified every beat either as normal, supraventricular premature beat (SVPB), VPB, AF, paced, artifact, or unknown. The rhythm was then confirmed or corrected by the expert. The software labeled also atrial flutters as AF, which were corrected after the visual inspection by the expert.

Out of the 40 patients, 14 had continuous AF during the recording period, 20 had normal sinus rhythm with premature contractions, 4 had continuous atrial flutter, 1 had continuous atrial flutter with atrial tachycardia, and one patient had a very noisy ECG reference. The patient with the very noisy ECG reference was excluded from the analysis, because no classification could be made based on the reference data.

For developing the classification models, the dataset was divided into two parts: a training set and a test set. The training set was used for training of the model and the test set was kept as unseen data to test the classification performance. The split was made by assigning 75% of the patients to the training set and 25% to the test set based on the rhythm, with the aim to have similar rhythm distributions in both datasets. These are presented in Table I based on the percentage of beats in every class and the number of patients having that rhythm. The atrial tachycardia was included in the class of AFL. The patient characteristics in both datasets are presented in Table II.

B. Preprocessing and Pulse Detection

The raw PPG data was uploaded from the data logging device and processed offline. As a preprocessing step prior to pulse detection, the PPG signal was downsampled from 128 Hz to 64 Hz and bandpass filtered to range from 0.3 to 5 Hz. The pulses were detected by identifying the fiducial points, i.e. the troughs, in the waveform by detecting the local minima. The time difference between two consecutive fiducial points was calculated to obtain the IPIs. The IPI time series were used to match the pulses to the labeled ECG beat times. The method for detecting the pulses and synchronizing the PPG and ECG beat information is described more in detail in [17]. Examples of 30-second segments of the PPG waveforms and the corresponding IPIs for different rhythm types are presented in Fig. 1 and Fig. 2. Fig. 2 shows differences in the rhythm characteristics between different AFL subjects.

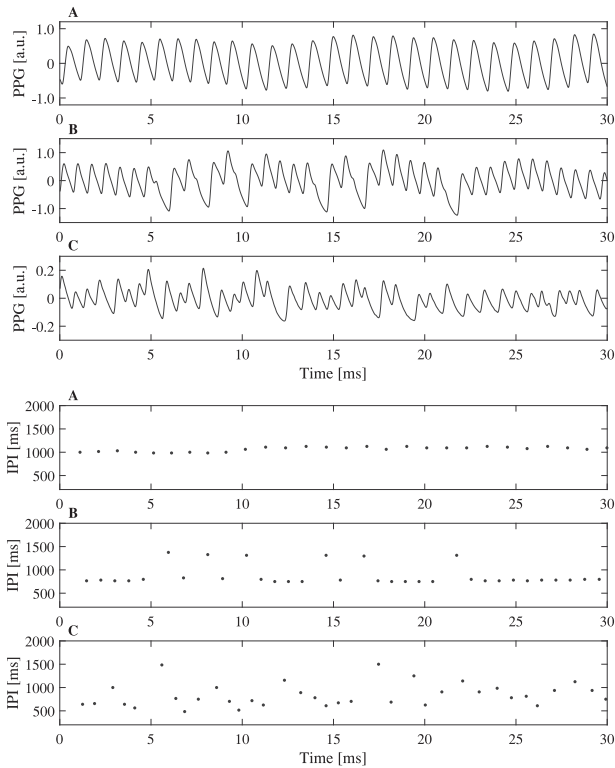


Fig. 1. PPG waveforms and the corresponding IPIs of sinus rhythm (A), sinus rhythm with SVPBs (B), and atrial fibrillation (C).

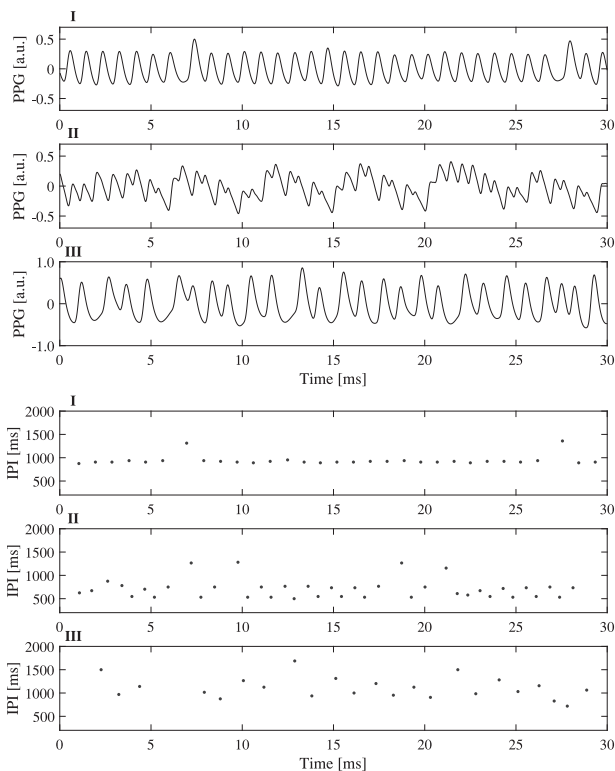


Fig. 2. PPG waveforms and the corresponding IPIs from three different atrial flutter subjects (I-III).

