

An artificial intelligence–based model for prediction of atrial fibrillation from single-lead sinus rhythm electrocardiograms facilitating screening

Tove Hygrell ^{1*}†, Fredrik Viberg ^{1†}, Erik Dahlberg ², Peter H. Charlton ³,
Katrín Kemp Gudmundsdóttir ¹, Jonathan Mant ³, Josef Lindman Hörnlund ²,
and Emma Svennberg ⁴

¹Department of Clinical Sciences, Karolinska Institutet, Danderyd University Hospital, Stockholm SE-182 88, Sweden; ²Modulai AB, Stockholm, Sweden; ³Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; and ⁴Department of Medicine, Karolinska Institutet, Karolinska University Hospital, Huddinge, Stockholm, Sweden

Received 19 September 2022; accepted after revision 25 January 2023; online publish-ahead-of-print 7 March 2023

Aims

Screening for atrial fibrillation (AF) is recommended in the European Society of Cardiology guidelines. Yields of detection can be low due to the paroxysmal nature of the disease. Prolonged heart rhythm monitoring might be needed to increase yield but can be cumbersome and expensive. The aim of this study was to observe the accuracy of an artificial intelligence (AI)-based network to predict paroxysmal AF from a normal sinus rhythm single-lead ECG.

Methods and results

A convolutional neural network model was trained and evaluated using data from three AF screening studies. A total of 478 963 single-lead ECGs from 14 831 patients aged ≥ 65 years were included in the analysis. The training set included ECGs from 80% of participants in SAFER and STROKESTOP II. The remaining ECGs from 20% of participants in SAFER and STROKESTOP II together with all participants in STROKESTOP I were included in the test set. The accuracy was estimated using the area under the receiver operating characteristic curve (AUC). From a single timepoint ECG, the artificial intelligence–based algorithm predicted paroxysmal AF in the SAFER study with an AUC of 0.80 [confidence interval (CI) 0.78–0.83], which had a wide age range of 65–90+ years. Performance was lower in the age-homogenous groups in STROKESTOP I and STROKESTOP II (age range: 75–76 years), with AUCs of 0.62 (CI 0.61–0.64) and 0.62 (CI 0.58–0.65), respectively.

Conclusion

An artificial intelligence–enabled network has the ability to predict AF from a sinus rhythm single-lead ECG. Performance improves with a wider age distribution.

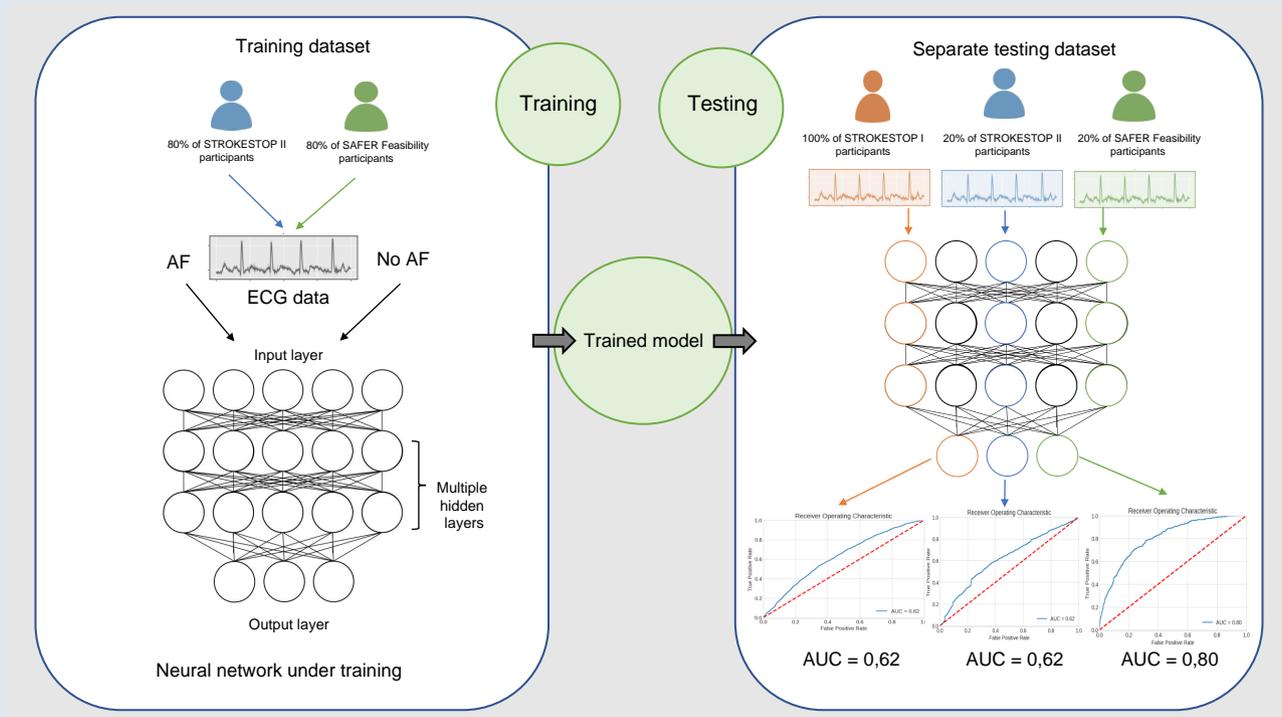
* Corresponding author. Tel: +46 73 768 6964. E-mail address: tove.hygrell@regionstockholm.se

† Shared first authorship.

