European Heart Journal - Digital Health (2024) **00**, 1–7
https://doi.org/10.1093/ehjdh/ztae047

ORIGINAL ARTICLE

Artificial intelligence–based electrocardiogram analysis improves atrial arrhythmia detection from a smartwatch electrocardiogram

Laurent Fiorina [1,*], Pascale Chemaly [1], Joffrey Cellier[1], Mina Ait Said[1], Charlène Coquard[1], Salem Younsi[1], Fiorella Salerno[1], Jérôme Horvilleur[1], Jérôme Lacotte[1], Vladimir Manenti[1], Audrey Plesse[2], Christine Henry [2], and Baptiste Lefebvre [2]

[1]Ramsay Santé, Institut Cardiovasculaire Paris Sud, Hôpital privé Jacques Cartier, 6 avenue du Noyer Lambert, 91 300 Massy, France; and [2]Cardiologs®, 100 rue Réaumur, 75002 Paris, France

Received 22 January 2024; revised 29 April 2024; accepted 29 May 2024; online publish-ahead-of-print 6 July 2024

Aims

Smartwatch electrocardiograms (SW ECGs) have been identified as a non-invasive solution to assess abnormal heart rhythm, especially atrial arrhythmias (AAs) that are related to stroke risk. However, the performance of these tools is limited and could be improved with the use of deep neural network (DNN) algorithms, particularly for specific populations encountered in clinical cardiology practice.

Methods and results

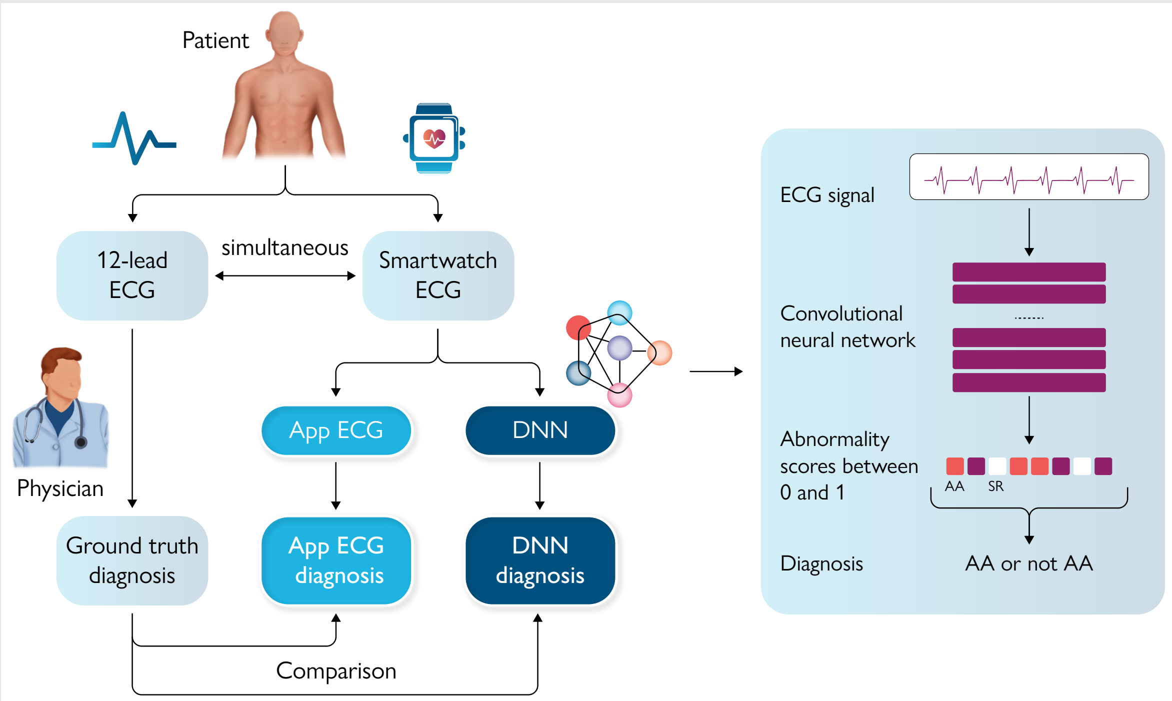
A total of 400 patients from the electrophysiology department of one tertiary care hospital were included in two similar clinical trials (respectively, 200 patients per study). Simultaneous ECGs were recorded with the watch and a 12-lead recording system during consultation or before and after an electrophysiology procedure if any. The SW ECGs were processed by using the DNN and with the Apple watch ECG software (Apple app). Corresponding 12-lead ECGs (12L ECGs) were adjudicated by an expert electrophysiologist. The performance of the DNN was assessed vs. the expert interpretation of the 12L ECG, and inconclusive rates were reported. Overall, the DNN and the Apple app presented, respectively, a sensitivity of 91% [95% confidence interval (CI) 85–95%] and 61% (95% CI 44–75%) with a specificity of 95% (95% CI 91–97%) and 97% (95% CI 93–99%) when compared with the physician 12L ECG interpretation. The DNN was able to provide a diagnosis on 99% of ECGs, while the Apple app was able to classify only 78% of strips (22% of inconclusive diagnosis).

