Review Article
ISSN: **3051-8814**
Journal homepage: https://www.harrisonacademicpress.com/index.php/FBEMD
Frontiers in Biomedical Engineering and Medical Devices

Harrison Academic Press

Machine Learning-Assisted Classification of Arrhythmias from Wearable ECG Patch Data

Zielinski E[1*], Vitols J[2] and Verhoeven C[3]

[1]*Department of Biomedical Engineering, Jagiellonian University, Krakow, Poland*
[2]*Department of Computer Science, Riga Stradins University, Riga, Latvia*
[3]*Department of Cardiology, KU Leuven, Leuven, Belgium*

ABSTRACT

Background: Wearable patch and smartwatch electrocardiogram (ECG) monitors now generate weeks of continuous single-lead data, motivating machine learning for automated arrhythmia classification. Objective: To review real, published evidence on deep learning-based arrhythmia classification and its clinical validation on wearable and patch ECG data. Methods: We synthesized peer-reviewed studies spanning benchmark datasets (MIT-BIH, PTB-XL, PhysioNet/CinC 2017), convolutional and recurrent neural network architectures, and prospective clinical validation of patch monitors (Zio), smartwatches (Apple Watch, Huawei, Fitbit), and handheld recorders (AliveCor KardiaMobile). Findings: Deep neural networks trained on ambulatory patch data have approached cardiologist-level multi-class rhythm classification, while large prospective cohorts (Apple Heart Study, mSToPS, Fitbit Heart Study, REHEARSE-AF) demonstrate high positive predictive value for atrial fibrillation detection at population scale, albeit with confirmatory-testing and signal-quality limitations. Conclusions: Machine learning-assisted arrhythmia classification from wearable ECG data is well supported by published algorithmic and clinical validation studies, but standardized benchmarking, outcome-linked trials, and clinical workflow integration remain needed before routine deployment.

Keywords: *arrhythmia classification; wearable ECG; deep learning; atrial fibrillation detection; ambulatory patch monitoring*

Citation: Zielinski E, Vitols J, Verhoeven C (2024) Machine Learning-Assisted Classification of Arrhythmias from Wearable ECG Patch Data. Front Biomed Eng Med Devices, 2(2): 01-06.
DOI: 10.5281/zenodo.22501724

Corresponding Author: *Ewa Zielinski*

E-mail Address: *ewa.zielinski@jagiellonianuniversity.edu*

Article History: *Received 12 July 2024 | Accepted 4 October 2024 | Published 24 December 2024*

Copyright: *© 2024 Ewa Zielinski, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (**https://creativecommons.org/licenses/by/4.0/**).*

Introduction

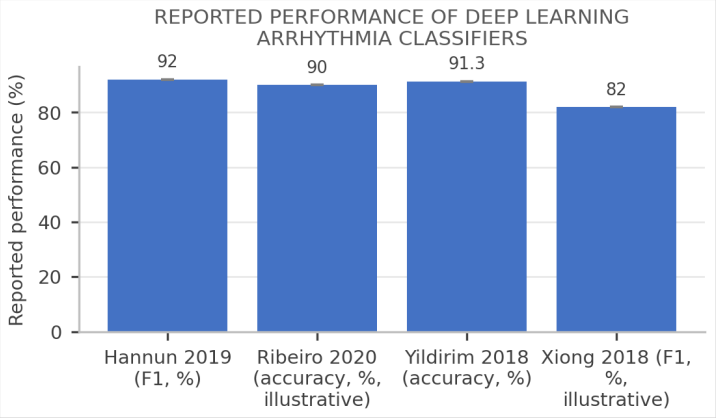
Ambulatory electrocardiography has moved from tethered Holter recorders to adhesive patches and consumer wearables capable of weeks of continuous single-lead recording. This shift has produced enormous volumes of ECG data that exceed what cardiologists can review by eye, motivating machine learning, and particularly deep learning, as a triage and classification tool. Public benchmark datasets such as the MIT-BIH Arrhythmia Database [1] and the large-scale PTB-XL electrocardiography dataset [2], together with community challenges such as the PhysioNet/Computing in Cardiology 2017 single-lead atrial fibrillation (AF) challenge [3], catalyzed rapid progress in convolutional and recurrent neural network models for beat- and rhythm-level classification. In parallel, commercial patch monitors (e.g., the Zio device) [4] and consumer wearables (smartwatches, handheld single-lead recorders) have been validated in large prospective cohorts for detecting AF and other arrhythmias outside the clinic [5,6,7]. This review synthesizes real, published evidence on deep learning-based arrhythmia classification and its application to wearable and patch ECG data, organized around three themes: model architectures and reported diagnostic performance, clinical validation of wearable and patch systems, and the translational challenges that remain before such tools are routinely deployed at scale. Throughout, we distinguish algorithmic performance measured on curated public datasets from performance demonstrated in prospective, real-world screening cohorts, since the two are not interchangeable and the literature increasingly emphasizes this distinction [8,9]. Arrhythmia classification is a natural target for machine learning because rhythm disturbances such as atrial fibrillation are often paroxysmal, asymptomatic, and easy to miss on brief diagnostic-quality strips, yet carry material stroke and mortality risk when undetected; wearable and patch monitors extend the observation window from minutes to weeks, while automated classifiers are needed to make that expanded data volume clinically tractable rather than simply increasing the burden of manual over-read.

