ESC European Society of Cardiology

European Heart Journal - Quality of Care and Clinical Outcomes (2022) **8**, 391–397
https://doi.org/10.1093/ehjqcco/qcab094

REVIEW

What is next for screening for undiagnosed atrial fibrillation? Artificial intelligence may hold the key

Ramesh Nadarajah [1,2,3,*], Jianhua Wu [2,4], Alejandro F. Frangi [1,2,5,6], David Hogg [7], Campbell Cowan [3] and Chris P. Gale [1,2,3]

[1]Leeds Institute for Cardiovascular and Metabolic Medicine, University of Leeds, 6 Clarendon Way, Leeds LS2 9DA, UK; [2]Leeds Institute of Data Analytics, University of Leeds, Leeds, UK; [3]Department of Cardiology, Leeds Teaching Hospitals NHS Trust, Leeds, UK; [4]School of Dentistry, University of Leeds, Leeds, UK; [5]Centre for Computational Imaging and Simulation Technologies in Biomedicine (CISTIB), School of Computing, University of Leeds, Leeds, UK; [6]Alan Turing Institute, London, UK; and [7]School of Computing, University of Leeds, Leeds, UK

Received 23 November 2021; editorial decision 7 December 2021; accepted 10 December 2021; online publish-ahead-of-print 23 December 2021

Atrial fibrillation (AF) is increasingly common, though often undiagnosed, leaving many people untreated and at elevated risk of ischaemic stroke. Current European guidelines do not recommend systematic screening for AF, even though a number of studies have shown that periods of serial or continuous rhythm monitoring in older people in the general population increase detection of AF and the prescription of oral anticoagulation. This article discusses the conflicting results of two contemporary landmark trials, STROKESTOP and the LOOP, which provided the first evidence on whether screening for AF confers a benefit for people in terms of clinical outcomes. The benefit and efficiency of systematic screening for AF in the general population could be optimized by targeting screening to only those at higher risk of developing AF. For this purpose, evidence is emerging that prediction models developed using artificial intelligence in routinely collected electronic health records can provide strong discriminative performance for AF and increase detection rates when combined with rhythm monitoring in a clinical study. We consider future directions for investigation in this field and how this could be best aligned to the current evidence base to target screening in people at elevated risk of stroke.

Keywords Atrial fibrillation • Screening • Stroke • Prediction model • Artificial intelligence

Introduction

Atrial fibrillation (AF) is the most common cardiac arrhythmia worldwide and is associated with a five-fold increased risk of stroke.[1,2] Contemporary estimates suggest that 15% of people with AF are undiagnosed, of whom over half are at moderate to high risk of stroke.[3] However, current European guidelines advocate opportunistic screening with pulse palpation in individuals aged over 65 years instead of a more systematic intensive approach.[2] In this focused review, we critically appraise the current evidence base for systematic screening for AF and consider how to make systematic screening more efficient by defining a population at higher risk of incident AF, including through the use of artificial intelligence.

What is the evidence base for screening for atrial fibrillation?

Screening for AF has received enthusiastic support from several sources.[4] Diagnosis can be challenging due to its often asymptomatic and paroxysmal nature. Asymptomatic or subclinical AF (SCAF) detected by cardiac implanted devices in the ASSERT study was associated with a 2.5-fold increase in the risk of stroke compared with no AF,[5] and treatment of asymptomatic AF with oral anticoagulation has been associated with a reduced risk of stroke and death compared with no antithrombotic therapy.[6,7] Several studies have shown that serial (STROKESTOP II and REHEARSE-AF) or continuous (mSToPS and SCREEN-AF) rhythm monitoring in older people

* Corresponding author. Tel: +44 (0) 113 343 3241, Email: r.nadarajah@leeds.ac.uk, Twitter: @Dr_R_Nadarajah

with vascular risk factors or elevated N-terminal pro B-type natriuretic peptide (NT-proBNP) leads to a significantly higher detection rate of previously undiagnosed AF and more frequent initiation of oral anticoagulation compared with routine standard of care.[8–11] However, a paucity of studies reporting on hard clinical endpoints has led to hesitancy to recommend systematic screening for AF in international guidelines.[2,12]