Conclusion

In this study, by including patients from a cardiology department, a DNN-based algorithm applied to an SW ECG provided an accurate diagnosis for AA detection on virtually all tracings, outperforming the SW algorithm.

* Corresponding author. Tel: +33160136647, Email: laurent_fiorina@yahoo.fr

Graphical Abstract



This outline illustrates the methodology used to assess the performance of the DNN and ECG App compared with the ground truth. AA, atrial arrhythmia; App, application; DNN, deep neural network.

Keywords Deep learning • Artificial intelligence • Atrial fibrillation • Smartwatch • Wearable

Introduction

Atrial fibrillation (AF) is the most common arrhythmia in adults, posing both a clinical and an economic burden for society, with prevalence estimated to increase from 3 to 5 million presently to over 17 million in 2050 in the USA and Europe.[1–3]

Atrial fibrillation is associated with an increased risk of stroke and heart failure. An early detection is essential to provide an appropriate treatment to reduce those risks and to manage associated comorbidities[2,4,5] and could enable the initiation of anticoagulation treatment.[6,7] As AF can be asymptomatic, new screening tools such as the smartwatch (SW) are being increasingly adopted and can be helpful for an early diagnosis,[8,9] as reflected by new recommendations in the 2020 European Society of Cardiology guidelines.[6]

Wearable devices such as SWs, with specific algorithms to detect AF, have shown sensitivities ranging from 60 to 98%[10,11] excluding inconclusive electrocardiograms (ECGs) which represents from 10 to 30% of ECGs where no diagnosis is provided.[11] Current guidelines highlight the need for screening silent AF, using mobile health technology (especially SW devices) and artificial intelligence (AI).[7] Those devices could be used to extend the limited monitoring period of the usually used ambulatory ECG monitors.[5,12] Therefore, wearables with diagnostic capabilities become essential in both screening and monitoring patients with a suspicion of AF.

There have been concerns raised regarding the required accuracy for identifying AF with SWs to prevent unnecessary patient anxiety and healthcare utilization.[13] Improving and validating current automated ECG interpretation is crucial. This study aims to evaluate the performance of a new AI-based solution in interpreting ECG data collected from SWs in a cardiovascular patient population at a tertiary hospital.

Methods

We evaluated the performance of our new deep neural network (DNN)-based algorithm using SW ECG data compared with the gold standard diagnosis made by expert physicians on the corresponding 12-lead ECG (12L ECG) recorded simultaneously with the SW ECG. We also assessed the performance of the Apple Watch algorithms (v1 and v2) on the same SW ECG tracings.

The AI Watch studies (AI Watch and AI Watch 2) shared the same design. They were prospective, non-randomized, and monocentric studies to evaluate the ability of the Cardiologs (CDL) algorithm, a DNN-based algorithm[14] [Cardiologs RPM Platform 1.0.0 (beta) with algorithm (a.k.a. cardiolib package) 2.1.44], in detecting atrial arrhythmia (AA) from Apple Watch (v1 and v2, respectively) ECGs compared with 12L ECGs interpreted by physicians. Both devices are CE- and FDA-approved.

Those studies were carried out according to the principles of the Declaration of Helsinki, approved by the local ethics committee (Comité de Protection des Personnes Nord Ouest IV, Lille, France) and registered at ClinicalTrials.gov (NCT04792905, NCT05045456). The subjects provided written consent.

Study aims

The primary objective of the studies was to evaluate the performance of the AI-based algorithm to detect AF from an SW ECG, while using the physician interpretation of the 12L ECG as reference. In a second step, the diagnostic performance of the AI-based algorithm was reported comparatively to the SW manufacturer's algorithm.

Study intervention

In total, 400 patients above 22 years old who presented at the Cardiovascular Institute Paris Sud for scheduled ablation, cardioversion,

Table 1 Clinical and demographic characteristics

	All subjects ($n = 400$)
Sex	
Male, *n* (%)	283 (70.8)
Female, *n* (%)	117 (29.2)
Age (year), mean (SD)	64.8 (14.6)
BMI (kg/m^2), mean (SD)	27.4 (5.0)
Cardiac pathologies	
Hypertension, *n* (%)	177 (44.2)
Diabetes, *n* (%)	62 (15.5)
Ischaemic cardiomyopathy, *n* (%)	65 (16.2)
Non-ischaemic cardiomyopathy, *n* (%)	30 (7.5)
Stroke, *n* (%)	20 (5.0)
Heart failure, *n* (%)	19 (4.8)
Other, *n* (%)	55 (13.8)
Admission reason	
Consultation, *n* (%)	180 (45.0)
Hospitalization, *n* (%)	220 (55.0)
Hospitalization reasons	
Heart failure, *n* (%)	111 (27.8)
Cardioversion, *n* (%)	45 (11.3)
Ablation of atrial fibrillation, *n* (%)	19 (4.8)
Ablation of atrial flutter, *n* (%)	24 (6.0)
Ablation of accessory pathways, *n* (%)	9 (2.2)
Ablation of slow pathways, *n* (%)	7 (1.8)
Ablation of PVCs, *n* (%)	1 (0.2)
Other, *n* (%)	4 (1.0)
Antiarrhythmic drug	
Yes, *n* (%)	232 (58.0)
AA at the time of recording	
Unavailable, *n* (%)	7 (1.8)
AA, *n* (%)	134 (33.5)
No AA, *n* (%)	259 (64.7)

AA, atrial arrhythmia; BMI, body mass index; PVC, pre-mature ventricular contraction; SD, standard deviation.

cardiac electrophysiological exploration, or regular consultation were screened for enrolment. Patients with an implanted pacemaker, defibrillator, or cardiac resynchronization device were excluded.