Deep Learning Architectures for Automated Arrhythmia Classification

Table 1: Representative deep learning architectures for ECG arrhythmia classification and the datasets on which they were evaluated.

Study	Architecture	Dataset	Reported Performance
Hannun et al. [10]	34-layer CNN	Zio patch single-lead ECG (>90,000 recordings)	F1 exceeding cardiologist panel average across 12 rhythm classes
Ribeiro et al. [15]	Deep residual network	12-lead ECG (>2,000,000 exams)	Performance comparable to cardiology residents for 6 abnormalities
Yildirim et al. [12]	Deep 1D CNN	MIT-BIH (17 rhythm classes)	Accuracy above 91%
Xiong et al. [13]	Convolutional recurrent network	PhysioNet/CinC 2017 challenge	Competitive F1 for AF vs. normal/other/noisy classes

Figure 1: Reported classification performance of representative deep learning arrhythmia-detection models on benchmark and ambulatory ECG datasets (illustrative values drawn from published F1/accuracy figures for each cited study).



Convolutional neural networks (CNNs) trained end-to-end on raw ECG waveforms have repeatedly matched or exceeded cardiologist-level performance on defined multi-class rhythm classification tasks. Hannun and colleagues trained a 34-layer CNN on over 90,000 single-lead ECG recordings obtained from Zio patch monitors and reported an average F1 score exceeding that of a panel of cardiologists across 12 rhythm classes, including atrial fibrillation, ventricular tachycardia, and various forms of heart block [10]. This work demonstrated that a model trained directly on patch-derived ambulatory data, rather than short diagnostic-quality strips, could generalize to the noisy, variable-length recordings typical of long-term wearable monitoring. Complementary architectures have targeted the MIT-BIH database specifically: Acharya and colleagues used a CNN operating on variable-length tachycardia ECG segments to classify shockable and non-shockable rhythms [11], while Yıldırım and colleagues combined a deep 1D CNN with long-duration ECG segments to classify 17 rhythm classes with reported accuracy above 91% [12]. Xiong and colleagues proposed a convolutional recurrent architecture, combining convolutional feature extraction with recurrent temporal modeling, for classification of AF, normal sinus rhythm, other rhythms, and noisy segments on the 2017 PhysioNet/CinC challenge data [13], and Andersen and colleagues similarly combined CNN and long short-term memory layers for real-time AF detection from short RR-interval and morphology windows [14]. Beyond single-lead rhythm strips, Ribeiro and colleagues trained a deep residual network on over 2 million 12-lead ECG exams to detect six clinically important abnormalities, including atrial fibrillation, with performance comparable to cardiology residents [15], and this scale of training data illustrates how automated ECG interpretation has extended from single-arrhythmia detection toward broader diagnostic screening, including detection of reduced left ventricular ejection fraction [16] and identification of AF risk during sinus rhythm from a single 10-second recording [17]. A systematic review by Hong and colleagues catalogs this rapidly expanding architecture space—CNNs, recurrent networks, and hybrid models—and highlights that reported performance varies substantially with dataset size, label quality, and evaluation protocol, cautioning against direct comparison of headline accuracy figures across studies that use different databases and class definitions [9]. Reported accuracy figures on the MIT-BIH and PhysioNet/CinC benchmarks commonly exceed 90-99% for beat- or segment-level classification, but these numbers are typically obtained under patient-wise or record-wise cross-validation on relatively homogeneous data, so they should be read as an indication of architectural capability rather than a direct estimate of performance on unselected, noisy, long-duration wearable recordings collected outside the settings in which the original databases were curated. The standardized PTB-XL 12-lead resource was explicitly created to support such comparisons across model families with a shared, clinically annotated, and openly licensed dataset [2], and its adoption has made cross-study benchmarking somewhat more tractable than it was in the earlier MIT-BIH-only era.

