A new smart wristband equipped with an artificial intelligence algorithm to detect atrial fibrillation

Erdong Chen, MD,* Jie Jiang, MD,* Rui Su, MM,† Meng Gao, MB,* Sainan Zhu, MD,‡ Jing Zhou, MD,* Yong Huo, MD*

*From the *Department of Cardiology, Peking University First Hospital, Beijing, China, †Huami (Beijing) Information Technology Co. Ltd, Beijing, China, and ‡Department of Statistics, Peking University First Hospital, Beijing, China.*

BACKGROUND Detection of atrial fibrillation (AF) occurrence over a long duration has been a challenge in the screening and follow-up of AF patients. Wearable devices may be an ideal solution.

OBJECTIVE The purpose of this study was to measure the sensitivity, specificity, and accuracy of a recently developed smart wristband device that is equipped with both photoplethysmographic (PPG) and single-channel electrocardiogram (ECG) systems and an AF-identifying, artificial intelligence (AI) algorithm, used in the short term.

METHODS Use of the Amazfit Health Band 1S, which records both PPG and single-channel ECG data, was assessed in 401 patients (251 normal individuals and 150 ECG-diagnosed AF patients).

RESULTS ECG and PPG readings could not be judged in 15 and 18 subjects, respectively. Subjects who were unable to be judged were defined as either false negative or false positive. The sensitivity, specificity, and accuracy of wristband PPG readings were 88.00%, 96.41%, and 93.27%, respectively, and those of wristband ECG readings were 87.33%, 99.20%, and 94.76%, respectively. When the original wristband ECG records were judged by physicians, the sensitivity, specificity, and accuracy were 96.67%, 98.01%, and 97.51%, respectively.

CONCLUSION This promising new combination of PPG, ECG, and AI algorithm has the potential to facilitate AF detection.

KEYWORDS Artificial intelligence; Atrial fibrillation; Electrocardiogram; Photoplethysmography; Wearable electronic device

(Heart Rhythm 2020;17:847–853)

Introduction

Atrial fibrillation (AF) is one of the most common chronic diseases, with an estimated total incidence of about 3% in adults older than 20 years. AF increases the incidence of heart failure, stroke, and dementia.[1] The diagnosis of AF depends on electrocardiographic (ECG) records. However, detection rates of AF based on noninvasive, standard 12-lead ECG and 24-hour ambulatory ECG recordings are much lower than those recorded by cardiovascular implantable electronic devices.[2] The ideal recording method should be noninvasive, continuous, affordable, and comfortable.

Wristbands have recently become a widely used health management tool, with large numbers of people wearing them on a daily basis. As such, wristbands equipped with photoplethysmographic (PPG) sensors would provide a low-power, convenient method for screening for heart rhythm abnormalities. Wristband wearable devices combined with an artificial intelligence (AI) algorithm may be an efficient method for detecting AF.[3,4] However, PPG is not able to identify atrial activity (P wave) and cannot be used to diagnose AF; confirmation by ECG is still required. A piece of equipment that incorporates PPG monitoring with ECG recordings to confirm AF would facilitate detection of the condition in daily life.[5]

The objective of this study was to evaluate the sensitivity, specificity, and accuracy of a recently developed smart wristband that is equipped with both PPG and ECG systems and uses an AF-identifying AI algorithm. We hypothesized that this new combination of features would provide high levels of sensitivity, specificity, and accuracy compared to those provided by 12-lead ECG.

Methods

Hardware and algorithm

The Amazfit Health Band 1S (Huami Technology, Anhui, China) is a wearable wristband device that combines a single-channel ECG recorder with a high-precision PPG optical sensor. The wristband synchronizes either periodically

This work was sponsored by Anhui Huami Information Technology Co Ltd, where Ms. Rui Su served as Medical Director. All other authors have reported that they have no conflicts relevant to the contents of this paper to disclose. **Address reprint requests and correspondence:** Dr Jie Jiang, Department of Cardiology, Peking University First Hospital, Beijing 100034, China. E-mail address: jiangjie417@vip.163.com.