Patients who met inclusion criteria and consented to participate in the study were provided SWs (Apple Watch Series 4, Apple Inc., CA, USA) that were paired via Bluetooth to smartphones (iPhone SE 1st generation, Apple Inc.) equipped with the Apple (APL) algorithm (ECG App 1.0 and 2.0, respectively; Apple Inc.) and the CDL platform (Cardiologs, Paris, France).

Once enrolled, the patients underwent a 12L ECG of 10 s, followed immediately by a 30 s SW ECG taken from the left wrist. Those recordings were considered near simultaneous. In case of ablation or cardioversion, the recordings were obtained before and/or after treatment. The SW provided its automatic diagnosis for each SW ECG strip (APL algorithm). Both 12L and SW ECG strips were transferred to the secure CDL server and were then automatically analysed by the DNN-based algorithm. The DNN-based algorithm classified each strip as ‘sinus rhythm’ (SR), and AF or atrial flutter (AFL) or atrial tachycardia (AT) was grouped under the term ‘AAs’ or ‘inconclusive’. All 12L ECGs were anonymized and allocated to cardiac Electrophysiologists (EPs) who interpreted independently disjoint sets of strips and assigned a diagnosis (SR, AA, or inconclusive) such that each strip was reviewed by one EP. Atrial fibrillation, atrial flutter, and AT were considered a single disease state for all interpretations.

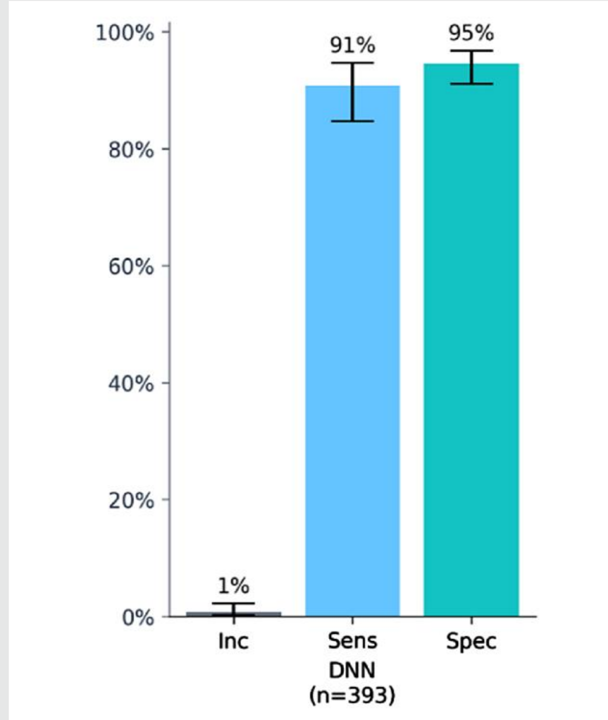


Figure 1 A deep neural network performance. A graph representing the performance of the deep neural network on the per-protocol population (sensitivity and specificity without the inconclusive). The deep neural network correctly diagnosed atrial arrhythmia with 91% sensitivity (95% confidence interval 85–95%) and 95% specificity (95% confidence interval 91–97%). DNN, deep neural network; Inc, inconclusive; Sens, sensitivity; Spec, specificity.

To assess the performance of the DNN-based algorithm in appropriately identifying AA, the automated CDL interpretations of SW ECGs were compared with the reference, i.e. the physician interpretations of 12L ECGs on the population included in both studies (AI Watch and AI Watch 2). The ECG App V1 and V2 diagnoses were also compared with the reference and the results compared with that of the DNN-based algorithm.

Intention-to-diagnose vs. per-protocol approach

The results were evaluated using an intention-to-diagnose approach,[15] which takes all the SW ECG recordings into account, including those with an inconclusive diagnosis. The inconclusive SW ECGs were relabelled in this analysis using the following rules: (i) as SR if the corresponding 12L ECG has been labelled as AA by the EP, to be considered as a false negative; and (ii) as AA in the opposite case, to be considered as a false positive. All 12L ECGs labelled as inconclusive by electrophysiologists were excluded from the analysis. This method has been validated[15] to assess the performance of diagnostic tests in real-world conditions and adopted by similar studies.[16,17]

In addition, the results have been provided with a more common per-protocol approach that excludes inconclusive results. This second method assesses the performances in ideal recording conditions.