© The Author(s) 2023. Published by Oxford University Press on behalf of the European Society of Cardiology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Graphical Abstract



Keywords

Artificial intelligence • Atrial fibrillation • Screening • Intermittent ECG

What's new?

- Artificial intelligence may be used to identify individuals that can benefit from prolonged screening for paroxysmal AF using single-lead ECG recordings.
- Using a single timepoint single-lead ECG machine learning can identify individuals at risk of undetected paroxysmal AF, with increasing performance in age-diverse populations.

Introduction

Atrial fibrillation (AF) is the most common cardiac arrhythmia. One of the most feared complications in patients with AF is ischaemic stroke. Untreated AF is associated with a five-fold increase in stroke risk; initiation of stroke-protective oral anticoagulation treatment is therefore of importance in most individuals with AF.¹ However, the arrhythmia is often asymptomatic and intermittent, making it challenging to identify with a single electrocardiogram (ECG). The European Society of Cardiology recommends that systematic screening for AF should be considered for high-risk patients.²

Several strategies have been used to reduce the numbers needed to screen and still detect AF in those with a high risk of stroke. The CHA₂DS₂-VASc score is used to predict stroke risk but has been used to direct screening efforts towards individuals with an increased risk of having AF.³ Measuring plasma levels of the biomarker N-terminal B-type natriuretic peptide (NT-proBNP) can be helpful in identifying individuals at high risk of previously undetected AF. Using

NT-proBNP with a cut-off of ≥ 125 ng/L (with a sensitivity of 75% and specificity of 37%) decreased the number of patients that needed prolonged ECG screening in a large screening study for AF.^{4,5} Nevertheless, venipuncture is invasive, relatively expensive, and impractical.

Growing attention is paid to artificial intelligence (AI) within the field of medicine, since it has the potential to automate human tasks and identify patterns beyond human capabilities.^{6,7} It has been shown that an AI-enabled network can be used to predict AF from a normal sinus rhythm 12-lead ECG.⁸ Although, when studied, AI has not been shown to improve prediction for stroke, major bleeding, or death in patients with AF.⁹

Developing a machine-learning algorithm designed to predict AF from a single-lead ECG instead of a standard 12-lead ECG could potentially provide an important contribution to simplified screening for AF. Single-lead ECGs can be easily obtained using clinical or consumer devices, and without clinical supervision outside of healthcare institutions. In the current study, we set out to study the accuracy of an AI-based network to predict paroxysmal AF from a normal sinus rhythm single-lead ECG.

Methods

Study population

The ECGs used in our analysis derive from three prospective screening studies: STROKESTOP I, STROKESTOP II, and SAFER. STROKESTOP I was a large AF screening study randomizing all 75- and 76-year olds living in Stockholm and Halland region in a 1:1 fashion into

an AF screening group and a control group. In the STROKESTOP I study, all participants in the screening group performed intermittent ECG recordings for 2 weeks.^{10,11} In STROKESTOP II, using a similar randomization of all 75–76-year olds in Stockholm, stepwise screening was performed. All randomized to screening had NT-proBNP levels measured, and only in the case of elevated levels (≥ 125 ng/L), prolonged screening was performed using intermittent ECG recordings.^{5,12} In the SAFER Feasibility Study, ISRCTN 16939438, general practitioners across England invited patients aged 65 years or over to take part in screening using intermittent ECG for 1–4 weeks. A comparison of the screening studies included is shown in [Supplementary material online, Table S1](#).

Patients with a previous AF diagnosis, or an inconclusive diagnosis, were removed. All participants with AF on their index ECG were also removed from the study.

Ethics

The study complies with the Declaration of Helsinki. The protocol was approved by the regional ethics committee in Stockholm (2011/1363-31, 2020-01211, 2015/2079-31, 2020-01436) and London Central NHS Research Ethics Committee (18/LO/2066). All participants provided informed consent before inclusion in the three screening studies.