TABLE II
PATIENT CHARACTERISTICS

Training set	AF	AFL	Other
Age, years	66 ± 11	63 ± 7	69 ± 12
BMI, kg/m ²	31.2 ± 7.0	30.3 ± 5.6	28.6 ± 6.1
Min HR, bpm	43 ± 11	42 ± 27	49 ± 8
Mean HR, bpm	79 ± 16	78 ± 28	71 ± 12
Max HR, bpm	136 ± 19	131 ± 26	108 ± 17
Record length per patient, h	23.1 ± 1.1	23.4 ± 0.2	23.5 ± 1.0
Total record length, h	208	70	353
Test set			
Age, years	76 ± 6	70 ± 4	62 ± 16
BMI, kg/m ²	24.3 ± 3.3	28.3 ± 3.6	26.9 ± 3.3
Min HR, bpm	38 ± 17	47 ± 4	49 ± 9
Mean HR, bpm	80 ± 20	84 ± 3	72 ± 14
Max HR, bpm	144 ± 43	159 ± 15	119 ± 14
Record length per patient, h	22.6 ± 0.5	23.2 ± 0.1	23.8 ± 0.4
Total record length, h	68	46	119

The values are presented as mean ± standard deviation over subjects.
AF = atrial fibrillation, AFL = atrial flutter, HR = heart rate.

C. Modeling Architecture

In this study, two models were developed: an AF classification model based on the traditionally used IPI variability patterns (benchmark model), and a multi-rhythm model to classify AF, AFL, and other rhythms. Fig. 3 shows the block diagram of the two models. The benchmark model takes as input only features computed from IPI series, whereas the multi-rhythm model uses the IPI series, PPG waveform, and accelerometer data as input for the feature computation.

The PPG, accelerometer, and IPI time series data were segmented in 30-second non-overlapping windows for computing the features and every window was labeled based on the rhythm. The beat labels from ECG were used as the ground truth. When the majority of beats were AF, the window was labeled as AF. This was the same for AFL. When there were any SVPBs in the window, and no AF or AFL, the window was labeled as SVPB, and the same for VPB. The windows were labeled as sinus rhythm when they did not contain any of the previously mentioned rhythms. If more than half of the beats were labeled as artifact, the window was discarded from the analysis.

For binary classification, which was done with the benchmark model ((1) in Fig. 3), the classification was performed in three different ways. First, all the windows labeled as AF were considered as one class, and AFL, NSR, SVPB, and VPB together as another class, i.e. non-AF rhythms. Second, AFL was considered as the same class with AF, and finally, AFL was completely removed from the analysis. For the multi-rhythm classification ((2) in Fig. 3), AFL was separated as its own class, and NSR, SVPB, and VPB, treated as one class to which we refer as 'other'. Premature beats were not separated as their own class because there were too few examples compared to AF, AFL, and NSR. In addition, they usually are not considered to require medical treatment.

D. Features

The features to be used for the classification were calculated for every 30-second window in two stages: first for the

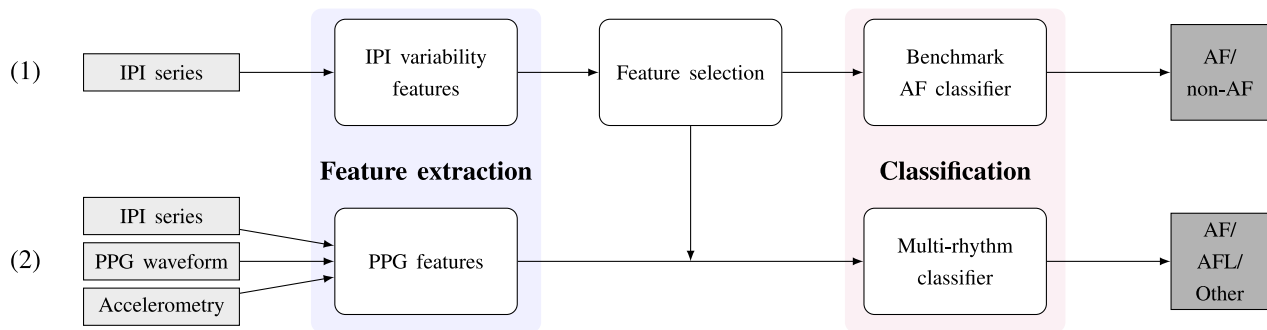


Fig. 3. Block diagram of the workflow for (1) the benchmark AF classifier and (2) the multi-rhythm classifier.

TABLE III
FEATURES FOR RHYTHM CLASSIFICATION

IPI-features	PPG waveform features	Accelerometer features
Shannon Entropy (ShEn)	Kurtosis (K)	Standard deviation*
Normalized Root Mean Square of Successive Differences (nRMSSD)	Hjorth mobility (\mathcal{H}_1)	Maximum absolute value*
pNN40	Hjorth complexity (\mathcal{H}_2)	
pNN70	Spectral Purity Index (SPI)	
Sample Entropy (sampEn)	Spectral Entropy (SE)	
Coefficient of Sample Entropy (CoSEn)	Maximum pulse amplitude	
Mean Markov score AF/AFL - IPIs	Minimum pulse amplitude	
Mean Markov score AF/AFL - PPPIs	Mean pulse amplitude	
Maximum IPI	Standard deviation of pulse amplitudes	
Minimum IPI		
Median IPI		
Standard deviation of IPIs		

*From the norm of the accelerations on the three axis.

