

# UCSF

## UC San Francisco Previously Published Works

### Title

Photoplethysmography based atrial fibrillation detection: a review

### Permalink

<https://escholarship.org/uc/item/64j2745n>

### Journal

npj Digital Medicine, 3(1)

### ISSN

2398-6352

### Authors

Pereira, Tania

Tran, Nate

Gadhoumi, Kais

et al.

### Publication Date

2020-01-10

### DOI

10.1038/s41746-019-0207-9

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

## REVIEW ARTICLE OPEN

## Photoplethysmography based atrial fibrillation detection: a review

Tania Pereira<sup>1\*</sup>, Nate Tran<sup>1</sup>, Kais Gadhomi<sup>1</sup>, Michele M. Pelter<sup>1</sup>, Duc H. Do<sup>2</sup>, Randall J. Lee<sup>3</sup>, Rene Colorado<sup>4</sup>, Karl Meisel<sup>4</sup> and Xiao Hu<sup>1,5,6,7</sup>

Atrial fibrillation (AF) is a cardiac rhythm disorder associated with increased morbidity and mortality. It is the leading risk factor for cardioembolic stroke and its early detection is crucial in both primary and secondary stroke prevention. Continuous monitoring of cardiac rhythm is today possible thanks to consumer-grade wearable devices, enabling transformative diagnostic and patient management tools. Such monitoring is possible using low-cost easy-to-implement optical sensors that today equip the majority of wearables. These sensors record blood volume variations—a technology known as photoplethysmography (PPG)—from which the heart rate and other physiological parameters can be extracted to inform about user activity, fitness, sleep, and health. Recently, new wearable devices were introduced as being capable of AF detection, evidenced by large prospective trials in some cases. Such devices would allow for early screening of AF and initiation of therapy to prevent stroke. This review is a summary of a body of work on AF detection using PPG. A thorough account of the signal processing, machine learning, and deep learning approaches used in these studies is presented, followed by a discussion of their limitations and challenges towards clinical applications.

*npj Digital Medicine* (2020)3:3; <https://doi.org/10.1038/s41746-019-0207-9>

## INTRODUCTION

Atrial fibrillation (AF) is an abnormal cardiac rhythm characterized by a disorganized atrial activity. AF is recognized in the electrocardiogram (ECG) as an irregularly irregular rhythm lasting more than 30 s, with no discernible P-waves preceding the QRS complex.<sup>1</sup> AF prevalence is age, gender, and race dependent.<sup>2</sup> It is particularly high in the elderly population, reaching 10–17% in subjects 80 years and older.<sup>3</sup> In addition, AF is more prevalent in males and in the white population.<sup>3</sup> AF is associated with significant morbidity and mortality. One in five strokes is associated with AF and one-third of cardiac arrhythmias hospitalizations are due to AF-related complications. AF has been associated with a twofold increase in the risk of death.<sup>4</sup> Additionally, the aging population in the US and worldwide is leading to a markedly increasing AF prevalence<sup>3,5</sup>.

The high prevalence of asymptomatic AF has significant clinical implications on the diagnosis and management of AF.<sup>6</sup> Intermittent ECG evaluation during clinical visits has a low likelihood of detecting paroxysmal AF. Continuous monitoring would increase the chances of AF detection, thereby allowing appropriate primary and secondary stroke prevention strategies to reduce the high morbidity and mortality of stroke.

For patients with acute ischemic stroke or transient ischemic attack, approximately 10% will have new AF detected during their hospital admission.<sup>7–9</sup> Continuous ECG monitoring for 30 days is recommended in case of an embolic stroke of undetermined cause (cryptogenic).<sup>9</sup> Novel non-intrusive approaches for cardiac rhythm monitoring can potentially enable early and accurate detection of asymptomatic paroxysmal AF and create a shift in AF management.<sup>10,11</sup> Especially for asymptomatic AF cases, new tools

that allow the AF detection will help make the appropriate clinical decisions.<sup>10</sup>

Photoplethysmography (PPG) has emerged as a low-cost and non-intrusive modality for continuous monitoring of heart rate. A variety of wearable devices offer PPG-based monitoring, including smartphones and smartwatches. A photoplethysmogram is a pulse pressure signal resulting from the propagation of blood pressure pulses along arterial blood vessels. Measured on the periphery, it carries rich information about the cardiac activity, cardiovascular condition, the interaction between parasympathetic and sympathetic nervous systems, and hemoglobin level.<sup>12</sup> Many physiological parameters can be derived from PPG, including oxygen saturation, heart rate, blood pressure, and cardiac output.<sup>13</sup> These capacities of PPG open the door to develop new ambulatory diagnosis tools enabling early screening of heart conditions, including arrhythmia.<sup>14</sup>

This review provides an account of the approaches used in PPG-based AF detection. A brief overview of the technology behind PPG is first presented, followed by a summary of methods and algorithms developed for PPG-based AF detection. Recognizing the importance of using PPG to detect AF at scale, the motivation of this review is to guide the future development of algorithms towards clinical-grade applications.

## PHOTOPLETHYSMOGRAPHY

PPG signal

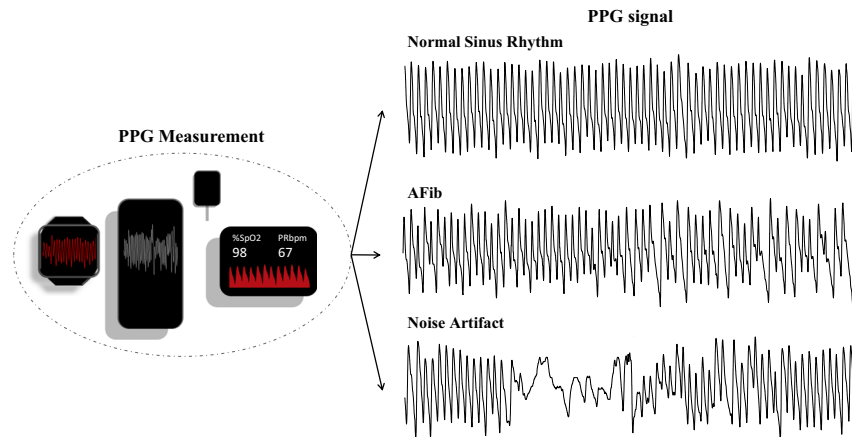
PPG waveform is generated during a cardiac cycle and typically measured at a peripheral site. Therefore, it is essentially a pulse pressure waveform that originates from the heart contraction and propagates through the vascular tree. As blood flow is controlled

<sup>1</sup>Department of Physiological Nursing, University of California, San Francisco, CA, USA. <sup>2</sup>David Geffen School of Medicine, University of California, Los Angeles, CA, USA.

<sup>3</sup>Cardiovascular Research Institute, Department of Medicine, Institute for Regeneration Medicine, University of California, San Francisco, CA, USA. <sup>4</sup>Department of Neurology, School of Medicine, University of California, San Francisco, CA, USA. <sup>5</sup>Department of Neurosurgery, School of Medicine, University of California, Los Angeles, CA, USA.

<sup>6</sup>Department of Neurological Surgery, University of California, San Francisco, CA, USA. <sup>7</sup>Institute of Computational Health Sciences, University of California, San Francisco, CA, USA.

\*email: [taniapereira10@gmail.com](mailto:taniapereira10@gmail.com)



**Fig. 1** PPG signal acquired using a wearable device and typical waveforms representing NSR, AF, and noise artifact.

by neural, cardiac, and respiratory interactions, various physiological parameters could theoretically be extracted from analyzing a PPG signal.<sup>15</sup> For this reason, the PPG signal has rich information about physiological conditions.<sup>13</sup>

PPG waveforms have typical morphological components corresponding to landmark events in the cardiac cycle. During the contraction of the left ventricle, blood is ejected out of the heart and propagates along the arterial tree, this corresponds to the initial positive slope of a PPG pulse. The systolic peak marks the maximum of the waveform. A decrease in amplitude following the systolic peak is marked by a local minimum, or the dicrotic notch, which corresponds to the closing of aortic valves separating the systolic and diastolic phases. In some cases, a third peak following the dicrotic notch can be identified. It corresponds to a reflected component of the forward wave from various reflection sites including vessel bifurcations.<sup>16</sup>

#### Clinical parameters

One primary clinical application of PPG is arterial blood oxygen saturation (SpO<sub>2</sub>) estimation through pulse oximetry.<sup>17</sup> SpO<sub>2</sub> is defined as the percentage of oxygen saturation in the arterial blood, which can be measured by the ratio of oxygenated hemoglobin concentration to the total hemoglobin concentration, with a normal range between 97% and 98%.<sup>18</sup> Recently, new applications of PPG have emerged for the continuous estimation of valuable cardiovascular parameters in ambulatory settings. Heart rate, blood pressure, and respiratory rate could be closely monitored for fitness or health assessment.<sup>19</sup> Advanced diagnostic applications of PPG were also envisaged. Cardiac function, arterial stiffness, autonomic nervous system (ANS) responses, and apnea are among conditions that could potentially be detected or evaluated using PPG.