1547-5271/

https://doi.org/10.1016/j.hrthm.2020.01.034

or actively with a smartphone application via a Bluetooth connection. The PPG signals are collected automatically by the photoelectric sensor on the inner side of the wristband, at a sampling frequency of 50 Hz. The ECG signals are collected via metal sensors on the outer and inner sides of the wristband for left- or right-handed individuals, respectively (Figure 1). The choice of whether to wear the wristband on the individual's left or right hand is preset in the application. The sampling frequency for ECG was 250 Hz. When the user touches the outer metal area with the opposite hand, the wristband automatically initiates ECG recording. A single-lead ECG is recorded for 60 seconds each time the recording is initiated. For example, when the user wears the wristband on the left hand and the right hand is on the outer side of the ring, this is equivalent to limb lead I (hereafter referred to as "wristband ECG").

The RealBeats Artificial Intelligence Biological Data Engine (Huami Technology) was developed using a deep convolution neural network (SEResNet), trained by 21,618 tagged ECGs (4734 of which were AF cases) and a test set of 8518 tagged sources of ECG data (241 of which were AF cases). The sensitivity and specificity of the test set were 93.36% and 99.75%, respectively. Training and testing of the AI algorithm are discussed in Supplemental Material 1.

The PPG signal was acquired for 71 seconds. If the built-in algorithm detected a suspected AF, the 71-second PPG signal was repeated to determine whether the AF would be identified again. The final output of the algorithm was "AF" if 2 consecutive tests within 3 minutes were judged as "AF." If the first or repeated determinations were "not AF" rhythms, the output of the algorithm was "not AF." If signal quality was too poor or requirements to make a determination were not met, the output was "unable to judge." The peak-to-peak interval data were transmitted to the smartphone application via Bluetooth, providing original data. The wristband recorded ECG signals for 60 seconds after the device was triggered by touching of its outer electrode. If contact with the electrode was disconnected during the acquisition process, the recording stopped and was restarted. After notch filtering and moving-average baseline filtering, the ECG waveform signal data were transmitted to the smartphone application via Bluetooth. ECG data were transmitted to an Internet server using a smartphone for AI algorithm-mediated determinations.

Study population

This study was conducted among inpatients and outpatients of the Cardiovascular Department, Peking University First Hospital. All patients were >18 years old and had a stable heart rhythm at the time of the study.

Exclusion criteria included situations in which both upper limbs prevented effective use of the wristband, such as (1) bilateral upper limb disabilities; (2) wrist color abnormalities; (3) severe occlusive vascular disease of the upper limbs; and (4) significant edema of the upper limbs. Patients with an implanted pulse generator also were excluded.

The study was approved by the institutional ethics committee of Peking University First Hospital. Written informed consent was obtained from all patients. The study has been registered in the China Clinical Trials Registry (chictr.org.cn, identifier: ChiCTR1900024808).

Workflow

Patients were screened and ECG/PPG data were collected from either hospital wards or clinics. After each patient signed the informed consent form, a standard 12-lead ECG was recorded with the subject in the supine position. Diagnosis of the 12-lead ECG was confirmed by a senior ECG physician. Patients designated as "AF" by 12-lead ECG were selected to be part of the AF-positive group, and patients who were AF-negative were included as members of the control group.

Patients were seated in a quiet environment, and PPG signals were recorded by the wristband for 3 minutes. After the PPG signals were acquired and the automatic judgment was completed, patients actively triggered the wristband to record an ECG for 60 seconds. PPG records were evaluated using the AI algorithm in the wristband. ECG records were transmitted to a smartphone and evaluated using an AI algorithm on an Internet server. After the patient's information was removed from the original wristband ECG records, 2 ECG physicians evaluated the records separately. Subjects with an inconsistent diagnosis were judged blindly by an

Figure 1 Wristband used in the study and the mechanism by which electrocardiographic recordings can be manually initiated.

electrophysiological specialist. The 2 consistent conclusions were used to confirm whether the subject's diagnosis should be "AF" or "not AF."

