JAMA Cardiology | **Original Investigation**

Passive Detection of Atrial Fibrillation Using a Commercially Available Smartwatch

Geoffrey H. Tison, MD, MPH; José M. Sanchez, MD; Brandon Ballinger, BS; Avesh Singh, MS; Jeffrey E. Olgin, MD; Mark J. Pletcher, MD, MPH; Eric Vittinghoff, PhD; Emily S. Lee, BA; Shannon M. Fan, BA; Rachel A. Gladstone, BA; Carlos Mikell, BS; Nimit Sohoni, BS; Johnson Hsieh, MS; Gregory M. Marcus, MD, MAS

IMPORTANCE Atrial fibrillation (AF) affects 34 million people worldwide and is a leading cause of stroke. A readily accessible means to continuously monitor for AF could prevent large numbers of strokes and death.

OBJECTIVE To develop and validate a deep neural network to detect AF using smartwatch data.

DESIGN, SETTING, AND PARTICIPANTS In this multinational cardiovascular remote cohort study coordinated at the University of California, San Francisco, smartwatches were used to obtain heart rate and step count data for algorithm development. A total of 9750 participants enrolled in the Health eHeart Study and 51 patients undergoing cardioversion at the University of California, San Francisco, were enrolled between February 2016 and March 2017. A deep neural network was trained using a method called heuristic pretraining in which the network approximated representations of the R-R interval (ie, time between heartbeats) without manual labeling of training data. Validation was performed against the reference standard 12-lead electrocardiography (ECG) in a separate cohort of patients undergoing cardioversion. A second exploratory validation was performed using smartwatch data from ambulatory individuals against the reference standard of self-reported history of persistent AF. Data were analyzed from March 2017 to September 2017.

MAIN OUTCOMES AND MEASURES The sensitivity, specificity, and receiver operating characteristic C statistic for the algorithm to detect AF were generated based on the reference standard of 12-lead ECG–diagnosed AF.

RESULTS Of the 9750 participants enrolled in the remote cohort, including 347 participants with AF, 6143 (63.0%) were male, and the mean (SD) age was 42 (12) years. There were more than 139 million heart rate measurements on which the deep neural network was trained. The deep neural network exhibited a C statistic of 0.97 (95% CI, 0.94-1.00; $P < .001$) to detect AF against the reference standard 12-lead ECG–diagnosed AF in the external validation cohort of 51 patients undergoing cardioversion; sensitivity was 98.0% and specificity was 90.2%. In an exploratory analysis relying on self-report of persistent AF in ambulatory participants, the C statistic was 0.72 (95% CI, 0.64-0.78); sensitivity was 67.7% and specificity was 67.6%.

CONCLUSIONS AND RELEVANCE This proof-of-concept study found that smartwatch photoplethysmography coupled with a deep neural network can passively detect AF but with some loss of sensitivity and specificity against a criterion-standard ECG. Further studies will help identify the optimal role for smartwatch-guided rhythm assessment.

Editorial

Supplemental content and Audio

Author Affiliations: Division of Cardiology, Department of Medicine, University of California, San Francisco (Tison, Sanchez, Olgin, Lee, Fan, Gladstone, Mikell, Marcus); Cardiogram Incorporated, San Francisco, California (Ballinger, Singh, Sohoni, Hsieh); Department of Epidemiology and Biostatistics, University of California, San Francisco (Pletcher, Vittinghoff).

Corresponding Author: Gregory M. Marcus, MD, MAS, Division of Cardiology, Department of Medicine, University of California, San Francisco, 505 Parnassus Ave, M1180B, San Francisco, CA 94143-0124 (marcusg@medicine.ucsf.edu).

A trial fibrillation (AF) affects up to 34 million people worldwide, and patients with AF exhibit a higher risk of severe health consequences, including death and stroke.[1-3] Atrial fibrillation is often asymptomatic and thus can remain undetected until a thromboembolic event occurs.[1,4] Earlier detection of AF would enable the use of anticoagulation therapy known to mitigate the risk of stroke and other thromboembolic complications.[1,4-9] Data from implantable cardiac monitors has demonstrated that years of monitoring may be required to detect clinically significant AF.[10,11]

An accessible means to continuously monitor for AF could have a valuable clinical effect by enabling AF detection, such as in patients with recurrent AF after ablation or pharmacologic cardioversion.[12,13] Various types of wearable sensors, ranging from fitness trackers to smartwatches, can measure heart rate and have exhibited rapid adoption among the general population, providing an opportunity for highly scalable AF detection.[10,14]

Deep neural networks are a type of machine learning algorithm that has shown high accuracy in performing pattern recognition from noisy, complex inputs, such as speech[15] and image recognition,[16] including in medical applications, such as detection of diabetic retinopathy[17] and skin cancer.[18] We sought to develop and validate a deep neural network to detect AF from smartwatch data.