This year two studies, STROKESTOP and the LOOP, advanced the evidence base by assessing the net benefits of systematic screening on clinical outcomes.[13,14] The STROKESTOP study was a randomized controlled trial (RCT) performed in two regions of Sweden. People aged 75 and 76 years, with no exclusions applied, were randomized 1:1 to an invitation to screening or routine care. Those who participated in screening were instructed on using a handheld single-lead electrocardiogram (ECG) to record ECGs twice daily for 2 weeks. Of 28 768 participants randomly assigned, 13 979 were invited to screening, of whom 7165 (51.3%) participated in screening and 13 996 made up the control group, with a median follow-up of 6.9 years. The LOOP study was an RCT conducted in four centres in Denmark. Individuals without known AF aged 70–90 years, with at least one additional stroke risk factor, were randomly assigned 1:3 to implantable loop recorder (ILR) monitoring or routine care. Of 6004 individuals randomly assigned, 1501 had ILR monitoring and 4503 made up the control group, with a median follow-up of 5.4 years. Treatment with oral anticoagulation was offered if AF was detected, as appropriate to the CHA_2D_S-VASC risk profile of these groups. In STROKESTOP, a small benefit in a composite outcome of ischaemic or haemorrhagic stroke, systemic embolism, bleeding leading to hospitalization, and all-cause death was demonstrated in the intervention group (4456 of 13 979; 5.45 events per 100 years) compared with the control group [4616 of 13 996; 5.68 events per 100 years; hazard ratio 0.96 (95% confidence interval, CI, 0.92–1.00)]. In contrast, the LOOP study found no significant reduction in the risk of stroke or systemic embolism between the ILR group [67 of 1501 (4.5%)] and the control group [251 of 4503 (5.6%); HR 0.80, 95% CI 0.61–1.05; $P = 0.11$].

Why do the results of the two RCTs differ? In the first instance, the sample size and the number of events were much smaller in the LOOP study, reflected in the wide confidence intervals for effect size, which may have left the study underpowered. Furthermore, in the LOOP study, an episode of 6 min of AF on continuous monitoring was sufficient for diagnosis and consideration of anticoagulation. In the ASSERT study, individuals with a duration of SCAF greater than 24 h were found to be at increased risk of stroke compared with those without AF but those with SCAF under 24 h in duration were not found to be at increased risk.[5,15] It is possible that the AF episodes diagnosed in STROKESTOP were more likely to be of longer duration and hence conferred elevated stroke risk and thus had a greater benefit from oral anticoagulation. The threshold of SCAF duration detected on continuous monitoring that would benefit from oral anticoagulation is under evaluation in the ongoing ARTESiA (NCT01938248) double-blind RCT that includes participants with stroke risk factors and an episode of SCAF of at least 6 min duration. Enrolled patients are randomized 1:1 to aspirin or apixaban, with a composite primary outcome of stroke and systemic embolism and a safety outcome of clinically overt major bleeding.

Indeed, STROKESTOP and the LOOP suggest that a more systematic screening approach does not cause harm in terms of a significant increase in bleeding events. Nevertheless, health anxiety to the individual elicited by additional investigations or side effects from treatments started due to an AF diagnosis were not investigated. A cost-effectiveness analysis of STROKESTOP, presented at the European Society of Cardiology (ESC) Congress 2021, suggested that a systematic approach may be cost effective within 3 years.[16] Even so, the efficiency of systematic AF screening could be maximized by targeting individuals at higher risk of incident AF.[17] By way of example, single-time-point screening of a general population of age at least 65 years detects undiagnosed AF in 1.4%,[18] but in STROKESTOP a higher age bracket of 75 and 76-years allied to twice daily ECG recordings over 2 weeks increased the detection rate to 3.0%,[19] and further restricting this group to only those with at least one additional stroke risk factor increased the detection rate to 7.4%.[20]

Could prediction models for incident atrial fibrillation provide more efficient screening?

In many countries, a large proportion of the general population is registered in primary care,[21,22] which provides an ideal setting for screening with nursing support and a direct link with a practitioner capable of prescribing oral anticoagulation.[4] People registered in primary care have a corresponding routinely collected primary care electronic health record (EHR), from which they can be selected for screening. A prediction model utilizing this information could more accurately discriminate people into a higher risk category than screening based on age alone. Several models have been developed or validated for the prediction of incident AF in community-based EHRs using traditional regression techniques,[23] but provide only moderate discriminative performance,[24–26] commonly use variables that may be missing in routinely collected community-based records (such as measures of height, weight, and blood pressure),[22] or give risk prediction over 5 or 10 years, which is difficult to translate into an investigational priority in the immediate future.[22,26]