IPI = inter-pulse intervals, PPPI = pulse-to-pulse-to-pulse interval.

benchmark AF classifier using only IPI-features, and second for the multi-rhythm classifier with the aim of improving overall accuracy and classify both AF and AFL separately. All the features used in the analysis are summarized in Table III.

Prior to the feature computation, IPIs that were shorter than 200 ms and longer than 2200 ms were excluded from the analysis. This was done because pulses could be sometimes falsely detected or missed, leading to incorrect IPIs. For more robust feature calculation, the features were calculated only if the window contained 20 or more IPIs.

For the benchmark AF classifier, six features characterizing the variability or entropy of the IPI series, which have been used for AF classification in the literature [22], [24], [30], [31], were computed:

- Shannon Entropy (ShEn)
- Normalized Root Mean Square of Successive Differences (nRMSSD)
- pNN40 and pNN70
- Sample Entropy (sampEn)
- Coefficient of Sample Entropy (CoSEn)

We have previously studied the discriminative power of these features individually for AF detection with PPG and compared that to ECG [17]. In the current paper, these features were used to build the first model to classify AF either with or without AFL.

For the multi-rhythm classifier, additional features were calculated in order to improve the detection accuracy and enable separate AFL classification. The features were derived either directly from the PPG waveform, from the IPI series, or from the accelerometer data. The features of the PPG waveform were computed from the signal segments after subtracting the mean value and dividing by its standard deviation. The new feature set consisted in total of 16 features, which will be described here.

One category consists of features that have been considered to measure the signal quality of PPG. A feature that has been used for motion artifact detection is kurtosis [32]. It is a statistical measure describing the tails of the distribution defined as

$$K = \frac{E(x - \mu)^4}{\sigma^4}, \quad (1)$$

where μ and σ are the mean and standard deviation of x , respectively, and $E(n)$ is the expected value of the quantity n .

In the PPG analysis domain, Hjorth descriptors called mobility and complexity [33], \mathcal{H}_1 and \mathcal{H}_2 , have previously been used for analyzing the quality of the signal [34]. The descriptors have initially been developed for electroencephalogram (EEG) analysis [33] and represent the mean frequency and half the bandwidth, respectively. The descriptors are calculated from spectral moments of the signal, the n th order spectral moment

being defined as

$$\bar{\omega}_n = \int_{-\pi}^{\pi} \omega^n S(e^{j\omega}) d\omega, \quad (2)$$

where $S(e^{j\omega})$ is the power spectrum. From the moments with different orders, $\mathcal{H}_1(n)$ is defined as

$$\mathcal{H}_1(n) = \sqrt{\frac{\bar{\omega}_2(n)}{\bar{\omega}_0(n)}}, \quad (3)$$

and $\mathcal{H}_2(n)$ as

$$\mathcal{H}_2(n) = \sqrt{\frac{\bar{\omega}_4(n)}{\bar{\omega}_2(n)} - \frac{\bar{\omega}_2(n)}{\bar{\omega}_0(n)}}. \quad (4)$$

The spectral moments were implemented in this work in the time domain according to [35].

A similar feature to $\mathcal{H}_2(n)$, which also has its origin in EEG analysis, is Spectral Purity Index (SPI) [35], [36]:

$$\text{SPI}(n) = \frac{\bar{\omega}_2^2(n)}{\bar{\omega}_4(n)\bar{\omega}_0(n)}. \quad (5)$$

Later SPI has been used also for ECG analysis in detecting false alarms of ventricular tachycardia and fibrillation/flutter [37], [38]. Recently, the method was also applied to PPG signals in order to distinguish ventricular arrhythmias from sinus rhythm and AF [23]. In addition, it has been considered as a signal quality metric for PPG [14].

Shannon Entropy in the time domain was included as a feature in the first stage to study the variability of the IPI series, but it can be also extended to the frequency domain. Spectral Entropy (SE) [39] measures the spectral complexity of the time series and can be calculated as

$$\text{SE} = \sum_f p_f \log \left(\frac{1}{p_f} \right), \quad (6)$$

where p_f is the power spectral density normalized with the total spectral power. Previously, Fallet *et al.* [23] have studied SE for discrimination of AF and ventricular arrhythmias from PPG.

In addition to the features from the frequency domain, additional IPI features were introduced. Modeling RR intervals as a Markov process has been used for AF detection initially by Moody and Mark [40]. In the model, each RR interval was considered to be either short, regular, or long. When every RR interval was assigned to one of these states, transitions between states and transition probability matrices for different rhythms could be calculated. From the transition probabilities, a score that represents the likelihood of the rhythm can be derived. With a similar approach having more states, a good performance in detecting AF from the IPI series has been recently shown [16]. In the current paper, the Markov model was used in order to distinguish between AF and AFL.

The transition probability matrices to model the Markov process were calculated for AF and AFL from the IPI series. These gave the score S that indicates whether the rhythm is

more likely to be AF than AFL:

$$S_{ij} = \log \left(\frac{p_{ij}^{AFL}}{p_{ij}^{AF}} \right), \quad (7)$$

where p_{ij}^{AFL} is the probability that after an interval belonging to state i , an interval belonging to state j occurs during AFL, and p_{ij}^{AF} is the same for AF.