Methods

Study Design
This study used 3 cohorts to achieve 3 goals: (1) algorithm development and training using a remote cohort; (2) external validation of AF detection in an in-person cardioversion cohort; and (3) an exploratory analysis to perform ambulatory AF detection in a third cohort. All participants provided written informed consent prior to enrollment. This study was approved by the University of California, San Francisco, institutional review board.

Data Collection
We used the publicly available Cardiogram mobile application (Cardiogram Inc) to access data from standard, commercially available Apple Watches (Apple Inc). Heart rate data obtained from the photoplethysmography sensor and step count data from the accelerometer were accessed by the Cardiogram application and input into a deep neural network. The frequency of heart rate recordings depended on whether the watch was in standby mode, during which heart rates were obtained every 5 minutes, or Workout mode, during which heart rates were continuously recorded and obtained as 5-second averages.

Algorithm Development and Training (Remote Cohort)
A large, geographically diverse remote cohort was used to perform deep neural network development and training (eFigures 1 and 2 in the Supplement). Data gathering and study management were performed using the Health eHeart Study infrastructure (http://www.health-eheartstudy.org), an online

Key Points

Question How well can smartwatch sensor data analyzed by a deep neural network identify atrial fibrillation?

Findings In this cohort study of 51 participants presenting for cardioversion, a commercially available smartwatch was able to detect atrial fibrillation with high accuracy. Among 1617 ambulatory individuals who wore a smartwatch, those with self-reported atrial fibrillation were correctly classified with moderate accuracy.

Meaning These data support the proof of concept that a commercially available smartwatch coupled with a deep neural network classifier can passively detect atrial fibrillation.

cohort study coordinated at the University of California, San Francisco. The Health eHeart Study enrolls English-speaking adults 18 years and older with an active email address recruited through academic institutions, lay press, social media, promotional events, and, for the current study, the Cardiogram mobile application. From March 2, 2016, to March 15, 2017, Health eHeart Study participants with an Apple Watch were sent an invitation to participate, and individuals providing informed consent were enrolled. Demographic information was self-reported.

Algorithm Development
We used a purpose-built neural network consisting of 8 layers, each of which had 128 hidden units, for a total of 564 227 parameters (eFigure 3 in the Supplement), that transformed raw sensor measurements—heart rate and step counts—into a sequence of scores corresponding to probabilities that a participant was in AF at each time interval. This deep neural network was trained sequentially using a semisupervised approach, described below. Our primary classification task was to passively detect AF from Apple Watch data while in Workout mode. The network was trained using an unsupervised approach, which we call heuristic pretraining, using Google's TensorFlow framework.[19] This training used 57 675 person-weeks of unlabeled data from the remote cohort (n = 6682) to compute several representations approximating R-R intervals (ie, time between heartbeats). Modeled after a heuristic previously used for AF detection,[20] we calculated the average absolute difference between successive heart rate measurements across window sizes of 5 seconds, 30 seconds, 5 minutes, and 30 minutes.

Evaluation and Validation of the Algorithm (Cardioversion Cohort)
To validate the neural network, we enrolled consecutive consenting patients with AF presenting to the University of California, San Francisco, for electrical cardioversion or pharmacologic cardioversion (using dofetilide during hospitalization with continuous monitoring) between March 24, 2016, and February 24, 2017. Patients with atrial arrhythmias other than AF at the time of enrollment, ventricular pacing, or a ventricular-assist device were excluded.

Table 1. Baseline Characteristics of Remote Cohort Participants[a]

Characteristic	No. (%)		P Value
	Atrial Fibrillation (n = 347)[b]	No Atrial Fibrillation (n = 8216)	
Age, mean (SD), y	55.7 (14.2)	41.5 (11.9)	<.001
Unknown	3 (0.9)	52 (0.6)	.41
Sex			
Male	263 (75.8)	5132 (62.5)	
Female	73 (21.0)	2122 (25.8)	<.001
Unknown	11 (3.2)	962 (11.7)	
Race/ethnicity			
White	307 (88.5)	5925 (72.1)	
Black	12 (3.5)	248 (3.0)	<.001
Asian	10 (2.9)	386 (4.7)	
Other/unknown	18 (5.2)	1657 (20.2)	
Hispanic			
Yes	14 (4.0)	873 (10.6)	
No	319 (91.9)	6284 (76.5)	<.001
Unknown	14 (4.0)	1059 (12.9)	
Education			
High school or less	27 (7.8)	497 (6.0)	
Some college	68 (19.6)	1408 (17.1)	
Bachelor's degree	81 (23.3)	1752 (21.3)	<.001
Postgraduate	126 (36.3)	1523 (18.5)	
Other/unknown	45 (13.0)	3036 (37.0)	
Other medical conditions			
Hypertension	178 (51.3)	3016 (36.7)	.01
Diabetes	40 (11.5)	624 (7.6)	.04
Coronary artery disease	53 (15.3)	468 (5.7)	.006
Myocardial infarction	34 (9.8)	257 (3.1)	.007
Congestive heart failure	36 (10.4)	117 (1.4)	.002
Stroke	36 (10.4)	198 (2.4)	.004
Sleep apnea	93 (26.8)	1379 (16.8)	.01
Implanted heart device	37 (10.7)	128 (1.6)	.002
Congenital heart disease	23 (6.6)	176 (2.1)	.01
Chronic obstructive pulmonary disease	20 (5.8)	223 (2.7)	.03
Peripheral vascular disease	17 (4.9)	96 (1.2)	.009

[a] Baseline characteristics of remote cohort participants after excluding those who did not complete medical history surveys (n = 1187).