Sample size and statistical analysis

The sample size required for determination of PPG function was estimated. The estimation was performed considering a specificity target value of 95%, an estimated sensitivity with respect to PPG function of 99%, unilateral $\alpha = 0.025$, and power of 80%. As a result, 145 patients were estimated to be required to assess PPG in patients with AF. Because an estimated 40% of patients were expected to present with AF, the total number of cases for consideration would be 363.

After the 2 sample size assessments were independently evaluated, we determined that, considering a 10% dropout rate, a total of 400 patients would be required. We planned to enroll 150 patients presenting with AF as the case group and 250 patients with a heart rhythm other than AF as the control group.

Automatic diagnosis of AF by PPG and wristband ECG may result in a determination of "unable to judge" if signal noise is too large. The number and percentage of cases determined to fall within the "unable to judge" category was determined. A 4-fold table was created, with diagnoses of "AF" or "not AF" using 12-lead ECG and diagnoses of "AF" or "not AF" using the wristband algorithm. Cases "unable to judge" were defined as "true positives and true negatives" or "false positives and false negatives," which were calculated and described separately. Because we assumed that the highly frequent and intermittent wristband samplings were close to continuous ECG monitoring, we regarded "false positives and false negatives" as worst case scenarios and used stricter criteria to produce a more convincing result.[6] The sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV) of the diagnosis of both PPG and EEG using the wristband algorithm were calculated. The Wald method was used to determine 95% confidence intervals (95% CIs). These calculations were performed twice for cases that were "unable to judge," which were then either included in or excluded from the analysis. The difference between wristband AI ECG judgments and human judgments was described as the matching rate when identifying AF, which is a means to evaluate the consistency of the determination of positive results.

The wristband is designed to use PPG for screening, and users with suspected AF are able to request recording of a single-lead ECG to confirm whether the episode is AF. We sought to analyze the PPG and ECG results as a complete procedure. Based on our design, the wristband reminds the user to perform ECG recording only when the PPG result is "AF," and only when the ECG result is "AF" will the user be informed that he or she may have "AF." The result of "unable to judge" does not lead to a reminder for ECG examination or notification of suspected "AF," so we adopted the use of "false positive and false negative" for statistical analysis. In this study, patients with a PPG result of "AF" were considered positive if the12-lead ECG result also was "AF," that is, a 12-lead ECG result that was also true was a true positive, and a 12-lead ECG result that was "not AF" was a false positive. Any 1 of the 2 steps with a "not AF" or "unable to judge" result from PPG and ECG was regarded as negative, but with a 12-lead ECG result of "AF" was a false negative.

Results

We enrolled 401 patients from May 16, 2019, to June 17, 2019 (150 cases and 251 controls). Details of the clinical characteristics are given in Supplemental Material 2. All subjects were evaluated using a 12-lead ECG and a wristband test that included both ECG and PPG. Among the subjects, 15 (3.74%) could not be evaluated via ECG using the wristband (13 in the case group and 2 in the control group). Eighteen subjects (4.49%) could not be evaluated via PPG using the wristband (11 in the case group and 7 in the control group).

Evaluation of wristband efficacy

Results of PPG and wristband ECG judgments and comparison with 12-lead ECG results are given in Table 1. For cases "unable to judge" defined as "true positives and true negatives," the sensitivity, specificity, accuracy, PPV, and NPV of wristband PPG are given in Table 2 ("Unable to judge" as "TPTN") and of wristband ECG in Table 3 ("Unable to judge" as "TPTN"). For cases "unable to judge" defined as "false positives or false negatives," the sensitivity, specificity, accuracy, PPV, and NPV of wristband PPG are given in Table 2 ("Unable to judge" as "FPFN") and of wristband ECG in Table 3 ("Unable to judge" as "FPFN"). The results using 95% confidence intervals (95% CIs) for PPG and ECG are also given in Tables 2 and 3, respectively. The sensitivity, specificity, accuracy, PPV, and NPV of the combination of wristband PPG and ECG with corresponding 95% CIs are given in Table 4.