Intermittent electrocardiogram

In all three AF screening studies, ECGs were recorded at home using a validated single-lead hand-held device, Zenicor ECG-2 (Zenicor Medical Systems, Stockholm, Sweden).¹³ A recording is performed by placing the thumbs on two electrodes, and each ECG registration of lead I has a duration of 30 s. In STROKESTOP I, the participants were instructed to record ECGs twice daily. In STROKESTOP II and SAFER, participants were instructed to record ECGs four times daily. In STROKESTOP I, all ECGs were manually annotated. In STROKESTOP II and SAFER, a validated computerized algorithm was used to discriminate between sinus rhythm and potential arrhythmias.¹⁴ All ECGs deemed as abnormal by the algorithm were reviewed by trained nurses and medical doctors.

Participants were classified into the AF group if they had at least one ECG with AF for 30 s. In addition, in STROKESTOP I, participants with two ECG recordings with >10 s of AF activity were regarded as having AF and included in the AF group.

The artificial intelligence model

Deep learning is a type of machine learning where neural networks with multiple layers are used to learn a relation between a set of inputs and a set of outputs. In the current study, we used a specialized type of neural network called a convolutional neural network (CNN). The connectivity pattern between the neurons in a CNN makes it a suitable choice for signal processing and image recognition and classification. The model used was a CNN with a binary cross entropy loss function and optimized to separate patients with AF diagnosed during the study from patients with no AF diagnosis. The CNN architecture used has similarities with the ReSE-2-Multi architecture developed by Lee *et al.*¹⁵ This architecture was modified to accommodate the longer length of 1 kHz 30 s ECGs and had seven blocks each consisting of two 1D convolutions, with kernel size three, plus a squeeze-and-excite step and residual connections. The three final block outputs were concatenated before a fully connected layer and a final linear classifier layer. Hyperparameter optimization was performed using Bayesian optimization.¹⁶ Further details on model architecture and hyperparameters can be found in supplementary technical details, [Supplementary material online, Figure S1](#) and [Table S2](#), respectively.

Handling of electrocardiograms

Raw data of 30 s long ECG recordings sampled at 500 Hz were used. Following upsampling to 1 kHz, baseline wander was removed and pre-filtering was applied, smoothing the ECGs to reduce noise inherent from the measurement process using the ECG Parser software from Cardiolund (Lund, Sweden).¹⁴ To eliminate noise caused by poor equipment handling in the beginning and end of the registration, the initial 5000 samples (5 s) and final 5400 samples (5.4 s) were removed. The resulting data set subsequently consisted of 20 000 samples long (20 s) ECGs. In the training set, 15 s random crops from the 20 s window were used,

whereas in the test set, the ECGs were cropped to the final 15 s of the original 20 s. Electrocardiograms that the ECG Parser software deemed to be of poor quality were removed because of a lack of clear ECG signal. In addition, ECGs tagged by the software as containing irregular sequences were removed due to correlation with intermittent AF. The training and testing samples were then normalized individually by first min–max scaling them to the interval [0, 1] and then shifting them to have Mode 0.

Multiple ECGs from the participants were used for training and testing of the AI model. Electrocardiograms were placed in the positive class if they came from a patient that was diagnosed with AF during the study, and in the negative class otherwise. When combining the studies, the results from each study were equally weighted. In addition, to mimic a screening setting, a sensitivity analysis was performed using only the first ECG recorded for each patient.

Data sets for training and evaluation

To construct the model training set, ECGs belonging to 80% of the patients from SAFER and 80% of the patients from STROKESTOP II were used. The remaining 20% of the patients from these two studies were placed in the test set, together with all ECGs from STROKESTOP I ([Figure 1](#)). Stratified sampling was employed to ensure equal class distribution in the training and testing data sets. We chose to train the algorithm solely on two of the data sets, completely excluding the data from STROKESTOP I to minimize the risk of overfitting of the model. The reason for choosing STROKESTOP I was that this population was age homogenous, with the least selection bias. To avoid problems due to the class imbalance, a fully balanced training data set was constructed by downsampling the number of negative class examples to the number of positive examples. Given that the number of ECGs per patient varies in both the training and testing data sets, a uniform weight of one was assigned to each patient, which was then distributed across their respective ECGs. In the training process, the training set was first split into 80% training and 20% validation sets. The split was done on patient level, and stratified sampling was employed to ensure class distributions. Model performance results from the validation set were used to tune the hyperparameters.

Statistics

The model was developed using TensorFlow and Keras in Python. Statistical evaluation of the model was done using scikit-learn 0.23. The model was developed and trained on a workstation with 4 RTX 2080 Ti Nvidia graphics processing units. Matplotlib 3.3.3. in Python 3 was used to generate graphs of area under the receiver operating characteristic curve (AUC), sensitivity, and specificity. Sensitivity was set to 75% by thresholding the model output score and specificity, and F1 scores were calculated. Specificity at increasing sensitivity levels for AF detection in the different test groups was also explored. In addition, in order to measure the robustness of the model classifications on a patient level, Cohen's kappa was computed by comparing the classification for the index ECG with the classifications for the k th ECGs of index k ($k = 1, 2, \dots, 5$). Cohen's kappa measures the inter-classifier reliability between two classifiers, correcting for agreement happening by random chance, which makes it a more robust metric than the more simplistic per cent overlap agreement. The range of Cohen's kappa is $[-1, 1]$ where 1 corresponds to perfect agreement, 0 to random agreement, and -1 to perfect disagreement. Uncertainty estimations were calculated using non-parametric bootstrapping.

