

Original Paper

# Deep Neural Networks for Automatic Atrial Fibrillation Detection Using Long-Term Ambulatory Electrocardiography: Retrospective Diagnostic Accuracy Study

Jagdeep Sedha<sup>1,2\*</sup>, MD; Jukka A Lipponen<sup>3\*</sup>, MSc, PhD; Antti Aho<sup>1</sup>, MD; Helena Jäntti<sup>4</sup>, MD, PhD; Onni E Santala<sup>1,4</sup>, MD, PhD; Tomi P Laitinen<sup>1,5</sup>, MD, PhD; Tiina M Laitinen<sup>3</sup>, MSc, PhD; Jari Halonen<sup>1,2</sup>, MD, PhD; Mika P Tarvainen<sup>3</sup>, MSc, PhD; Eemu-Samuli Seljola<sup>1,6,7</sup>, MD, PhD; Noora S Naukkarinen<sup>1</sup>, MD; Olli Rantula<sup>1,2</sup>, MD, MSc; Tuomas T Rissanen<sup>1,8</sup>, MD, PhD; Juha E K Hartikainen<sup>1,2</sup>, MD, PhD; Tero J Martikainen<sup>9</sup>, MD, PhD

<sup>1</sup>School of Medicine, Faculty of Health Sciences, University of Eastern Finland, Kuopio, Finland

<sup>2</sup>Heart Center, Kuopio University Hospital, Kuopio, Finland

<sup>3</sup>Department of Technical Physics, University of Eastern Finland, Kuopio, Finland

<sup>4</sup>Department of Anesthesiology and Intensive Care, North Karelia Central Hospital, Joensuu, Finland

<sup>5</sup>Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio, Finland

<sup>6</sup>Mikkeli Central Hospital, Mikkeli, Finland

<sup>7</sup>University of Helsinki, Helsinki, Finland

<sup>8</sup>Heart Center, North Karelia Central Hospital, Joensuu, Finland

<sup>9</sup>Department of Prehospital Emergency Care, Acute Services, Kuopio, Finland

\*these authors contributed equally

## Corresponding Author:

Jagdeep Sedha, MD

School of Medicine, Faculty of Health Sciences, University of Eastern Finland

Yliopistonranta 8

Kuopio 70210

Finland

Phone: 358 443358523

Email: [jagdeep881@gmail.com](mailto:jagdeep881@gmail.com)

## Abstract

**Background:** Atrial fibrillation (AF), the most prevalent cardiac arrhythmia, affects 2% to 4% of the global adult population and is associated with an increased risk of stroke. Early diagnosis of AF and atrial flutter (AFL) is crucial due to their association with stroke risk and the challenge posed by their often asymptomatic and episodic nature. Traditional electrocardiogram (ECG) interpretation requires substantial expert input and can be challenging, especially with poor-quality ECGs.

**Objective:** This study aimed to evaluate the performance of a deep neural network (DNN) model in detecting AF/AFL from a large, heterogeneous set of long-term ambulatory ECG recordings, including clinical data collected over 6 months at a university hospital, and assess its effectiveness in a setting reflecting the diversity and complexity of real-world clinical data.

**Methods:** The research combined public datasets totaling 10,248 patients, ECG data from our previous studies (648 patients), and authentic long-term ECG recordings from 4346 patients at Kuopio University Hospital for development of the DNN model. Its clinical accuracy and generalizability were assessed using a separate test dataset consisting of 1039 pseudonymized long-term ECG recordings from 1010 patients, all thoroughly reviewed and annotated by experts.

**Results:** The DNN model demonstrated high effectiveness, achieving 96.4% sensitivity and more than 99.99% specificity for time-level AF and AFL detection. At the recording level, it identified AF and AFL with 100% sensitivity and 98.77% specificity, producing false positives in only 1.2% (11/897) of recordings, of which 81.8% (9/11) had other non-AF/AFL arrhythmias. The model maintained high performance across diverse patient characteristics, including varying ages, comorbidities, coexisting arrhythmias, and poor-quality ECG recordings.

**Conclusions:** The results demonstrate that the proposed DNN model may support automated screening for AF and AFL in long-term ambulatory ECG recordings and may reduce manual review workload in clinical practice.

**Keywords:** atrial fibrillation; deep neural networks; electrocardiogram; electrocardiogram analysis; ECG analysis; cardiac arrhythmia detection; artificial intelligence; AI; machine learning

## Introduction

Atrial fibrillation (AF) is the most prevalent cardiac arrhythmia, affecting 2% to 4% of the global adult population [1]. AF, as well as atrial flutter (AFL), is associated with an increased risk of stroke and substantial health care costs [1]. The prevalence of AF and AFL is predicted to increase 2.3-fold by 2060 due to the aging population [1-4]. The diagnosis of AF and AFL requires electrocardiogram (ECG) documentation. However, AF and AFL are often intermittent and asymptomatic. Thus, standard 24- or 48-hour ECG recordings are often insufficient for detecting AF and AFL, making early detection and prevention challenging [5-7].

AF is characterized by disorganized atrial electrical activity, resulting in an irregular ventricular response and absence of discrete P waves on surface ECG recordings [8-11]. In contrast, AFL is a more organized atrial arrhythmia, typically driven by a macroreentrant circuit and associated with regular flutter waves [12]. Although AF and AFL share common risk factors and may coexist, their ECG manifestations differ in ways that are relevant for automated rhythm detection [9-11].

From an automated ECG analysis perspective, AFL presents a particular challenge due to variable atrioventricular (AV) conduction. Irregular ventricular responses in AFL can closely resemble those of AF on surface ECG, leading to frequent misclassification by both rule-based and deep learning-based methods [13-15]. As a result, approaches relying primarily on RR interval irregularity may struggle to distinguish AFL with variable AV conduction from AF [6-8].

AF may be classified as paroxysmal or persistent based on episode duration and termination characteristics [16,17]. Paroxysmal AF is intermittently present and, therefore, more likely to be missed in long-term ambulatory ECG recordings, whereas persistent AF is continuously detectable [16,17].

In addition to true atrial arrhythmias, several cardiac rhythm disturbances such as supraventricular ectopic beats (SVEBs), AV conduction abnormalities, and atrial tachycardias may mimic AF or AFL on surface ECG, increasing the risk of false-positive detections in automated ECG analysis [18,19].

Recently, longer ECG recordings lasting a week or even longer have become available owing to the emergence of novel ECG recording devices such as wearable single-channel ECG devices [20,21]. However, longer ECG recordings typically also encompass 10% to 30% of noisy data and artifacts [22,23] arising from motion-related interference or muscular activity [20,21]. Consequently, contemporary analysis techniques require substantial manual effort from health care personnel, particularly when analyzing noisy ECG data [20,21]. Given the growing incidence of cardiac arrhythmias and the increasing duration of recordings, there

is an unmet demand for automated approaches capable of accurate detection of cardiac arrhythmias [20,21].

In contrast to the traditional stepwise method of automated arrhythmia analysis involving signal preprocessing, feature extraction, and classification [24], our study focused on the application of deep neural networks (DNNs). These networks offer comprehensive, end-to-end processing of raw ECG data [21], potentially uncovering new diagnostic features from large volumes of unprocessed data.

Several end-to-end DNN models have been proposed for ECG-based AF detection [25-34] and arrhythmia analysis [7, 21]. Most of these models have been developed and evaluated using small public databases with a limited number of recordings. This limitation hampers the development of models that are generalizable, resulting in poor performance when applied to unseen data. While a few recent studies have applied DNNs to long-term ambulatory recordings, they have not assessed performance at the full recording or patient level in real-world clinical datasets [35,36].

Furthermore, only a few models have been designed and tested using a substantial amount of ambulatory data, which are often affected by various artifacts, other rhythm disorders, and noise that can mimic AF [25,37,38]. The application of these DNN models for AF screening in ambulatory, free-living conditions remains an open challenge [11]. Even the most advanced DNN methods developed with substantial amounts of ECG data [28-30,35,36] still face a significant issue: a high rate of false positives when applied to patient-operated ambulatory ECGs collected under real-world conditions.