[b] Atrial fibrillation was determined by participant self-report.

A 12-lead electrocardiogram (ECG) was obtained, and an Apple Watch (paired with an iPhone [Apple Inc] preloaded with the Cardiogram application) was applied to the participant's wrist for at least 20 minutes in Workout mode. Patients undergoing electrical cardioversion remained supine during the study. A study coordinator time-stamped the moment of cardioversion and removed the Apple Watch after at least 20 minutes following cardioversion. Patients receiving dofetilide were monitored for conversion to sinus rhythm on cardiac telemetry, after which a 12-lead ECG and at least 20 minutes of Apple Watch data were obtained. Apple Watch data were used as inputs to the neural network, and rhythm diagnoses were determined by 12-lead ECGs overread by board-certified cardiac electrophysiologists.

Second Validation: An Exploratory Analysis

Using a subset of the remote cohort, we performed an exploratory analysis for a second classification task: detecting AF from ambulatory data. In this remote cohort subset, the reference standard AF diagnosis was limited to self-reported persistent AF. The heuristic pretrained network underwent a second, supervised training phase using labels of AF or no AF obtained from AliveCor devices. The AliveCor Kardia (AliveCor Inc) is a smartphone-based device equipped with 2 electrodes that enables remote participants to obtain a single-lead ECG. The first 200 remote cohort participants who self-reported an arrhythmia diagnosis were mailed a device and instructed to use it at least once per day while wearing the Apple Watch in Workout mode; 183 (91.5%) were ultimately successfully connected. A validated AliveCor algorithm approved by the US Food and Drug Administration demonstrating high specificity for AF detection[21,22] was used to label AliveCor recordings. Supervised training of the network used 6338 mobile ECGs from the 183 participants, including 625 recordings from 100 participants who exhibited AF. Inputs included more than 139 million Apple Watch sensor measurements; unlabeled inputs were retained using output masking.

Table 2. Baseline Characteristics of Cardioversion Cohort Participants

Baseline Characteristics	Total, No. (%) (n = 51)
Age, mean (SD), y	66.1 (10.7)
Male	43 (84)
White	39 (76)
Medical characteristics	
Hypertension	43 (84)
Hyperlipidemia	31 (61)
Current smoking	7 (14)
Diabetes	13 (25)
Coronary artery disease	8 (16)
Left ventricular ejection fraction >55%	37 (72)
Obstructive sleep apnea	13 (25)
Prior cardioversion	30 (59)
Stroke/transient ischemic attack	11 (21)

For this ambulatory validation, we held out a subcohort of participants randomly selected from the remote cohort not used for algorithm development. Those with pacemakers and implanted cardioverter/defibrillators, those who reported that their AF "comes and goes," and those with inadequate heart rate measurements (averaging 8 hours or less per day) were excluded. Participants completed online questionnaires regarding demographic characteristics and medical history. The self-reported diagnosis of persistent AF by a health care professional was the primary outcome for this exploratory analysis.

Statistical Analysis

Normally distributed continuous variables are presented as means with standard deviations and were compared using t tests, and continuous variables with skewed distributions are presented as medians with interquartile ranges and were compared using the Wilcoxon rank-sum test. Categorical variables were compared using the χ^2 test or Fisher exact test. Test characteristics with 95% CIs and receiver operating characteristic curves with C statistics were calculated using standard techniques. Unadjusted and multivariable logistic regression was used to determine odds ratios.

In the validation cardioversion cohort, we estimated that 76 participants would provide a margin of sampling error of 5% for a sensitivity and specificity of 95%. We used an early stopping rule for efficacy, testing the joint null hypothesis that either the sensitivity or specificity would be less than 85%. The P value for the joint test was calculated by comparing the minimum of the sample sensitivity and specificity to its simulated null distribution, with both parameters set to 0.85. We planned to compare the resulting 2-tailed P values after 50 participants were enrolled and, if needed, after 76 participants to critical α values (as specified by the O'Brien-Fleming stopping rule[23]) of .0031 and .0490, respectively. Otherwise, for all other aspects of the study, a 2-tailed P value < .05 was considered statistically significant.