Results

Of the 7165 participants from STROKESTOP I, $n = 617$ were excluded due to a previous AF diagnosis, or an inconclusive diagnosis. From the STROKESTOP II, $n = 663$ out of 6869 were excluded, and for SAFER 68 of 2146.

In total, 535 295 intermittent ECGs belonging to 14 832 AF screening participants were identified. Of those, 56 332 ECGs were excluded, the majority due to irregular sequences or baseline disturbances. The unbalanced data set consisted of 248 964 ECGs from 6658 patients. After training set class balancing, 14 464 ECGs from 3623 patients remained, which were used in training. It was found that adding more negative examples in training did not improve the model. The test set consisted of 229 999 ECGs from 8173 participants ([Figure 1](#)).

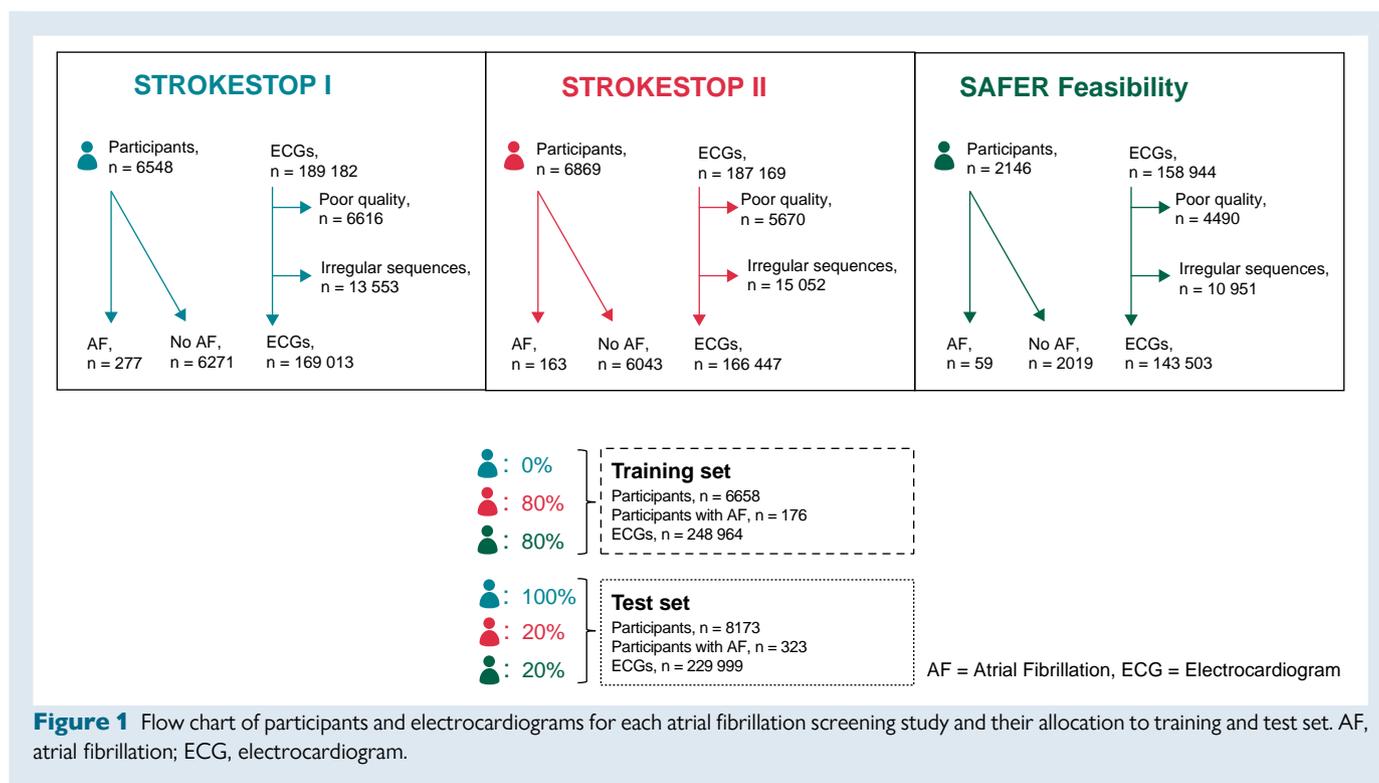


Figure 1 Flow chart of participants and electrocardiograms for each atrial fibrillation screening study and their allocation to training and test set. AF, atrial fibrillation; ECG, electrocardiogram.