In this study, we used an extensive set of authentic clinical data and specifically investigated the capability of a DNN model for automatic identification of patients with AF and AFL based on 24- and 48-hour ECG recordings while highlighting pertinent ECG segments exceeding 30 seconds for physician review.

## Methods

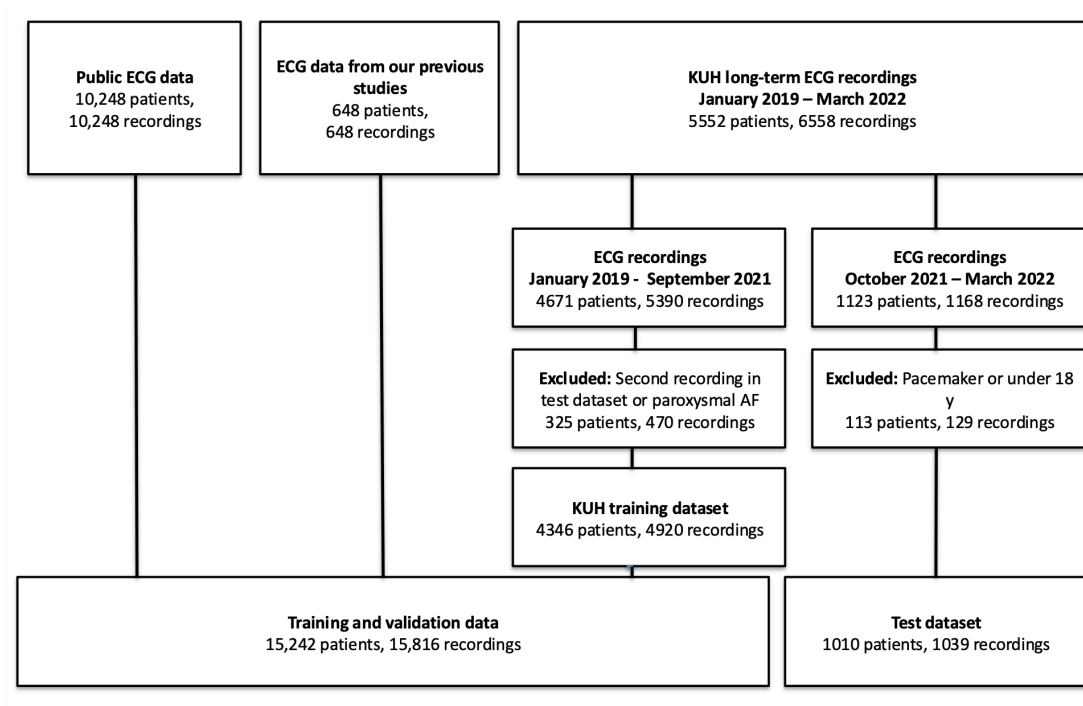
### *Training and Validation Data*

The training and validation dataset comprised 10,896 recordings from 10,896 patients. Of these, 10,248 recordings were gathered from a public ECG dataset, while the remaining 648 recordings were obtained from our previous studies designed for training and validation of machine learning methods for arrhythmia detection (Figure 1, Table 1). The flowchart illustrating the training and validation data is presented in the Results section. Public datasets provided AF and AFL labels that were used as is, whereas ECG recordings from previous studies were annotated by trained physicians. Additionally, the training and validation dataset included samples from long-term ECG recordings at Kuopio University Hospital (KUH). The KUH training dataset encompasses

recordings from January 2019 to September 2021 (4920 recordings from 4346 patients). Clinical ECG data from KUH were labeled according to standard clinical practice and analyzed using the Medilog Darwin software (Schiller). During the clinical review process, arrhythmic events, including AF and AFL, were identified and verified by experienced ECG analysis specialists. In addition, ectopic beats were classified as ventricular ectopic beats or SVEBs and verified during expert review. ECG segments in which

all 3 channels were too noisy for reliable identification of QRS complexes were labeled as nonanalyzable. Recordings from patients with paroxysmal AF were excluded from the training and validation data and moved to the KUH test dataset (see below). In addition, patients who were identified in the KUH test dataset were excluded from the KUH training and validation dataset. Across all training datasets, including public datasets, the total number of patients was 15,242, of whom 3180 (20.9%) had AF or AFL.

**Figure 1.** Study flowchart for electrocardiogram (ECG) recordings illustrating the division into training and validation data and a separate test dataset, with exclusions noted. Patient counts in the period-specific KUH subsets are not mutually exclusive because some patients had ECG recordings in both time periods. These patients are counted in both period-specific boxes but only once in the total KUH patient count. AF: atrial fibrillation; KUH: Kuopio University Hospital.



**Table 1.** Overview of training and validation datasets detailing the number of electrocardiogram (ECG) recordings, the number of patients per dataset, and the duration of ECG recordings for each dataset.

Database	Number of recordings (n=15,816), n (%)	Number of patients (n=15,242), n (%)	Recording duration
KUH <sup>a</sup> training dataset	4920 (31.1)	4346 (28.5)	24 h
ECG data from our previous studies			
Arrhythmia detection and 1-time ECG [39,40]	461 (2.9)	461 (3.0)	0.5-1 h
AFIB24h <sup>b</sup> dataset [41]	187 (1.2)	187 (1.2)	24 h
Public ECG data			
Long-Term AF <sup>c</sup> Database [42]	84 (0.5)	84 (0.6)	24 h
Chapman University dataset [43]	10,000 (63.2)	10,000 (65.6)	10 s
Smart Health for Assessing the Risk of Events via ECG dataset Database [44]	139 (0.9)	139 (0.9)	24 h
MIT <sup>d</sup> -BIH <sup>e</sup> Atrial Fibrillation Database [45]	25 (0.2)	25 (0.2)	10 h

<sup>a</sup>KUH: Kuopio University Hospital.

<sup>b</sup>AFIB24h: Atrial Fibrillation Detection: 24 Hour Study.

<sup>c</sup>AF: atrial fibrillation.

<sup>d</sup>MIT: Massachusetts Institute of Technology.

<sup>e</sup>BIH: Beth Israel Hospital.

Altogether, the training and validation data contained ECG samples from 15,242 patients obtained using various ECG

recorders and from different geographical areas. A randomly selected 20% of the training dataset recordings were used as

a validation dataset to perform DNN hyperparameter tuning. The training, validation, and test datasets had separate sets of patients. The initial split was performed at the recording level. Subsequently, all recordings from patients appearing in the test set were removed from the validation or training sets, and recordings from patients appearing in the validation set were removed from the training set. As a result, the final training, validation, and test datasets had mutually exclusive patient groups, whereas the test dataset could contain multiple recordings from the same patient. Public datasets were derived from independently collected cohorts in the United States, China, and Italy, whereas the local dataset was collected at KUH. Given the independent origins of these datasets, patient overlap between sources was not expected, thereby reducing the risk of data leakage across training, validation, and test sets.

### Test Dataset

Clinical accuracy and the generalizability of the method were tested using 1039 retrospective pseudonymized long-term ECG recordings from 1010 patients (test dataset) conducted at KUH between October 2021 and March 2022. The exclusion criteria were pacemaker use ( $n=24$ , 2.3% of the recordings) and patients under 18 years ( $n=105$ , 10.1% of the recordings). Arrhythmias underwent electrophysiologist review including arrhythmia classification identified during clinical practice analyzed using the Medilog Darwin software. In addition, segments whose 3 channels were too noisy for identification of QRS complexes were assessed as nonanalyzable. ECG recordings in the test datasets underwent an additional rhythm annotation procedure where researchers verified and corrected (if needed) on a second-by-second basis all AF and AFL segments. Rhythm annotations made during clinical practice were used as a starting point. The AF class combined both AF and AFL rhythms. Medication data were obtained through automated extraction from the hospital electronic health record system at KUH.

### Development of the DNN Model

We developed a convolutional DNN model specifically designed for the detection of AF and AFL. We selected a residual network as the core architecture, incorporating shortcut connections between every second convolutional layer to address the vanishing gradient problem [46,47]. A similar residual network has been successfully applied to an ECG arrhythmia classification task before [21].