In addition, as a new feature, the same procedure was used to calculate a score when using pulse-to-pulse-to-pulse intervals (PPPI) instead of IPIs. PPPIs were calculated as the time difference between two fiducial points by skipping one pulse in between, the length of the interval corresponding to the sum of two consecutive IPIs. The patterns formed by PPPIs are different from IPIs and this can be especially helpful for distinguishing when irregularity is due to an alternating IPI-pattern, such as during AFL, instead of due to AF. Both IPI and PPPI series were normalized and the scores filtered according to [40], having as the coefficient $k = 0.25$. The number of states for the used model was 12. Because a score is produced for every interval, a mean of the scores in the window was taken.

To include information reflecting the heart rate, the maximum, minimum, median, and standard deviation of the IPIs were included as features. In addition, from the PPG waveform, maximum, minimum, mean, and standard deviation of the pulse amplitude, i.e. the difference between the peak and the onset of the pulse, were calculated.

Finally, from the accelerometer data the norm of the accelerations on the three axis was calculated. From the norm, the standard deviation and the maximum absolute value in each window were included to the feature set.

E. Feature Selection

Feature selection was employed to select the optimal set of IPI variability features for the benchmark AF classification models. Based on our previous work [17], all the six features reflecting variability or entropy of the IPI sequence have individually a strong discriminative power in AF classification. Moreover, all these features try to capture relatively similar information and could be redundant. Therefore, the Minimal-Redundancy-Maximal-Relevance (mRMR) criterion [41] was selected as the method to rank the features.

The mRMR method tries to find features that maximize the mean value of mutual information between all individual features and the target class, i.e. the maximal relevance. However, it is likely that these features have a large dependency on each other. Therefore, the minimal redundancy criterion is added, and it is based on the mutual information between the individual features. The balance between the two criteria is optimized by finding the set that maximizes the difference between the maximal relevance and minimal redundancy.

The calculations were made with the implementation provided by the authors of [41] on [42]. The method requires discretization of the features and that was made by having three states when using thresholds at mean \pm standard deviation [42].

F. Classifiers

The benchmark AF classifier (Fig. 3 (1)) for the binary classification using the selected IPI-features as input was a generalized logistic regression model. The probability for the 30-second window to contain AF was given by the function:

$$y_{AF}(t) = \frac{e^{X(t) \cdot b}}{1 + e^{X(t) \cdot b}}, \quad (8)$$

where t is the index of the window, $X(t)$ a vector containing the feature values for the window at time t , and b a vector of the model coefficients.

The threshold for the probability was selected as the one that maximizes Youden index J [43] defined as

$$J = \frac{TP}{TP + FN} + \frac{TN}{TN + FP} - 1, \quad (9)$$

where TP are true positives, FN false negatives, TN true negatives, and FP false positives.

The multi-rhythm classifier (Fig. 3 (2)) to perform the classification into AF, AFL, and other, was a Random Forest (RF) model [44] taking as input the selected IPI-features and the additional 16 PPG, IPI, and accelerometry features. RFs are ensembles of decision trees that are grown in parallel by selecting a random subset of features to grow each tree. The final classification is based on a majority vote given by the classifications of each individual tree. RFs have a few beneficial characteristics, such as that they are relatively robust to noise and outliers, and can give useful internal estimates about the error and importance of the variables. The latter is an advantage with the small dataset because performing a feature selection for the multi-class problem is more challenging than for the binary case.

G. Cross-Validation

For selecting the number of features for the benchmark AF classifier, leave-one-subject-out cross-validation with the training set was used. The performance was calculated by using the data of one subject for testing and the data of the remaining subjects for training. The number of features went from one to six by adding the features in the order given by the mRMR. The number of features reaching the highest accuracy was selected.

The multi-rhythm classifier was also evaluated with the training set by using 10-fold cross-validation. In order to maintain an equal distribution of the three classes in every fold, the dataset was divided in 10 sets of equal size that were stratified by the classes but not by patients because of the number of AFL patients. The training and testing was performed 10 times with each of the 10 sets serving as a test set once.

III. RESULTS

A. Training Set

The mRMR selection method on IPI features was used for each of the three class divisions. For AF vs. non-AF classification gave the following ranking independent whether AFL was included in non-AF rhythms or completely excluded from the

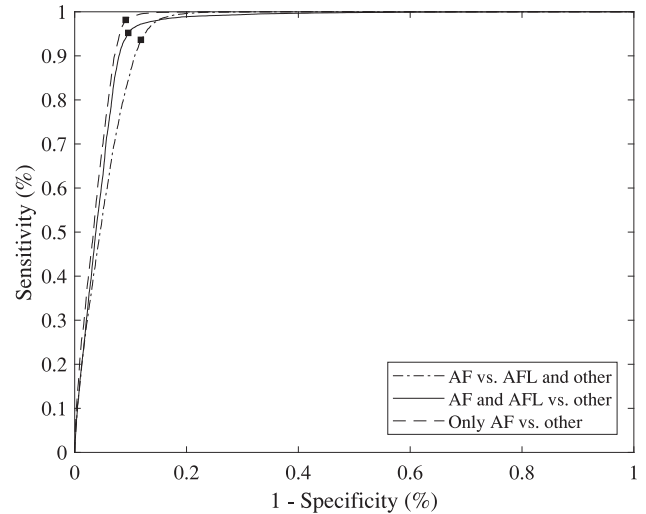


Fig. 4. Receiver Operating Characteristics of the binary classification models with three IPI-features. The square depicts the operating point defined by the Youden index.