Description of Cardiologs artificial intelligence

The CDL algorithm uses two different neural networks, one for wave detection and one dedicated to rhythm classification. The wave detector is a convolutional neural network with a U-net architecture[18] composed of 11 convolutional layers and 6 residual blocks (800 000 parameters in total). This network uses the ECG signal as input and outputs the onsets and offsets of P waves, QRS complexes, and T waves. For the heart rhythm, a

DNN with a Visual Geometry Group (VGG)-like architecture[19] is used, and it consists of 4M parameters and 13 convolutional layers, followed by 3 fully connected layers. The rhythm predictor outputs the presence of multiple labels (SR, AF, AFL, other heart abnormalities, noise). Both neural networks were trained and validated using a data set of more than 1M ECGs from an anonymized data set, which had previously been adjudicated by physicians or certified ECG technicians. This data set contains 12L ECGs and Holter ECGs acquired from North American, European, and Asian Independent Diagnostic Testing Facilities, hospitals, and public data sets. The DNNs also support by design the analysis of single-lead ECGs. Training was achieved using stochastic gradient descent, early stopping, and dropout[20] to avoid overfitting. The neural networks were implemented using the Keras framework with a TensorFlow backend (Google, CA, USA) on K-80 graphics processing units (Nvidia, CA, USA).

Statistical analysis

Quantitative variables were described by mean (standard deviation) and median (1st and 3rd quartiles). Categorical variables were described by number and percentage. Sensitivity and specificity were calculated with 95% confidence intervals (95% CIs) for SW-automated interpretations, both by the CDL and the APL algorithms, compared with 12L physician interpretations. Positive predictive values, negative predictive values, and Cohen's kappa coefficients were assessed for inter-observer agreement and are provided in Supplementary material. Cohen's kappa coefficients >0.80 were considered to represent excellent agreement. No hypothesis testing was used; however, 95% CIs were systematically provided when appropriate. The statistical analyses were performed with Python programming language (Python Software Foundation, https://www.python.org).

Results

Population characteristics

A total of 400 patients were enrolled from April 2021 to February 2022: 200 patients from AI Watch between April 2021 to July 2021, and 200 patients from AI Watch 2 between November 2021 and February 2022.

Four subjects were excluded from AI Watch 2 because they had pacemakers. Three subjects were excluded from the analysis as they had missing data (2 12L and 1 AW ECGs missing), resulting in 393 patients with simultaneous 12L and SW ECGs included in the analyses (see Supplementary material online, *Figure S1*).

Demographics and clinical characteristics are summarized in *Table 1*. The subjects were on average 64.8 ± 14.6 years old, and 29.2% were women. Among the 400 study participants, 55.0% were hospitalized for cardioversion (11.3%), ablation (13.2%), or heart failure (30.5%). An AA was present at the time of ECG recordings for 33.5% of the subjects.

Performances of the Cardiologs deep neural network–based algorithm

The DNN-based algorithm provided a conclusive diagnosis in 99% of the ECGs (inconclusives, $n = 3$; 1%). Using the per-protocol approach, it correctly diagnosed AA with 91% sensitivity (95% CI 85–95%) and 95% specificity (95% CI 91–97%) when compared with physician 12L ECG interpretations (*Figure 1*; performance metrics are given in Supplementary material online, *Table S2*).

Comparison of Apple with Cardiologs for smartwatch electrocardiogram

To assess the performance of AA detection by the CDL algorithm compared with the APL algorithm embedded by default in the SW, the abilities of each algorithm (DNN, ECG App v1, ECG App v2) were compared. On the intend-to-diagnose population, i.e. including the inconclusive ECGs, the DNN, ECG App v1, and ECG App v2 algorithms correctly identified AA with a sensitivity of 89% (95% CI 82–93%), 37% (95% CI 25–50%), and 62% (95% CI 52–72%), respectively,