Table 1 Specificity at different sensitivity levels

Sensitivity	All, weighted, specificity (95% CI)	STROKESTOP I, specificity (95% CI)	STROKESTOP II, specificity (95% CI)	SAFER Feasibility, specificity (95% CI)
0.75	45 (45–46)	41 (41–41)	38 (36–40)	70 (69–70)
0.80	38 (38–39)	36 (35–36)	32 (31–34)	64 (63–65)
0.90	23 (22–24)	21(21–22)	16(14–17)	49(49–50)

CI, confidence interval.

From a single timepoint ECG, the AI-based algorithm predicted current paroxysmal AF with an AUC of 0.62 [95% confidence interval (CI) 0.61–0.64] in STROKESTOP I. The algorithm was naive to the ECGs from STROKESTOP I, as they were not used to train the algorithm. As the AI model is constructed to be used as an interim step in identifying individuals that might benefit from screening, we aimed for a high sensitivity. Specificity at increasing sensitivity levels for AF detection in the different test groups is shown in *Table 1*. With a 75% sensitivity for AF detection, the specificity was 41% (95% CI 0.38–0.44) with an F1 score of 0.099 (95% CI 0.092–0.106) in STROKESTOP I. The ability of the AI model to predict AF from a sinus rhythm ECG was also evaluated for the two other screening studies separately. The outcome in STROKESTOP II was similar to STROKESTOP I, with an AUC of 0.62 (95% CI 0.58–0.65), F1 score of 0.074 (95% CI 0.063–0.088), and specificity of 38% (95% CI 0.33–0.44) when sensitivity was set to 75%. The AI model performed best on the ECGs from the SAFER data set, where the age distribution was wider, with an AUC of 0.80 (95% CI 0.78–0.83), F1 score of 0.097 (95% CI 0.072–0.133), and specificity of 71% (95% CI 0.65–0.77) at 75% sensitivity. Combining the test sets for the different studies gave an AUC of 0.66 (95% CI 0.64–0.67), F1 score of 0.080 (95% CI 0.072–0.087), and a specificity of 46% (95% CI 0.43–0.49) at sensitivity 75%. These results are shown in *Figure 2*.

In a sensitivity analysis, the ability to predict AF using only the index ECG of the patient was used to mimic a true screening situation. There was no difference in the predictive abilities in the algorithm for STROKESTOP I with an AUC of 0.63 (95% CI 0.60–0.66) and F1 score of 0.099 (95% CI 0.087–0.113), nor for STROKESTOP II with an AUC of 0.61 (95% CI 0.54–0.69) and F1 score of 0.083 (95% CI 0.056–0.115) or SAFER with an AUC of 0.88 (95% CI 0.76–0.96) and F1 score of 0.186 (95% CI 0.045–0.360). Similar performance was measured when performing a sensitivity analysis based on the second sinus rhythm ECG (see *Supplementary material online, Table S3*). In order to quantify patient-level classification robustness, Cohen's kappa was computed by comparing the index ECG classification against the classifications of the k th ECG ($k = 1, 2, \dots, 5$). The results range from 0.273 (0.179–0.365) for $k = 5$ in STROKESTOP II to 0.396 (0.258–0.543) for $k = 3$ in SAFER (see *Supplementary material online, Table S4*).

Discussion

The main finding of the present study is that an AI-enabled single-lead ECG algorithm can predict AF from normal sinus rhythm with an acceptable accuracy for an age-homogenous cohort (AUC 0.62), but