Changes in blood volume are synchronous with the heart beats, such synchrony is manifested by the concordance of inter-beat intervals (RR intervals) measured in PPG and time-synchronized ECG.<sup>20</sup> Heart rate variability (HRV) is an indirect measurement of ANS, and it has also been considered as a surrogate parameter of the interaction between the brain and cardiovascular system.<sup>21</sup> HRV metrics can be derived from analyzing RR intervals in time and/or frequency domain as well as using nonlinear dynamic analysis approaches.<sup>22</sup> Respiratory rate is one of the fundamental vital signs and can be determined from the time–frequency representation of a PPG signal.<sup>23</sup>

Some hemodynamic parameters such as augmentation index (AIx) and pulse wave velocity (PWV) are important biomarkers of arterial stiffness, which is a direct cause of hypertension and a major risk factor for cardiovascular events such as myocardial

infarction and stroke. Both AIx and PWV could be derived from PPG,<sup>24,25</sup> Subendocardial Viability Ratio (SEVR %) and Ejection Time Index (ETI) are two hemodynamic parameters used in the evaluation of cardiac workload that can be estimated with PPG analysis.<sup>25</sup> Additionally, some studies claim that arterial blood pressure could be estimated using advanced analysis of PPG.<sup>17</sup>

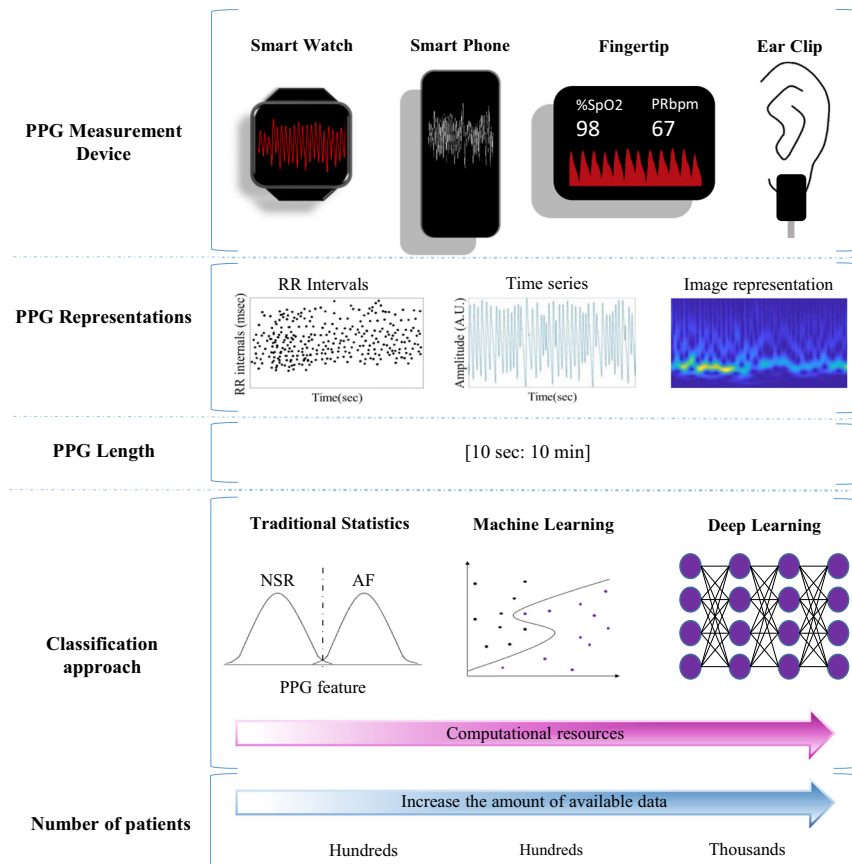
#### Modes of PPG measurement

A PPG signal has two main components: a quasi-static direct current (DC) component, which represents light reflected/transmitted from static arterial blood, venous blood, skin and tissues; and pulsatile alternate current (AC) component which arises from modulation in light absorption due to changes in arterial blood volume. PPG measurement can be carried out using two modes: transmission and reflectance. In transmission mode, the light transmitted through the medium is detected by a photodetector (PD), which is positioned in the opposite site of the light source. The sensor must be located on the body at a site where transmitted light can be detected. The measurement site is limited to the extremities of the body, such as the fingertip or earlobe. The greatest disadvantage of the transmission mode is the location of the device that can interfere with daily routine movements.<sup>26</sup> In reflectance mode, the PD detects light that is back scattered or reflected from tissues, bone, and/or blood vessels, which means the light source and PD are positioned on the same side. Unlike the transmission mode, the measurement sites are not restricted to any particular location, which facilitates a user-friendly monitoring approach. The wrist, forearm, ankle, and forehead are common measurement sites.<sup>27</sup>

Since the basic form of PPG technology requires only a few optoelectronic components (a light source and a PD: to measure the variations on the light reflected/transmitted by the tissues), it can be easily and inexpensively incorporated in various digital devices such as watches, smartphones, or wearables.<sup>28</sup> The ubiquitous availability of PPG in a wide range of wearable digital devices has motivated the search for new applications and the development of novel biomedical solutions.

#### PPG-BASED AF DETECTION

In a PPG signal, AF is manifested as varying pulse-to-pulse intervals and pulse morphologies. On the other hand, a normal sinus rhythm (NSR) is recognizable through regularly spaced PPG pulses with similar morphologies between consecutive pulses. Recognizing an arrhythmia in a PPG signal can sometimes be challenging in the presence of artifacts. Common sources of artifacts are motion and poor sensor contacts. Artifacts can be misinterpreted as



**Fig. 2** Overview of the main features extracted from PPG signals used in the studies reviewed (see Tables 1–3). SpO2 oxygen saturation, PRbpm pulse rate (beats per minute).

physiological abnormalities. Motion artifacts can be identified using accelerometry data. Most modern wearable devices include accelerometry sensors that measure acceleration forces along different spatial directions. It is a common practice to discard PPG contaminated with an artifact. Figure 1 depicts samples of PPG with NSR, AF, and artifact.

ECG remains the gold standard for the electrophysiological definition and recognition of arrhythmias,<sup>1</sup> including AF diagnosis.<sup>29</sup> In a recent study, new deep learning approaches achieved cardiologist-level AF detection of 12 types of arrhythmia (F1 score = 0.84 vs F1 score = 0.78) when 91,232 single-lead ECGs from 53,549 patients were analyzed.<sup>30</sup> Compared to ECG, PPG-based AF detection is more challenging but also rewarding in situations where longer monitoring time and lower cost beyond what ECG offers is needed, e.g., screening AF at scale.

Recent advances in sensor technologies and wearable devices have increased the role that a PPG-based solution could play in the assessment of health status. Electronics capable of recording PPG signals with relatively high signal-to-noise ratio (SNR) may warrant reliable PPG monitoring and screening of arrhythmia.<sup>11,31</sup>

In a typical AF detection algorithm, features (temporal, spectral, or morphological) are extracted from the acquired PPG signal and analyzed by the detection algorithm to inform if an AF rhythm is detected. In some approaches, image representation of the temporal waveform has been considered. The derived image would then be analyzed using conventional image processing or artificial intelligence-based methods (Fig. 2).<sup>32–34</sup> Traditionally, prominent features were derived from the tachogram (RR intervals) since it is a reliable measure of heart beats.<sup>35</sup> Realizing that PPG waveforms may carry physiological information beyond heart rate, new features beyond RR intervals were derived.<sup>36</sup> The use of PPG time series and their images representation (e.g. raw

plot of the signal, fast Fourier transform spectrum, or wavelet spectrogram—represented in the Fig. 2 in PPG representation part) were used with promising results in the detection of physiological events,<sup>32,37,38</sup> Images for PPG representation in Fig. 2 is a general depiction of the format types of information used by the different algorithms.