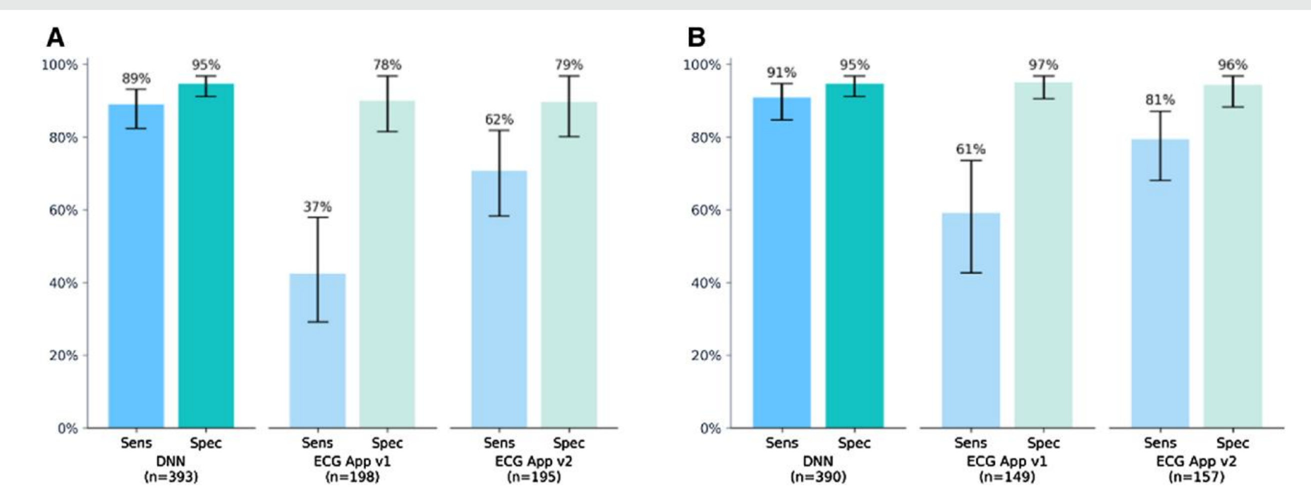


Figure 2 A comparison of a deep neural network and an ACG App. These graphs illustrate the performance of the deep neural network, the ECG App v1, and ECG App v2 compared with physician's interpretations. (*A*) On the intend-to-diagnose population, i.e. including the inconclusive deep neural network, App v1, and Appv2 algorithms correctly identified atrial arrhythmia with a sensitivity of 89% (95% confidence interval 82–93%), 37% (95% confidence interval 25–50%), and 62% (95% confidence interval 52–72%), respectively, and a specificity of 95% (95% confidence interval 91–97%), 78% (95% confidence interval 71–84%), and 79% (95% confidence interval 71–86%), respectively. (*B*) On the per-protocol population, i.e. excluding the inconclusive electrocardiograms, the three algorithms correctly identified atrial arrhythmia with a sensitivity of 91% (95% confidence interval 85–95%), 61% (95% confidence interval 44–75%), and 81% (95% confidence interval 69–89%), respectively, and a specificity of 95% (95% confidence interval 91–97%), 97% (95% confidence interval 93–99%), and 96% (95% confidence interval 90–98%), respectively. DNN, deep neural network; Sens, sensitivity; Spec, specificity.

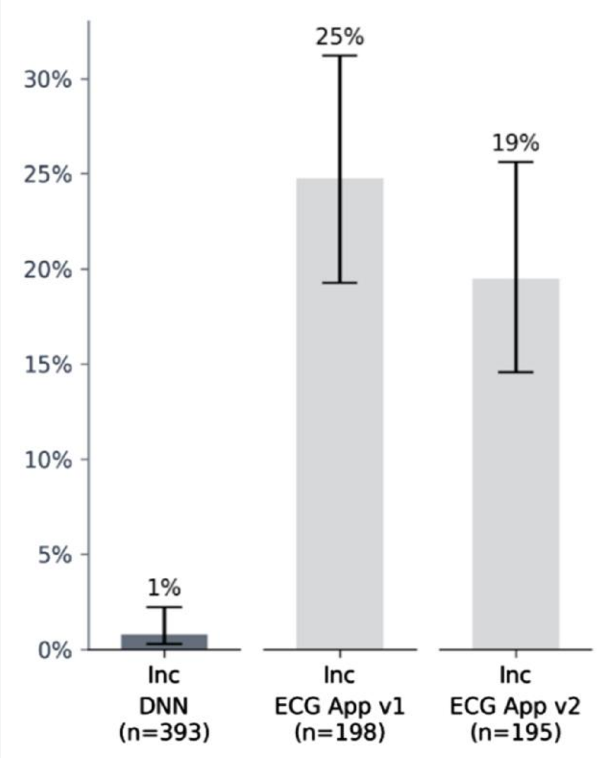


Figure 3 A comparison of deep neural network and ECG App inconclusive rates. The graph illustrates the rate of inconclusive diagnoses for the deep neural network, the ECG App v1 and ECG App v2: 1% (95% confidence interval 0–2%) vs. 25% (95% confidence interval 19–31%) and 19% (95% confidence interval 15–26%), respectively. DNN, deep neural network; Inc, inconclusive; Sens, sensitivity; Spec, specificity.

and a specificity of 95% (95% CI 91–97%), 78% (95% CI 71–84%), and 79% (95% CI 71–86%), respectively (*Figure 2A*; Supplementary material online, *Table S1*).

On the per-protocol population, i.e. excluding the inconclusive ECGs, the three algorithms correctly identified AA with a sensitivity of 91% (95% CI 85–95%), 61% (95% CI 44–75%), and 81% (95% CI 69–89%), respectively, and a specificity of 95% (95% CI 91–97%), 97% (95% CI 93–99%), and 96% (95% CI: 90–98%), respectively (*Figure 2B*; Supplementary material online, *Table S2*).

The number of recordings with an inconclusive diagnosis was significantly lower for the CDL algorithm than for the APL algorithm for both versions of ECG, App v1 and v2: 1% (95% CI 0–2%) vs. 25% (95% CI 19–31%) and 19% (95% CI 15–26%), respectively (*Figure 3*). Note that all the 12L ECGs were found interpretable by the physicians.

Discussion

Mobile health technology is a fast-evolving field that offers new solutions to reduce health-related risks ranging from descriptive monitoring tools to digital diagnostics.[21] The ubiquity of smartphones, combined with the wide spread of chronic diseases, makes the promise of mobile health particularly attractive. More recently, the coronavirus disease 2019 (COVID-19) pandemic further revealed the potential of wearable devices in times of health crisis.[22,23] Nonetheless, the clinical validity of these new technologies is generally lacking, and more rigorous studies are needed to assess effectiveness and safety.

Mobile health technologies often combine hardware with software. For complex problems, such as the identification of ECG rhythms and in particular the detection of AA, elaborated algorithms, such as machine learning–based algorithms, are often found more effective; however, safety can be a concern if they do not generalize well on the global population.