Machine learning for prediction of incident atrial fibrillation

Over recent years, interest has grown in the use of machine learning (ML) on routinely collected data for the prediction of incident AF.[27] Powerful models have been created using this methodology on ECGs in sinus rhythm and data on hospital outpatient clinics,[28,29] but these are less useful for community-based screening programmes. Three studies have derived prediction models in community-based EHRs using supervised ML methods.[30–32] Random forests were used in all three studies,[30–32] and neural networks in two.[30,32] The random forest model is ensemble learning that combines multiple decision trees where each tree structure, using computationally selected parameters, can differentiate features step by step by creating appropriate splits.[31,32] In neural networks, layers of neurons share weighted directed connections with each neuron

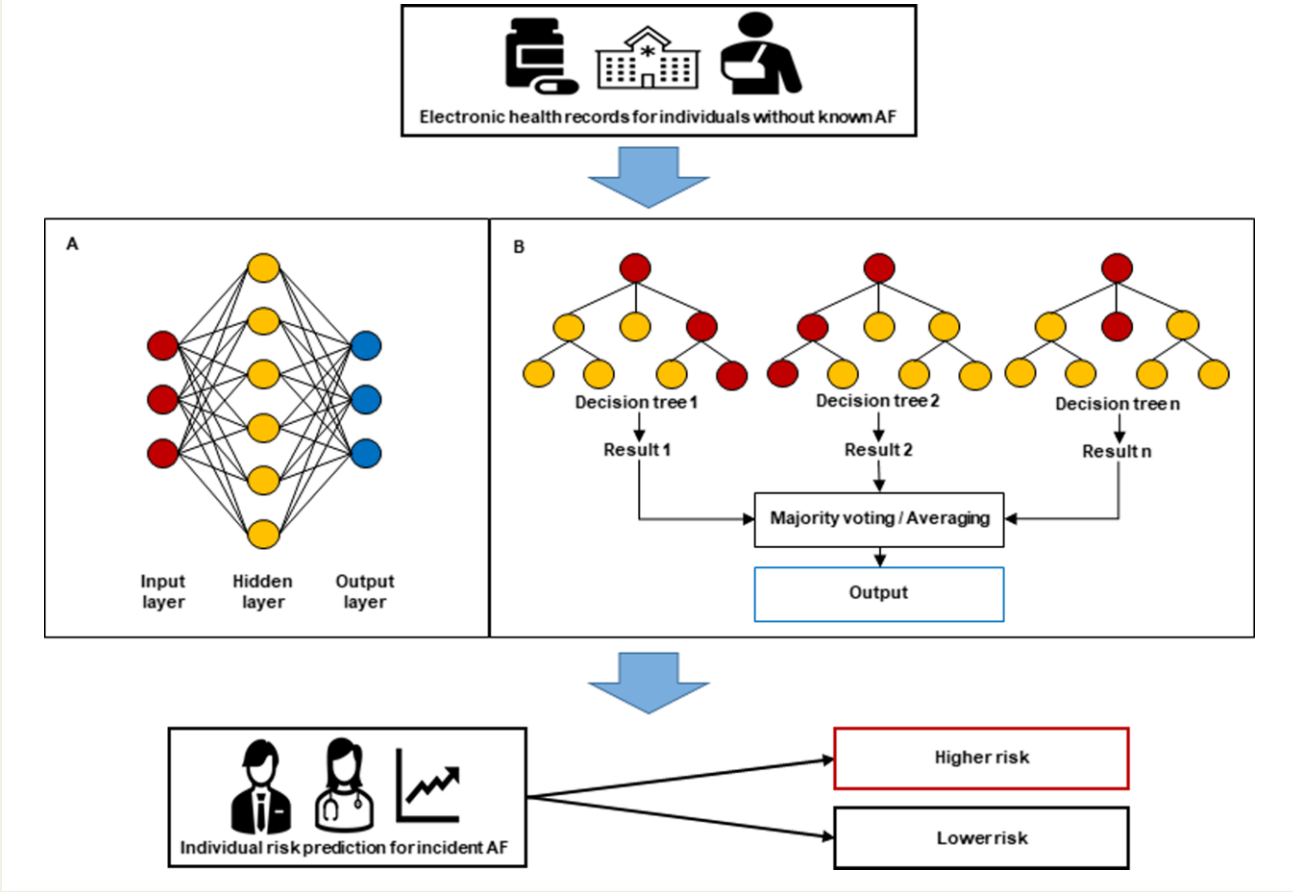

Figure 1 A schematic representation of a 'shallow' neural network (*A*) and a random forest (*B*) model processing an electronic health record to provide risk prediction for incident atrial fibrillation. In (*A*), the data are passed between layers by weighted directed connections, with each neuron responsible for combining inputs via a propagation function and generating outputs. In (*B*), the data are sequentially classified in decision trees through a flowchart-like structure. Random forests construct an ensemble output from many decision trees.