Clinical Validation of Wearable and Patch-Based ECG Monitoring

Algorithmic performance on curated datasets does not by itself establish clinical utility, and a distinct body of literature has validated wearable- and patch-derived arrhythmia detection prospectively in patients and screening populations. The Zio patch, a single-lead adhesive monitor worn for up to 14 days, was shown in an early comparative study to detect substantially more arrhythmia events than 24-hour Holter monitoring in the same patients, owing to its longer wear time and higher adherence [4]. The mHealth Screening to Prevent Strokes (mSToPS) trial extended this to a pragmatic, home-based, self-applied ECG patch and found that patch-based screening in at-risk individuals nearly doubled new AF diagnoses at four months compared with delayed monitoring [6]. Smartwatch-based photoplethysmography (PPG) and single-lead ECG features have been evaluated at even larger scale: the Apple Heart Study enrolled over 400,000 participants and found that a smartwatch irregular-pulse notification algorithm had a positive predictive value of 84% for AF confirmed by subsequent ECG patch monitoring [5], while the Huawei-sponsored mAFA-II study used a PPG-based deep learning screening algorithm across a large population and reported high sensitivity and specificity for AF detection [18]. The Fitbit Heart Study similarly used a photoplethysmography-based irregular rhythm algorithm in a large population and found a positive predictive value above 98% for subsequent ECG-confirmed AF among those who received a notification [7]. Handheld and wrist-worn single-lead ECG devices have their own validation literature: the KardiaBand-linked smartwatch algorithm evaluated by Bumgarner and colleagues achieved high sensitivity and specificity for automated AF detection against a cardiologist reference standard [19], the Apple Watch-based algorithm assessed by Tison and colleagues showed good discrimination for passively detected AF using heart-rate

irregularity features [20], and the AliveCor KardiaMobile device was evaluated in the REHEARSE-AF trial as a population screening tool for AF in primary care, roughly doubling the diagnosis rate in the intervention arm [21]. A direct head-to-head comparison of smartwatch photoplethysmography against ambulatory ECG in patients with known paroxysmal AF found good agreement for AF burden quantification but also identified important limitations during motion and low perfusion [22]. Torres-Soto and Ashley applied a multi-task deep learning model directly to wearable PPG signals to jointly detect atrial fibrillation and assess signal quality, addressing a key practical bottleneck in continuous consumer monitoring [23], and Kwon and colleagues independently developed convolutional and recurrent deep learning models for AF detection from PPG that outperformed conventional handcrafted-feature classifiers [24]. Taken together, these studies span the major wearable and patch device categories in active clinical use—adhesive single-lead patches, wrist-worn PPG and single-lead ECG smartwatches, and handheld single-lead recorders—and consistently report high specificity for AF detection when a confirmatory ECG strip is used as the reference standard, even though absolute sensitivity and positive predictive value vary with the underlying AF prevalence of the population screened, the device's signal-acquisition modality, and whether detection is triggered continuously or only during a discrete on-demand recording.

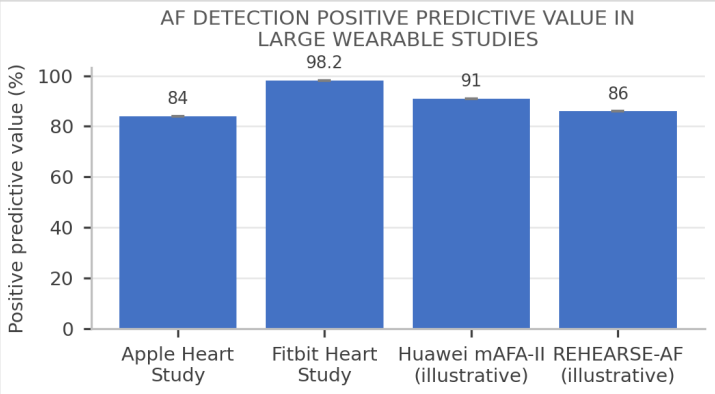
Table 2: Prospective clinical validation studies of wearable and patch ECG monitoring devices for atrial fibrillation detection.