Table 1 PPG and wristband ECG results compared with 12-lead ECG results

	12-Lead ECG		
	Not AF	AF	All
Wristband PPG			
Not AF	242 (60.35)	7 (1.75)	249 (62.09)
AF	2 (0.50)	132 (32.92)	134 (33.42)
Unable to judge	7 (1.75)	11 (2.74)	18 (4.49)
All	251 (62.59)	150 (37.41)	401 (100)
Wristband ECG			
Not AF	249 (62.09)	6 (1.50)	255 (63.59)
AF	0 (0)	131 (32.67)	131 (32.67)
Unable to judge	2 (0.50)	13 (3.24)	15 (3.74)
All	251 (62.59)	150 (37.41)	401 (100)

Values are given as n (%) of 401 total cases.

AF = atrial fibrillation; ECG = electrocardiography; PPG = photoplethysmography.

Table 2 Sensitivity, specificity, accuracy, PPV, and NPV of PPG results

	"Unable to judge" as "TPTN"	"Unable to judge" as "FPFN"
Sensitivity	95.33 (90.51–97.89)	88.00 (81.75–92.36)
Specificity	99.20 (96.95–99.97)	96.41 (93.23–98.20)
Accuracy	97.76 (95.75–98.88)	93.27 (90.35–95.37)
PPV	98.62 (94.60–99.76)	93.62 (87.87–96.85)
NPV	97.27 (94.21–98.80)	93.08 (89.10–95.73)

Values are given as % (95% CI). 95% CI values were calculated using the modified Wald method.

CI = confidence interval; FPFN = false positives and false negatives; NPV = negative predictive value; PPG = photoplethysmography; PPV = positive predictive value; TPTN = true positives and true negatives.

The original records of the wristband ECG with patient identity information removed subsequently were judged by ECG physicians and an electrophysiologist. Compared with the 12-lead ECG, the sensitivity, specificity, and accuracy of their assessments were 96.67%, 98.01%, and 97.51%, respectively.

Retrospective evaluations of cases

We retrospectively inspected some of cases to evaluate original PPG and ECG records. Case MZ030 was recorded in an outpatient environment. Twelve-lead ECG confirmed sinus rhythm. P, R, and T waves could be clearly recognized on wristband ECG. This finding suggested that the wristband has the ability to collect clear ECG signal even in an environment with complex electromagnetic signals (Figure 2).

We retrospectively inspected all cases with inconsistency or deemed "unable to judge." Because PPG signals were converted to peak intervals, only the regularity of peak intervals can be observed by retrospective inspection. The wristband ECG records retained more information, as discussed in the following cases.

In case MZ270, the wristband ECG correctly judged the case as "not AF" by AI, whereas physicians misjudged the case as "AF." In retrospect, upon evaluation of the wristband ECG, it was obvious that the R wave was irregular, and the P-wave display was ambiguous. Distinguishing clear P waves before R waves was difficult. The 12-lead ECG produced different P waves and occasionally also showed a long pause, suggesting frequent premature atrial complexes, some of which may be nonconducted. The recording shows that the AI algorithm can distinguish potential rhythm patterns, especially premature beats (Figure 3).

Table 3 Sensitivity, specificity, accuracy, PPV, and NPV of wristband ECG results

	"Unable to judge" as "TPTN"	"Unable to judge" as "FPFN"
Sensitivity	96.00 (91.36–98.34)	87.33 (80.98–91.82)
Specificity	100 (98.18–100)	99.20 (96.95–99.97)
Accuracy	98.50 (96.69–99.39)	94.76 (92.09–96.59)
PPV	100 (96.76–100)	98.50 (94.12–99.74)
NPV	97.67 (94.74–99.05)	92.91 (88.97–95.56)

Values are given as % (95% CI). 95% CI values were calculated using the modified Wald method.