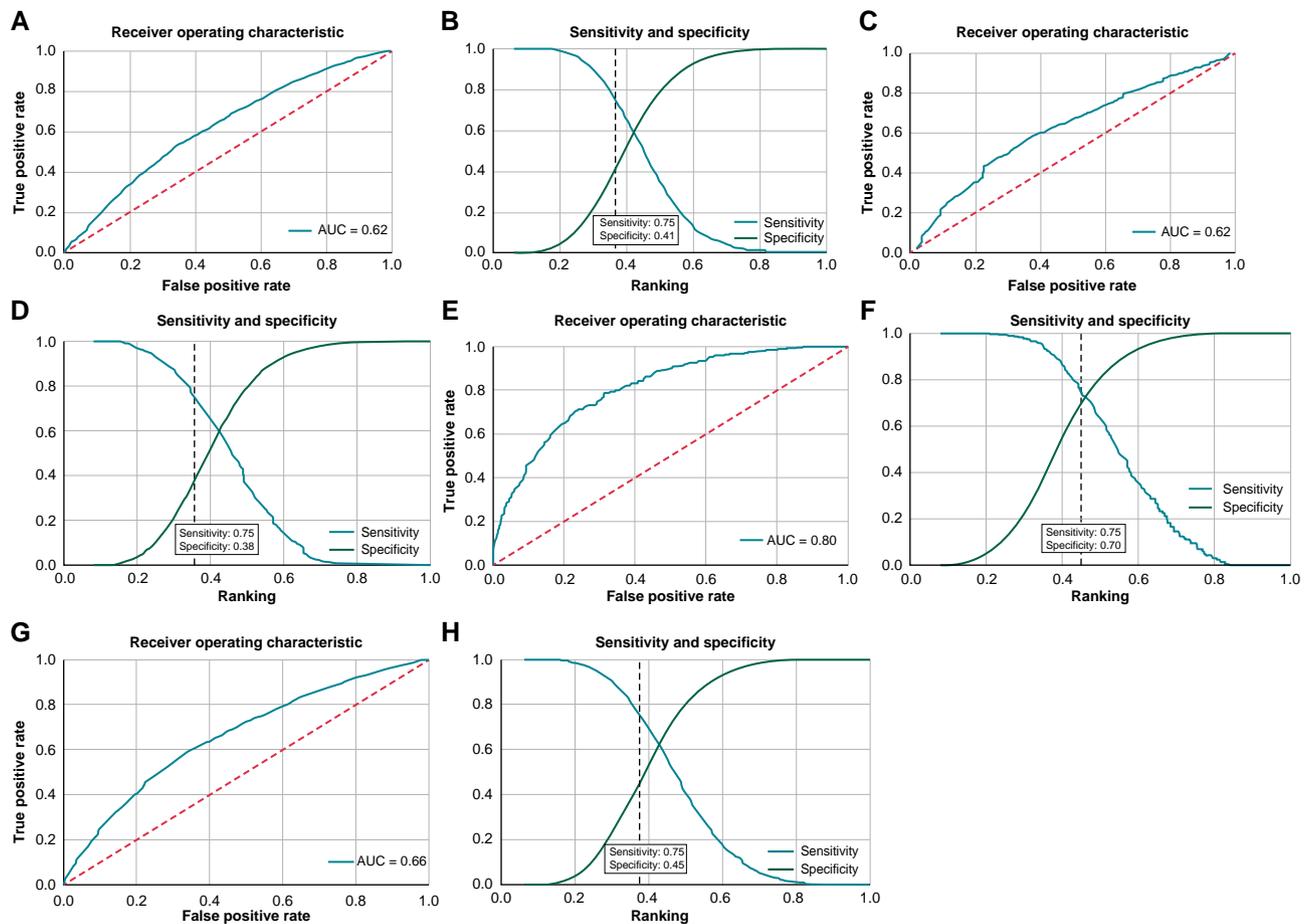


Figure 2 (A) The receiver operating characteristic (ROC) curve displaying performance of the AI model in STROKESTOP I. (B) Sensitivity and specificity for atrial fibrillation (AF) detection during evaluation of the artificial intelligence (AI) model in STROKESTOP I. The dotted vertical line indicates the specificity at 75% sensitivity. (C) The ROC curve displaying performance of the AI model in STROKESTOP II. (D) Sensitivity and specificity for AF detection during evaluation of the AI model in STROKESTOP II. The dotted vertical line indicates the specificity at 75% sensitivity. (E) The ROC curve displaying performance of the AI model in SAFER. (F) Sensitivity and specificity for AF detection during evaluation of the AI model in SAFER. The dotted vertical line indicates the specificity at 75% sensitivity. (G) The ROC curve displaying performance of the AI model in all combined studies. (H) Sensitivity and specificity for AF detection during evaluation of the AI model in all combined studies. The dotted vertical line indicates the specificity at 75% sensitivity.

with substantial improvements if an age-diverse cohort is studied (AUC 0.80).

Our results are in line with an earlier study conducted by Attia *et al.*,⁸ in which an AI-enabled 12-lead ECG was able to detect presence of AF during normal sinus rhythm with an AUC of 0.87 for a single ECG. We have several explanations for obtaining lower AUC values in the current study. Comorbidities may influence the outcome with regard to algorithm performance. In Attia's study, the 12-lead ECGs were obtained in clinical practice for a medical indication, which may create referral bias towards ECG abnormalities. In systematic screening programmes like STROKESTOP I, STROKESTOP II, and SAFER, there might be less bias, but still participants tend to be healthier than non-participants.¹⁷ In Attia's study, all patients aged 18 years or older with at least one 12-lead normal sinus rhythm ECG performed at the Mayo Clinic during a specific time span were included, leading to heterogeneity in terms of age, while in our study, the developed AI algorithm was tested on ECGs

from patients aged exclusively ≥ 65 and mainly 75–76 years. It is well known that AF is highly correlated with age, and previous work has shown that ECGs can predict heart age.¹⁸ Consequently, it is conceivable that previously identified age-related patterns, in addition to AF-specific changes in the ECGs, could explain the high accuracy. This could also potentially explain why our algorithm performed better in SAFER, including patients with a wider age distribution, compared with that in STROKESTOP I and II. The results indicate that the method developed in this work has a comparable performance as the model developed on a 12-lead ECG when patients in the same age group are compared.¹⁹ A 12-lead ECG contains more data compared with the single-lead ECG possibly making modelling more precise. Using a single-lead ECG in ambulatory settings makes the model more susceptible to noise compared with a multi-channel lead ECG performed in the office setting. Comparing the reported F1 scores to the results of Attia *et al.*,⁸ the lower reported values are due to the lower prevalence of AF in the test data set.