being responsible for combining inputs via a propagation function and generating outputs to be passed to the next layer (*Figure 1*).[27] These ML methods incorporated a large number of variables, of a variety of data types, and demonstrated strong discriminative performance in the derivation data sets, though there has been limited external validation (*Table 1*).[23,33]

The PULsE-AI study, presented at the ESC Congress in 2021, extended AF risk prediction using ML methods into the clinical setting. The investigators implemented a neural network ML model, which had been derived and validated in the UK-based Clinical Practice Research Datalink (CPRD) and Discover EHR data sets, in routinely collected primary care EHRs. Individuals defined by the model at higher risk for developing AF were invited for serial rhythm monitoring.[34] A total of 23 745 participants from six general practices were randomized, with 1880 participants defined as higher risk. Among higher risk participants, 906 were in the intervention arm, of whom 255 (28.1%) consented to diagnostic testing, with 974 in the control group. In the intention-to-treat analysis, AF or related arrhythmias were diagnosed in 5.63% (51/906) and 4.97% (48/974) of the participants in the intervention and control arms, respectively [odds ratio (OR) 1.15; 95% CI 0.77–1.73; $P = 0.486$]. In a subgroup analysis of higher risk participants who accepted the diagnostic testing ('treatment received'), twice as many were diagnosed with AF or related arrhythmias compared with higher risk participants in the control arm [9.41% (24/255) vs. 4.93% (48/973), respectively; OR 2.23, 95% CI 1.31–3.73, $P = 0.003$]. Thus, risk stratification by ML embedded in community-based EHRs allied to serial rhythm monitoring can lead to an increased rate of AF detection compared with routine care.

However, questions remain about whether the implementation of this model would lead to a change in clinical outcomes. First, how rates of prescription of anticoagulation were affected was not sought in the protocol. One must consider that the model was created using data on people as young as 30 years of age,[30] a large proportion of whom may not have stroke risk factors. Second, a complete data set of height, weight, body mass index, and systolic and diastolic blood pressure is required for the model to stratify an individual's risk of undiagnosed AF. A complete data set for these values is only recorded in a minority of primary care EHRs,[22,35,36] therefore potentially limiting the population to whom the model can be applied. Third, the response rate to invitation for diagnostic testing was much lower than that of STROKESTOP (28.1% vs. 53.8%),

Table 1 Examples of the use of random forests and neural networks in community-based electronic health records for prediction of incident atrial fibrillation

Study aim	EHR cohort (country)	Number of variables	Data type used	AUROC
Neural networks				
Derivation	CPRD (UK)	100	Demographics, diagnoses, medications, observations	0.827
External validation	DISCOVER (UK)	100	Demographics, diagnoses, medications, observations	0.870
Derivation	NHIS-NSC (KR)	22	Demographics, diagnoses, medications, observations, laboratory measurements, socioeconomic status	0.813
Random forests				
Derivation	CPRD (UK)	24	Demographics, diagnoses, medications, observations	0.812
Derivation	NHIRD (TW)	19	Demographics, diagnoses, follow-up duration (years), mean CHA_2DS_2-VASC score	0.948
Derivation	NHIS-NSC (KR)	22	Demographics, diagnoses, medications, observations, laboratory measurements, socioeconomic status	0.838

AUROC, area under receiver operating characteristic; CHA_2DS_2-VASc, congestive heart failure, hypertension, age >75 (2 points), stroke/transient ischaemic attack/thromboembolism (2 points), vascular disease, age 65–74, sex category; CPRD, Clinical Practice Research Datalink; EHR, electronic health record; KR, Republic of Korea; NHIRD, National Health Insurance Research Database; NHIS-NSC, National Health Insurance Service-based National Sample Cohort; TW, Taiwan; and UK, United Kingdom.

even though people had been identified as higher risk of developing AF.[19] This may reflect the younger cohort that was analysed by the prediction model and invited to screening, who may have a lower risk profile for sequelae from AF and a weaker mandate for anticoagulation, and thus perceive a lower incentive to attend for screening (especially if they are asymptomatic).