In the following sections, we review studies of PPG-based AF detection. A body of white papers and peer-reviewed works indexed by PubMed, Scopus, IEEE Xplore, and Web of Science up to June 2019 was selected based on the following search expression: (PPG “OR” Photoplethysmography) “AND” (atrial fibrillation “OR” AF “OR” AFib) “AND” (detection “OR” recognition). Each study is reviewed with respect to the size, the number of patients, and recording settings of data analyzed, the PPG device and site of recording, the AF detection algorithm, and its performance. Figure 2 summarizes the main features examined in these studies, described with more details in Tables 1–3.

#### Performance metrics

AF detection algorithms can be evaluated using several performance metrics. It is common for many studies to report sensitivity, specificity, and accuracy. Sensitivity is defined as the probability to detect true AF events, while the specificity measures the proportion of actual Non-AF instances correctly identified as such. Accuracy is a balanced metric of sensitivity and specificity. The accuracy of an AF detection algorithm is its ability to differentiate between AF and Non-AF cases.<sup>39</sup> Generally, accuracy is the most common reported metric, along with the area under the curve (AUC) of the receiver operating characteristic (ROC). A ROC for differentiating AF vs Non-AF is generated by plotting sensitivity vs (1-specificity) at different classification thresholds. AUC is a

**Table 1.** Studies on photoplethysmography-based AF detection using statistical analysis approaches.

Author (year) [ref.]	Number of patients	Dataset features	Age of population	Length PPG segments	Measurement device	Acquisition conditions	Input data	Methodology	Performance results for rhythms detection
Lee et al. (2013) <sup>91</sup>	74	74 prior and after cardioversion + Public databases (MIT-BIH AF + MIT-BIH NSR + MIT-BIH Arrhythmia Database)	–	2 min	Video camera of smartphone	Inpatient	RR times series features Time varying coherence functions and Shannon entropy	Derived threshold values of features for best ROC	Acc = 0.9645, Sen = 0.9716, Sp = 0.9539
Nemati et al. (2016) <sup>52</sup>	46	15 with AF 31 non-symptomatic	–	3.5 to 8.5 min	Wrist-worn device Samsung Simband	Inpatient	RR times series features sample entropy with the embedding dimensions $m = 1$ , and 2 (SampEn1 and SampEn2), standard deviation, robust standard deviation	Elastic Net logistic model	Acc = 0.95, Sen = 0.97, Sp = 0.94, AUC = 0.99
Bonomi et al. (2016) <sup>50</sup>	16	4 with AF, 1 atrial flutter, 11 NSR	65.2 ± 14.0	30 s	Wrist-wearable sensor—Philips Cardio and Motion Monitoring Module	Outpatient—continuous measurement	RR times series features	First-order 11-state Markov model	Sen = 0.97 ± 0.02, Sp = 0.99 ± 0.03
J. Eckstein et al. (2016) <sup>92</sup>	80	40 with AF 40 Non-AF	80 ± 8 75 ± 7	5 min	Video camera of smartphone	In- and outpatient checkpoint	RR times series features RMSSD and SD1/SD2 index extracted from the Poincare plot	Derived threshold values of features for best ROC	AUC = 0.931, Sen = 0.950, Sp = 0.950
D. McManus et al. (2016) <sup>93</sup>	121	98 with AF 15 with PAC 15 with PVC	66	2 min	Video camera of smartphone	Inpatient	RR times series features RMSSD, Shannon Entropy, Poincare plot	Derived threshold values of features for best ROC	AF: Acc = 0.951, Sen = 0.970, Sp = 0.935, PAC: Acc = 0.955, Sen = 0.667, Sp = 0.980, PVC: Acc = 0.960, Sen = 0.733, Sp = 0.976
Shashikumar et al. (2017) <sup>51</sup>	98	45 with AF, 53 with other rhythms (ARR)	–	30 s	Wrist-worn device Samsung Simband	Inpatient	PPG image spectral representation of wavelet transform—features obtained from the CNIN + RR times series features (sample entropy, standard deviation, robust version of the standard deviation, min and the max of the sample entropy features)	Elastic net logistic model	Acc = 0.918, AUC = 0.95
T. Conroy et al. (2017) <sup>94</sup>	77 Test: 34	44 healthy subjects, 33 with AF, 13 healthy subjects, 21 with AF	38 ± 12, 64 ± 11, 45 ± 17, 68 ± 11	5 min	Single earlobe PPG sensor	Inpatient	RR times series features. Coefficient of variation, standard deviation, average of	Derived threshold values of features for best ROC	Acc = 0.952, Sen = 0.909, Sp = 0.909

Table 1 continued

Author (year) [ref.]	Number of patients	Dataset features	Age of population	Length PPG segments	Measurement device	Acquisition conditions	Input data	Methodology	Performance results for rhythms detection
Tang et al. (2017) <sup>48</sup>	666 stroke patients	150 with AF, 516 Non-AF	74.5 ± 12.8, 66.3 ± 14.8	1 min, 2 min, 10 min	Bedside monitor	Inpatient	the difference in beat-to-beat: pNN35 RR times series features. Time domain: mean, standard deviation, and RMSSD. Frequency domain: low-frequency range (LF), power in the high-frequency range (HF), and the ratio of LF and HF. Nonlinear analytical methods: Shannon entropy, and turning point ratio	Logistic regression analysis	1-min: AUC = 0.949, 2-min: AUC = 0.972, 10-min: AUC = 0.973
Bashar et al. (2018) <sup>64</sup>	200	-	Older than 45 years old	30 s	Video camera of smartphone	Outpatient—checkpoint	Noise/movement detection: Variable frequency complex demodulation. AF detection: RR times series features (RMSSD, Shannon entropy and sample entropy)	Noise/movement detection: threshold. AF detection: Support vector machines	Acc = 0.9116
Chong et al. (2018) <sup>65</sup>	99	88 patients with AF prior and after cardioversion 11 health subjects	-	2 min	Video camera of smartphone	Outpatient—checkpoint	Noise/movement detection: Signal slope changes, turning point ratio changes, and kurtosis change. AF detection: RMSSD and Shannon Entropy (ShE)	Noise/movement detection: threshold. AF detection: thresholds	Acc = 0.9667, Sen = 0.9765, Sp = 0.9714
Tarniceriu et al. (2018) <sup>49</sup>	29	15 NSR, 14 with AF	67.5 ± 10.7, 74.8 ± 8.3	20 consecutive RR	Wrist-worn device PulseOn Ltd.	Inpatient	RR times series features	Markov model	Sen = 0.9845, Sp = 0.9913
H.M. de Morree et al. (2018) <sup>74</sup>	27	8 AF 19 non-AF	69 ± 101, 67 ± 13	120 s	Wrist-worn device Philips Cardio	Outpatient—continuous measurement	RR times series features Shannon entropy, RMSSD, normalized RMSSD, pNN40, pNN70, sample entropy, and coefficient of sample entropy	Derived threshold values of features for best ROC	Acc = 0.981, Sen = 0.984, Sp = 0.980

RR R to R interval, NSR normal sinus rhythm, ARR other arrhythmias, VA ventricular arrhythmias, AUC area under the curve, Acc accuracy, Sen sensitivity, Sp specificity, PVC premature ventricular contractions, PAC premature atrial contraction, pNN35/pNN40/pNN70 percentage of differences of successive RR that exceeded 35 or 40 or 70 ms by the total number of RR intervals