Abbreviations as in Tables 1 and 2.

Table 4 Sensitivity, specificity, accuracy, PPV, and NPV of combination of wristband PPG and ECG results (correspond to FPFN)

	Combination results
Sensitivity	80.00 (72.52–85.90)
Specificity	96.81 (93.58–98.51)
Accuracy	90.52 (87.24–93.04)
PPV	93.75 (87.66–97.06)
NPV	89.01 (84.54–92.35)

Values are given as % (95% CI). 95% CI values were calculated using the modified Wald method.

Abbreviations as in Tables 1 and 2.

In case BF044, the 12-lead ECG indicated sinus rhythm. The wristband ECG was disturbed by the occurrence of environmental noise. Only the locations of R waves could be recognized dimly. This case showed the highest degree of signal disturbance in the study. Human judgment misinterpreted the case as "AF," whereas the AI diagnostic results of wristband ECG and PPG correctly identified the case as "not AF." These results show that the AI algorithm performs well even in locations with signal interference (Figure 4).

Discussion

Our study suggests the sensitivity, specificity, and accuracy of the wristband equipped with PPG, ECG and AI algorithm have a satisfactory performance in the short term.. Previously, the diagnosis of AF relied on ECG evidence. The recording of long-term, single-lead ECG is an effective screening method for paroxysmal AF, especially in asymptomatic AF and cryptogenic stroke patients.[1] The early diagnosis of asymptomatic AF before occurrence of the event remains challenging. In this regard, low-cost wearable electronic devices have great potential to facilitate the long-term screening and monitoring of a vast number of healthy people or those at risk for AF.

Use of portable ECG equipment to record intermittently, frequently, and regularly can theoretically approach the effect of continuously recorded ECG data.[7] In the Halland Hospital Halmstad single-center study[8] and the larger STROKESTOP study,[9] the feasibility and effectiveness of intermittent, self-manipulated ECG measurements were discussed. The STROKESTOP study determined that regular, intermittent ECG recorded twice daily can effectively identify AF in a high-risk population and can be used to suggest anticoagulant treatment strategies.

PPG technology is currently used to judge AF by evaluating the irregularity of pulse peak intervals. Because the optical signal sensors are imperceptible by the patient and have low power consumption, they are suitable for recording continuous or highly frequent and intermittent data. AI algorithms can be used to process PPG signals, and the