Deep neural networks for disease prediction

Deep learning is a subset of ML methods using multilayered neural networks (deep neural networks, DNNs) to model increasingly complex relationships by adding many 'hidden' layers before the output layer (*Figure 2*).[37] As information passes through the layers of neurons (represented by the nodes in the network), the activation of individual neurons in the network is computed as a weighted sum (with added bias value) of its input neurons from the previous layer and passed through a non-linear activation function.[38] Before becoming operational, the weights and bias values of the DNN must be adjusted to give optimal performance on a training data set of example inputs and outputs. During training, the DNN is repeatedly shown examples of the data and correct answers (supervised learning) with backpropagation underpinning efficient learning.[39] The DNN computes the discrepancy between the output produced for each patient EHR and the correct answer (also known as loss function) and back propagates it through the network to compute and adjust all weights to reduce this discrepancy. The adjustment is made following the gradient of loss function, i.e. along the direction that reduces the discrepancy by an amount proportional to the magnitude of this discrepancy.[39]

Several DNN architectures have demonstrated exceptional discriminative performance for disease prediction in the UK-based routinely collected primary care database provided by the CPRD (*Table 2*),[40,41] which is broadly representative of the UK population by age, sex, and ethnicity.[21] Interestingly, the transformer architecture, a recently proposed multilayer network in which the computation in each layer is more involved than for the regular networks outlined above,[42] achieved an area under the receiver operating characteristic (AUROC) of 0.901 for prediction of incident AF diagnosis in the next 6 months in a supplementary analysis in Li *et al.*[40] This is superior to the performance achieved by the shallow neural network ML model used in the PULsE-AI study when developed in the same database (0.827) (*Table 1*).[30]

DNNs possess a number of advantageous characteristics for clinical disease prediction in EHRs. First, they can use as much information as available in an EHR data set to generate abstract concept and patient representations (unsupervised learning), which may then be used for prediction.[43] Accordingly, they can uncover novel associations and are more scalable than regression and other ML techniques, which require domain and expert knowledge to manually engineer features.[40] Second, different variants of DNNs—convolutional neural networks (CNNs), multilayer recurrent neural networks (RNNs), and transformers—are well suited to capture information on the sequential order of visits and intervisit duration,[40,41] which may better model the temporality of EHR data, a person's evolving health status, and disease pathogenesis. Third, the data types used by the aforementioned DNN models in studies thus far—diagnoses, medications, and demographics—are available with high completeness in EHR (as opposed to say observations and laboratory measurements), meaning these models could be implemented on a greater proportion of patient records. One