The model processes a 10-second segment of raw ECG data as input, consisting of 3 channels sampled at 125 Hz, and requires no additional patient-related information. The model classifies the 10-second ECG input into 3 categories: AF or AFL, non-AF/AFL, or nonanalyzable. Segment-level classification was performed using the maximum of the output without class-specific probability thresholds. The number of convolutional filters was doubled after every 4 convolutional layers [21]. Prior to each convolutional layer,

batch normalization and a rectified linear unit activation were applied [48].

The network weights were initialized using the initialization method by He et al [49]. For optimization, the Adam optimizer [50] was used with a batch size of 128 and an initial learning rate of  $10^{-3}$ . Additionally, dropout with a probability of 0.1 was used to prevent overfitting [51]. The model was trained using a categorical cross-entropy loss. To improve training efficiency, the learning rate was reduced by a factor of 10 if the validation loss did not improve for 2 consecutive epochs.

Model hyperparameters—including the number of convolutional layers (search space: 17 to 37), number of filters (8 to 32), filter size, and dropout rate (0 to 0.4)—were optimized using grid search. The model with the lowest validation loss during training was selected for final evaluation.

The final optimized network consists of 37 convolutional layers. The convolutional layers initially use 16 filters with a filter length of 12, and the number of filters increases after every 4 layers. A schematic overview of the DNN architecture is presented in Figure S1 in [Multimedia Appendix 1](#).

The model output was subsequently postprocessed to account for the temporal continuity of detected rhythms. AF detections shorter than 30 seconds were excluded as clinically insignificant in accordance with current European Society of Cardiology guidelines defining clinically relevant AF episodes as lasting at least 30 seconds [2]. Closely spaced, clinically significant AF and AFL segments were merged into unified episodes using a 4-minute window to enhance detection sensitivity and provide more meaningful information for clinical interpretation.

### DNN Model Evaluation

All 1039 long-term ambulatory recordings in the test dataset were segmented into fixed, nonoverlapping 10-second windows and input into the DNN model. Each segment was scaled to a fixed range between  $-1$  and  $1$  to reduce amplitude variability across recordings. The DNN produced 1 rhythm classification per 10-second segment (AF or AFL, non-AF/AFL, or nonanalyzable) based on signal quality and rhythm characteristics, with no segments excluded from analysis. As ECG recordings are annotated to true classes second by second, model accuracy can be assessed at the time and recording level. Time-level accuracy of AF and AFL detection was evaluated by comparing annotations to the closest prediction and deemed to be correctly classified if the annotation and prediction matched. At the recording level, patients with AF or AFL were deemed correctly detected (true positive) if there was at least one correctly detected 30-second AF or AFL segment. Patients without AF or AFL were correctly identified (true negative) if the DNN model did not predict AF or AFL in any 30-second sequence for

the patient. Evaluation at the recording level is useful for assessing how the DNN model can be applied to the analysis of online ECG recordings and notify health care personnel of whether AF or AFL is detected from the ongoing recording. The time-level evaluation is similar in clinical applications such as offline ambulatory ECG analysis, where it is critical to identify the onset and offset of an AF or AFL segment.

## Ethical Considerations

This study was approved by the ethics committee of the Northern Savo Hospital District (2061/2020). This study was conducted according to the principles of the Declaration of Helsinki. No informed consent from patients was needed due to the retrospective nature of the study. All data were pseudonymized before analyses. This study used retrospective ECG data owned by KUH and was conducted as part of institutional research activities. The manuscript and supplementary materials do not include images or other content that would allow for the identification of individual participants.

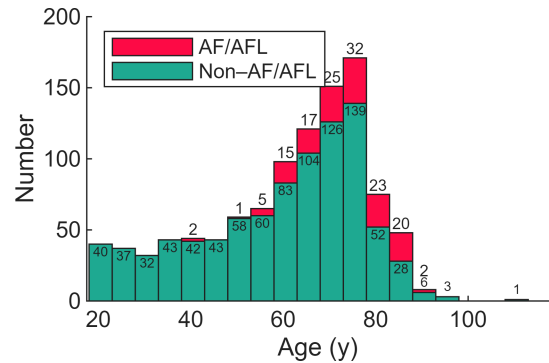
## Results

### Overview

The final test dataset consisted of 1039 three-channel ECG recordings ranging from 24 to 48 hours in duration from 1010 unique individuals, as shown in the [Figure 1](#) flowchart. The mean age of the patients was 60.37 (SD 17.75; range 18-105) years, and 48% (485/1010) were men. AF or AFL was present in 13.7% (142/1039) of the ECG recordings, of which 16.2% (23/142) exhibited both AF and AFL and non-AF/AFL arrhythmias. [Figure 2](#) shows the age distribution and rhythm distribution of the test dataset. In total, there were 25,072 hours of ECG data, with 3028.8 (12.1%) hours of AF

and AFL. AF and AFL were most prevalent in individuals aged 60 to 90 years. In addition, 0.4% (92.9/25,072) of hours of data were deemed nonanalyzable.

**Figure 2.** Age distribution of the test set (n=1039) recordings. The red bars denote patients with atrial fibrillation (AF) or atrial flutter (AFL), and the green bars denote patients with non-AF/AFL.



Data on medication were available for 66.6% (673/1010) of the dataset patients ([Table 2](#)). Notably, oral anticoagulants were used in 51.5% (50/97) of patients in the AF and AFL group. Antihypertensive drugs were the most used medication in both the AF and AFL group and non-AF/AFL group. Non-AF/AFL rhythms are shown in [Table 3](#). Cardiologist statements on arrhythmias were available for 88.1% (915/1039) of the recordings. Notably, combined cases of ventricular ectopic beats (>1000) and SVEBs (>1000) accounted for 28.2% (258/915) of the non-AF/AFL arrhythmias. Other significant observations included supraventricular tachycardias, which were present in 17.8% (163/915) of patients, and sinus bradycardia, which was present in 7.3% (67/915) of patients.

**Table 2.** Medication use in test dataset recordings.

	AF <sup>a</sup> or AFL <sup>b</sup> (n=97 recordings), n (%)	Non-AF/AFL (n=576 recordings), n (%)
Data available	97 (68.3)	576 (64.2)
No medication	0 (0)	38 (6.6)
Antiarrhythmic drugs	4 (4.1)	9 (1.6)
Antihypertensive drugs	73 (75.3)	417 (72.4)
Anticoagulants	50 (51.5)	59 (10.2)
Antithrombotic drugs	8 (8.2)	101 (17.5)
Dyslipidemia	34 (35.1)	205 (35.6)
Diabetes oral drugs	11 (11.3)	56 (9.7)
Insulin	0 (0)	19 (3.3)

<sup>a</sup>AF: atrial fibrillation.

<sup>b</sup>AFL: atrial flutter.

**Table 3.** Prevalence of non-atrial fibrillation or atrial flutter arrhythmias in the test dataset categorized by type and count, with corresponding percentages of the total cases. The table includes data on ventricular and supraventricular ectopic beats (SVEBs), atrioventricular blocks (AVBs), sinus bradycardia, and other arrhythmias identified in the patient population.

Finding	Recordings (N=1039), n (%)
Data available	915 (88.1)
VEB <sup>a</sup> —>100	341 (37.3)
VEB—>1000	167 (18.3)
VEB—>5000	71 (7.8)
VEB—>10,000	37 (4.0)
SVEB—>100	271 (29.6)
SVEB—>1000	91 (9.9)
SVEB—>5000	23 (2.5)
SVEB—>10,000	7 (0.8)
LBBB <sup>b</sup>	11 (1.2)
RBBB <sup>c</sup>	10 (1.1)
AVB I	66 (7.2)
AVB II	33 (3.6)
AVB III	0 (0)
Junctional rhythm	8 (0.9)
SVT <sup>d</sup>	163 (17.8)
VT <sup>e</sup>	49 (5.4)
Sinus bradycardia	67 (7.3)
Sinus pause	6 (0.7)

<sup>a</sup>VEB: ventricular ectopic beat.