**Table 2.** Studies on photoplethysmography based AF detection using ML approaches.

Author (year) [ref.]	Number of patients	Dataset features	Age of population	Length PPG segments	Measurement device	Acquisition conditions	Input data	Methodology	Performance results for rhythms detection
Shan et al. (2016) <sup>56</sup>	468 stroke patients	-	-	2 min	Bedside monitor	Inpatient	RR times series features. Mean, median, standard deviation, RMSSD, power in very low-frequency range, low frequency (LF), high frequency (HF), ratio of power in LF and HF (LF/HF), multi-scale entropy, Shannon entropy, Turning point ratio	Support vector machines	AUC = 0.971, Sen = 0.942, Acc = 0.957
M. Lemay et al. (2016) <sup>57</sup>	20	-	-	10 s	PPG wrist-based device	Inpatient	RR times series features mean, minimum, median, and interquartile range of RR	Support vector machine	Acc = 0.9385
V. Corino et al. (2017) <sup>58</sup>	70	30 with AF, 9 ARR, 31 NSR	40 ± 17, 76 ± 9, 65 ± 15	2 min	Empatica E4 wristband	Inpatient	RR times series features 24 features: spectral analysis, variability and irregularity analysis, PPG-waveform features	KNN classifier	NSR: Sen = 0.773, Sp = 0.928. AF: Sen = 0.754, Sp = 0.963. ARR: Sen = 0.758, Sp = 0.768
T. Schack et al. (2017) <sup>59</sup>	326	20 with AF, 294 of NSR, 12 of Noise	-	20 s	Video camera of smartphone	Inpatient	RR times series features Mean, median, standard deviation and the mean absolute deviation (MAD); RMSSD; normalized RMSSD; Shannon entropy. PPG-waveform features: mean, median, SD and MAD, crest time, peak rise height, fall height, waveform width, cross-correlation of consecutive pulse segments, very low frequency, low frequency, high frequency and quotients of these spectral powers	Support vector machines	Perfect detection of AF
S. Fallet et al. (2019) <sup>60</sup>	17	AF NSR ventricular arrhythmias (VA)	57 ± 13	10 s	Wrist-worn device	Inpatient	RR times series features: mean, standard deviation, minimum, maximum, RMSSD. PPG-waveform features: adaptive organization index, variance of the slope of the phase difference, permutation entropy, spectral entropy, fractional spectral radius, and spectral purity index).	Decision trees	AF vs NSR: Acc = 0.981, Sen = 0.997, Sp = 0.924. AF vs VA: Acc = 0.959, Sen = 0.981, Sp = 0.887. AF vs (NSR&VA): Acc = 0.950, Sen = 0.962, Sp = 0.928

RR R to R interval, NSR normal sinus rhythm, ARR other arrhythmias, VA ventricular arrhythmias, AUC area under the curve, Acc accuracy, Sen sensitivity, Sp specificity, PVC premature ventricular contractions, PAC premature atrial contraction

**Table 3.** Studies on photoplethysmography based AF detection using DL approaches.

Author (year) [Ref]	Number of patients	Dataset features	Age of population	Length PPG segments	Measurement device	Input data	Acquisition conditions	Methodology	Performance results for rhythms detection
Aliamiri and Shen(2018) <sup>75</sup>	19	—	—	30 s	Samsung gear device	PPG segment	—	Quality classification: CNN AF detection: Convolution- Recurrent Hybrid Model (CRNN)	Acc = 0.9819, AUC = 0.9967
Tison et al. (2018) <sup>72</sup>	Train: 9750. Test set 1: 51. Test set 2: 1617	Train: 347 with AF, 8216 no AF	Train: 42 ± 12. Test set 1: 66.1 ± 10.7	5 s	Wrist-worn device Apple watch	RR times series features	Test set 1: Inpatient Test. set 2: Outpatient checkpoint	Neural network of 8-layers	Test set 1: AUC = 0.97, Sen = 0.980, Sp = 0.902. Test set 1: AUC = 0.72, Sen = 0.677, Sp = 0.676
M. Poh et al. (2018) <sup>73</sup>	Train: 3373. Test: 1013	Train: Public databases (MIMIC-III critical care database + Vortal dataset from healthy volunteers + IEEE-TBME PPG Respiratory Rate Benchmark dataset)	Test group: 68.4 ± 12.2	17 s	Smartphone	PPG segment	Outpatient	CNN architecture with six dense blocks	Overall Acc = 0.961. Noise: Sen = 0.970, Sp = 1. NSR: Sen = 0.991, Sp = 0.982. ARR: Sen = 0.722, Sp = 0.988. AF: Sen = 0.976, Sp = 0.965
I. Gotlibovych et al. (2018) <sup>97</sup>	Train: 42. Test: 11	Train: 29 with AF 13 NSR. Test: 7 with AF 4 NSR	37–85 years old	—	Wrist-worn prototype fitness tracker device	PPG segment	Inpatient + Outpatient NSR: asleep continuous measurement	Convolutional-recurrent neural network	AUC = 0.999, Sen = 0.999, Sp = 0.998
M. Voisin et al. (2019) <sup>81</sup>	81	Train + validation: 19 AF/32 NSR. Test: 10 AF/20 NSR	—	30 s	Wrist-worn device Samsung Simband	PPG segment	Outpatient—continuous measurement	1D ResNet	AUC = 0.949
S. Shashikumar et al. (2018) <sup>32</sup>	Train: 2850. Test: 97	Test: 44 AF and 53 ARR	Train: 47 ± 25 years. Test: 18–89 years old	30 s	Wrist-worn device Samsung Simband	ECG image-based-training. PPG image-based - test	Outpatient—checkpoint	Bidirectional Recurrent Neural Network (BRNN) with transfer learning	AUC = 0.97, AUC <sub>Cpr</sub> = 0.97, Sp = 1.0, Acc = 0.95
S. Kwon et al. (2019) <sup>96</sup>	75	—	63 ± 7.8	30 s	PPG fingertip	PPG segment	Inpatient	1-Dimensional convolutional neural network. Recurrent neural network	Acc = 0.9758, AUC = 0.998. Acc = 0.9715, AUC = 0.996

RR R to R interval, NSR normal sinus rhythm, ARR other arrhythmias, VA ventricular arrhythmias, AUC area under the curve, Acc accuracy, Sen sensitivity, Sp specificity, PVC premature ventricular contractions, PAC premature atrial contraction, AUC<sub>Cpr</sub> area under the precision-recall curve



measure of how well AF cases ranked higher than Non-AF cases. Since AF has a low prevalence it is generally required that PPG-based AF detectors show high precision (positive predictive value). Rather than reporting the AUC, the area under the precision-recall curve (AUPRC) is an alternative metric suitable for highly imbalanced data (i.e. low prevalence).<sup>40</sup> In general, any reported performance metric should take into account the low prevalence of AF and be evaluated on an independent test dataset.

#### AF detection studies

Studies were split into three groups based on the approaches undertaken to build an AF detector: traditional statistical analysis, machine learning (ML), and deep learning (DL) methods. In traditional statistical analysis, statistical metrics are derived from PPG signals, and classification thresholds were estimated to distinguish between AF and Non-AF segments. ML techniques call for the extraction of pre-selected features, a process that can be quite manual, labor-intensive, and can usually benefit from incorporating complex physiological knowledge. An ML classifier is then built upon extracted features from training data samples. DL approaches require less manual feature engineering than conventional ML since DL incorporates automatic features representation process of input data. Recently, there was a significant focus on DL methods driven notably by technological advancement in computational power and the acclaimed success in computer vision applications.<sup>41,42</sup>

#### Statistical analysis approaches

Statistical models for AF detection are built using the thresholds for a set of features extracted from the RR-interval time series of well-annotated and publicly available ECG databases, such as MIT-BIH atrial fibrillation, MIT-BIH normal sinus rhythm, or MIT-BIH arrhythmia database.<sup>43–45</sup> Specifically, features were first extracted from the RR-interval time series of pre-annotated ECG waveforms. The histograms of each feature were analyzed respectively with or without the presence of AF and other cardiac rhythms in order to define the threshold that best separates the rhythm classes. These thresholds were then applied to the same RR time series-based features extracted from PPG signals.<sup>46,47</sup> Other arrhythmias (i.e., premature ventricular contractions, and premature atrial contraction) could also be detected similarly in a sequence of binary classifications.<sup>36</sup>

Other statistical approaches can also be applied to classify between AF and Non-AF such as logistic regression.<sup>48</sup> Logistic regression models use the logistic function, instead of a straight line or a hyperplane, to fit output the probability between 0 and 1 (corresponding to Non-AF and AF). Markov model is another statistical tool that could be used for AF detection. RR-interval time series features are used in this model to define the distributions that best fit the data, and the probability for various rhythms can be drawn from these distributions.<sup>44,49,50</sup> Elastic net is a regularization method for regression and classification models. Elastic net performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the logistic regression model. Regularization approaches were successfully applied to improve the performance of AF detection.<sup>51,52</sup>

Table 1 summarizes a selection of PPG-based AF detection studies which used statistical models. Different study aspects are shown to depict the patient population and datasets used, the features and methods, the context (inpatient vs outpatient), and the performance results.