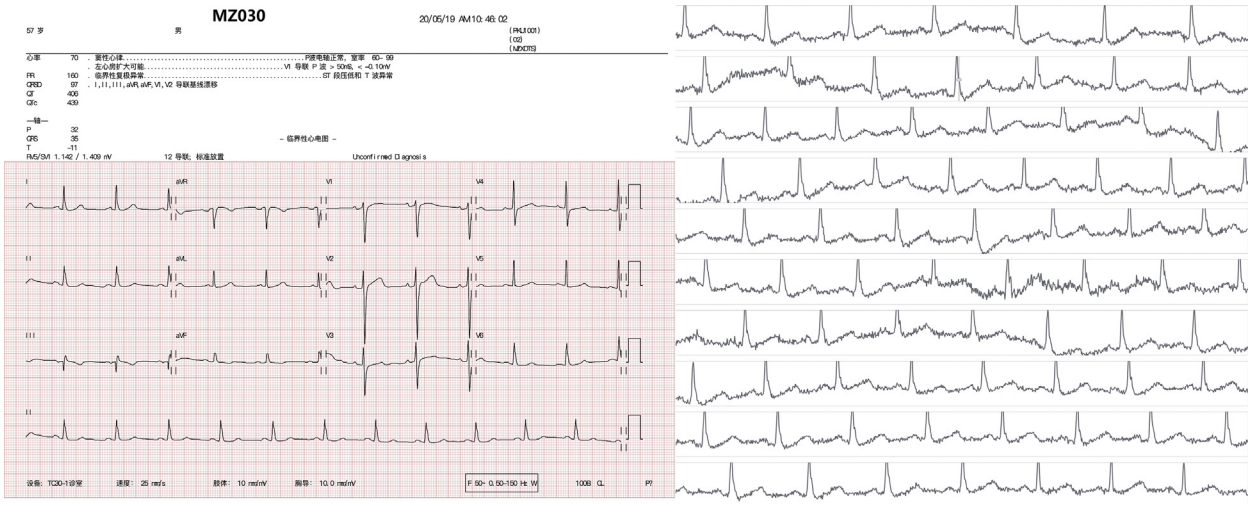

Figure 2 Case MZ030 was confirmed as sinus rhythm by 12-lead ECG. P, R, and T waves can be clearly recognized by analysis of wristband ECG data. ECG = electrocardiogram.

occurrence of AF can be detected with good sensitivity and specificity.[10]

PPG and ECG technologies can be combined to take advantage of the low power consumption of PPG technology, which is ideal for screening. Single-lead ECG provides higher specificity and clear ECG recordings for physicians to review.[11] However, the resulting volume of ECG data volume is gigantic, and the noise component is much higher than from standard 12-lead or patch ECG records. The combined health and economic benefits of these artificially read ECG records are poor.[12] Wearable devices that can be widely popularized are capable of processing large amounts of data daily at a low cost.[13] AI can help to achieve this goal. The wearable device assessed in this study used an AI algorithm to process PPG and ECG signals, partially achieved by automatic recognition of arrhythmia, and established a bridge between consumer health electronic products and clinical practice. The wristband tested in this study does not require frequent communication between the device and smartphone, thus reducing power consumption and increasing the time devoted to continuous recording of data.

The application of AI in the automatic identification of AF from PPG and ECG data has been discussed for many years. Several well-known wearable equipment manufacturers have attempted to make progress in this area, mainly using PPG data, but the results have not yet been satisfactory. In the Apple Heart Study, the coincidence rate of PPG and patch ECG in the diagnosis of AF was low, suggesting that PPG had poor specificity in the diagnosis of AF under the conditions tested.[14] In the HUAWEI Heart Study, the practice of screening AF with PPG seems to show even lower sensitivity.[15] These 2 companies both have recently developed a smartwatch with combined PPG and ECG. This combination strategy seems to provide better resolution than PPG records