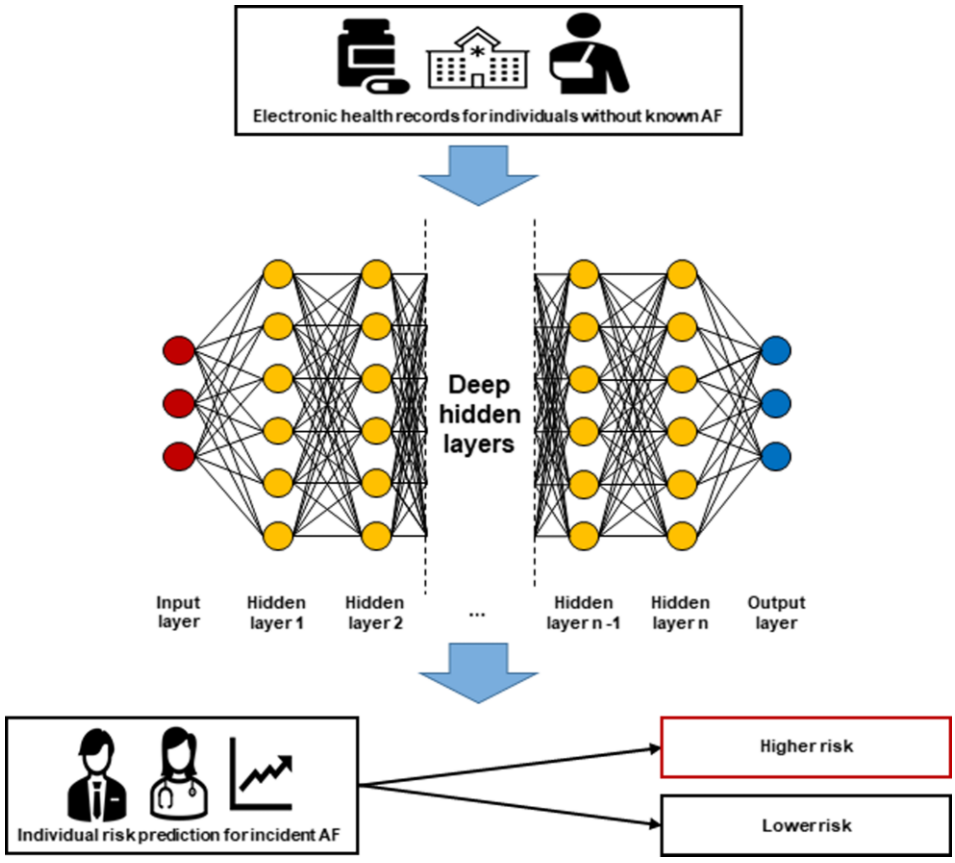

Figure 2 A schematic representation of a deep neural network processing an electronic health record to provide risk prediction for incident atrial fibrillation. The data pass through multiple interconnected layers of neurons allowing the modelling of complex relationships.

Table 2 Deep learning architectures that have been applied in the Clinical Practice Research Datalink data set for disease prediction tasks

Deep learning architecture	Model	Data type used	AUROC
Heart failure diagnosis in next 6 months			
Autoencoder	eNRBM	Demographics, diagnoses, medications	0.920
Stacked denoising autoencoder	Deep Patient	Demographics, diagnoses, medications	0.947
Convolutional neural network	Deepr	Demographics, diagnoses, medications	0.949
Recurrent neural network	RETAIN	Demographics, diagnoses, medications	0.950
Diagnoses at next visit			
Transformer	BEHRT	Demographics, diagnoses	0.954

AUROC, area under receiver operating characteristic.

drawback of DNNs is that they are commonly a 'black box' model where due to their multilayer non-linear structure their predictions are not traceable by humans.[44] The random forest method can show the importance of variables used in its predictions and so may be more likely to be 'trusted' by healthcare professionals and explainable to people when implemented at scale within a screening programme.[45]

Improving the applicability of prediction models for screening for atrial fibrillation

The excellent discriminative performance of DNNs and ML models for prediction of incident AF has been demonstrated in data sets containing all people greater than 16 years of age,[41] 18 years

of age,[31,32] or 30 years of age.[30,33] This is very different from the population that derived benefit from systematic screening for AF in STROKESTOP, participants of age 75 or 76 years.[14] To increase the efficiency of systematic screening, prediction models for incident AF should be developed within the cohort of people who have been proven to derive benefit from screening, especially as targeting screening by age alone is relatively profligate—~67 people aged at least 75 years will need to be screened to identify one new case of AF.[46]

Having said this, channelling a prediction model to differentiate risk of incident AF just for screening in individuals aged at least 75 years may be overly narrow and deny the possibility of an essential diagnosis to younger cohorts with vascular morbidities that increase their risk of developing both AF and stroke. A balance could be struck in developing a prediction model that provides risk stratification in a population similar to the STROKESTOP population, while not inadvertently displaying 'reverse ageism', i.e. a prediction model that can differentiate risk of incident AF *among people with an elevated risk of stroke*, irrespective of age.

In either case, limiting prediction model development to a smaller and more homogeneous population increases the risk of overfitting (as the number of outcome events compared with the number of candidate predictor parameters to be assessed is relatively smaller) and failure, especially for ML models that are notoriously 'data hungry'.[47] Nonetheless, achieving this challenge will likely provide the most incremental benefit by targeting screening within a population that is most likely to derive benefit.

Conclusions

The evidence base for systematic screening for AF continues to evolve, amid widespread enthusiasm and with some conflicting findings.[4] The growth in consumer-facing devices that can detect AF coupled with the strong desire to avoid stroke will continue to fuel the discourse on screening for AF.[48] Undeniably, the STROKESTOP, the LOOP, and the PULsE-AI studies have moved the needle forward. Novel innovation may hold the key to unlocking the maximum benefit of AF screening and make the arguments for translation from trials to clinical practice more convincing.

Acknowledgements

We acknowledge the tremendous help of Katerina Davidson in formatting the figure.

Funding

British Heart Foundation Clinical Research Training Fellowship (FS/20/12/34789 to R.N.); the Royal Academy of Engineering Chair in Emerging Technologies (CiET1819\19 to A.F.F.).