Fig. 5. The distributions of the rhythm classes of the false positive AF detections (left) and false negative AF and AFL detections (right) with IPI-models in the training set.

analysis: pNN70, sampEn, ShEn, CoSEn, pNN40, nRMSSD, starting from the most relevant one. When AF and AFL vs. other classification was considered, only pNN70 and pNN40 switched order with each other in aforementioned ranking. Based on the leave-one-subject-out cross-validation, the best accuracy was obtained by combining the first three features. The Receiver Operating Characteristics (ROC) curve of the models combining these three features are presented in Fig. 4 when AFL is either included in or excluded from non-AF rhythms, or completely excluded from the analysis. The results are calculated with 43.6% of the data after the windows not containing a sufficient number of IPIs or the reference was labeled as artifact were excluded. The median (lower - upper quartiles) coverage per patient was 47.0% (29.6 – 56.2)%.

Fig. 4 shows the operating points for the three binary AF classifiers selected by maximizing J . The model AF vs. AFL and other had sensitivity of 93.6% and specificity of 88.2%. Misclassification of AF occurred primarily in presence of SVPB, and AFL, i.e. other arrhythmia generated from the atria, as shown in Fig. 5 on the left. The presence of VPB caused relatively little

TABLE IV

CONFUSION MATRIX OF THE CLASSIFICATIONS WITH THE MULTI-RHYTHM CLASSIFIER OF THE TRAINING SET

		Predicted label			N. windows
		AF	AFL	Other	
True label	AF	97.2%	0.7%	2.2%	11468
	AFL	9.9%	86.3%	3.8%	2261
	Other	3.0%	0.0%	97.0%	21281

TABLE V

CLASSIFICATION PERFORMANCE OF THE MULTI-RHYTHM CLASSIFIER WITH THE TRAINING SET

Rhythm type	Sensitivity	Specificity	PPV	Accuracy
AF	97.2%	96.4%	92.8%	96.7%
AFL	86.3%	99.7%	95.5%	98.9%
Other	97.0%	97.6%	98.5%	97.2%

TABLE VI

CONFUSION MATRIX OF THE CLASSIFICATIONS WITH THE MULTI-RHYTHM CLASSIFIER OF THE TEST SET

		Predicted label			N. windows
		AF	AFL	Other	
True label	AF	97.6%	0.7%	1.6%	3264
	AFL	1.7%	84.5%	13.8%	2793
	Other	1.8%	0.1%	98.1%	6747

false detections of AF. The model that classified AF and AFL vs. other had better sensitivity and specificity (95.2% and 90.4%). With this model the false negative classifications were mainly due to AFL, as shown on the right of Fig. 5. The best performance for binary classification with sensitivity of 98.2% and specificity of 90.9%, was obtained when AFL was not present in the data.

The multi-rhythm classifier was a RF model consisting of 100 trees and classified the 30-second windows into AF, AFL, or other rhythm. The results calculated with cross-validation are listed in Tables IV and V. Table IV is the confusion matrix of the classifications and in Table V the results are presented as one class vs. all the rest in terms of sensitivity, specificity, positive predictive value (PPV), and accuracy.

B. Test Set

The results of the test set were first calculated with the benchmark AF models. The test set had a coverage of 41.6%, with the remaining windows excluded due to insufficient IPIs or the ECG reference being labeled as artifact. The median (lower – upper quartiles) proportion of windows included for each subject was 42.2% (31.1–50.1)%.

For AF vs. AFL and other classification, the sensitivity and specificity were 96.1% and 96.2%, respectively. When AF and AFL were included in the same class against the rest, the sensitivity decreased to 58.9% and specificity to 92.5%. The best classification performance was again obtained when AFL was completely excluded from the analysis, the sensitivity being 99.1% and specificity 95.4%.

The results of the multi-rhythm model are presented in Tables VI and VII. The sensitivity and specificity for detecting

TABLE VII

CLASSIFICATION PERFORMANCE OF THE MULTI-RHYTHM CLASSIFIER WITH THE TEST SET

Rhythm type	Sensitivity	Specificity	PPV	Accuracy
AF	97.6%	98.2%	95.0%	98.1%
AFL	84.5%	99.7%	98.7%	96.4%
Other	98.1%	92.8%	93.8%	95.6%

AF were 97.2% and 98.2%, respectively. AF and other rhythms were rarely detected as AFL, and the specificity for AFL was 99.7%. The most false detections are with AFL detected as other rhythm.

IV. DISCUSSION

This is the first study showing that both AF and AFL detection is possible from PPG data in daily life. We demonstrated that many of the false positive classifications of a benchmark AF classification model were due to instances of AFL. When considering AFL to be classified with AF as the same class, it was often missed by the benchmark model. The presented multi-rhythm classification algorithm showed much improved performance, particularly making less false positive AF detections, when trained to classify both AF as well as AFL. As the results with the benchmark AF models show, the rhythm characteristics when using the IPI-features did not belong to either of the binary classes, i.e. AF and other, and decreased the classification performance. Therefore, considering AFL as a separate class can be beneficial also in terms of improving AF detection and not only for giving classification for the rhythm type itself.