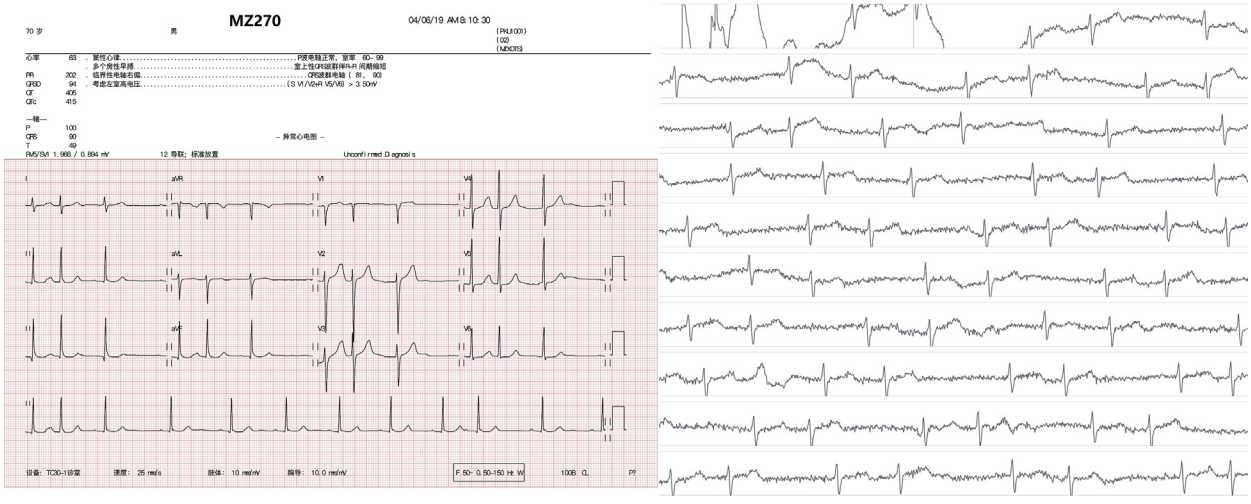

Figure 3 In case MZ270, the wristband ECG data were correctly judged as "not AF" using artificial intelligence, whereas physicians misjudged ECG data as "AF." On wristband ECG, the R wave was obviously irregular, and the P-wave display was ambiguous. The 12-lead ECG manifests different P waves and occasional long pauses, suggesting frequent premature atrial complexes, some of which may be nonconducted. AF = atrial fibrillation; ECG = electrocardiogram.

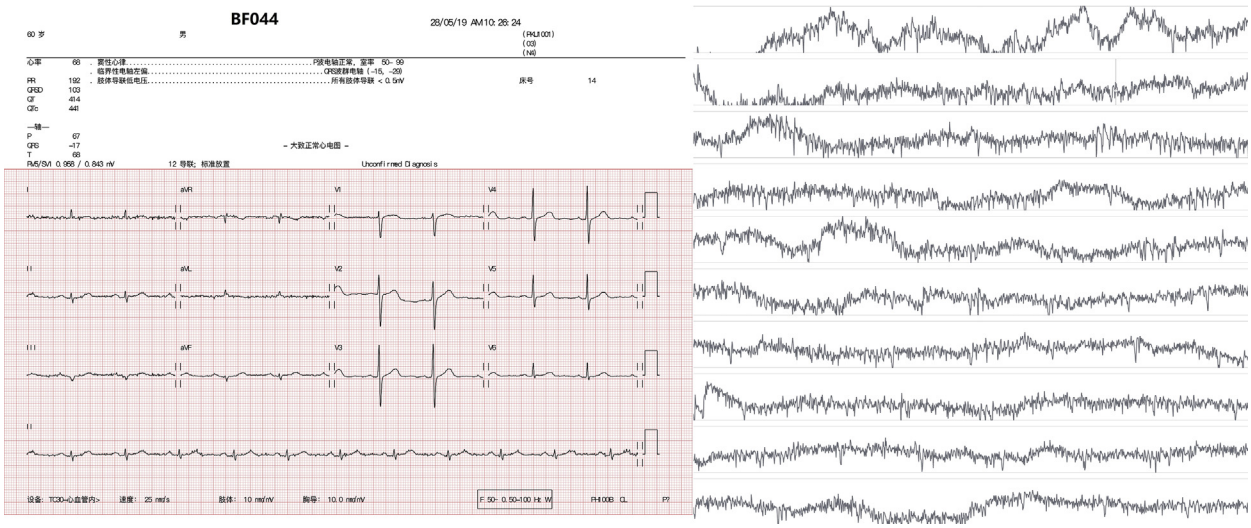

Figure 4 In case BF044, the wristband ECG was disturbed dramatically. The R wave could be recognized dimly. Human judgment incorrectly determined that the data were consistent with "AF," whereas the artificial intelligence diagnostic results of wristband ECG and PPG correctly identified data as "not AF." AF = atrial fibrillation; ECG = electrocardiogram; PPG = photoplethysmography.

only. However, because the data used in the 2 studies were all from continuous recording, we cannot compare the PPG performance of those devices with the wristband tested in this study.

In this study, the wristband ECG used all ECG waveform patterns rather than R-wave peak values for AI recognition. The AI algorithm evaluated in this study was completely trained and tested by large-scale real user data, both PPG and ECG, which could make the algorithm more credible.