<sup>b</sup>LBBB: left bundle branch block.

<sup>c</sup>RBBB: right bundle branch block.

<sup>d</sup>SVT: supraventricular tachycardia.

<sup>e</sup>VT: ventricular tachycardia.

### Time-Level AF and AFL Detection

In the time-level classification (confusion matrix), the DNN model correctly classified 2918.2 hours as AF or AFL out of 3028.8 true hours of AF and AFL, resulting in 96.4% sensitivity for time-level AF and AFL detection (Table 4).

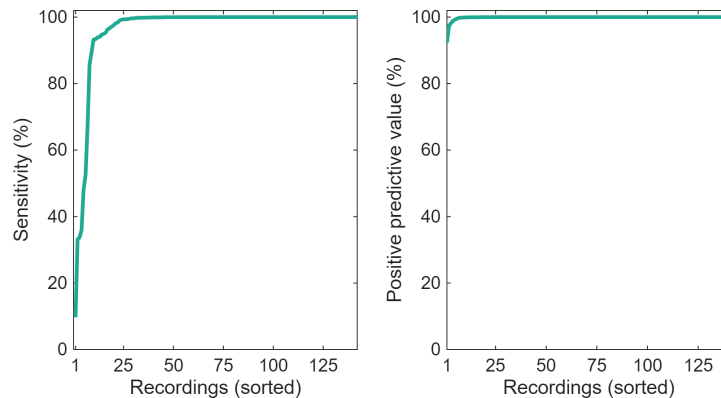
The DNN misclassified 1.0 hours of non-AF/AFL data as AF or AFL and 0.1 hours of nonanalyzable ECG data as AF or AFL resulting in over 99.99% specificity, 99.96% positive predictive value (PPV), and 99.5% negative predictive value (NPV) for time-level AF and AFL detection.

**Table 4.** Comparative analysis of true vs deep neural network (DNN)-detected hours for atrial fibrillation (AF) and atrial flutter (AFL), non-AF/AFL, and nonanalyzable categories.

	DNN-detected AF and AFL (h)	DNN-detected non-AF/AFL (h)	DNN-detected nonanalyzable (h)
True AF and AFL (h)	2918.2	109.1	1.5
True non-AF/AFL (h)	1.0	21,925.0	24.3
True nonanalyzable (h)	0.1	1.5	91.3

In the time-level analysis, the time-level sensitivity of the DNN model for detecting AF and AFL was greater than 90% in 93.7% (133/142) of AF and AFL recordings. For 2.8% (4/142) of the recordings, the sensitivity ranged between 50% and 90%, and for 3.5% (5/142) of the recordings, it was less

than 50% (Figure 3). A detailed breakdown of the time-level analysis revealed that all 9 recordings with less than 90% time-level sensitivity had AFL rhythm with a 2:1, 3:1, or 4:1 (mainly 2:1) conduction ratio where part of the F waves were merged into the QRS complexes.

**Figure 3.** Time-level sensitivity and positive predictive value for atrial fibrillation and atrial flutter detection.

### Recording-Level AF and AFL Detection

In the recording-level detection, the DNN model detected correctly at least one AF or AFL segment of more than 30 seconds in all 142 recordings containing AF or AFL. Thus, the recording-level AF and AFL detection sensitivity was 100%.

Of the 897 recordings without AF or AFL, the DNN model generated false-positive AF or AFL predictions for

11 (1.2%), with recording-level specificity of 98.77% and PPV of 92.81%. Of these 11 recordings, 9 (81.8%) showed arrhythmias: 3 contained frequent atrial premature beats, 3 contained acute rhythm changes (eg, junctional rhythm to normal SR), and 3 contained 1- or 2-degree AV block. A total of 0.2% (2/897) of the recordings showed noisy ECGs. Recording-level sensitivity, specificity, PPV, and NPV are reported in Table 5, with 95% CIs calculated using the Wilson score method.

**Table 5.** Recording-level detection performance of the deep neural network model.

Metric	Estimate (%; 95% CI)
Sensitivity	100 (97.37-100)
Specificity	98.77 (97.82-99.31)
Positive predictive value	92.81 (87.59-95.94)
Negative predictive value	100 (99.57-100)

Nonanalyzable sections of the ECG were identified in 11.1% (115/1039) of the recordings during clinical adjudication. The total duration of nonanalyzable ECGs was 92.9 hours; most of it (58.6 hours) was derived from 0.6% (6/1039) of the recordings, where lead failure or improper skin electrode contact corrupted the ECG. The DNN model correctly classified 91.3 of the 92.9 hours of nonanalyzable ECG as nonanalyzable. In addition, the DNN model labeled 25.8 hours of data as nonanalyzable that were deemed to be analyzable in the clinical workflow. A common factor between these segments was 2 channels exhibiting large ECG artifacts, whereas only 1 channel displayed visible R peaks.

## Discussion

### Principal Findings

This study provides a comprehensive demonstration of a deep learning approach to perform automatic AF and AFL detection from a substantial dataset encompassing more than 1000 long-term ambulatory ECG recordings from real-life clinical patients. Our DNN model successfully identified all patients exhibiting AF and AFL. Notably, it generated false AF or AFL alarms in only 1.2% (11/897) of recordings of patients who did not have AF or AFL, with 81.8% (9/11) of the false alarms due to other true arrhythmias.

This demonstrates that an end-to-end DNN approach has the potential to significantly enhance the precision of algorithmic interpretation for long-term ambulatory ECG analysis. We also emphasize the importance of using a sufficiently extensive test dataset in this study, allowing for the comprehensive evaluation of the DNN model's capabilities across a diverse range of patients, including those with other arrhythmias, poor-quality ECGs, additional heart diseases, and various conditions. The training and test datasets included patients receiving a variety of rate-controlling and antiarrhythmic medications that may influence RR interval variability. The model was intentionally trained on heterogeneous real-world data without medication-specific inputs, aiming to achieve robust AF and AFL detection independent of pharmacological modulation of heart rate dynamics.

The standard approach to automated ECG interpretation involves several steps, including preprocessing the signal; extracting features; selecting or reducing these features; and, finally, classifying them [24]. In contrast, DNNs offer a fundamentally different method by allowing a single algorithm to handle all these tasks in an end-to-end manner. This means that the DNN can process raw ECG data directly and provide diagnostic probabilities without the need for specific feature extraction [21]. If trained with sufficient raw data, DNNs could potentially learn all significant features previously identified manually as well as discover

new features through a data-driven approach. Indeed, our study at the recording level showed very high sensitivity and specificity for AF and AFL detection at 100% and 98.77%, respectively. The clinical relevance of these advancements is substantial given the significant amount of unprocessed clinical data used to train the DNN algorithm. The automated detection system may not only streamline the diagnostic process but could also drastically reduce manual work, enhancing efficiency in clinical workflows. Furthermore, its capability to provide real-time alerts in online measurements with an exceedingly low rate of false alarms emphasizes its potential in real-world applications.

### Comparison to Prior Work

Many prior studies examining the performance of DNN models and other methodologies for AF and AFL detection have often relied on limited ECG datasets such as the IRIDIA-AF dataset or Long-Term AF Database. These datasets typically consist of ECG recordings from a relatively small patient group [35,36,39,40]. Even DNN models developed using extensive ambulatory ECG datasets [28-30] still face a significant challenge: a high rate of false positives when applied to patient-operated ambulatory ECGs collected in real-world settings, as evidenced by a relatively low specificity of less than 99.0%. In contrast, we analyzed entire 24- to 48-hour recordings using an end-to-end DNN model that highlights clinically relevant AF and AFL segments of 30 seconds or longer for review. We evaluated its performance on a large single-center test set of 1039 recordings from real clinical patients, offering a more comprehensive and representative evaluation of DNN performance under real-world conditions. By leveraging a large and diverse training dataset collected from various ECG recorders, our model achieved comparable sensitivity while significantly improving specificity and PPV. These results highlight the potential of DNN models trained on substantial and heterogeneous datasets for improving AF detection in clinical practice.