Combining information from the PPG signal, IPIs, and accelerometer improved the classification accuracy and enabled discrimination of AFL from AF and the other rhythm types. Adding features from the PPG waveform helped in detecting AFL compared to using only IPI-information. In previous work, a comparison of PPG pulses gave different results in a patient suffering from a regular form of typical AFL than in patients with AF or other rhythms [45]. Moreover, some of the features derived from the PPG waveform have been used to discriminate ventricular arrhythmias [23]. The PPG waveform characteristics have been also studied in the context of force-interval relationship during AF [46] and mechanical alternans [47]. This could indicate that the PPG waveform itself contains information about different rhythms and cardiac function.

Discriminating AFL from AF has been also possible based on RR interval series when a multilevel model of the atrioventricular node was used [48]. In the current paper, we included the Markov model approach to process the IPI and PPPI series in order to classify these two rhythms. Thus, the RR intervals or IPIs also contain valuable information when processed in an adequate manner.

The validity of the windows to be analyzed was judged based on the number of pulses detected in that window, but no further signal quality analysis was made. The method detecting the pulses already considers the body acceleration and therefore the

number of pulses indirectly already reflects this. In addition, some of the features included in the RF model have been developed or used previously for PPG signal quality assessment, and information of the body acceleration was also given as an input. This may have improved the classification of the more noisy segments.

The study suffers from some limitations. In the dataset, all the patients that suffered from AF, had it continuously. For datasets with inpatient rhythm variability, data measured before and after electrical cardioversion have been used in different studies to investigate AF detection with PPG. This reflects a hospital setting but not ambulatory monitoring, which is where the wrist-worn applications can really add value. Studies that show AF detection with PPG in ambulatory setting are very few. Shen *et al.* [49] had measurements from 3 to 8 hours and mention subjects with rhythms that change over time. However, these include AF and eight other rhythms and the proportion of paroxysmal AF remains unclear. Sološenko *et al.* [20] presented in their study a dataset measured for approximately 22 hours per subject in cardiac rehabilitation. Similarly to our study, all the AF subjects had continuous AF.

The number of AFL subjects in the dataset was less compared to the other groups. AFL can have different rhythm characteristics depending on the type and therefore can vary between subjects. The results between the training set and test set in AFL classification differ slightly due to the different types of AFL cases, e.g. very regular AFL being misclassified as other rhythm in the test set. Yet, the results remain relatively similar when compared to the benchmark AF models.

Premature beats were not separately classified in this work because of the small number compared to other beat types nor were they suppressed in order to reduce false positive detections. Furthermore, the classification was done in windows which makes identifying individual beats difficult. As Fig. 5 shows, premature beats had an impact on false positives when only IPI-based features were used in the classification. Adding PPG-waveform and accelerometry features helped in improving specificity by also reducing false positives caused by premature beats and not only by AFL. Understanding the effect on the classification accuracy caused by higher burden of premature beats and variability in their beat patterns between patients remains for future research.

The dataset was divided into training and test set in order to leave some of the recordings untouched while developing the models. The split was done by patients and no data from a patient assigned to the training set ended up in the test set. Therefore, it was not possible to match the rhythm class distributions completely between the two sets. The characteristics of the sets may, therefore, not be entirely comparable, which is reflected especially in the results of the benchmark AF model.

The choice to use RF was made because of the class distribution and the ability of RF to give information about the feature importance. The model has many advantages, but one drawback is poor interpretability. For clinical applications, more transparent options of models may be more suitable when larger datasets are available.

The selected approach was based on calculating the features in 30-second non-overlapping windows. The features based on the IPIs require a sufficient amount of data for the calculations, and missed pulses caused the data to be discarded, thereby reducing coverage to approximately 45%. The coverage of our approach could be improved by reducing the number of IPIs required per window or by using overlapping windows. However, the effects on the classification performance should be studied. Sološenko *et al.* [20] used an approach to classify every beat separately, which resulted in a higher coverage (89.2%) during 24-hour measurements. However, the sensitivity for AF with this coverage was only 72.0%. When 50% of the data was judged as analyzable, the sensitivity increased to 97.2% and specificity was 99.6%.

V. CONCLUSION

In this study, we demonstrated that PPG and acceleration measurements at the wrist can be used to discriminate between AF, AFL, and other rhythms in daily life. We showed that with an AF vs. non-AF model and AF and AFL vs. other rhythms model that used only information derived from inter-pulse intervals, the false detections were for a large part caused by AFL. The multi-rhythm model included more information from the wrist measurement, such as features from the PPG waveform and accelerometer data. This model was not only able to improve the overall performance of AF detection, but could also classify AFL with high accuracy. The results of this study indicate that the PPG signal contains sufficient information, derived both from the waveform and IPIs, to accurately classify between AF, AFL, and other rhythms. Thus, PPG could provide promising means to detect AFL along with AF.

ACKNOWLEDGMENT

The authors would like to thank N. Sturkenboom (MD), L. Verborg (MD), R. Eerdeken (MD), L. van den Heuvel, the personnel of the Holter department, and the Holter analysts of the Catharina Hospital for their help and contributions in the data collection.