Conflict of Interest: All the authors have completed the ICMJE uniform disclosure form at www.icjme.org/col_disclosure.pdf and declare support from the British Heart Foundation for the submitted work. R.N., J.W., D.H., C.C., and C.P.G. are currently conducting the FIND-AF study (Future Innovations in Novel Detection of Atrial Fibrillation, NCT04657900), investigating the use of artificial intelligence to predict incident AF in routinely collected primary care electronic health records. C.P.G. reports personal fees from AstraZeneca, Amgen, Bayer, Boehringer Ingelheim, Daiichi Sankyo, Vifor, Pharma, Menarini, Wondr Medical, Raisio Group, and Oxford University Press, and has received educational and research grants from BMS, Abbott Inc., the British Heart Foundation, National Institute of Health Research, Horizon 2020, and the European Society of Cardiology, outside the submitted work.

Author contributions

R.N. and C.P.G. conceived the idea of the review. R.N. drafted the manuscript. J.W., C.C., A.F.F., D.H., and C.P.G. critically reviewed the manuscript and R.N. revised the manuscript for final submission. All authors have approved the final draft of the manuscript. The corresponding author attests that all listed authors meet the authorship criteria and that no others meeting the criteria have been omitted.

Data availability

No new data were generated or analysed in support of this research.

References

1. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. *Stroke* 1991;**22**:983–988.
2. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomstrom-Lundqvist C *et al.* 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association of 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* 2020;**42**:373–498.
3. Turakhia MP, Shafrin J, Bognar K, Trocio J, Abdulsattar Y, Wiederkehr D *et al.* Estimated prevalence of undiagnosed atrial fibrillation in the United States. *PLoS One* 2018;**13**:e0195088.
4. Freedman B, Camm J, Calkins H, Healey JS, Rosenqvist M, Wang J *et al.* Screening for atrial fibrillation: a report of the AF-SCREEN International Collaboration. *Circulation* 2017;**135**:1851–1867.
5. Van Gelder IC, Healey JS, Crijns HJ, Wang J, Hohnloser SH, Gold MR *et al.* Duration of device-detected subclinical atrial fibrillation and occurrence of stroke in ASSERT. *Eur Heart J* 2017;**38**:1339–1344.
6. Martinez C, Katholing A, Freedman SB. Adverse prognosis of incidentally detected ambulatory atrial fibrillation. *Thromb Haemost* 2014;**112**:276–286.
7. Freedman B, Martinez C, Katholing A, Rietbrock S. Residual risk of stroke and death in anticoagulant-treated patients with atrial fibrillation. *JAMA Cardiol* 2016;**1**:366–368.
8. Gladstone DJ, Wachter R, Schmalstieg-Bahr K, Quinn FR, Hummers E, Ivers N *et al.* Screening for atrial fibrillation in the older population: a randomized clinical trial. *JAMA Cardiol* 2021;**6**:558–567.
9. Halcox JP, Wareham K, Cardew A, Gilmore M, Barry JP, Phillips C *et al.* Assessment of remote heart rhythm sampling using the AliveCor heart monitor to screen for atrial fibrillation: the REHEARSE-AF study. *Circulation* 2017;**136**:1784–1794.
10. Steinhubl SR, Waalen J, Edwards AM, Ariniello LM, Mehta RR, Ebner GS *et al.* Effect of a home-based wearable continuous ECG monitoring patch on detection of undiagnosed atrial fibrillation: the mSToPS randomized clinical trial. *JAMA* 2018;**320**:146–155.
11. 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. *EP Europace* 2020;**22**:24–32.
12. Curry SJ, Krist AH, Owens DK, Barry MJ, Caughey AB, Davidson KW *et al.* Screening for atrial fibrillation with electrocardiography: US Preventive Services Task Force recommendation statement. *JAMA* 2018;**320**:478–484.
13. Svendsen JH, Diederichsen SZ, Højberg S, Krieger DW, Graff C, Kronborg C *et al.* Implantable loop recorder detection of atrial fibrillation to prevent stroke (The LOOP Study): a randomised controlled trial. *Lancet North Am Ed* 2021;**398**:1507–1516.
14. Svennberg E, Friberg L, Frykman V, Al-Khalili F, Engdahl J, Rosenqvist M. Clinical outcomes in systematic screening for atrial fibrillation (STROKESTOP): a