A notable exception in long-term ambulatory ECG rhythm analysis are the studies by Zhang et al [38] and Fiorina et al [37], who evaluated DNN-based AF screening in large-scale hospital scenarios. Both DNN methods were developed using substantial ambulatory ECG datasets, yielding very promising AF screening results with a sensitivity of 99% and a specificity exceeding 98%. However, both studies focused primarily on screening patients with AF, and the capability of the DNNs to detect AF and AFL at the time level was not reported [37,38]. Importantly, our results provide strong evidence that state-of-the-art DNN models can identify patients with AF and AFL and detect AF and AFL segments in ambulatory ECGs with high temporal accuracy while effectively handling corrupted or nonanalyzable segments and maintaining a very low false positive rate regardless of the presence of other arrhythmias.

The DNN model exhibited a notably high AF and AFL time-level detection sensitivity (>90%) for 93.7% (133/142) of the patients. Upon detailed analysis, it was observed that the lower time-level sensitivity in the remaining 6.3% (9/142) of the recordings could be attributed to a specific 2:1 AFL

rhythm pattern where every second P wave merged into the QRS complex. This implies that, despite its overall efficiency, the DNN model may encounter challenges in accurately identifying specific ECG patterns, as demonstrated by the 2:1 AFL rhythm. Models incorporating longer temporal context, such as convolutional neural network–bidirectional long short-term memory or transformer-based architectures, may detect transitions in AFL conduction patterns (eg, 3:1 to 2:1) and improve detection of stable 2:1 AFL segments. However, training such models would require hundreds of recordings with paroxysmal AFL and varying conduction patterns to properly capture temporal and morphological variability.

Another point of discussion is the model's performance on recordings without AF or AFL. With a very high recording-level specificity of 98.77%, the model generated false-positive AF and AFL predictions for only 1.2% (11/897) of the recordings. In fact, 81.8% (9/11) of these false-positive cases were other true arrhythmias such as frequent ectopic beats, acute rhythm changes, and 1- or 2-degree AV block. It is also notable that only a very small percentage of these conditions caused false-positive predictions. Our model also efficiently identified nonanalyzable sections of the ECG, defined as segments in which QRS complexes could not be reliably visually verified during clinical review. No separate quantitative noise index was used. Initially, the DNN model classified only 0.1% (25.8/25072.0) of hours of data as nonanalyzable, which were marked to be analyzable in the clinical workflow. False-positive nonanalyzable labeling exceeding 10 minutes occurred in 37 of 1039 recordings. Further in-depth analysis uncovered that a significant portion of these data, totaling 10 hours, originated from seven 24-hour recordings where 2 out of 3 ECG channels manifested significant noise or lead failure. Regarding clinical consequences, AF and AFL masking due to nonanalyzable labeling was limited. As shown in the confusion matrix, approximately 0.1% (1.5/3028.8) of hours of true AF and AFL detections were classified as nonanalyzable, whereas a markedly larger proportion of true AF and AFL detections (109.1/3028.8, 3.6%) was misclassified as non-AF/AFL.

In the test dataset, the prevalence of AF and AFL was 13.7% (142/1039), which is lower than that reported in some real-world Holter monitoring populations [52] but substantially higher than that reported in opportunistic screening settings, where prevalence may be as low as 1% [53]. At this prevalence level, the model achieved a PPV of 92.81% (95% CI 87.59%-95.94%) and an NPV of 100% (95% CI 99.57%-100%; Table 5). As predictive values are inherently prevalence dependent, the high NPV observed in this study reflects the moderate prevalence of AF and AFL in the test set and indicates strong ability to rule out arrhythmia in this context. In lower-prevalence screening populations, PPV would be expected to decrease despite unchanged sensitivity and specificity, underscoring the importance of population-specific evaluation.

Subgroup analyses by age, sex, noise burden, and non-AF arrhythmia burden are presented in Table S1 in [Multimedia Appendix 1](#). These analyses showed that AF and AFL detection sensitivity was lowest in recordings with a lower

burden of non-AF arrhythmias, whereas both sensitivity and PPV were lowest in recordings with a high burden of non-AF arrhythmias. This was expected as other atrial and ventricular arrhythmias represent the most challenging confounders for differentiating AF and AFL in long-term ambulatory ECG recordings.

## **Methodological Considerations and Limitations**

Despite the promising results, our study had certain limitations. First, we excluded patients with pacemakers from the KUH dataset (Figure 1). The pacemaker inherently maintains a steady rhythm during AF and AFL. This stabilization renders the recognition of AF and AFL from a surface ECG exceedingly challenging. Pacemaker-related artifacts are a common source of false-positive AF and AFL detections in clinical practice, and their exclusion may lead to higher specificity than in a true all-comer population. Therefore, the reported performance metrics should be interpreted in the context of this limitation.

Second, the test dataset was derived from a single center (KUH), resulting in limited ethnic diversity and limited variation in ECG devices. Although the training and validation datasets included ECG recordings from multiple systems (including Schiller, Bittium Faros, GE HealthCare MUSE 12-lead ECG, and other ambulatory 3- and 4-lead recorders), the model may still be sensitive to device-specific artifacts, which could constrain external validity and generalizability across different clinical settings, recording systems, and populations; therefore, future multicenter evaluations are warranted.

Third, patients with paroxysmal AF were excluded from the KUH training and validation datasets. The model was primarily trained on persistent AF segments, which may differ to some extent from paroxysmal AF episodes. However, because the model operates on independent

10-second ECG segments without temporal context, AF and AFL detection is based on segment-level rhythm characteristics, which limits the impact of this difference.

Fourth, the model demonstrated reduced time-level sensitivity in AFL with fixed 2:1 AV conduction ratio, which represents a clinically relevant limitation. In this rhythm pattern, atrial activity may be partially obscured by the QRS complexes, making reliable discrimination from surface ECG challenging. Consequently, this specific AFL subtype constitutes a potential blind spot for automated screening.

Fifth, from a clinical safety perspective, the proposed DNN model should be regarded as a screening and decision support tool rather than a stand-alone diagnostic system. This is particularly important given the limited interpretability of the DNN models as learned features cannot be directly linked to established physiological markers; therefore, negative automated findings should be interpreted in conjunction with clinical judgment, especially when clinical suspicion of arrhythmia remains high.

Sixth, the 4-minute merge window used to ensure continuity of longer AF and AFL segments may lead to overestimation of AF burden, particularly when multiple short episodes occur close in time.

## **Conclusions**

The results of this study affirm the potential of using a DNN model for AF and AFL diagnostics in ambulatory settings with high sensitivity and specificity. When trained using an ample amount of data, the DNN model demonstrates the ability to distinguish ambulatory ECG changes caused by motion artifacts from AF and AFL with a low false positive rate. The presented automated detection system may streamline the diagnostic process and clinical workflows with a low false alarm rate, supporting its clinical applicability.

---

## **Acknowledgments**

Generative artificial intelligence tools were used for language editing and spelling correction. No generative artificial intelligence tools were used for scientific content generation or data analysis and interpretation, and all scientific responsibility remains with the authors.

---

## **Funding**

This work was supported by the Research Committee of the Kuopio University Hospital Catchment Area for the State Research Funding (projects 5101137 and 507T044; Kuopio, Finland). JS and JAL received research support from the Finland State Research Fund. JS received support from the University of Eastern Finland (Matti and Vappu Mauko foundation).

---

## **Data Availability**

The data underlying this article are not publicly available due to the privacy of the individuals who participated in the study and applicable data protection regulations. The data supporting the findings of this study will be shared on reasonable request to the corresponding author. No synthetic data were generated in this study. The code and trained model weights are not publicly available.