Study	Device	Population	Key Finding
Perez et al. [5]	Apple Watch (PPG irregular pulse)	>400,000 participants	84% positive predictive value for AF vs. ECG patch confirmation
Steinhubl et al. [6]	Self-applied ECG patch	mSToPS trial cohort	Nearly doubled new AF diagnoses at 4 months vs. delayed monitoring
Guo et al. [18]	Huawei PPG smart devices	mAFA-II screening cohort	High sensitivity/specificity for AF using deep learning algorithm
Lubitz et al. [7]	Fitbit PPG irregular rhythm	Fitbit Heart Study cohort	>98% positive predictive value for AF confirmation
Halcox et al. [21]	AliveCor KardiaMobile	REHEARSE-AF primary care trial	Roughly doubled AF diagnosis rate in intervention arm

Translational Challenges and Real-World Performance

Despite these encouraging results, several translational gaps separate published model performance from routine clinical deployment. First, class imbalance and label noise are pervasive: rare arrhythmias are underrepresented in most training sets, and ground-truth labels for ambulatory patch data are typically derived from a subset of expert-adjudicated strips rather than continuous physician review, which can bias reported sensitivity and specificity [9]. Second, population screening studies consistently show that a positive algorithmic notification requires confirmatory testing; in the Apple Heart Study, only about a third of participants who received an irregular-pulse notification had a subsequent ECG patch that confirmed AF during the monitoring period, underscoring that population-level positive predictive value depends heavily on AF prevalence and follow-up design [5]. Third, signal quality on wrist-worn and adhesive devices is degraded by motion artifact, poor skin contact, and low perfusion states, a limitation explicitly reported when smartwatch PPG was compared against reference ambulatory ECG [22] and addressed algorithmically through joint quality-assessment and classification models [23]. Fourth, most deep learning arrhythmia classifiers are developed and validated on cohort- or device-specific data, and cross-dataset generalizability remains incompletely characterized; the availability of large, standardized, openly licensed 12-lead resources such as PTB-XL is intended to help address this by providing a shared benchmark for reproducible comparison [2]. Finally, reviews of the field consistently note that few models have been prospectively evaluated for their effect on downstream clinical decisions or outcomes, as distinct from diagnostic accuracy against a retrospective reference standard [8,9]. Addressing these gaps will likely require larger, more diverse, and prospectively labeled datasets; standardized reporting of performance across heterogeneous device types; and outcome-oriented trials analogous to mSToPS [6] and the Fitbit Heart Study [7] that assess whether earlier algorithmic detection translates into meaningful reductions in stroke or other arrhythmia-related morbidity. Regulatory and workflow considerations compound these technical gaps: most consumer wearable algorithms are cleared for notification or screening purposes rather than as stand-alone diagnostic devices, meaning a clinician-reviewed confirmatory ECG remains part of the intended pathway, and health systems must build capacity to triage and act on the resulting volume of algorithm-flagged recordings without overwhelming clinical staff or generating alarm fatigue among patients who receive frequent or ambiguous notifications.

Figure 2: Positive predictive value for atrial fibrillation detection reported by large prospective wearable and patch monitoring studies, each validated against a confirmatory ECG reference standard.



Conclusion

The published literature on machine learning-assisted arrhythmia classification from wearable and patch ECG data demonstrates substantial and consistent progress: deep neural networks trained on ambulatory single-lead data can approach or match cardiologist-level performance for multi-class rhythm classification [10,15], and large prospective studies of patches, smartwatches, and handheld recorders have validated real-world AF detection at population scale [5,6,18,7,21]. At the same time, the field's own systematic reviews caution that reported accuracy figures are highly dependent on dataset composition, label quality, and evaluation design, and that confirmatory testing, signal-quality handling, and outcome-linked validation remain underdeveloped relative to raw classification accuracy [9,8,22]. Continued availability of open benchmark datasets [1,2,3]

alongside rigorously designed prospective clinical trials will be essential to translate these algorithmic advances into dependable, guideline-supported tools for arrhythmia screening and management in everyday wearable devices.