In this work, we aimed to assess whether the interpretation of SW ECGs by the CDL algorithm could accurately and reliably differentiate SR from AA in patients likely to be found in AF. We compared these automated interpretations with physician-interpreted 12L ECGs and found a clear agreement between the two. Atrial arrhythmia was correctly identified with 91% sensitivity and 95% specificity by the CDL algorithm. Only 3 out of 393 SW ECG were interpreted as inconclusive linked to a poor recording quality (i.e. noisy signal). This demonstrates the capability of the CDL algorithm to interpret SW ECGs for AA detection.

For almost a quarter of all SW ECGs (25 and 19% for AI Watch and AI Watch 2, respectively), the ECG App algorithms returned inconclusive diagnoses. The corresponding 12L ECGs were interpreted as 48 SR and 39 AA (55 and 45%, respectively), including 27 AF, 11 AFL, and 1 AT. In other words, the likelihood of an inconclusive diagnosis is slightly higher when the patient is known to be in AA rather than in SR.

The ECG App algorithms returned 25 false negatives (13 and 12 for v1 and v2, respectively; example provided in *Figure 4*). The detailed interpretation of the physicians revealed that they corresponded to 10 AF, 14 AFL, and 1 AT, while the proportion of patients found at the time of recording with these heart rhythms were 25, 8, and 1%, respectively. In this case, AFL was reportedly more challenging to detect than AF.

Most of the studies related to AF detection excluded patients with an inconclusive diagnosis (i.e. 'unclassified', 'no analysis') from the sensitivity and specificity analysis. Concerns have been raised that excluding them may lead to the exclusion of those patients with high heart rates or ambiguous results that are more relevant to AA detection.[24] This is particularly true when many patients are excluded from the analysis. Importantly, the DNN-based algorithm evaluated in this study succeeded in reaching a diagnosis on virtually all episodes (99%), with a performance equivalent to the one reached by the SW-embedded algorithm, but on 78% of the ECG tracings only. Note that by allowing more inconclusive diagnosis recordings (i.e. threshold adjustment), the sensitivity and specificity would likely increase for the DNN-based algorithm.

The performance differences between the DNN-based algorithm of Cardiologs and the algorithm embedded in the SW can be attributed to technical constraints. An algorithm with cloud-based processing on servers allows for greater computational power compared with the one embedded in an SW. However, an embedded algorithm requires less energy and can remain effective for simple analyses.

In our daily practice, an increasing number of patients have access to SW ECGs. Therefore, the enhancement of automatic interpretation through a new, more accurate DNN-based algorithm paves the way for improved diagnosis and, consequently, better outcomes for patients with rhythm disorders.

A potential clinical impact of this research is that a routine application of a DNN-guided SW ECG analysis could offer significant promise in AA diagnosis and reduce inconclusive diagnosis with potentially less SW ECG over-reading for physicians or technicians. The low number of false negatives will most likely prevent any delay in the initiation of anticoagulation treatment. The limited number of false positives should help decrease patient anxiety[25] and healthcare utilization. Expanding the use of SWs for monitoring purposes may require a secure and robust platform upon which SW ECG recordings can be received, stored, analysed by the DNN, and possibly visualized by physicians. Beyond AF, further clinical studies are needed to assess the performance of the CDL algorithm for the detection of other rhythm disorders.

Those studies have several limitations. Initially, the study cohort originated predominantly from a single tertiary centre, resulting in a