---

## **Authors' Contributions**

Conceptualization: JAL, JS, TM, HJ, JH, OES, ESS, TTR, MPT, TPL, TML

Data curation: JS, JAL, AA

Formal analysis: JAL, JS

Investigation: JS, JAL

Methodology: JAL

Software: JAL

Supervision: JAL, JEKH, TJM, OES

Writing—original draft: JS, JAL

Writing—review and editing: JS, JAL, AA, HJ, OES, TPL, TML, JH, MPT, ESS, NSN, OR, TTR, JEKH, TJM

### Conflicts of Interest

JAL, TTR, TJM, HJ, JH, OES, ESS, and MPT are shareholders of a company (Heart2Save) that designs electrocardiogram-based software for medical equipment. JAL, MPT, OES, and HJ report personal fees from Heart2Save. The retrospective electrocardiogram data used in this study are owned by Kuopio University Hospital. The deep learning model was developed at Kuopio University Hospital as part of institutional research activities, and no data sharing or intellectual property agreements related to this study were in place with Heart2Save at the time of model development. All other authors declare no other conflicts of interest.

### Multimedia Appendix 1

Deep neural network (DNN) model architecture, performance evaluation before temporal postprocessing, and analyses across demographic and signal-quality subgroups.

[\[PDF File \(Adobe File\), 280 KB-Multimedia Appendix 1\]](#)

### References

1. Chugh SS, Havmoeller R, Narayanan K, et al. Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. *Circulation*. Feb 25, 2014;129(8):837-847. [doi: [10.1161/CIRCULATIONAHA.113.005119](https://doi.org/10.1161/CIRCULATIONAHA.113.005119)] [Medline: [24345399](https://pubmed.ncbi.nlm.nih.gov/24345399/)]
2. Hindricks G, Potpara T, Dagres N, 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*. Feb 1, 2021;42(5):373-498. [doi: [10.1093/eurheartj/ehaa612](https://doi.org/10.1093/eurheartj/ehaa612)] [Medline: [32860505](https://pubmed.ncbi.nlm.nih.gov/32860505/)]
3. Alberts MJ, Eikelboom JW, Hankey GJ. Antithrombotic therapy for stroke prevention in non-valvular atrial fibrillation. *Lancet Neurol*. Dec 2012;11(12):1066-1081. [doi: [10.1016/S1474-4422\(12\)70258-2](https://doi.org/10.1016/S1474-4422(12)70258-2)] [Medline: [23153406](https://pubmed.ncbi.nlm.nih.gov/23153406/)]
4. Ruff CT, Giugliano RP, Braunwald E, et al. Comparison of the efficacy and safety of new oral anticoagulants with warfarin in patients with atrial fibrillation: a meta-analysis of randomised trials. *Lancet*. Mar 15, 2014;383(9921):955-962. [doi: [10.1016/S0140-6736\(13\)62343-0](https://doi.org/10.1016/S0140-6736(13)62343-0)] [Medline: [24315724](https://pubmed.ncbi.nlm.nih.gov/24315724/)]
5. Zhu H, Cheng C, Yin H, et al. Automatic multilabel electrocardiogram diagnosis of heart rhythm or conduction abnormalities with deep learning: a cohort study. *Lancet Digit Health*. Jul 2020;2(7):e348-e357. [doi: [10.1016/S2589-7500\(20\)30107-2](https://doi.org/10.1016/S2589-7500(20)30107-2)] [Medline: [33328094](https://pubmed.ncbi.nlm.nih.gov/33328094/)]
6. Steinberg JS, Varma N, Cygankiewicz I, et al. 2017 ISHNE-HRS expert consensus statement on ambulatory ECG and external cardiac monitoring/telemetry. *Heart Rhythm*. Jul 2017;14(7):e55-e96. [doi: [10.1016/j.hrthm.2017.03.038](https://doi.org/10.1016/j.hrthm.2017.03.038)] [Medline: [28495301](https://pubmed.ncbi.nlm.nih.gov/28495301/)]
7. Kirchhof P, Bax J, Blomstrom-Lundquist C, et al. Early and comprehensive management of atrial fibrillation: executive summary of the proceedings from the 2nd AFNET-EHRA consensus conference 'research perspectives in AF'. *Eur Heart J*. Dec 2009;30(24):2969-2977c. [doi: [10.1093/eurheartj/ehp235](https://doi.org/10.1093/eurheartj/ehp235)] [Medline: [19535417](https://pubmed.ncbi.nlm.nih.gov/19535417/)]
8. Wigginton JG, Agarwal S, Bartos JA, et al. Part 9: adult advanced life support: 2025 American Heart Association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care. *Circulation*. Oct 21, 2025;152(16\_suppl\_2):S538-S577. [doi: [10.1161/CIR.0000000000001376](https://doi.org/10.1161/CIR.0000000000001376)] [Medline: [41122884](https://pubmed.ncbi.nlm.nih.gov/41122884/)]
9. Fuster V, Rydén LE, Cannom DS, et al. 2011 ACCF/AHA/HRS focused updates incorporated into the ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines developed in partnership with the European Society of Cardiology and in collaboration with the European Heart Rhythm Association and the Heart Rhythm Society. *J Am Coll Cardiol*. Mar 15, 2011;57(11):e101-e198. [doi: [10.1016/j.jacc.2010.09.013](https://doi.org/10.1016/j.jacc.2010.09.013)] [Medline: [21392637](https://pubmed.ncbi.nlm.nih.gov/21392637/)]
10. Correction to: 2023 ACC/AHA/ACCP/HRS guideline for the diagnosis and management of atrial fibrillation: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation*. Jun 11, 2024;149(24):e1413. [doi: [10.1161/CIR.0000000000001263](https://doi.org/10.1161/CIR.0000000000001263)] [Medline: [38857333](https://pubmed.ncbi.nlm.nih.gov/38857333/)]
11. Ko D, Chung MK, Evans PT, Benjamin EJ, Helm RH. Atrial fibrillation: a review. *JAMA*. Jan 28, 2025;333(4):329-342. [doi: [10.1001/jama.2024.22451](https://doi.org/10.1001/jama.2024.22451)] [Medline: [39680399](https://pubmed.ncbi.nlm.nih.gov/39680399/)]
12. Johner N, Namdar M, Shah DC. Typical atrial flutter: a practical review. *J Cardiovasc Electrophysiol*. Oct 2025;36(10):2677-2691. [doi: [10.1111/jce.16542](https://doi.org/10.1111/jce.16542)] [Medline: [39764740](https://pubmed.ncbi.nlm.nih.gov/39764740/)]