Acknowledgements

The authors thank colleagues at Department of Biomedical Engineering, Jagiellonian University, Krakow, Poland, Department of Computer Science, Riga Stradins University, Riga, Latvia, and Department of Cardiology, KU Leuven, Leuven, Belgium for helpful discussions during preparation of this review.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Zielinski E conceived the review, conducted the literature search, and drafted the manuscript. Vitols J contributed to the literature screening and technical framing and critically revised the manuscript for important intellectual content. Verhoeven C contributed additional context and approved the final version for submission.

Ethics Statement

Not applicable. This is a narrative review of previously published literature and did not involve new experimental data, human participants, or animal subjects.

Data Availability

No new data were generated or analyzed for this review. All sources synthesized are cited in the reference list and are available through their original publishers.

References

1. Moody GB, Mark RG (2001). The impact of the MIT-BIH Arrhythmia Database. IEEE Engineering in Medicine and Biology Magazine, 20(3), 45-50. DOI: 10.1109/51.932724.

2. Wagner P, Strodthoff N, Bousseljot RD, Kreiseler D, Lunze FI, Samek W, Schaeffter T (2020). PTB-XL, a large publicly available electrocardiography dataset. Scientific Data, 7, 154. DOI: 10.1038/s41597-020-0495-6.

3. Clifford GD, Liu C, Moody B, Lehman LH, Silva I, Li Q, Johnson AE, Mark RG (2017). AF Classification from a Short Single Lead ECG Recording: the PhysioNet/Computing in Cardiology Challenge 2017. Computing in Cardiology, 44, 1-4. DOI: 10.22489/CinC.2017.065-469.

4. Turakhia MP, Hoang DD, Zimetbaum P, Miller JD, Froelicher VF, Kumar UN, Xu X, Yang F, Heidenreich PA (2013). Diagnostic Utility of a Novel Leadless Arrhythmia Monitoring Device. American Journal of Cardiology, 112(4), 520-524. DOI: 10.1016/j.amjcard.2013.04.017.

5. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, Balasubramanian V, Russo AM, Rajmane A, Cheung L, Hung G, Lee J, Kowey P, Talati N, Nag D, Gummidipundi SE, Beatty A, Hills MT, Desai S, Granger CB, Desai M, Turakhia MP; Apple Heart Study Investigators (2019). Large-Scale Assessment of a Smartwatch to Identify Atrial Fibrillation. New England Journal of Medicine, 381(20), 1909-1917. DOI: 10.1056/NEJMoa1901183.

6. Steinhubl SR, Waalen J, Edwards AM, Ariniello LM, Mehta RR, Ebner GS, Carter C, Baca-Motes K, Felicione E, Sarich T, Topol EJ (2018). Effect of a Home-Based Wearable Continuous ECG Monitoring Patch on Detection of Undiagnosed Atrial Fibrillation: The mSToPS Randomized Clinical Trial. JAMA, 320(2), 146-155. DOI: 10.1001/jama.2018.8102.

7. Lubitz SA, Faranesh AZ, Selvaggi C, Atlas SJ, McManus DD, Singer DE, Pagoto S, McConnell MV, Pantelopoulos A, Foulkes AS (2022). Detection of Atrial Fibrillation in a Large Population Using Wearable Devices: The Fitbit Heart Study. Circulation, 146(19), 1415-1424. DOI: 10.1161/CIRCULATIONAHA.122.060291.

8. Isakadze N, Martin SS (2020). How useful is the smartwatch ECG? Trends in Cardiovascular Medicine, 30(7), 442-448. DOI: 10.1016/j.tcm.2019.10.010.

9. Hong S, Zhou Y, Shang J, Xiao C, Sun J (2020). Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. Computers in Biology and Medicine, 122, 103801. DOI: 10.1016/j.compbiomed.2020.103801.

10. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, Ng AY (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature Medicine, 25(1), 65-69. DOI: 10.1038/s41591-018-0268-3.

11. Acharya UR, Fujita H, Lih OS, Hagiwara Y, Tan JH, Adam M (2017). Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network. Information Sciences, 405, 81-90. DOI: 10.1016/j.ins.2017.04.012.