**Machine learning approaches.** ML has been used for AF detection with interesting results. ML techniques require extensive domain expertise to design features suitable for a comprehensive representation of PPG waveforms and the detection of class-

differentiating patterns. Features commonly extracted from PPG time series are morphological descriptors, time domain statistics, frequency domain statistics, nonlinear measures, wavelet based measures, and cross-correlation measures.<sup>53–60</sup>

There were generally three main ML approaches used in the reviewed studies: k-nearest neighbors (KNN), support vector machine (SVM), and decision trees (DT). KNN classification is a relatively simple clustering technique where a sample is classified by a plurality vote of its neighbors and assigned to the class based on the most common class among its  $k$  closest neighbors.<sup>61</sup>

SVM finds a hyperplane that separates two classes with a high margin that maximizes the distances between nearest data points from each class. SVMs prove to be successful in nonlinear classification problems by mapping non-separable features into a higher dimensional space, a procedure known as the kernel trick which uses kernel functions such as Radial Basis Function (RBF) or polynomial.<sup>62</sup>

In DT approaches, the training set is continuously split according to a chosen feature. A feature tree can be explained by two entities, namely decision nodes and leaves. The leaves are the decisions or final outcomes. And the decision nodes are where the data is split.<sup>63</sup> The objective is to find in each decision node of the tree, the best attribute allowing to diminish as much as possible the overlapping of classes. The classification starts from the root, and it evaluates the relative attribute and it takes the branch corresponding to the outcome. This process is repeated until a leaf is encountered and a sample is assigned the class labeling the leaf.

Some studies used a combination of threshold-based and ML approaches. For example, the thresholds of some features were first used to exclude poor pulses, then an ML model was built for the detection of AF in the clean pulses.<sup>64,65</sup> Table 2 is a chronological summary of the selected ML studies and the reported performance results. All the studies reported in the Table 2 were based in short length of PPG segment, with maxima of 2 min.

**Deep learning approaches.** DL has recently emerged as a powerful method for the detection of abnormalities in physiological signals, encouraging applications of arrhythmia detection from ECG and PPG, including AF detection. Unlike ML, deep learning models automatically learn feature representations, sparing the tedious task of feature crafting. DL uses a neural network, a set of interconnected layers of computational nodes. The most common DL approaches used for AF detection are based on Convolutional Neural Networks (CNN). CNN was applied in automatic feature extraction and in classification problems. Some studies used CNNs only for automatic feature extraction.<sup>51</sup> In one study, an aggregated model of two serially connected CNNs was proposed, where the former detects clean segments from which the latter identifies instances of AF.<sup>51</sup> Some DL models were trained with hybrid input data (i.e., time series and images) in order to capture a wide range of features spanning more than one domain.

Training a DL model from scratch requires a large amount of labeled training data and generally poses a major constraint in biomedical applications due to the limited amount of labeled data. A possible solution to overcome this limitation is transfer learning where the task is to fine-tune a sophisticated pre-trained DL.<sup>66</sup> The required number of layers and the complexity of fine-tuning depend on specific applications.<sup>67</sup> In ref. <sup>32</sup> a pre-trained ECG-based CNN model was used to detect AF from PPG segments by fine-tuning the network using a small set of labeled PPG segments.

Table 3 is a chronological summary of the selected DL studies and the reported performance results. Notably, all studies were based on relatively short segments (less than 10 min), mainly due to the low yield of AF events in PPG data. Table 3 is a

chronological summary of the selected DL studies and the reported performance results. Notably, all studies were based on relatively short segments (<10 min), mainly due to the low yield of AF events in PPG data. Many of recent studies were developed and tested using ambulatory data (outpatients). Such data setting is the closest to real conditions where developed solutions would be applied.<sup>68</sup> Training and testing algorithms on real-world data is crucial to assess their true performance and evaluate their readiness for commercial use. For this reason, most of the studies in Table 3 are based on outpatient data.

Recently, large-scale AF screening studies have been performed by the leading tech companies, in order to test the performance of the most recent PPG-based smart devices.

In Huawei Heart study a continuous PPG was monitored in a cohort of 187,912 patients using a smart-device in ambulatory conditions. Participants were monitoring for at least 14 days with a wristband or wristwatch and a 60-s PPG signal was continuously measured at every 10 min.<sup>69</sup> Only individuals identified with “suspected AF” were assessed by using ECG. Results showed 87.0% of cases were AF with positive predictive value (PPV) of 91.6%.<sup>69</sup> The algorithm for AF detection was not described; however, morphology and frequency analysis of the pulse waveform was used to identify the AF events according to a previous publication.<sup>70</sup>

Apple Heart Study enrolled 419,093 participants in a prospective study.<sup>71</sup> PPG was monitored by Apple Watch, a subsequent ambulatory ECG patch was used for cases initially identified as AF using a proprietary PPG-based algorithm. Therefore, this study was not included in this review.

Both studies claim they demonstrate the ability of a smart devices to screen AF events. Due to the study design of both studies, they cannot assess the sensitivity of PPG-based AF detection, since only participants identified with AF events (irregular pulse notification) received an ECG monitoring system.

## MAIN CURRENT CHALLENGES

While the performance results summarized in Tables 1–3 suggest that PPG can be an alternative to ECG for AF detection, it remains that in real-world applications, PPG-based AF detection could be limited by a number of factors.

### Other cardiac arrhythmias

The presence of different cardiac rhythms within a recording poses a challenge for AF detection. For some of the works reviewed, control data only contain NSR.<sup>46,47,72</sup> Statistical methods for AF detection are usually limited to distinguishing between AF and NSR since other arrhythmias will likely present a distribution with mixed characteristics from both distributions of AF and NSR cases. ML and DL methods are generally more robust than statistical methods in distinguishing AF from a variety of rhythms. With DL expected to be on the rise for the next years, AF detection can benefit from new and more reliable DL algorithms. Such advancement will likely require very large training datasets of comprehensive cardiac rhythmic morphologies. New algorithms are expected to be better in discriminating AF and AF-mimic rhythms leading to high and perhaps clinical-grade performance levels.<sup>30</sup> New algorithms are also expected to demonstrate robustness to noise, inter- and intra-subject variabilities, and to other variables currently posing issues to existing methods.

### Corrupted signals

A significant limitation with PPG signals is motion artifacts. Artifacts in PPG signals can generate fluctuations and distortions that complicate the detection of cardiac components. Sources of motion artifact usually include movement of the PPG sensors either on the skin or loose contact with the skin, and deformation

of the illuminated tissue volume by dynamic variations in sensor contact-pressure. Specifically, for AF detection, these problems present a major challenge. PPG signals with AF and motion artifacts can both have similar characteristics of irregular pulse-to-pulse intervals. This situation can lead sinus rhythm signals corrupted by motion artifacts to be incorrectly detected as AF and vice versa.<sup>65,73</sup> The most common approach to deal with this issue is to simply discard the corrupted segments and use only the clean parts of PPG signals.<sup>49,50,58,60,64,74</sup> Some of the works followed a two-step approach: first, to identify motion artifacts by using accelerometer data, or by performing PPG signal quality assessment; and second, to perform AF detections with only good quality signals.<sup>51,75</sup> This often implies loss, and in some cases, a huge part of the signals acquired. One study shows that almost 40% of collected PPG signals were reported unreliable.<sup>50</sup> To overcome this limitation, one can improve the wearables in order to be more robust in detecting motion artifacts and to develop powerful methods to recover poor signals. Recent progress in flexible and stretchable sensors could help to enhance the SNR. Novel flexible sensors for transmission and reflection-mode pulse oximetry show a higher SNR due to a reduction in ambient noise.<sup>76</sup>

Algorithms such as the independent component analysis (ICA),<sup>77</sup> Kalman filtering, wavelet denoising,<sup>78</sup> and empirical mode decomposition<sup>79,80</sup> were proposed for removing artifacts in PPG signal; however, these techniques were mainly proposed for scenarios with weak noise.

Ideally, AF detection models can continuously and accurately detect AF episodes in PPG collected in an ambulatory setting, without discarding PPG segments and being robustness to motion artifacts.<sup>81</sup> Other DL techniques can be interesting options to solve artifact issues. For example, Generative Adversarial Networks (GAN) is a technique that combines two different neural networks together into a single pipeline—generative and discriminative neural networks. GANs have been used to recover information from images<sup>82</sup> and biomedical signals,<sup>83</sup> and can be an interesting option to recover the PPG signal.