For comparison with the deep neural network, we concurrently analyzed input heart rate data using 2 statistical techniques used previously for AF detection: root mean square of successive difference (RMSSD),[20] reflecting heart rate vari-

ability, and Shannon entropy (ShE),[24] which characterizes rhythm complexity. Odds ratios were computed in R, version 3.4.0 (The R Foundation), and area under the receiver operating characteristic curve was computed using the scikit-learn, version 0.19.1, package in Python (Python Software Foundation).

Results

Remote Cohort

We enrolled 9750 Health eHeart Study participants with an Apple Watch who downloaded the Cardiogram mobile application, completed an intake survey, and linked their Cardiogram account (eFigures 1 and 2 in the Supplement). **Table 1** shows the baseline demographic characteristics of the remote cohort. Those with AF were more likely to be older, male, and white and were more likely to have completed more education and exhibited concurrent clinical comorbidities.

Cardioversion Cohort Validation

After 50 patients were enrolled, the early stopping rule was applied for efficacy given evidence that neither sensitivity nor specificity was less than 85% ($P < .001$). One additional patient was enrolled during this interim analysis and was included in the final data set. Of these 51 patients, 43 (84%) underwent electrical cardioversion (all successful) and 8 (16%) converted with dofetilide (**Table 2**). Fifty-three total hours of Workout mode data were obtained.

Figure 1 shows a t-distributed stochastic neighbor embedding for AF and sinus rhythm.[25] The C statistic for AF detection for the deep neural network was 0.97 (95% CI, 0.94-1.00; $P < .001$) (**Figure 2**A). At an operating threshold set to the high sensitivity of 98.0%, the specificity was 90.2% to detect ECG-diagnosed AF. In comparison, the C statistic for RMSSD using the same data was 0.91 (95% CI, 0.85-0.97) and for ShE was 0.86 (95% CI, 0.79-0.93). Normalizing all tracings by heart rate did not meaningfully reduce the accuracy of the neural network, yielding a C statistic of 0.96 (95% CI, 0.94-0.97).

Analysis of Ambulatory AF Detection

In this ambulatory subcohort, 1617 participants were available, 64 (4%) of whom reported having persistent AF. From these participants, 18.5 million of 27.9 million heart rate measurements (66.3%) were obtained while in Workout mode. The C statistic for the deep neural network to detect an individual with persistent AF was 0.72 (95% CI, 0.64-0.78) (Figure 2B). The C statistics for RMSSD using the same data was 0.45 (95% CI, 0.38-0.52) and for ShE was 0.48 (95% CI, 0.40-0.55). Using an operating cut point for the neural network that maximizes the sum of sensitivity (67.7%) and specificity (67.6%), those with neural network–predicted AF had an unadjusted odds ratio of 3.95 (95% CI, 3.02-5.17; $P < .001$) for persistent AF. After adjusting for age, sex, race/ethnicity, hypertension, diabetes, heart failure, and coronary artery disease, neural network–predicted AF remained significantly associated, with an adjusted odds ratio of 1.98 (95% CI, 1.48-2.65; $P = .02$) for persistent AF. **Table 3** shows the performance characteristics of