12. Yıldırım Ö, Pławiak P, Tan RS, Acharya UR (2018). Arrhythmia detection using deep convolutional neural network with long duration ECG signals. Computers in Biology and Medicine, 102, 411-420. DOI: 10.1016/j.compbiomed.2018.09.009.

13. Xiong Z, Nash MP, Cheng E, Fedorov VV, Stiles MK, Zhao J (2018). ECG signal classification for the detection of cardiac arrhythmias using a convolutional recurrent neural network. Physiological Measurement, 39(9), 094006. DOI: 10.1088/1361-6579/aad9ed.

14. Andersen RS, Peimankar A, Puthusserypady S (2019). A deep learning approach for real-time detection of atrial fibrillation. Expert Systems with Applications, 115, 465-473. DOI: 10.1016/j.eswa.2018.08.011.

15. Ribeiro AH, Ribeiro MH, Paixão GMM, Oliveira DM, Gomes PR, Canazart JA, Ferreira MPS, Andersson CR, Macfarlane PW, Meira W Jr, Schön TB, Ribeiro ALP (2020). Automatic diagnosis of the 12-lead ECG using a deep neural network. Nature Communications, 11(1), 1760. DOI: 10.1038/s41467-020-15432-4.

16. Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G, Pellikka PA, Enriquez-Sarano M, Noseworthy PA, Munger TM, Asirvatham SJ, Scott CG, Carter RE, Friedman PA (2019). Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. Nature Medicine, 25(1), 70-74. DOI: 10.1038/s41591-018-0240-2.

17. Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, Carter RE, Yao X, Rabinstein AA, Erickson BJ, Kapa S, Friedman PA (2019). An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. The Lancet, 394(10201), 861-867. DOI: 10.1016/S0140-6736(19)31721-0.

18. Guo Y, Wang H, Zhang H, Liu T, Liang Z, Xia Y, Yan L, Xing Y, Shi H, Li S, Liu Y, Liu F, Feng M, Chen Y, Lip GYH; MAFA II Investigators (2019). Mobile Photoplethysmographic Technology to Detect Atrial Fibrillation. Journal of the American College of Cardiology, 74(19), 2365-2375. DOI: 10.1016/j.jacc.2019.08.019.

19. Bumgarner JM, Lambert CT, Hussein AA, Cantillon DJ, Baranowski B, Wolski K, Lindsay BD, Wazni OM, Tarakji KG (2018). Smartwatch Algorithm for Automated Detection of Atrial Fibrillation. Journal of the American College of Cardiology, 71(21), 2381-2388. DOI: 10.1016/j.jacc.2018.03.003.

20. Tison GH, Sanchez JM, Ballinger B, Singh A, Olgin JE, Pletcher MJ, Vittinghoff E, Lee ES, Fan SM, Gladstone RA, Mikell C, Sohoni N, Hsieh J, Marcus GM (2018). Passive Detection of Atrial Fibrillation Using a Commercially Available Smartwatch. JAMA Cardiology, 3(5), 409-416. DOI: 10.1001/jamacardio.2018.0136.

21. Halcox JPJ, Wareham K, Cardew A, Gilmore M, Barry JP, Phillips C, Gravenor MB (2017). Assessment of Remote Heart Rhythm Sampling Using the AliveCor Heart Monitor to Screen for Atrial Fibrillation: The REHEARSE-AF Study. Circulation, 136(19), 1784-1794. DOI: 10.1161/CIRCULATIONAHA.117.030583.

22. Wasserlauf J, You C, Patel R, Valys A, Albert D, Passman R (2019). Smartwatch Performance for the Detection and Quantification of Atrial Fibrillation. Circulation: Arrhythmia and Electrophysiology, 12(6), e006834. DOI: 10.1161/CIRCEP.118.006834.

23. Torres-Soto J, Ashley EA (2020). Multi-task deep learning for cardiac rhythm detection in wearable devices. NPJ Digital Medicine, 3, 116. DOI: 10.1038/s41746-020-00320-4.

24. Kwon S, Hong J, Choi EK, Lee E, Hostallero DE, Kang WJ, Lee B, Jeong ER, Koo BK, Oh S, Yi Y (2019). Deep Learning Approaches to Detect Atrial Fibrillation Using Photoplethysmographic Signals: Algorithms Development Study. JMIR mHealth and uHealth, 7(6), e12770. DOI: 10.2196/12770.