### Data annotation

The methods used for AF classification presented in Tables 1–3 are based on supervised methods where data were labeled, using a ground truth (often based on simultaneously recorded ECG). Manual annotation is usually very time consuming, expensive, and laborious. Other annotation methods, such as the use of other devices to identify ground truth, are usually not reliable since the patient needs to use more devices, in a continuous and synchronized way and there are no “perfect” devices with perfect sensitivity and specificity in detecting AF. Active learning is a strategy that aims to ease the data labeling process by automatically deciding which instances should be labeled. Instead of annotating an entire dataset, active learning minimizes manual labeling effort, creating annotated datasets in an efficient way,<sup>84,85</sup>

Another disadvantage of the manual annotation is inherent inter-rater variability reflected in inter-rater disagreements coefficients (such as Kappa scores). Ideally, heuristic rules defining the annotation process are precise enough to warrant an ideal inter-rater agreement. However, in AF detection as in most physiological patterns, data complexity and pattern variability among and within patients render high inter-rater agreement difficult to reach. Such disagreement leads to a labeling noise that translates into a bias that the AF detection algorithm needs to deal with. Mislabeled training samples can potentially affect the performance of supervised classifiers.<sup>86</sup> Generally, relatively small proportions of noisy labels are tolerated in large training datasets but high proportions of noisy labels can severely degrade the performance of a classifier.<sup>87,88</sup> At the data level, mislabeling can be mitigated using voting from multiple annotators. At the

algorithm level, some DL models were shown to be resistant to relatively small levels of label noise.<sup>87</sup>

### Other challenges

DL algorithms have started to be applied in PPG-based AF detection and in general, showed superior performance. However, the computational resources required for low-latency real-time inference can be a huge engineering challenge to actually deploy the DL algorithm on wearable devices. Wearables have limited energy and constrains for battery size and heat dissipation. Due to these limitations, some of PPG devices were combined with computation-intensive applications on smartphones.<sup>89</sup>

In addition, the interpretability of black-box DL algorithms can be a challenge for doctors to perform diagnosis based on results from those algorithms. Many machine learning algorithms used in the clinical field are essentially black boxes, which make predictions without giving any clinical explanation. The clinical community points out that the accuracy achieved by the ML algorithm is not enough for accepting ML-enabled technology.<sup>90</sup> Explainable ML algorithms are emerging with interpretable models, which can give information about which aspect of input data contributes more to the final prediction. The most recent approaches for AF detection showed that the model seems to focus on systolic and diastolic peaks and slopes.<sup>81</sup>

A general limitation in the medical field and also in the studies of AF detection is the fact that models take into consideration a small group of patients from one medical center. The models should be developed using data from multiple medical centers in order to ensure that all the population heterogeneities were represented in the sample used for the study.

### CONCLUSIONS

A review of statistical and machine learning approaches applied to AF detection using PPG is presented. Although PPG has proven to be a good alternative to ECG for ambulatory real-time and continuous detection of AF, there are challenges remaining to be solved that are currently limiting the expansion of PPG-based AF detection beyond consumer wearables and its application in other clinical applications. Recent advances in computational power and the advent of powerful deep learning algorithms capable of solving complex pattern recognition problems have led to new AF detection methods proving a significant improvement in accuracy, robustness, and reliability compared to earlier approaches. Whether deep learning will ultimately lead to clinical-grade performance levels of PPG-based AF detection remains a question. Wearable devices powered by sophisticated algorithms offering precise and continuous AF detection will provide an excellent opportunity to screen AF at scale as demonstrated in the recent Apple and Huawei studies.

Received: 26 July 2019; Accepted: 22 November 2019;

Published online: 10 January 2020

### REFERENCES

- Kirchhof, P. et al. 2016 ESC guidelines for the management of atrial fibrillation developed in collaboration with EACTS. *Eur. Heart J.* **37**, 2893–2962 (2016).
- Lane, D. A. & Lip, G. Y. H. Stroke prevention in atrial fibrillation: can we do better?. *Nat. Rev. Cardiol.* **13**, 511–512 (2016).
- Zoni-Berisso, M., Lercari, F., Carazza, T. & Domenicucci, S. Epidemiology of atrial fibrillation: European perspective. *Clin. Epidemiol.* **6**, 213–220 (2014).
- Hollis, I. B., Reed, B. N., Rodgers, J. E., Cook, A. M. & Danelich, I. M. Clinical update on the management of atrial fibrillation. *Pharmacother. J. Hum. Pharmacol. Drug Ther.* **33**, 422–446 (2013).
- Singh, D. et al. Worldwide epidemiology of atrial fibrillation. *Circulation* **129**, 837–847 (2013).

- Rho, R. W. & Page, R. L. Asymptomatic atrial fibrillation. *Prog. Cardiovasc. Dis.* **48**, 79–87 (2005).
- Hart, R. G. et al. Independent predictors of stroke in patients with atrial fibrillation: The Stroke Risk in Atrial Fibrillation Working Group. *Neurology* **69**, 546–554 (2007). <https://doi.org/10.1212/01.wnl.0000267275.68538.8d>.
- Flint, A. C., Banki, N. M., Ren, X., Rao, V. A. & Go, A. S. Detection of paroxysmal atrial fibrillation by 30-day event monitoring in cryptogenic ischemic stroke. *Stroke* **43**, 2788–2790 (2012).
- Kernan, W. N. et al. Guidelines for the prevention of stroke in patients with stroke and transient ischemic attack. *Stroke* **45**, 2160–2236 (2014).
- Boriani, G. & Padeletti, L. Management of atrial fibrillation in bradyarrhythmias. *Nat. Rev. Cardiol.* **12**, 337–349 (2015).
- Trayanova, N. From genetics to smart watches: developments in precision cardiology. *Nat. Rev. Cardiol.* **16**, 72–73 (2019).
- Kavsaoğlu, A. R., Polat, K. & Hariharan, M. Non-invasive prediction of hemoglobin level using machine learning techniques with the PPG signal's characteristics features. *Appl. Soft Comput. J.* **37**, 983–991 (2015).
- Pereira, T. et al. Pulse pressure waveform estimation using distension profiling with contactless optical probe. *Med. Eng. Phys.* **36**, 1515–1520 (2014).
- Gadhoumi, K., Keenan, K., Colorado, R., Meisel, K. & Hu, X. A Statistical Comparative Study of Photoplethysmographic Signals in Wrist-Worn and Fingertip Pulse-Oximetry Devices. *Comput. Cardiol. Computing in Cardiology Conference (CinC)*. pp. 1–4 (Maastricht, Netherlands, 2018). <https://doi.org/10.22489/CinC.2018.316>.
- Njoum, H. & Kyriacou, P. A. Investigation of finger reflectance photoplethysmography in volunteers undergoing a local sympathetic stimulation. *J. Phys. Conf. Ser.* **450**, 012012 (2013).
- Nara, S., Kaur, M. & Lal Verma, K. Novel notch detection algorithm for detection of Dicrotic notch in PPG signals. *Int. J. Comput. Appl.* **86**, 36–39 (2014).
- Elgendi, M. et al. The use of photoplethysmography for assessing hypertension. *npj Digit. Med.* **2**, 60 (2019). <https://doi.org/10.1038/s41746-019-0136-7>.
- Nitzan, M., Romem, A. & Koppel, R. Pulse oximetry: fundamentals and technology update. *Med. Dev. Evid. Res.* **7**, 231–239 (2014).
- Moraes, J. L. et al. Advances in photoplethysmography signal analysis for biomedical applications. *Sensors (Switzerland)* **18**, 1–26 (2018).
- Bilgaiyan, A., Sugawara, R., Elsannah, F., Shim, C., Affiq, M. & Hattori, R. Optimizing performance of reflectance-based organic photoplethysmogram (PPG) sensor. *Proc. SPIE 10738, Organic and Hybrid Sensors and Bioelectronics XI*. **1073808**, 7 (2018). <https://doi.org/10.1117/12.2321060>.
- Ernst, G. Heart-rate variability—more than heart beats? *Front. Public Heal.* **5**, 240 (2017). <https://doi.org/10.3389/fpubh.2017.00240>.
- Pereira, T., Almeida, P. R., Cunha, J. P. S. & Aguiar, A. Heart rate variability metrics for fine-grained stress level assessment. *Comput. Methods Programs Biomed.* **148**, 71–80 (2017).
- Pirhonen, M., Peltokangas, M. & Vehkaoja, A. Acquiring respiration rate from photoplethysmographic signal by recursive Bayesian tracking of intrinsic modes in time-frequency spectra. *Sensors* **18**, 1693 (2018).
- Pereira, T., Correia, C. & Cardoso, J. Novel methods for pulse wave velocity measurement. *J. Med. Biol. Eng.* **35**, 555–565 (2015).
- Pereira, T. et al. Characterization of optical system for hemodynamic multi-parameter assessment. *Cardiovasc. Eng. Technol.* **4**, 87–97 (2013).
- Tamura, T., Maeda, Y., Sekine, M. & Yoshida, M. Wearable photoplethysmographic sensors—past and present. *Electronics* **3**, 282–302 (2014).
- Ghamari, M. A review on wearable photoplethysmography sensors and their potential future applications in health care. *Int. J. Biosens. Bioelectron.* **4**, 195–202 (2018).
- Allen, J. Photoplethysmography and its application in clinical physiological measurement. *Physiol. Meas.* **28**, R1–R39 (2007).
- Slocum, J., Sahakian, A. & Swiryn, S. Diagnosis of atrial fibrillation from surface electrocardiograms based on computer-detected atrial activity. *J. Electrocardiol.* **25**, 1–8 (1992).
- Hannun, A. Y. et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat. Med.* **25**, 65–69 (2019). <https://doi.org/10.1038/s41591-018-0268-3>.
- Chon, K. H. & McManus, D. D. Detection of atrial fibrillation using a smartwatch. *Nat. Rev. Cardiol.* **15**, 657–658 (2018).
- Shashikumar, S. P., Shah, A. J., Clifford, G. D. & Nemati, S. Detection of paroxysmal atrial fibrillation using attention-based bidirectional recurrent neural networks. In *KDD '18 Proc. of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. pp. 715–723, Preprint at <http://arxiv.org/1805.09133> [q-bio.NC] (London, UK, 2018).
- Liang, Y., Chen, Z., Ward, R. & Elgendi, M. Photoplethysmography and deep learning: enhancing hypertension risk stratification. *Biosensors* **8**, 1–13 (2018).
- Cicone, A. & Wu, H. T. How nonlinear-type time-frequency analysis can help in sensing instantaneous heart rate and instantaneous respiratory rate from photoplethysmography in a reliable way. *Front. Physiol.* **8**, 1–17 (2017).