Figure 1. T-Distributed Stochastic Neighbor Embedding Visualization of the Deep Neural Network's Last Layer Using Data From the Cardioversion Cohort
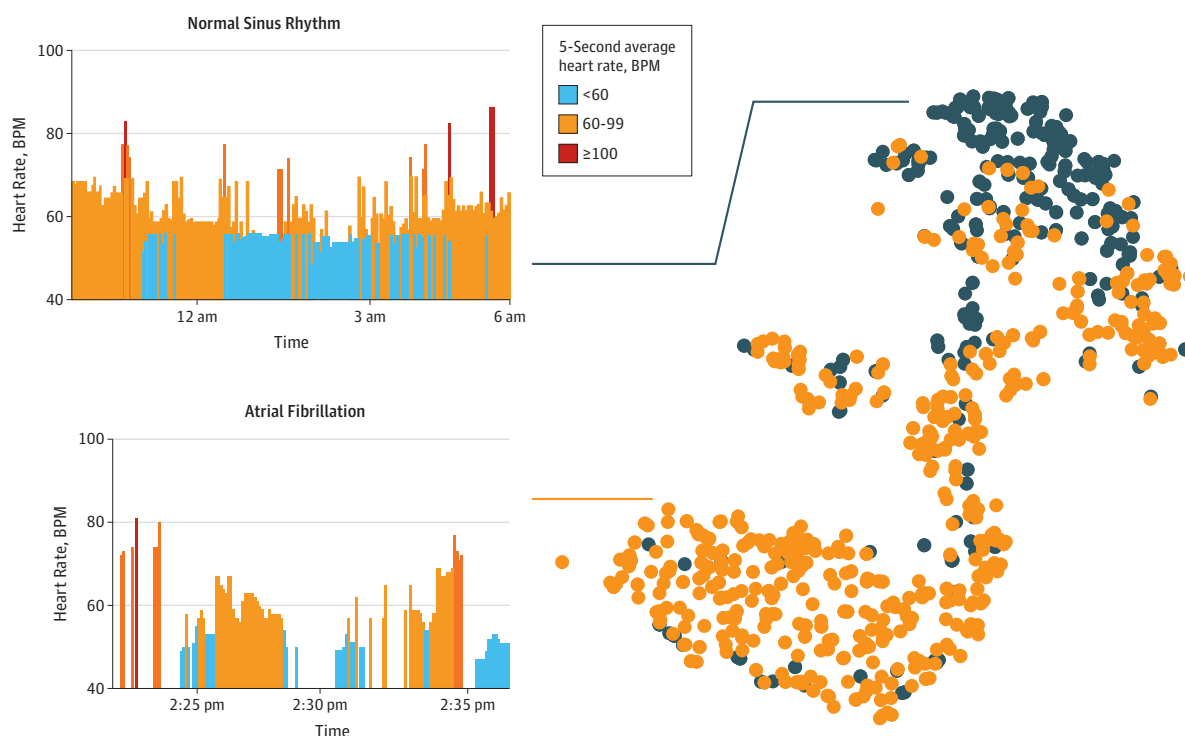
This t-distributed stochastic neighbor embedding, which is a technique that assists in visualizing high-dimensional data in 2 dimensions, depicts the deep neural network's internal representation of the data derived from the last recurrent layer of the neural network. Each point represents a 10-minute segment of data from our validation (cardioversion) data set; orange points represent atrial fibrillation segments (precardioversion) and blue points represent normal sinus rhythm segments (postcardioversion). The neural network has largely clustered atrial fibrillation from normal sinus rhythm segments, as depicted when plotted on 2 dimensions (axes) that were chosen arbitrarily. Most points classified as normal sinus rhythm are in the upper part of the visualization, while atrial fibrillation points are separated in alternate clusters. The upper inset shows an example of raw smartwatch heart rate data associated with normal sinus rhythm, and the lower inset shows raw atrial fibrillation smartwatch data; each vertical bar represents a 5-second average heart rate color-coded by beats per minute (BPM; blue, <60; orange, 60-99; red, ≥100).

the algorithm in both validation cohorts. In the setting of the low (4%) AF prevalence in the remote cohort, the positive predictive value was low.
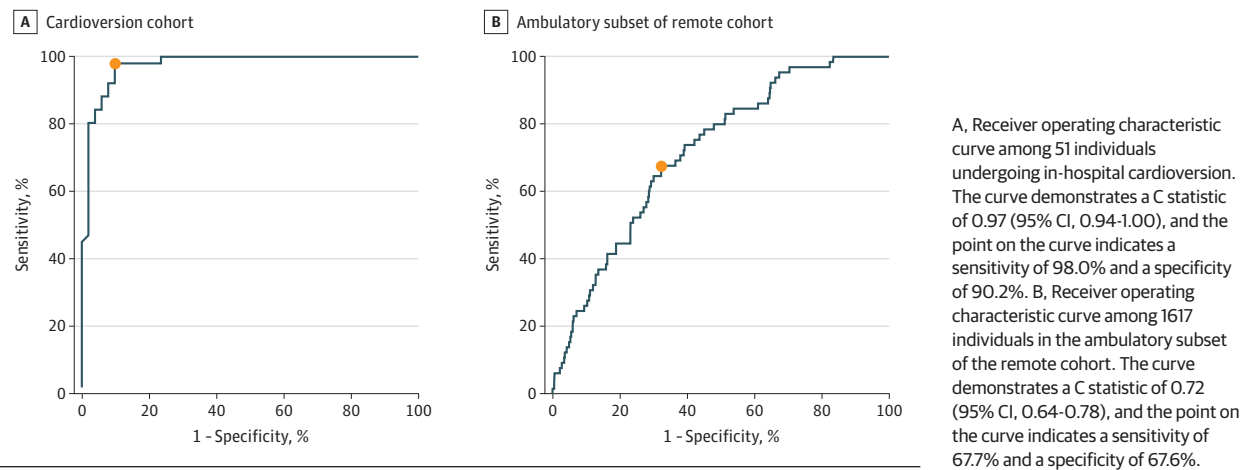
Discussion

We demonstrate that a commercially available smartwatch can passively detect AF using a readily available mobile application using a deep neural network. In external validation using the standard 12-lead ECG as the reference, algorithm performance achieved a C statistic of 0.97. The passive detection of AF from free-living smartwatch data has substantial clinical implications. Importantly, the accuracy of detecting self-reported AF in an ambulatory setting was more modest (C statistic of 0.72). Although the deep neural network's AF classification in the exploratory analysis exhibited higher odds ratios than other measured AF risk factors, this proof-of-concept experiment likely demonstrates the challenges of accurately detecting ambulatory arrhythmia among constantly mobile individuals in natural environments.