Specifically, proper F1 score comparison depends on equal prevalence.

The ability to select individuals that are at risk for having undetected AF from a single-lead sinus rhythm ECG could have important implications for AF screening. A single-lead ECG is inexpensive and easy to use. The algorithm could be used as a first selection step in AF screening to decide which patients would benefit most from prolonged AF screening using intermittent ECG. In such a setting, high sensitivity is more important than specificity. As there is no optimal method available for the selection of individuals who would benefit the most from AF screening, a combined approach using also population discriminators could increase specificity.

Over the last 20 years, several efforts have been made to predict AF by analysing P-wave morphology and other ECG features, but no results have been strong enough to be deemed usable in clinical practice.²⁰ Atrial fibrillation activity causes myocyte changes that lead to inflammation resulting in atrial fibrosis.²¹ The tachycardia caused by AF leads to decreased contractility in the atria that eventually leads to atrial dilatation.²² Atrial fibrosis and enlargement as well as distorted movement patterns in the left atrium caused by low-amplitude electrical activity may give rise to subtle ECG changes that can be discerned by the AI model, but not by the human eye or the earlier used methods.²³

A limitation of our study is that intermittent ECGs only monitor heart rhythm for a brief time period. Hence, it is possible that there are patients deemed as free of AF in our study in whom a more continuous ECG method would have detected AF. This introduces false negatives in the labels and causes noise in the model.

The most important limitation of the study is the low number of individuals diagnosed with AF in the ECG database, in particular from STROKESTOP II and SAFER. With more positive findings, the model might have improved. Data from STROKESTOP II were collected with a bias towards high-risk patients (with NT-proBNP ≥ 125 ng/L), which may negatively impact the ability of the algorithm to predict AF in patients without structural heart disease. Nonetheless, the algorithm performed equally well in the test set consisting of ECGs collected from STROKESTOP I, where the blood test had not been used to pre-screen patients, as in the STROKESTOP II study. The inclusion of only elderly individuals, within two regions of northern Europe, may affect the external validity of the study. However, our cohorts were well balanced from a gender perspective.

Conclusion

In conclusion, an AI-enabled network can predict paroxysmal AF from a sinus rhythm single-lead ECG with reasonable accuracy in an age-homogenous group. In a screening programme, the algorithm may be used as an interim step to identify individuals that might benefit from prolonged screening. This would reduce the number of individuals requiring prolonged screening and increase feasibility.

Supplementary material

Supplementary material is available at *Europace* online.

Funding

The project was funded by Vinnova, Sweden's innovation agency (grant to Zenicor Medical Systems AB). In addition, the project received funding from the Swedish Heart-Lung Foundation and The Center for Innovative Medicine – CIMED. The study also received a research grant from the Swedish Research Council, Dnr 2022-01466. E.S. is supported by the Stockholm County Council (clinical researcher appointment). The SAFER study was funded by the National Institute for Health and Care Research

(NIHR), grant number RP-PG-0217-20007, and by the NIHR School for Primary Care Research. P.H.C. is supported by the British Heart Foundation), grant number FS/20/20/34626. Zenicor Medical Systems AB was involved in planning and gathering data for the study, but all analysis work was performed by an independent contractor (Modulai AB), and have not read nor participated in writing the manuscript.

Conflict of interest: J.M. is on an Academic Advisory Board for Bristol Myers Squibb–Pfizer. K.K.G. has received lecture/consultation fees from Roche Diagnostics, Pfizer, and Boehringer Ingelheim. E.S. has received lecture fees from Bayer, Bristol Myers Squibb–Pfizer, Boehringer Ingelheim, Johnson & Johnson, and Merck Sharp & Dohme. The remaining authors have no conflicts of interest to declare.

Data availability

Requests about the data can be submitted to the authors for consideration. Due to the sensitive nature of the health information contained, not all data can be provided.