13. Rogalewski A, Plümer J, Feldmann T, et al. Detection of atrial fibrillation on stroke units: comparison of manual versus automatic analysis of continuous telemetry. *Cerebrovasc Dis*. 2020;49(6):647-655. [doi: [10.1159/000511563](https://doi.org/10.1159/000511563)] [Medline: [33207338](https://pubmed.ncbi.nlm.nih.gov/33207338/)]
14. Avidan Y, Aker A, Sliman H, Weizman B, Danon A, Kassem S. Recognizing atrial flutter in the emergency department: challenges in diagnosis. *Am J Emerg Med*. Oct 2025;96:224-229. [doi: [10.1016/j.ajem.2025.07.013](https://doi.org/10.1016/j.ajem.2025.07.013)] [Medline: [40633488](https://pubmed.ncbi.nlm.nih.gov/40633488/)]
15. Domazetoski V, Gligoric G, Marinkovic M, et al. The influence of atrial flutter in automated detection of atrial arrhythmias - are we ready to go into clinical practice?. *Comput Methods Programs Biomed*. Jun 2022;221:106901. [doi: [10.1016/j.cmpb.2022.106901](https://doi.org/10.1016/j.cmpb.2022.106901)] [Medline: [35636359](https://pubmed.ncbi.nlm.nih.gov/35636359/)]
16. Michaud GF, Stevenson WG. Atrial fibrillation. *N Engl J Med*. Jan 28, 2021;384(4):353-361. [doi: [10.1056/NEJMcp2023658](https://doi.org/10.1056/NEJMcp2023658)] [Medline: [33503344](https://pubmed.ncbi.nlm.nih.gov/33503344/)]
17. Prystowsky EN, Padanilam BJ, Fogel RI. Treatment of atrial fibrillation. *JAMA*. Jul 21, 2015;314(3):278-288. [doi: [10.1001/jama.2015.7505](https://doi.org/10.1001/jama.2015.7505)] [Medline: [26197188](https://pubmed.ncbi.nlm.nih.gov/26197188/)]
18. Chaumont C, Suffee N, Gandjbakhch E, Balse E, Anselme F, Hatem SN. Epicardial origin of cardiac arrhythmias: clinical evidences and pathophysiology. *Cardiovasc Res*. Jun 22, 2022;118(7):1693-1702. [doi: [10.1093/cvr/cvab213](https://doi.org/10.1093/cvr/cvab213)] [Medline: [34152392](https://pubmed.ncbi.nlm.nih.gov/34152392/)]
19. van der Does LJ, Kharbanda RK, Teuwen CP, et al. Atrial ectopy increases asynchronous activation of the endo- and epicardium at the right atrium. *J Clin Med*. Feb 18, 2020;9(2):558. [doi: [10.3390/jcm9020558](https://doi.org/10.3390/jcm9020558)] [Medline: [32085601](https://pubmed.ncbi.nlm.nih.gov/32085601/)]
20. Parvaneh S, Rubin J, Babaeizadeh S, Xu-Wilson M. Cardiac arrhythmia detection using deep learning: a review. *J Electrocardiol*. 2019;57:S70-S74. [doi: [10.1016/j.jelectrocard.2019.08.004](https://doi.org/10.1016/j.jelectrocard.2019.08.004)]
21. Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med*. Jan 2019;25(1):65-69. [doi: [10.1038/s41591-018-0268-3](https://doi.org/10.1038/s41591-018-0268-3)] [Medline: [30617320](https://pubmed.ncbi.nlm.nih.gov/30617320/)]
22. Holgado-Cuadrado R, Plaza-Seco C, Lovisolo L, Blanco-Velasco M. Characterization of noise in long-term ECG monitoring with machine learning based on clinical criteria. *Med Biol Eng Comput*. Sep 2023;61(9):2227-2240. [doi: [10.1007/s11517-023-02802-5](https://doi.org/10.1007/s11517-023-02802-5)] [Medline: [37010711](https://pubmed.ncbi.nlm.nih.gov/37010711/)]
23. Everss-Villalba E, Melgarejo-Meseguer FM, Blanco-Velasco M, et al. Noise maps for quantitative and clinical severity towards long-term ECG monitoring. *Sensors (Basel)*. Oct 25, 2017;17(11):2448. [doi: [10.3390/s17112448](https://doi.org/10.3390/s17112448)] [Medline: [29068362](https://pubmed.ncbi.nlm.nih.gov/29068362/)]
24. Lyon A, Mincholé A, Martínez JP, Laguna P, Rodriguez B. Computational techniques for ECG analysis and interpretation in light of their contribution to medical advances. *J R Soc Interface*. Jan 2018;15(138):20170821. [doi: [10.1098/rsif.2017.0821](https://doi.org/10.1098/rsif.2017.0821)] [Medline: [29321268](https://pubmed.ncbi.nlm.nih.gov/29321268/)]
25. Kumar D, Peimankar A, Sharma K, Domínguez H, Puthusserypady S, Bardram JE. DeepAware: a hybrid deep learning and context-aware heuristics-based model for atrial fibrillation detection. *Comput Methods Programs Biomed*. Jun 2022;221:106899. [doi: [10.1016/j.cmpb.2022.106899](https://doi.org/10.1016/j.cmpb.2022.106899)] [Medline: [35640394](https://pubmed.ncbi.nlm.nih.gov/35640394/)]
26. Faust O, Shenfield A, Kareem M, San TR, Fujita H, Acharya UR. Automated detection of atrial fibrillation using long short-term memory network with RR interval signals. *Comput Biol Med*. Nov 1, 2018;102:327-335. [doi: [10.1016/j.compbimed.2018.07.001](https://doi.org/10.1016/j.compbimed.2018.07.001)] [Medline: [30031535](https://pubmed.ncbi.nlm.nih.gov/30031535/)]
27. Ebrahimi Z, Loni M, Daneshtalab M, Gharehbaghi A. A review on deep learning methods for ECG arrhythmia classification. *Expert Syst Appl*. Sep 2020;7:100033. [doi: [10.1016/j.eswax.2020.100033](https://doi.org/10.1016/j.eswax.2020.100033)]
28. Ben-Moshe N, Tsutsui K, Brimer SB, Zvuloni E, Sormmo L, Behar JA. RawECGNet: deep learning generalization for atrial fibrillation detection from the raw ECG. *IEEE J Biomed Health Inform*. Sep 2024;28(9):5180-5188. [doi: [10.1109/JBHI.2024.3404877](https://doi.org/10.1109/JBHI.2024.3404877)] [Medline: [38787663](https://pubmed.ncbi.nlm.nih.gov/38787663/)]
29. Chocron A, Oster J, Biton S, et al. Remote atrial fibrillation burden estimation using deep recurrent neural network. *IEEE Trans Biomed Eng*. Aug 2021;68(8):2447-2455. [doi: [10.1109/TBME.2020.3042646](https://doi.org/10.1109/TBME.2020.3042646)] [Medline: [33275575](https://pubmed.ncbi.nlm.nih.gov/33275575/)]
30. Tutuko B, Nurmaini S, Tondas AE, et al. AFibNet: an implementation of atrial fibrillation detection with convolutional neural network. *BMC Med Inform Decis Mak*. Jul 14, 2021;21(1):216. [doi: [10.1186/s12911-021-01571-1](https://doi.org/10.1186/s12911-021-01571-1)] [Medline: [34261486](https://pubmed.ncbi.nlm.nih.gov/34261486/)]
31. Chen X, Cheng Z, Wang S, et al. Atrial fibrillation detection based on multi-feature extraction and convolutional neural network for processing ECG signals. *Comput Methods Programs Biomed*. Apr 2021;202:106009. [doi: [10.1016/j.cmpb.2021.106009](https://doi.org/10.1016/j.cmpb.2021.106009)] [Medline: [33631641](https://pubmed.ncbi.nlm.nih.gov/33631641/)]
32. Fan X, Yao Q, Cai Y, Miao F, Sun F, Li Y. Multiscaled fusion of deep convolutional neural networks for screening atrial fibrillation from single lead short ECG recordings. *IEEE J Biomed Health Inform*. Nov 2018;22(6):1744-1753. [doi: [10.1109/JBHI.2018.2858789](https://doi.org/10.1109/JBHI.2018.2858789)] [Medline: [30106699](https://pubmed.ncbi.nlm.nih.gov/30106699/)]