35. Christov, I. et al. Ranking of the most reliable beat morphology and heart rate variability features for the detection of atrial fibrillation in short single-lead ECG. *Physiol. Meas.* **39**, 094005 (2018).
36. Chong, J. W., Esa, N., McManus, D. D. & Chon, K. H. Arrhythmia discrimination using a smart phone. *IEEE J. Biomed. Heal. Inform.* **19**, 815–824 (2015).
37. Reiss, A., Schmidt, P., Indlekofer, I. & Van Laerhoven, K. PPG-based heart rate estimation with time-frequency spectra. In *UbiComp '18 Proc. of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. pp. 1283–1292 (Singapore, Singapore, 2018).
38. Remya, R., Selvakumar, J. & Anburajan, M. Evaluation of hypotension using wavelet and time frequency analysis of photoplethysmography (PPG) signal. *Int. J. Comput. Sci. Inform. Security (CIC 2016 Special Issue)* **14**, 57–61 (2016).
39. Baratloo, A., Hosseini, M., Negida, A. & Ashal, G. El Part 1: Simple definition and calculation of accuracy, sensitivity and specificity. *Emergency* **3**, 48–49 (2015).
40. Saito, T. & Rehmsmeier, M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE* **10**, e0118432 (2015). <https://doi.org/10.1371/journal.pone.0118432>.
41. Szegedy, C., Vanhoucke, V., Shlens, J. & Wojna, Z. Rethinking the inception architecture for computer vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2818–2826. <https://doi.org/10.1109/CVPR.2016.308>, Preprint at <http://arXiv.org/1512.00567v3> (Las Vegas, NV, 2016).
42. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 770–778 <https://doi.org/10.1109/CVPR.2016.90>, Preprint at <http://arXiv.org/1512.03385v1> (Las Vegas, NV, 2016).
43. Goldberger AL, S. H. et al. PhysioBank, PhysioToolkit, and PhysioNet components of a new research resource for complex physiologic signals. *Circulation* **101**, e215–e220 (2000).
44. Moody G. & Mark, R. A new method for detecting atrial fibrillation using R-R intervals. *Comput. Cardiol.* **10**, 227–230 (1983).
45. Moody, G. B. & Mark, R. G. The impact of the MIT-BIH arrhythmia database. *IEEE Eng. Med. Biol. Mag.* **20**, 45–50 (2001).
46. Lee, J., Reyes, B. A., McManus, D. D., Mathias, O. & Chon, K. H. Atrial fibrillation detection using an iPhone 4S. *IEEE Trans. Biomed. Eng.* **60**, 203–206 (2013).
47. McManus, D. D. et al. A novel application for the detection of an irregular pulse using an iPhone 4S in patients with atrial fibrillation. *Hear. Rhythm* **10**, 315–319 (2013).
48. Tang, S. C. et al. Identification of atrial fibrillation by quantitative analyses of fingertip photoplethysmogram. *Sci. Rep.* **7**, 1–7 (2017).
49. Tarniceriu, A. et al. The accuracy of atrial fibrillation detection from wrist photoplethysmography: a study on post-operative patients. *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS* **2018**, 1–4 (2018).
50. Bonomi, A. et al. Atrial fibrillation detection using photo-plethysmography and acceleration data at the wrist. *Comput. Cardiol. Conference (CinC)*, 277–280 (Vancouver, BC, 2016).
51. Shashikumar, S. P., Shah, A. J., Li, Q., Clifford, G. D. & Nemati, S. A deep learning approach to monitoring and detecting atrial fibrillation using wearable technology. In *2017 IEEE EMBS Int. Conf. Biomed. Heal. Informatics, BHI 2017*, pp. 141–144. <https://doi.org/10.1109/BHI.2017.7897225>, (Orlando, FL, 2017).
52. Nemati, S. et al. Monitoring and detecting atrial fibrillation using wearable technology. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3394–3397. <https://doi.org/10.1109/EMBC.2016.7591456>, (Orlando, FL, 2016).
53. Pereira, T., Paiva, J. S., Correia, C. & Cardoso, J. An automatic method for arterial pulse waveform recognition using KNN and SVM classifiers. *Med. Biol. Eng. Comput.* **54**, 1049–1059 (2016).
54. Paiva, J. S., Cardoso, J. & Pereira, T. Supervised learning methods for pathological arterial pulse wave differentiation: a SVM and neural networks approach. *Int. J. Med. Inform.* **109**, 30–38 (2018).
55. Pereira, T. et al. A supervised approach to robust photoplethysmography quality assessment. In *IEEE J. Biomed. Health Inform.* 1–9 (2019). <https://doi.org/10.1109/JBHI.2019.2909065>.
56. Shan, S. M. et al. Reliable PPG-based algorithm in atrial fibrillation detection. In *Proc. In 2016 IEEE Biomedical Circuits and Systems Conference, (BioCAS)*, 340–343. <https://doi.org/10.1109/BioCAS.2016.7833801>, (Shanghai, 2016).
57. Lemay, M. et al. Wrist: located optical device for atrial fibrillation screening: a clinical study on twenty patients. *Comput. Cardiol.* **43**, 3–6 (2016).
58. Corino, V. D. A. et al. Detection of atrial fibrillation episodes using a wristband device. *Physiol. Meas.* **38**, 787–799 (2017).
59. Schack, T., Harb, Y. S., Muma, M. & Zoubir, A. M. Computationally efficient algorithm for photoplethysmography-based atrial fibrillation detection using smartphones. In *2017 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 104–108. <https://doi.org/10.1109/EMBC.2017.8036773>, (Seogwipo, 2017).
60. Fallet, S. et al. Can one detect atrial fibrillation using a wrist-type photoplethysmographic device? *Med. Biol. Eng. Comput.* **57**, 477 (2019). <https://doi.org/10.1007/s11517-018-1886-0>.
61. Md Isa, N. E., Amir, A., Ilyas, M. Z. & Razali, M. S. The performance analysis of K-nearest neighbors (K-NN) algorithm for motor imagery classification based on EEG signal. The International Conference on Emerging Electronic Solutions for IoT. *MATEC Web Conf. Vol. 140*, 01024 (Batu Feringghi, Malaysia, 2017).
62. Nalepa, J. & Kawulok, M. Selecting training sets for support vector machines: a review. *Artif. Intell. Rev.* **52**, 857–900 (2019). <https://doi.org/10.1007/s10462-017-9611-1>.
63. Yan-yan, S. O. N. G. & Ying, L. U. Decision tree methods: applications for classification and prediction. *Shanghai Arch. Psychiatry* **27**, 130–135 (2015).
64. Bashar, S. K. et al. Developing a novel noise artifact detection algorithm for smartphone PPG signals: preliminary results. In *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)* pp. 79–82. <https://doi.org/10.1109/BHI.2018.8333374>, (2018).
65. Chong, J. W. et al. Motion and noise artifact-resilient atrial fibrillation detection using a smartphone. In *IEEE J. Emerg. Sel. Top. Circuits Syst.* **8**, pp. 2156–3357. <https://doi.org/10.1109/JETCAS.2018.2818185>, (2018).
66. Xiao, R. et al. Monitoring significant ST changes through deep learning. *J. Electrocardiol.* **51**, S78–S82 (2018).
67. Tajbakhsh, N. et al. Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans. Med. Imaging* **35**, 1299–1312 (2016).
68. Trull, T. J. & Ebner-Priemer, U. Ambulatory assessment. *Annu Rev. Clin. Psychol.* **9**, 151–176 (2013).
69. Guo, Y. et al. Mobile health technology for atrial fibrillation screening using photoplethysmography-based smart devices: The HUAWEI Heart study. *J. Am. Coll. Cardiol.* **74**, 2365–2375 (2019).
70. Fan, Y. Y. et al. Diagnostic performance of a smart device with photoplethysmography technology for atrial fibrillation detection: Pilot study (PremAFA II registry) JMIR Mhealth Uhealth. *J. Med. Internet Res.* **7**, e11437 (2019). <https://doi.org/10.2196/11437>.
71. Turakhia, M. P. et al. Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartwatch: The Apple Heart Study. *Am. Heart J.* **207**, 66–75 (2019).
72. Tison, G. H. et al. Passive detection of atrial fibrillation using a commercially available smartwatch. *JAMA Cardiol.* **3**, 409–416 (2018). <https://doi.org/10.1001/jamacardio.2018.0136>.
73. Poh, M. Z. et al. Diagnostic assessment of a deep learning system for detecting atrial fibrillation in pulse waveforms. *Heart* **104**, 1921–1928 (2018).
74. de Morree, H. M. et al. Comparison between electrocardiogram- and photoplethysmogram-derived features for atrial fibrillation detection in free-living conditions. *Physiol. Meas.* **39**, 084001 (2018).
75. Aliamiri, A. & Shen, Y. Deep learning based atrial fibrillation detection using wearable photoplethysmography sensor. In *2018 IEEE EMBS Int. Conf. Biomed. Heal. Inform., BHI*, pp. 442–445. <https://doi.org/10.1109/BHI.2018.8333463>, (Las Vegas, NV, 2018).
76. Lochner, C. M. et al. A flexible organic reflectance oximeter array. *Proc. Natl Acad. Sci. USA* **115**, E11015–E11024 (2018).
77. Kim, B. S. & Yoo, S. K. Motion artifact reduction in photoplethysmography using independent component analysis. *IEEE Trans. Biomed. Eng.* **53**, 566–568 (2006).
78. Raghuram, M. et al. Dual-tree complex wavelet transform for motion artifact reduction of PPG signals. In *2012 IEEE International Symposium on Medical Measurements and Applications*, pp. 1–4. <https://doi.org/10.1109/MeMeA.2012.6226643>, (Budapest, 2012).
79. Sadrawi, M. et al. Ensemble empirical mode decomposition applied for PPG motion artifact. In *IECBES 2016—IEEE-EMBS Conf. Biomed. Eng. Sci (IECBES)*. pp. 266–269. <https://doi.org/10.1109/IECBES.2016.7843455>, (Kuala Lumpur, 2016).
80. Li, S., Liu, L., Wu, J., Tang, B. & Li, D. Comparison and noise suppression of the transmitted and reflected photoplethysmography signals. *Biomed. Res. Int.* **2018**, 4523593 (2018).
81. Voisin, M. et al. Ambulatory atrial fibrillation monitoring using wearable photoplethysmography with deep learning. In *KDD '19 Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. pp. 1909–1916, Preprint at <http://arXiv.org/1811.07774> (Anchorage, AK, USA, 2019).
82. Tripathi, S., San, U. C., Zachary, D., Lipton, C. & Nguyen, T. Q. Correction by projection: denoising images with generative adversarial networks. Preprint at <http://arXiv.org/1803.04477v1> (2018).
83. Gandhi, S., Oates, T., Mohsenin, T. & Hairston, D. In Phung, D. & Tseng, V. S. (eds) *Advances in Knowledge Discovery and Data Mining* Vol. 0939, 285–296 (Springer International Publishing, 2018).
84. Bressan, R. S., Camargo, G., Bugatti, P. H. & Saito, P. T. M. Exploring active learning based on representativeness and uncertainty for biomedical data classification. *IEEE J. Biomed. Health Inform.* **XX**, 1–7 (2018).