References

1. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham study. *Stroke* 1991;**22**:983–8.
2. 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). *Eur Heart J* 2021;**42**:373–498.
3. Saliba W, Gronich N, Barnett-Griness O, Rennert G. Usefulness of CHADS2 and CHA2DS2-VASc scores in the prediction of new-onset atrial fibrillation: a population-based study. *Am J Med* 2016;**129**:843–9.
4. Svennberg E, Henriksson P, Engdahl J, Hijazi Z, Al-Khalili F, Friberg L et al. N-terminal pro B-type natriuretic peptide in systematic screening for atrial fibrillation. *Heart (British Cardiac Society)* 2017;**103**:1271–7.
5. Kemp Gudmundsdottir K, Fredriksson T, Svennberg E, Al-Khalili F, Friberg L, Frykman V et al. Stepwise mass screening for atrial fibrillation using N-terminal B-type natriuretic peptide: the STROKESTOP II study. *Europace* 2020;**22**:24–32.
6. Schnabel RB, Marinelli EA, Arbelo E, Boriani G, Boveda S, Buckley CM et al. Early diagnosis and better rhythm management to improve outcomes in patients with atrial fibrillation: the 8th AFNET/EHRA consensus conference. *Europace* 2023;**25**:6–27.
7. Leclercq C, Witt H, Hindricks G, Katra RP, Albert D, Belliger A et al. Wearables, telemedicine, and artificial intelligence in arrhythmias and heart failure: proceedings of the European Society of Cardiology Cardiovascular Round Table. *Europace* 2022;**24**:1372–83.
8. Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ 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* 2019;**394**:861–7.
9. Loring Z, Mehrotra S, Piccini JP, Camm J, Carlson D, Fonarow GC et al. Machine learning does not improve upon traditional regression in predicting outcomes in atrial fibrillation: an analysis of the ORBIT-AF and GARFIELD-AF registries. *Europace* 2020;**22**:1635–44.
10. Svennberg E, Engdahl J, Al-Khalili F, Friberg L, Frykman V, Rosenqvist M. Mass screening for untreated atrial fibrillation: the STROKESTOP study. *Circulation* 2015;**131**:2176–84.
11. Engdahl J, Andersson L, Mirskaya M, Rosenqvist M. Stepwise screening of atrial fibrillation in a 75-year-old population: implications for stroke prevention. *Circulation* 2013;**127**:930–7.
12. Engdahl J, Svennberg E, Friberg L, Al-Khalili F, Frykman V, Kemp Gudmundsdottir K et al. Stepwise mass screening for atrial fibrillation using N-terminal pro b-type natriuretic peptide: the STROKESTOP II study design. *Europace* 2017;**19**:297–302.
13. Doliwa PS, Frykman V, Rosenqvist M. Short-term ECG for out of hospital detection of silent atrial fibrillation episodes. *Scand Cardiovasc J* 2009;**43**:163–8.
14. Svennberg E, Stridh M, Engdahl J, Al-Khalili F, Friberg L, Frykman V et al. Safe automatic one-lead electrocardiogram analysis in screening for atrial fibrillation. *Europace* 2017;**19**:1449–53.
15. Lee J, Kim T, Park J, Nam J. Raw waveform-based audio classification using sample-level CNN architectures. *arXiv* 2017. 1712.00866.
16. Mockus J. On Bayesian methods for seeking the extremum. In Marchuk GI (ed.), *Optimization Techniques IFIP Technical Conference: Novosibirsk*. Berlin: Springer, 1975, pp. 400–4.
17. Gudmundsdottir KK, Bonander C, Hygrel T, Svennberg E, Frykman V, Strömberg U et al. Factors predicting participation and potential yield of screening-detected disease

- among non-participants in a Swedish population-based atrial fibrillation screening study. *Prev Med* 2022;**164**:107284.
18. Attia ZI, Friedman PA, Noseworthy PA, Lopez-Jimenez F, Ladewig DJ, Satam G et al. Age and sex estimation using artificial intelligence from standard 12-lead ECGs. *Circ Arrhythm Electrophysiol* 2019;**12**:e007284.
 19. Noseworthy PA, Attia ZI, Carter RE, Yao X, Friedman PA. An AI-ECG algorithm for atrial fibrillation risk: steps towards clinical implementation - authors' reply. *Lancet* 2020;**396**:236–7.
 20. Filos D, Chouvarda I, Dakos G, Vassilikos V, Maglaveras N. Beat to beat wavelet variability in atrial fibrillation. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual International Conference*, Basel, 2011. pp. 953–6.
 21. Frustaci A, Chimenti C, Bellocci F, Morgante E, Russo MA, Maseri A. Histological substrate of atrial biopsies in patients with lone atrial fibrillation. *Circulation* 1997;**96**:1180–4.
 22. Sun H, Gaspo R, Leblanc N, Nattel S. Cellular mechanisms of atrial contractile dysfunction caused by sustained atrial tachycardia. *Circulation* 1998;**98**:719–27.
 23. Warraich HJ, Gandhavadi M, Manning WJ. Mechanical discordance of the left atrium and appendage: a novel mechanism of stroke in paroxysmal atrial fibrillation. *Stroke* 2014;**45**:1481–4.