33. Marinucci D, Sbröllini A, Marcantoni I, Morettini M, Swenne CA, Burattini L. Artificial neural network for atrial fibrillation identification in portable devices. *Sensors (Basel)*. Jun 24, 2020;20(12):3570. [doi: [10.3390/s20123570](https://doi.org/10.3390/s20123570)] [Medline: [32599796](https://pubmed.ncbi.nlm.nih.gov/32599796/)]
34. Tzou HA, Lin SF, Chen PS. Paroxysmal atrial fibrillation prediction based on morphological variant P-wave analysis with wideband ECG and deep learning. *Comput Methods Programs Biomed*. Nov 2021;211:106396. [doi: [10.1016/j.cmpb.2021.106396](https://doi.org/10.1016/j.cmpb.2021.106396)] [Medline: [34592687](https://pubmed.ncbi.nlm.nih.gov/34592687/)]
35. Krasteva V, Stoyanov T, Naydenov S, Schmid R, Jekova I. Detection of atrial fibrillation in Holter ECG recordings by ECHOView images: a deep transfer learning study. *Diagnostics (Basel)*. Mar 28, 2025;15(7):865. [doi: [10.3390/diagnostics15070865](https://doi.org/10.3390/diagnostics15070865)] [Medline: [40218215](https://pubmed.ncbi.nlm.nih.gov/40218215/)]
36. Rahman MM, Rivolta MW, Vaglio M, Maison-Blanche P, Badilini F, Sassi R. Residual-attention deep learning model for atrial fibrillation detection from Holter recordings. *J Electrocardiol*. 2025;89:153876. [doi: [10.1016/j.jelectrocard.2025.153876](https://doi.org/10.1016/j.jelectrocard.2025.153876)] [Medline: [39827741](https://pubmed.ncbi.nlm.nih.gov/39827741/)]
37. Fiorina L, Maupain C, Gardella C, et al. Evaluation of an ambulatory ECG analysis platform using deep neural networks in routine clinical practice. *J Am Heart Assoc*. Sep 20, 2022;11(18):e026196. [doi: [10.1161/JAHA.122.026196](https://doi.org/10.1161/JAHA.122.026196)] [Medline: [36073638](https://pubmed.ncbi.nlm.nih.gov/36073638/)]
38. Zhang P, Lin F, Ma F, et al. Automatic screening of patients with atrial fibrillation from 24-h Holter recording using deep learning. *Eur Heart J Digit Health*. 2023;4(3):216-224. [doi: [10.1093/ehjdh/ztad018](https://doi.org/10.1093/ehjdh/ztad018)] [Medline: [37265871](https://pubmed.ncbi.nlm.nih.gov/37265871/)]
39. Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet*. Sep 7, 2019;394(10201):861-867. [doi: [10.1016/S0140-6736\(19\)31721-0](https://doi.org/10.1016/S0140-6736(19)31721-0)] [Medline: [31378392](https://pubmed.ncbi.nlm.nih.gov/31378392/)]
40. Santala OE, Lipponen JA, Jäntti H, et al. Necklace-embedded electrocardiogram for the detection and diagnosis of atrial fibrillation. *Clin Cardiol*. May 2021;44(5):620-626. [doi: [10.1002/clc.23580](https://doi.org/10.1002/clc.23580)] [Medline: [33629410](https://pubmed.ncbi.nlm.nih.gov/33629410/)]
41. Santala OE, Halonen J, Martikainen S, et al. Automatic mobile health arrhythmia monitoring for the detection of atrial fibrillation: prospective feasibility, accuracy, and user experience study. *JMIR Mhealth Uhealth*. Oct 22, 2021;9(10):e29933. [doi: [10.2196/29933](https://doi.org/10.2196/29933)] [Medline: [34677135](https://pubmed.ncbi.nlm.nih.gov/34677135/)]
42. Petrutiu S, Sahakian AV, Swiryn S. Abrupt changes in fibrillatory wave characteristics at the termination of paroxysmal atrial fibrillation in humans. *Europace*. Jul 2007;9(7):466-470. [doi: [10.1093/europace/eum096](https://doi.org/10.1093/europace/eum096)] [Medline: [17540663](https://pubmed.ncbi.nlm.nih.gov/17540663/)]
43. Zheng J, Zhang J, Danioko S, Yao H, Guo H, Rakovski C. A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients. *Sci Data*. Feb 12, 2020;7(1):48. [doi: [10.1038/s41597-020-0386-x](https://doi.org/10.1038/s41597-020-0386-x)] [Medline: [32051412](https://pubmed.ncbi.nlm.nih.gov/32051412/)]
44. Melillo P, Izzo R, Orrico A, et al. Automatic prediction of cardiovascular and cerebrovascular events using heart rate variability analysis. *PLoS One*. 2015;10(3):e0118504. [doi: [10.1371/journal.pone.0118504](https://doi.org/10.1371/journal.pone.0118504)] [Medline: [25793605](https://pubmed.ncbi.nlm.nih.gov/25793605/)]
45. Moody G. A new method for detecting atrial fibrillation using R-R intervals. *Proc Comput Cardiol*. 1983;10:227-230. [doi: [10.3390/jcm12020687](https://doi.org/10.3390/jcm12020687)]
46. He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In: Leibe B, Matas J, Sebe N, Welling M, editors. *Computer Vision ECCV 2016*. Springer; 2016:630-645. [doi: [10.1007/978-3-319-46493-0\\_38](https://doi.org/10.1007/978-3-319-46493-0_38)]
47. Xiong Z, Nash MP, Cheng E, Fedorov VV, Stiles MK, Zhao J. ECG signal classification for the detection of cardiac arrhythmias using a convolutional recurrent neural network. *Physiol Meas*. Sep 24, 2018;39(9):094006. [doi: [10.1088/1361-6579/aad9ed](https://doi.org/10.1088/1361-6579/aad9ed)] [Medline: [30102248](https://pubmed.ncbi.nlm.nih.gov/30102248/)]
48. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. *Proceedings of the 32nd International Conference on Machine Learning*. 2015:448-456. URL: <https://proceedings.mlr.press/v37/ioffe15.html> [Accessed 2026-06-18]
49. He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. *2015 IEEE International Conference on Computer Vision (ICCV)*. 2015:1026-1034. URL: <https://ieeexplore.ieee.org/document/7410480> [Accessed 2026-06-18]
50. Kingma DP, Ba J. Adam: a method for stochastic optimization. *arXiv*. Preprint posted online on Dec 22, 2014. [doi: [10.48550/arXiv.1412.6980](https://doi.org/10.48550/arXiv.1412.6980)]
51. Srivastava N, Hinton GE, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 2014;15(1):1929-1958. URL: <https://dl.acm.org/doi/10.5555/2627435.2670313> [doi: [10.5555/2627435.2670313](https://doi.org/10.5555/2627435.2670313)]
52. Wu M, Lu Y, Yang W, Wong SY. A study on arrhythmia via ECG signal classification using the convolutional neural network. *Front Comput Neurosci*. 2021;14:564015. [doi: [10.3389/fncom.2020.564015](https://doi.org/10.3389/fncom.2020.564015)] [Medline: [33469423](https://pubmed.ncbi.nlm.nih.gov/33469423/)]
53. Lowres N, Olivier J, Chao TF, et al. Estimated stroke risk, yield, and number needed to screen for atrial fibrillation detected through single time screening: a multicountry patient-level meta-analysis of 141,220 screened individuals. *PLoS Med*. Sep 2019;16(9):e1002903. [doi: [10.1371/journal.pmed.1002903](https://doi.org/10.1371/journal.pmed.1002903)] [Medline: [31553733](https://pubmed.ncbi.nlm.nih.gov/31553733/)]

**Abbreviations**

**AF:** atrial fibrillation  
**AFL:** atrial flutter  
**AV:** atrioventricular  
**DNN:** deep neural network  
**ECG:** electrocardiogram  
**KUH:** Kuopio University Hospital  
**NPV:** negative predictive value  
**PPV:** positive predictive value  
**SVEB:** supraventricular ectopic beat

*Edited by Andrew Coristine; peer-reviewed by Hans Mautong, Sadhasivam Mohanadas; submitted 13.Sep.2025; final revised version received 12.Apr.2026; accepted 13.Apr.2026; published 30.Jun.2026*

*Please cite as:*

*Sedha J, Lipponen JA, Aho A, Jäntti H, Santala OE, Laitinen TP, Laitinen TM, Halonen J, Tarvainen MP, Seljola ES, Naukkarinen NS, Rantula O, Rissanen TT, Hartikainen JEK, Martikainen TJ*

*Deep Neural Networks for Automatic Atrial Fibrillation Detection Using Long-Term Ambulatory Electrocardiography: Retrospective Diagnostic Accuracy Study*

*JMIR Cardio 2026;10:e83714*

*URL: <https://cardio.jmir.org/2026/1/e83714>*

*doi: [10.2196/83714](https://doi.org/10.2196/83714)*

© Jagdeep Sedha, Jukka A Lipponen, Antti Aho, Helena Jäntti, Onni E Santala, Tomi P Laitinen, Tiina M Laitinen, Jari Halonen, Mika P Tarvainen, Eemu-Samuli Seljola, Noora S Naukkarinen, Olli Rantula, Tuomas T Rissanen, Juha E K Hartikainen, Tero J Martikainen. Originally published in JMIR Cardio (<https://cardio.jmir.org>), 30.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on <https://cardio.jmir.org>, as well as this copyright and license information must be included.