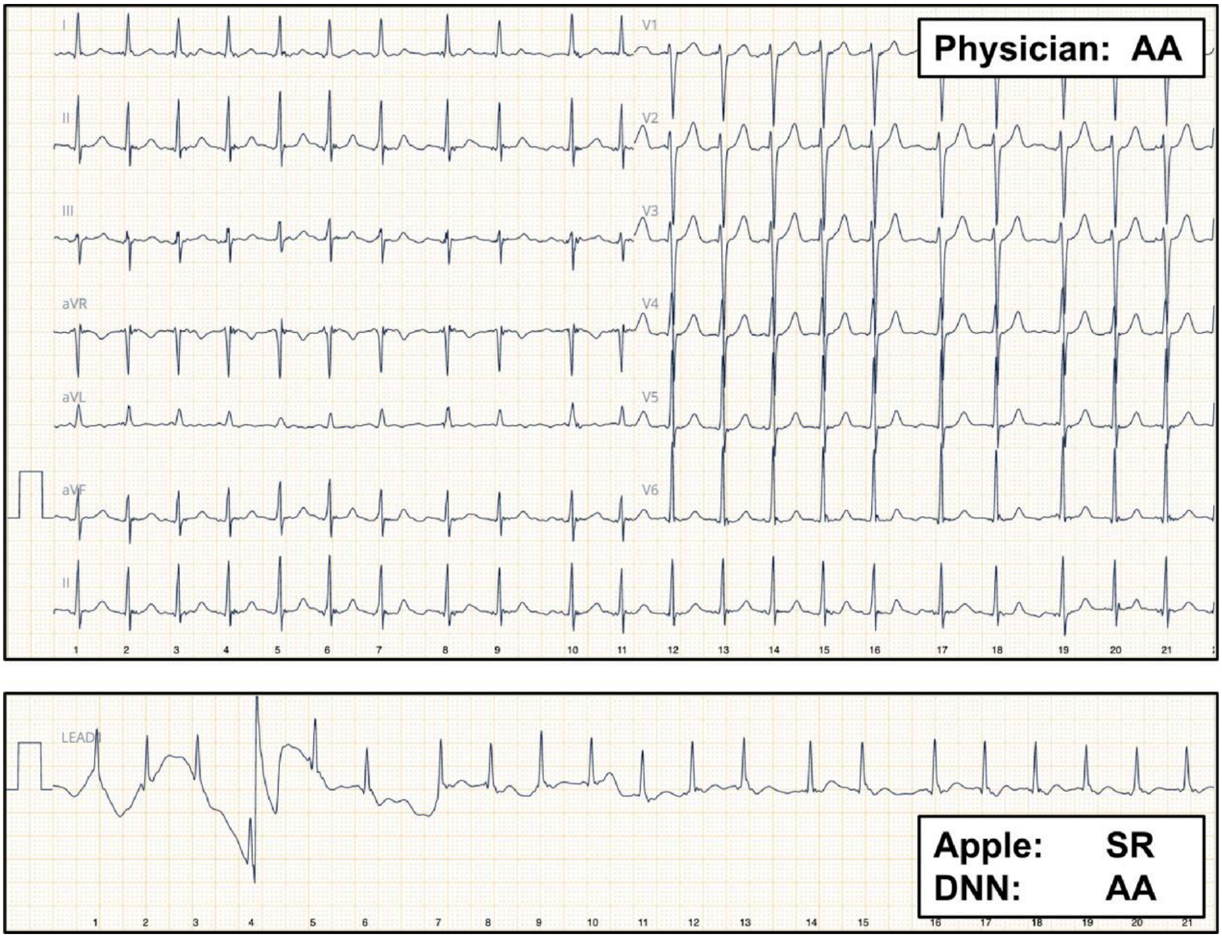


Figure 4 A example of misclassification from the ECG App. Here, ECG App interpreted the signal as sinus rhythm, whereas both the physician and the deep neural network classify this signal as an atrial arrhythmia. AA, atrial arrhythmias; DNN, deep neural network; SR, sinus rhythms.

monocentric representation. It is noteworthy that a significant proportion of the patients exhibited AFL, likely due to the high number of referrals for AFL ablation at this centre. However, it is important to emphasize that the intrinsic sensitivity and specificity of algorithms should not vary for the same pathology. Of note, only 29.2% of the enrolled patients were female. Participants were instructed on how to use the SW immediately prior to obtaining each recording. Their ability to record each strip was directly observed. As a result, the clarity of the recorded strip and the performance of the CDL algorithm may be less accurate in an outpatient or ambulatory setting. Note that only one SW model has been tested in this study, and the generalization to other SWs remains to be clinically validated. The performance of APL is dependent on the version of the ECG App. Finally, the CDL algorithm was developed for Holter ECGs, and better performances could be expected with a custom algorithm, i.e. one trained with SW ECG data.

We included persistent, paroxysmal, or permanent AF indiscriminately and solely assessed its presence at the time of the ECG recording. Further research is needed to explore potential variations in detection performance based on AF type in 30 s ECG recordings.

We acknowledge that the explainability of the model should be evaluated in future studies. The clinical acceptance of 'blackbox' DNN models in the absence of explainability remains to be studied in this application and could potentially hinder the adoption of AI-driven tools. However, we believe that efforts to significantly improve performances are an important factor in the acceptance of a new diagnostic tool by both patients and physicians.

Conclusions

Our findings demonstrate that using a DNN-based algorithm can improve the capacity to identify AA from SW ECG compared with the isolated utilization of the APL solution. The CDL AI platform has a low rate of inconclusive diagnosis, which makes it particularly suitable for AA detection in tertiary care hospitals.

While the embedded algorithms in the SW used in this study were able to identify a rhythm in only 78% of the recordings, the DNN-based algorithm managed to reach a diagnosis for 99% of the recording strips with similar performance. This could likely result in time savings for physicians, which, however, will have to be clinically evaluated.

Supplementary material

Supplementary material is available at *European Heart Journal – Digital Health*.

Acknowledgements

This clinical study would not have been possible without the partnership of its participants. The authors sincerely thank Christophe Gardella for useful comments and discussions on technical aspects. They also thank Leila Farid for her assistance in editing the final version of this manuscript.

Funding

This work was supported by funding from Cardiologs.

Conflict of interest: L.F. is a medical expert for the Cardiologs. A.P., Leila Farid, C.H., Christophe Gardella, and B.L. are or were employees of the Cardiologs. The remaining authors have no disclosures to report.

Data availability

The data underlying this article will be shared on reasonable request with the corresponding author.