Atrial fibrillation is the leading cause of stroke, and its detection is difficult because of its often asymptomatic nature

and paroxysmal frequency.[1,4,10,11,14] Readily accessible means to detect and screen for silent AF are needed. Even though monitors with automated capabilities, such as implantable loop recorders, can be used to detect AF, they are invasive, expensive, and inconvenient.[26,27] The ideal instrument for AF detection would be noninvasive and provide real-time, accurate AF detection in a passive fashion—specifically, not requiring the user to remember to perform some action and not limited to any one snapshot in time. Smartwatches are well positioned to accomplish these goals in a cost-efficient and resource-efficient fashion. Wearable technology has shown a steady increase in global usage,[28] and the smartwatch, most popular among all wearable sensors, is projected to reach 55 million global shipments by 2020.[29]

Prior efforts to automatically detect AF among free-living participants have predominantly used ambulatory blood pressure monitors,[30] although some recent studies[21,31] have used smartphones and wearable devices. Two studies[20,21] showed that AF can be detected using a photoplethysmography waveform obtained via the iPhone camera. Similar to the limitations of ambulatory blood pressure cuffs, smartphone-based data collection is limited in requiring active participa-

Figure 2. Accuracy of Detecting Atrial Fibrillation in the Cardioversion Cohort

A, Receiver operating characteristic curve among 51 individuals undergoing in-hospital cardioversion. The curve demonstrates a C statistic of 0.97 (95% CI, 0.94-1.00), and the point on the curve indicates a sensitivity of 98.0% and a specificity of 90.2%. B, Receiver operating characteristic curve among 1617 individuals in the ambulatory subset of the remote cohort. The curve demonstrates a C statistic of 0.72 (95% CI, 0.64-0.78), and the point on the curve indicates a sensitivity of 67.7% and a specificity of 67.6%.

Table 3. Performance Characteristics of Deep Neural Network in Validation Cohorts[a]

| Cohort | % | | | | |
	Sensitivity	Specificity	PPV	NPV	AUC
Cardioversion cohort (sedentary)	98.0	90.2	90.9	97.8	0.97
Subset of remote cohort (ambulatory)	67.7	67.6	7.9	98.1	0.72

Abbreviations: AUC, area under the receiver operating characteristic curve; NPV, negative predictive value; PPV, positive predictive value.

[a] In the cardioversion cohort, the atrial fibrillation reference standard was 12-lead electrocardiography diagnosis; in the remote cohort, the atrial fibrillation reference standard was limited to self-reported history of persistent atrial fibrillation.

tion from the participant (dependent on user adherence) and by the episodic nature of data obtained. A Samsung Simband (Samsung) exhibited high sensitivity and specificity for AF detection among 46 individuals.[32] However, validation in an external cohort was not performed, and these findings are tied to a single stand-alone device used for research that is not commercially available. To our knowledge, our study represents the first to use a deep neural network to passively detect AF using smartwatch data.

When tested against 12-lead ECG–diagnosed AF in our validation experiment, the deep neural network outperformed 2 conventional methods for the detection of AF.[20] Although the mean heart rate may differ between those in AF and sinus rhythm, our results were not meaningfully changed after heart rate data were normalized. This external validation demonstrates that the neural network can passively detect AF from smartwatch data with excellent performance characteristics obtained in sedentary individuals captured at high temporal resolution (ie, Workout mode). Even within these constraints, public health implications for AF screening may be broad because periods of sleep can provide long, uninterrupted periods of sedentary data, and it is technically feasible to enable high temporal resolution data collection at scheduled periods.

In light of the relative frequency of subclinical AF detected by implanted devices among patients at risk of stroke,[33-35] it is very likely that the widespread use of an accurate algorithm to detect AF among the large population continuously wearing smartwatches would result in a substantial increase in new AF diagnoses. While there may be increased

costs associated with the care of those patients, the potential reduction in stroke could ultimately provide cost savings.

Several factors make detection of AF from ambulatory data an inherently more difficult classification task: (1) the predominance of ambulatory heart rates are represented by non–Workout mode data sampled every 5 minutes, which translates to a significant loss of temporal resolution; (2) the variability in heart rate in an ambulatory population is significantly increased by a wide range of activities; and (3) heart rate sensor noise is increased with movement. Because the remote cohort data set was limited to using self-reported diagnoses of persistent AF rather than ECG-diagnosed AF as in our validation cohort, we considered this analysis exploratory. We did not know how many participants were actually in AF at the time the Apple Watch measurements were taken. Acknowledging these limitations, when using conventional algorithms to analyze the ambulatory data (RMSSD and ShE), C statistics were similar to predictors based on chance alone. In contrast, even after adjustment for conventional risk factors, the neural network's algorithm classified individuals with persistent AF. As technology develops, expected improvements in sensors and battery life will likely improve the temporal resolution of ambulatory heart rate measurements, enabling enhanced algorithm performance for disease detection.