REFERENCES

- [1] S. Björck, B. Palaszewski, L. Friberg, and L. Bergfeldt, "Atrial fibrillation, stroke risk, and warfarin therapy revisited," *Stroke*, vol. 44, no. 11, pp. 3103–3108, 2013. [Online]. Available: <https://www.ahajournals.org/doi/10.1161/STROKEAHA.113.002329>
- [2] M. Haim *et al.*, "Prospective national study of the prevalence, incidence, management and outcome of a large contemporary cohort of patients with incident nonvalvular atrial fibrillation," *J. Am. Heart Assoc.*, vol. 4, no. 1, pp. 1–12, 2015.
- [3] J. S. Healey *et al.*, "Subclinical atrial fibrillation and the risk of stroke," *New England J. Med.*, vol. 366, no. 2, pp. 120–129, 2012.
- [4] P. Kirchhof *et al.*, "2016 ESC guidelines for the management of atrial fibrillation developed in collaboration with EACTS," *Europace*, vol. 18, no. 11, pp. 1609–1678, 2016.
- [5] R. L. Page *et al.*, "2015 ACC/AHA/HRS guideline for the management of adult patients with supraventricular tachycardia: Executive summary," *J. Am. College Cardiology*, vol. 67, no. 13, pp. 1575–1623, 2015.
- [6] S. S. Bun, D. G. Latcu, F. Marchlinski, and N. Saoudi, "Atrial flutter: More than just one of a kind," *Eur. Heart J.*, vol. 36, no. 35, pp. 2356–2363, 2015.
- [7] S. C. Halligan, B. J. Gers, and R. D. Brown, "The natural history of lone atrial flutter," *Ann. Internal Med.*, vol. 40, no. 4, pp. 265–268, 2004.