85. Yang, L., Zhang, Y., Chen, J., Zhang, S. & Chen, D. Z. Suggestive annotation: a deep active learning framework for biomedical image segmentation. *Lecture Notes on Computer Science (including Subseries Lecture Notes on Artificial Intelligence, Lecture Notes on Bioinformatics)* 10435 (no. 1), 399–407 (MICCAI, 2017).
86. Goldgof, D. B. et al. Active cleaning of label noise. *Pattern Recognit.* **51**, 463–480 (2015).
87. Rolnick, D., Veit, A., Belongie, S. & Shavit, N. Deep learning is robust to massive label noise. Preprint at <http://arXiv.org/1705.10694> [cs.LG] (2017).
88. Hendrycks, D., Mazeika, M., Wilson, D. & Gimpel, K. Using trusted data to train deep networks on labels corrupted by severe noise. In *NIPS'18 Proceedings of the 32nd International Conference on Neural Information Processing Systems*. pp. 10477–10486, Preprint at <http://arXiv.org/1802.05300v4>, no. Nips (Montréal, Canada, 2018).
89. Xu, M., Qian, F. & Pushp, S. Enabling cooperative inference of deep learning on wearables and smartphones. Preprint at <http://arXiv.org/1712.03073>, 1–21 (2017).
90. The Lancet Respiratory Medicine. Opening the black box of machine learning *Lancet Respir. Med.* **6**, P801 (2018).
91. Lee, J., Nam, Y., McManus, D. D. & Chon, K. H. Time-varying coherence function for atrial fibrillation detection. *IEEE Trans. Biomed. Eng.* **60**, 2783–2793 (2013).
92. Eckstein, J. et al. Smart detection of atrial fibrillation. *Europace.* **19**, 753–757 (2016).
93. McManus, D. D. et al. PULSE-SMART: Pulse-Based Arrhythmia Discrimination Using a Novel Smartphone Application. *Journal of Cardiovascular Electrophysiology.* **27**, 51–57 (2016).
94. Conroy, T., Guzman J. H., Hall, B., Tsouri, G. & Couderc, J. P. Detection of atrial fibrillation using an earlobe photoplethysmographic sensor. *Physiol. Meas.* **38**, 1906–1918 (2017).
95. Schack, T., Harb, Y. S., Muma, M. & Zoubir, A. M. Computationally Efficient Algorithm for Photoplethysmography-Based Atrial Fibrillation Detection Using Smartphones. In *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 104–108 (2017).
96. Gotlibovych, I. et al. End-to-end deep learning from raw sensor data: Atrial fibrillation detection using wearables. Preprint at <https://arxiv.org/abs/1807.10707> (2018).
97. Kwon, S. et al. Deep Learning Approaches to Detect Atrial Fibrillation Using Photoplethysmographic Signals: Algorithms Development Study. *JMIR mHealth and uHealth.* **7**, e12770 (2019).

## ACKNOWLEDGEMENTS

This work is partially funded by NIH award R01NHLB128679.

## AUTHOR CONTRIBUTIONS

T.P. designed the manuscript, collected the studies, and wrote the first draft. N.T., K.G. and X.H. contributed to the structure of the manuscript, the discussion of the studies, and thoroughly revised the paper. M.M.P., D.H.D., R.J.L., R.C. and K.M. contributed with clinical insights. All authors reviewed the manuscript.

## COMPETING INTERESTS

The authors declare the following competing interests: R.J.L.: non-financial competing interest as an equity holder in Life Signal, Inc. and SentreHEART/Atricare, Inc. and co-founder of Indevor Corp. K.M.: non-financial competing interest as local investigator for Medtronic funded STROKE-AF study. X.H.: non-financial competing interest as a co-founder of Indevor Corp. Other authors declare no competing interests.

## ADDITIONAL INFORMATION

**Correspondence** and requests for materials should be addressed to T.P.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

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



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

© The Author(s) 2020