In contrast to methods used in other recent medical applications of deep neural networks,[17,18] we developed a semisupervised heuristic pretraining procedure that is not dependent on manual physician annotation of training data. Here, the deep neural network automatically learned features representative of a heuristic relevant to our specific task of AF de-

tection, namely an approximation of the R-R interval.[20] Deep neural networks pretrained using this approach can subsequently be trained in a supervised manner using a relatively small amount of labeled data to perform more specialized classification tasks, exemplified in our exploratory analysis. Deep neural networks are data-hungry algorithms, typically requiring tens of thousands of labeled examples for optimal performance.[17] However, labeled data at this scale is not readily available for most medical applications and is costly to label when it is possible to obtain. Our semisupervised method can be generalized to train data-efficient deep neural networks for other medical tasks, requiring significantly less labeled data than previously required.

Limitations

Our study has several limitations. Many participants initially contacted for the study did not complete surveys and link Cardiogram accounts, which may result in selection bias. This would likely not invalidate positive associations but would limit generalizability. All participants already owned a smartwatch or, among the patients undergoing cardioversion, had a coordinator provide assistance; therefore, it is possible these results would not generalize to less tech-savvy individuals. However, as demonstrated by the growing majority that now use the internet and own smartphones,[28,36] the regular use of smartwatches may become more mainstream with time. The training using the AliveCor devices relied on an automated algorithm. Although the algorithm is approved by the US Food and Drug Administration and has exhibited reasonable

accuracy in previous studies,[21,22] it is possible that manual over-reads may have enhanced algorithm performance in the exploratory ambulatory analysis. These data focused on individuals with a known history of AF. Therefore, we did not demonstrate an ability to identify a new diagnosis of the disease. Finally, despite the excellent test characteristics observed among sedentary patients undergoing cardioversion, the modest performance in the ambulatory scenario, a context more representative of the ultimate application of this technology, suggests that these data should be primarily interpreted as a proof of concept.

Conclusions

In conclusion, we showed that a commercially available smartwatch can passively identify AF against the criterion-standard 12-lead ECG among sedentary individuals when heart rate data are collected at a high temporal resolution. We also present an exploratory analysis demonstrating that our deep neural network substantially outperforms standard techniques to detect self-reported persistent AF from ambulatory data, albeit with modest accuracy in these free-living natural environments. Given the broad and growing use of smartwatches and ready accessibility of downloadable mobile applications, this approach may ultimately be applied to perform AF detection at large scale, ultimately leveraging common wearable devices to guide AF management and rhythm assessment.

REFERENCES

1. January CT, Wann LS, Alpert JS, et al; ACC/AHA Task Force Members. 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines and the Heart Rhythm Society. *Circulation*. 2014;130(23):e199-e267.

2. Chugh SS, Havmoeller R, Narayanan K, et al. Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. *Circulation*. 2014;129(8):837-847.

3. Xian Y, O'Brien EC, Liang L, et al. Association of preceding antithrombotic treatment with acute ischemic stroke severity and in-hospital outcomes among patients with atrial fibrillation. *JAMA*. 2017; 317(10):1057-1067. doi:10.1001/jama.2017.1371

4. Kernan WN, Ovbiagele B, Black HR, et al; American Heart Association Stroke Council, Council

on Cardiovascular and Stroke Nursing, Council on Clinical Cardiology, and Council on Peripheral Vascular Disease. Guidelines for the prevention of stroke in patients with stroke and transient ischemic attack: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2014;45(7):2160-2236.

5. Santangeli P, Di Biase L, Bai R, et al. Atrial fibrillation and the risk of incident dementia: a meta-analysis. *Heart Rhythm*. 2012;9(11):1761-1768.

6. Kalantarian S, Stern TA, Mansour M, Ruskin JN. Cognitive impairment associated with atrial fibrillation: a meta-analysis. *Ann Intern Med*. 2013; 158(5, pt 1):338-346.

7. Bansal N, Fan D, Hsu CY, Ordonez JD, Marcus GM, Go AS. Incident atrial fibrillation and risk of end-stage renal disease in adults with chronic kidney disease. *Circulation*. 2013;127(5):569-574.

8. Soliman EZ, Safford MM, Muntner P, et al. Atrial fibrillation and the risk of myocardial infarction. *JAMA Intern Med*. 2014;174(1):107-114. doi:10.1001/jamainternmed.2013.11912

9. Dukes JW, Marcus GM. Atrial fibrillation begets myocardial infarction. *JAMA Intern Med*. 2014;174 (1):5-7. doi:10.1001/jamainternmed.2013.11392

10. Albers GW, Bernstein RA, Brachmann J, et al. Heart rhythm monitoring strategies for cryptogenic stroke: 2015 Diagnostics and Monitoring Stroke Focus Group Report. *J Am Heart Assoc*. 2016;5(3): e002944.

11. Sanna T, Diener HC, Passman RS, et al; CRYSTAL AF Investigators. Cryptogenic stroke and underlying atrial fibrillation. *N Engl J Med*. 2014; 370(26):2478-2486.