Study limitations

All PPG and ECG data used in this study were collected while the users were inactive. ECGs are designed to be recorded while subjects are inactive, so the influence of movement on PPG signals cannot be ignored. In addition, PPG data were collected for only 3 minutes in this study, which may not be representative of continuous data. The use of PPG in dynamic recordings and in the longer term may not perform as well and requires further study. Therefore, we plan to perform a comparative study of ambulatory ECG and wearable devices to examine the diagnostic value of this algorithm for screening and diagnosing AF in patients as they go about activities of daily living.

This study assessed the use of a single device in a small population in a centralized laboratory. Evaluation of larger populations and for longer periods of continuous recording from real-world situations is needed. We are also currently planning a multicenter, large-scale clinical study to evaluate these wristbands and smartwatches that combine PPG and ECG.

Conclusion

The Amazfit Health Band 1S and RealBeats Artificial Intelligence Biological Data Engine used in this study provided good sensitivity, specificity, and accuracy in determining the presence of AF. The ability to use this device for screening AF requires further evaluation in larger populations for longer durations.

Appendix

Supplementary data

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.hrthm.2020.01.034.

References

1. Camm AJ, Kirchhof P, Lip GY, et al. Guidelines for the management of atrial fibrillation: the Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC). Eur Heart J 2010;31:2369–2429.
2. Osaka Y, Takigawa M, Takahashi A, et al. The proportion of asymptomatic recurrence after catheter ablation of atrial fibrillation in patients with a pacemaker for sick sinus syndrome. Indian Pacing Electrophysiol J 2017;17:125–131.
3. Nemati S, Ghassemi MM, Ambai V, et al. Monitoring and detecting atrial fibrillation using wearable technology. Conf Proc IEEE Eng Med Biol Soc 2016;3394–3397.
4. Wasserlauf J, You C, Patel R, Valys A, Albert D, Passman R. Smartwatch performance for the detection and quantification of atrial fibrillation. Circ Arrhythm Electrophysiol 2019;12:e006834.
5. Sološenko A, Petrėnas A, Paliakaitė B, Sörnmo L, Marozas V. Detection of atrial fibrillation using a wrist-worn device. Physiol Meas 2019;40:025003.
6. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open 2016;6:e012799.
7. Charitos EI, Stierle U, Ziegler PD, et al. A comprehensive evaluation of rhythm monitoring strategies for the detection of atrial fibrillation recurrence: insights from 647 continuously monitored patients and implications for monitoring after therapeutic interventions. Circulation 2012;126:806–814.
8. Olsson AS, Engdahl J. Detection of atrial fibrillation with intermittent handheld electrocardiogram in patients with ischemic stroke and transient ischemic attack. J Stroke Cerebrovasc Dis 2016;25:2648–2652.
9. Svennberg E, Engdahl J, Al-Khalili F, Friberg L, Frykman V, Rosenqvist M. Mass screening for untreated atrial fibrillation: the STROKESTOP study. Circulation 2015;131:2176–2184.
10. Dörr M, Nohturfft V, Brasier N, et al. The WATCH AF trial: SmartWATCHes for detection of atrial fibrillation. JACC Clin Electrophysiol 2019;5:199–208.
11. Svennberg E, Stridh M, Engdahl J, et al. Safe automatic one-lead electrocardiogram analysis in screening for atrial fibrillation. Europace 2016;19:1449–1453.

12. Aronsson M, Svennberg E, Rosenqvist M, et al. Cost-effectiveness of mass screening for untreated atrial fibrillation using intermittent ECG recording. Europace 2015;17:1023–1029.
13. Deering TF, Hindricks G, Marrouche NF. Digital health: present conundrum, future hope or hype? Heart Rhythm 2019;16:1303–1304.
14. Perez MV, Mahaffey KW, Hedlin H, et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med 2019;381:1909–1917.
15. Guo Y, Wang H, Zhang H, et al. Mobile health technology for atrial fibrillation screening using photoplethysmography-based smart devices: the HUAWEI Heart Study. J Am Coll Cardiol 2019;74. 2375–2375.