References

1. Wolowacz SE, Samuel M, Brennan VK, Jasso-Mosqueda J-G, Van Gelder IC. The cost of illness of atrial fibrillation: a systematic review of the recent literature. *Europace* 2011; **13**:1375–1385.
2. Lippi G, Sanchis-Gomar F, Cervellin G. Global epidemiology of atrial fibrillation: an increasing epidemic and public health challenge. *Int J Stroke* 2021;**16**:217–221.
3. Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra M, Benjamin EJ, *et al.* Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 study. *Circulation* 2014;**129**:837–847.
4. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham study. *Stroke* 1991;**22**:983–988.
5. Healey JS, Connolly SJ, Gold MR, Israel CW, Van Gelder IC, Capucci A, *et al.* Subclinical atrial fibrillation and the risk of stroke. *N Engl J Med* 2012;**366**:120–129.
6. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, *et al.* 2020 ESC guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): the Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. *Eur Heart J* 2021;**42**:373–498.
7. January CT, Wann LS, Calkins H, Chen LY, Cigarroa JE, Cleveland JC, *et al.* 2019 AHA/ACC/HRS focused update of the 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation. *J Am Coll Cardiol* 2019;**74**:104–132.
8. Verma A, Champagne J, Sapp J, Essebag V, Novak P, Skanes A, *et al.* Discerning the incidence of symptomatic and asymptomatic episodes of atrial fibrillation before and after catheter ablation (DISCERN AF): a prospective, multicenter study. *JAMA Intern Med* 2013;**173**:149–156.
9. Quirino G, Giammaria M, Corbucci G, Pistelli P, Turri E, Mazza A, *et al.* Diagnosis of paroxysmal atrial fibrillation in patients with implanted pacemakers: relationship to symptoms and other variables. *Pacing Clin Electrophysiol* 2009;**32**:91–98.
10. Prasitlumkum N, Cheungpasitporn W, Chokesuwattanaskul A, Thangjui S, Thongprayoon C, Bathini T, *et al.* Diagnostic accuracy of smart gadgets/wearable devices in detecting atrial fibrillation: a systematic review and meta-analysis. *Arch Cardiovasc Dis* 2021;**114**:4–16.
11. Mannhart D, Lischer M, Knecht S, du Fay de Lavallaz J, Strebel I, Serban T, *et al.* Clinical validation of 5 direct-to-consumer wearable smart devices to detect atrial fibrillation. *JACC Clin Electrophysiol* 2023;**9**:232–242.
12. Sanna T, Diener H-C, Passman RS, Di Lazzaro V, Bernstein RA, Morillo CA, *et al.* Cryptogenic stroke and underlying atrial fibrillation. *N Engl J Med* 2014;**370**:2478–2486.
13. Marcus GM. The Apple Watch can detect atrial fibrillation: so what now? *Nat Rev Cardiol* 2020;**17**:135–136.
14. Smith SW, Rapin J, Li J, Fleureau Y, Fennell W, Walsh BM, *et al.* A deep neural network for 12-lead electrocardiogram interpretation outperforms a conventional algorithm, and its physician overread, in the diagnosis of atrial fibrillation. *Int J Cardiol Heart Vasc* 2019;**25**:100423.
15. Schuetz GM, Schlattmann P, Dewey M. Use of 3×2 tables with an intention to diagnose approach to assess clinical performance of diagnostic tests: meta-analytical evaluation of coronary CT angiography studies. *BMJ* 2012;**345**:e6717.
16. Abu-Alrub S, Strik M, Ramirez FD, Moussaoui N, Racine HP, Marchand H, *et al.* Smartwatch electrocardiograms for automated and manual diagnosis of atrial fibrillation: a comparative analysis of three models. *Front Cardiovasc Med* 2022;**9**:836375.
17. Mannhart D, Lischer M, Knecht S, du Fay de Lavallaz J, Strebel I, Serban T, *et al.* Clinical validation of 5 direct-to-consumer wearable smart devices to detect atrial fibrillation: BASEL wearable study. *JACC Clin Electrophysiol* 2023;**9**:232–242.
18. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, eds. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. Lecture Notes in Computer Science.* Cham: Springer International Publishing; 2015. p234–241.
19. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556 [cs]. 2015. https://doi.org/10.48550/arXiv.1409.1556, preprint: not peer reviewed.
20. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 2014;**15**:1929–1958.
21. Sim I. Mobile devices and health. *N Engl J Med* 2019;**381**:956–968.
22. Quer G, Radin JM, Gadaleta M, Baca-Motes K, Ariniello L, Ramos E, *et al.* Wearable sensor data and self-reported symptoms for COVID-19 detection. *Nat Med* 2021;**27**: 73–77.
23. Maille B, Wilkin M, Million M, Rességuier N, Franceschi F, Koutbi-Franceschi L, *et al.* Smartwatch electrocardiogram and artificial intelligence for assessing cardiac-rhythm safety of drug therapy in the COVID-19 pandemic. The QT-logs study. *Int J Cardiol* 2021;**331**:333–339.
24. Wegner FK, Kochhäuser S, Ellermann C, Lange PS, Frommeyer G, Leitz P, *et al.* Prospective blinded evaluation of the smartphone-based AliveCor Kardia ECG monitor for atrial fibrillation detection: the PEAK-AF study. *Eur J Intern Med* 2020;**73**:72–75.
25. Rosman L, Gehi A, Lampert R. When smartwatches contribute to health anxiety in patients with atrial fibrillation. *Cardiovasc Digit Health J* 2020;**1**:9–10.