12. Ganesan AN, Shipp NJ, Brooks AG, et al. Long-term outcomes of catheter ablation of atrial fibrillation: a systematic review and meta-analysis. *J Am Heart Assoc*. 2013;2(2):e004549.

13. Szilágyi J, Marcus GM, Badhwar N, et al. Atrial fibrillation patients with isolated pulmonary veins: is sinus rhythm achievable? *J Cardiovasc Electrophysiol*. 2017;28(7):754-761.

14. Freedman B, Camm J, Calkins H, et al; AF-Screen Collaborators. Screening for atrial fibrillation: a report of the AF-SCREEN International Collaboration. *Circulation*. 2017;135(19):1851-1867.

15. Sak H, Senior A, Beaufays F. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. Paper presented at: Interspeech 2014; September 14-18, 2014; Singapore.

16. Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis*. 2015;115(3):211-252. doi:10.1007 /s11263-015-0816-y

17. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22): 2402-2410. doi:10.1001/jama.2016.17216

18. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017; 542(7639):115-118.

19. Abadi M, Agarwal A, Barham P, et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems. https://static.googleusercontent.com/media /research.google.com/en//pubs/archive/45166.pdf. Accessed December 1, 2017.

20. McManus DD, Lee J, Maitas O, et al. A novel application for the detection of an irregular pulse using an iPhone 4S in patients with atrial fibrillation. *Heart Rhythm*. 2013;10(3):315-319.

21. Chan PH, Wong CK, Poh YC, et al. Diagnostic performance of a smartphone-based photoplethysmographic application for atrial fibrillation screening in a primary care setting. *J Am Heart Assoc*. 2016;5(7):e003428.

22. Desteghe L, Raymaekers Z, Lutin M, et al. Performance of handheld electrocardiogram devices to detect atrial fibrillation in a cardiology and geriatric ward setting. *Europace*. 2017;19(1): 29-39.

23. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics*. 1979;35(3): 549-556.

24. Dash S, Chon KH, Lu S, Raeder EA. Automatic real time detection of atrial fibrillation. *Ann Biomed Eng*. 2009;37(9):1701-1709.

25. van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res*. 2008;9:2579-2605.

26. Mittal S, Rogers J, Sarkar S, et al. Real-world performance of an enhanced atrial fibrillation detection algorithm in an insertable cardiac monitor. *Heart Rhythm*. 2016;13(8):1624-1630.

27. Sanders P, Pürerfellner H, Pokushalov E, et al; Reveal LINQ Usability Investigators. Performance of a new atrial fibrillation detection algorithm in a miniaturized insertable cardiac monitor: results from the Reveal LINQ Usability Study. *Heart Rhythm*. 2016;13(7):1425-1430.

28. Smith A. Record shares of Americans now own smartphones, have home broadband. http://www.pewresearch.org/fact-tank/2017/01/12 /evolution-of-technology/. Accessed December 3, 2017.

29. Statista. Projected smart/basic wearables unit shipments worldwide 2014-2020.

https://www.statista.com/statistics/296565 /wearables-worldwide-shipments/. Accessed December 1, 2017.

30. Kane SA, Blake JR, McArdle FJ, Langley P, Sims AJ. Opportunistic detection of atrial fibrillation using blood pressure monitors: a systematic review. *Open Heart*. 2016;3(1):e000362.

31. Garabelli P, Stavrakis S, Po S. Smartphone-based arrhythmia monitoring. *Curr Opin Cardiol*. 2017; 32(1):53-57.

32. Nemati S, Ghassemi MM, Ambai V, et al. Monitoring and detecting atrial fibrillation using wearable technology. *Conf Proc IEEE Eng Med Biol Soc*. 2016;2016:3394-3397.

33. Reiffel JA, Verma A, Kowey PR, et al; REVEAL AF Investigators. Incidence of previously undiagnosed atrial fibrillation using insertable cardiac monitors in a high-risk population: the REVEAL AF Study. *JAMA Cardiol*. 2017;2(10): 1120-1127. doi:10.1001/jamacardio.2017.3180

34. Nasir JM, Pomeroy W, Marler A, et al. Predicting determinants of atrial fibrillation or flutter for therapy elucidation in patients at risk for thromboembolic events (PREDATE AF) study. *Heart Rhythm*. 2017;14(7):955-961.

35. Hohnloser SH, Capucci A, Fain E, et al; ASSERT Investigators and Committees. Asymptomatic Atrial Fibrillation and Stroke Evaluation in Pacemaker Patients and the Atrial Fibrillation Reduction Atrial Pacing Trial (ASSERT). *Am Heart J*. 2006;152(3): 442-447.

36. Poushter J. Smartphone ownership and internet usage continues to climb in emerging economies. http://www.pewglobal.org/2016/02/22 /smartphone-ownership-and-internet-usage -continues-to-climb-in-emerging-economies/. Accessed November 20, 2017.