- [8] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological Meas.*, vol. 28, pp. R1–R39, 2007.
- [9] J. van Andel, C. Ungureanu, R. Aarts, F. Leijten, and J. Arends, "Using photoplethysmography in heart rate monitoring of patients with epilepsy," *Epilepsy Behav.*, vol. 45, pp. 142–145, 2015. [Online]. Available: <http://dx.doi.org/10.1016/j.yebeh.2015.02.018>
- [10] I. Gotlibovych *et al.*, "End-to-end deep learning from raw sensor data: atrial fibrillation detection using wearables," 2018, *arXiv:1807.10707*.
- [11] C. Yang *et al.*, "Using PPG signals and wearable devices for atrial fibrillation screening," *IEEE Trans. Ind. Electron.*, vol. 66, no. 11, pp. 8832–8842, Nov. 2019. [Online]. Available: <https://ieeexplore.ieee.org/document/8600713/>
- [12] Y.-Y. Fan, Y.-G. Li, J. Li, W.-K. Cheng, and Z.-L. Shan, "Diagnostic performance of a smart device with photoplethysmography technology for atrial fibrillation detection: Pilot study (Pre-mAFA II Registry)," *JMIR mHealth uHealth*, vol. 7, no. 3, pp. 1–11, 2019.
- [13] M. Lemay, S. Fallet, P. Renevey, J. Sol, E. Pruvot, and J.-M. Vesin, "Wrist-located optical device for atrial fibrillation screening: A clinical study on twenty patients," in *Proc. Comput. Cardiology*, 2016, pp. 681–684.
- [14] S. Nemati *et al.*, "Monitoring and detecting atrial fibrillation using wearable technology," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2016, pp. 3394–3397. [Online]. Available: <http://ieeexplore.ieee.org/document/7591456/>
- [15] S. P. Shashikumar, A. J. Shah, Q. Li, G. D. Clifford, and S. Nemati, "A deep learning approach to monitoring and detecting atrial fibrillation using wearable technology," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Inform.*, 2017, pp. 141–144.
- [16] A. G. Bonomi *et al.*, "Atrial fibrillation detection using a novel cardiac ambulatory monitor based on photo-plethysmography at the wrist," *J. Am. Heart Assoc.*, vol. 7, no. 15, 2018. Art. no. e009351.
- [17] L. M. Eerikäinen *et al.*, "Comparison between electrocardiogram- and photoplethysmogram-derived features for atrial fibrillation detection in free-living conditions," *Physiological Meas.*, vol. 39, no. 8, 2018, Art. no. 084001.
- [18] A. Aliamiri and Y. Shen, "Deep learning based atrial fibrillation detection using wearable photoplethysmography sensor," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Inform.*, 2018, vol. 2018, pp. 442–445.
- [19] M. Voisin *et al.*, "Ambulatory atrial fibrillation monitoring using wearable photoplethysmography with deep learning," 2018, *arXiv:1811.07774*.
- [20] A. Sološenko, A. Petrénas, L. Sörnmo, B. Paliakaitė, and V. Marozas, "Detection of atrial fibrillation using a wrist-worn device," *Physiological Meas.*, vol. 40, 2019, Art. no. 025003.
- [21] A. Kashiwa *et al.*, "Performance of an atrial fibrillation detection algorithm using continuous pulse wave monitoring," *Ann. Noninvasive Electrocardiology*, vol. 24, no. 2, pp. 1–8, 2019.
- [22] V. D. A. Corino *et al.*, "Detection of atrial fibrillation episodes using a wristband device," *Physiological Meas.*, vol. 38, pp. 787–799, 2017.
- [23] S. Fallet *et al.*, "Can one detect atrial fibrillation using a wrist-type photoplethysmographic device?" *Med. Biol. Eng. Comput.*, vol. 57, no. 2, pp. 477–487, 2018.
- [24] J. W. Chong, D. D. McManus, and K. H. Chon, "Arrhythmia discrimination using a smart phone," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 3, pp. 1–4, May 2015. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6575493>
- [25] D. D. McManus *et al.*, "PULSE-SMART: Pulse-based arrhythmia discrimination using a novel smartphone application," *J. Cardiovascular Electrophysiology*, vol. 27, no. 1, pp. 51–57, 2016.
- [26] M. Z. Poh *et al.*, "Diagnostic assessment of a deep learning system for detecting atrial fibrillation in pulse waveforms," *Heart*, vol. 104, pp. 1921–1928, 2018.
- [27] M. Dörr *et al.*, "The WATCH AF trial: SmartWATCHes for detection of atrial fibrillation," *JACC Clin. Electrophysiology*, vol. 5, no. 2, pp. 199–208, 2019. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/30784691>
- [28] D. E. Krummen, G. K. Feld, and S. M. Narayan, "Diagnostic accuracy of irregularly irregular RR intervals in separating atrial fibrillation from atrial flutter," *Am. J. Cardiology*, vol. 98, no. 2, pp. 209–214, 2006.
- [29] F. Sartor, J. Gelissen, R. van Dinther, D. Roovers, G. B. Papini, and G. Coppola, "Wrist-worn optical and chest strap heart rate comparison in a heterogeneous sample of healthy individuals and in coronary artery disease patients," *BMC Sports Sci., Med. Rehabil.*, vol. 10, no. 1, pp. 1–10, 2018.
- [30] S. Dash, K. H. Chon, S. Lu, and E. A. Raeder, "Automatic real time detection of atrial fibrillation," *Ann. Biomed. Eng.*, vol. 37, no. 9, pp. 1701–1709, 2009.
- [31] D. E. Lake and J. R. Moorman, "Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices," *AJP, Heart Circulatory Physiology*, vol. 300, no. 1, pp. H319–H325, 2011. [Online]. Available: <http://ajpheart.physiology.org/cgi/doi/10.1152/ajpheart.00561.2010>
- [32] N. Selvaraj, Y. Mendelson, K. H. Shelley, D. G. Silverman, and K. H. Chon, "Statistical approach for the detection of motion/noise artifacts in photoplethysmogram," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2011, pp. 4972–4975.
- [33] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalography Clin. Neurophysiology*, vol. 29, no. 3, pp. 306–310, 1970.
- [34] E. Gil, J. María Vergara, and P. Laguna, "Detection of decreases in the amplitude fluctuation of pulse photoplethysmography signal as indication of obstructive sleep apnea syndrome in children," *Biomed. Signal Process. Control*, vol. 3, pp. 267–277, 2008.
- [35] L. Sörnmo and P. Laguna, *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. New York, NY, USA: Academic, 2005.
- [36] I. I. Goncharova and J. S. Barlow, "Changes in EEG mean frequency and spectral purity during spontaneous alpha blocking," *Electroencephalography Clin. Neurophysiology*, vol. 76, no. 3, pp. 197–204, 1990.
- [37] S. Fallet, S. Yazdani, and J.-M. Vesin, "False arrhythmia alarms reduction in the intensive care unit: A multimodal approach," *Physiological Meas.*, vol. 37, pp. 1217–1232, 2016.
- [38] L. M. Eerikäinen, J. Vanschoren, M. J. Rooijackers, R. Vullings, and R. M. Aarts, "Reduction of false arrhythmia alarms using signal selection and machine learning," *Physiological Meas.*, vol. 37, pp. 1204–1216, 2016.
- [39] I. A. Rezek and S. J. Roberts, "Stochastic complexity measures for physiological signal analysis," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 9, pp. 1186–1191, Sep. 1998.
- [40] G. B. Moody and R. G. Mark, "A new method for detecting atrial fibrillation using R-R intervals," *Comput. Cardiology*, vol. 10, 1983, pp. 227–230.
- [41] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.
- [42] H. Peng, "mRMR (minimum redundancy maximum relevance feature selection)," Accessed: Feb. 11, 2019. [Online]. Available: <http://home.penglab.com/proj/mRMR/>
- [43] W. J. Youden, "Index for rating diagnostic tests," *Cancer*, vol. 3, no. 1, pp. 32–35, 1950.
- [44] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.
- [45] G. B. Papini *et al.*, "Sinus or not: A new beat detection algorithm based on a pulse morphology quality index to extract normal sinus rhythm beats from wrist-worn photoplethysmography recordings," *Physiological Meas.*, vol. 39, no. 11, 2018, Art. no. 115007.
- [46] L. M. Eerikäinen *et al.*, "Force-interval relationships of the heart measured with photoplethysmography during atrial fibrillation," in *Proc. Comput. Cardiology Conf.*, 2018, vol. 45, pp. 1–4.
- [47] T. Besleaga *et al.*, "Detecting mechanical alternans utilizing photoplethysmography," in *Proc. Comput. Cardiology Conf.*, 2018, vol. 45, pp. 1–4.
- [48] E. P. Scholz *et al.*, "Discriminating atrial flutter from atrial fibrillation using a multilevel model of atrioventricular conduction," *Heart Rhythm*, vol. 11, no. 5, pp. 877–884, 2014. [Online]. Available: <http://dx.doi.org/10.1016/j.hrthm.2014.02.013>
- [49] Y. Shen, M. Voisin, A. Aliamiri, A. Avati, A. Hannun, and A. Ng, "Ambulatory atrial fibrillation monitoring using wearable photoplethysmography with deep learning," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2019, pp. 1909–1916.