multicentre, parallel group, unmasked, randomised controlled trial. *Lancet North Am Ed* 2021;**398**:1498–1506.
15. Jones NR, Taylor CJ, Hobbs FR, Bowman L, Casadei B. Screening for atrial fibrillation: a call for evidence. *Eur Heart J* 2020;**41**:1075–1085.
16. Lyth J, Svennberg E, Bernfort L, Aronsson M, Frykman V Friberg L *et al.*, eds. STROKESTOP: Cost-effectiveness of population screening for atrial fibrillation based on long-term follow-up data. *European Society of Cardiology Congress*; 2021.
17. Khurshid S, Healey JS, McIntyre WF, Lubitz SA. Population-based screening for atrial fibrillation. *Circ Res* 2020;**127**:143–154.
18. Lowres N, Neubeck L, Redfern J, Freedman SB. Screening to identify unknown atrial fibrillation. *Thromb Haemost* 2013;**110**:213–222.
19. 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**:2176–2184.
20. 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**:930–937.
21. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, Van Staa T *et al.* Data resource profile: clinical practice research datalink (CPRD). *Int J Epidemiol* 2015;**44**:827–836.
22. Himmelreich JC, Lucassen WA, Harskamp RE, Aussems C, van Weert HC, Nielen MM. CHARGE-AF in a national routine primary care electronic health records database in the Netherlands: validation for 5-year risk of atrial fibrillation and implications for patient selection in atrial fibrillation screening. *Open Heart* 2021;**8**:e001459.
23. Nadarajah R, Alsaeed E, Hurdus B, Aktaa S, Hogg D, Bates MG *et al.* Prediction of incident atrial fibrillation in community-based electronic health records: a systematic review with meta-analysis. *Heart*; doi:10.1136/heartjnl-2021-320036. Published online ahead of print 4 October 2021.
24. 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**:843–849.
25. Suenari K, Chao TF, Liu CJ, Kihara Y, Chen TJ, Chen SA. Usefulness of HATCH score in the prediction of new-onset atrial fibrillation for Asians. *Medicine (Baltimore)* 2017;**96**:e5597.
26. Li YG, Pastori D, Farcomeni A, Yang PS, Jang E, Joung B *et al.* A simple clinical risk score (C2HEST) for predicting incident atrial fibrillation in Asian subjects: derivation in 471,446 Chinese subjects, with internal validation and external application in 451,199 Korean subjects. *Chest* 2019;**155**:510–518.
27. Trayanova NA, Popescu DM, Shade JK. Machine learning in arrhythmia and electrophysiology. *Circ Res* 2021;**128**:544–566.
28. Hulme OL, Khurshid S, Weng L-C, Anderson CD, Wang EY, Ashburner JM *et al.* Development and validation of a prediction model for atrial fibrillation using electronic health records. *JACC Clin Electrophysiol* 2019;**5**:1331–1341.
29. 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 North Am Ed* 2019;**394**:861–867.
30. Hill NR, Ayoubkhani D, McEwan P, Sugrue DM, Farooqui U, Lister S *et al.* Predicting atrial fibrillation in primary care using machine learning. *PLoS One* 2019;**14**:e0224582.
31. Hu W-S, Hsieh M-H, Lin C-L. A novel atrial fibrillation prediction model for Chinese subjects: a nationwide cohort investigation of 682 237 study participants with random forest model. *EP Europace* 2019;**21**:1307–1312.
32. Kim I-S, Yang P-S, Jang E, Jung H, You SC, Yu HT *et al.* Long-term PM 2.5 exposure and the clinical application of machine learning for predicting incident atrial fibrillation. *Sci Rep* 2020;**10**:1–11.
33. Sekelj S, Sandler B, Johnston E, Pollock KG, Hill NR, Gordon J *et al.* Detecting undiagnosed atrial fibrillation in UK primary care: validation of a machine learning prediction algorithm in a retrospective cohort study. *Eur J Prev Cardiol* 2020;**28**:598–605.
34. Cohen AT, Hill NR, Groves L, Hurst M, Pollock KG, Arden C *et al.*, eds. Identification of undiagnosed atrial fibrillation using a machine learning risk prediction algorithm and diagnostic testing (PULsE-AI) in primary care: a multi-centre randomised controlled trial in England. *European Society of Cardiology Congress*; 2021.
35. Marston L, Carpenter JR, Walters KR, Morris RW, Nazareth I, Petersen I. Issues in multiple imputation of missing data for large general practice clinical databases. *Pharmacoepidemiol Drug Saf* 2010;**19**:618–626.
36. Verberne LD, Nielen MM, Leemrijse CJ, Verheij RA, Friele RD. Recording of weight in electronic health records: an observational study in general practice. *BMC Fam Pract* 2018;**19**:1–8.
37. Nadarajah R, Wu J, Frangi AF, Hogg D, Cowan C, Gale C. Predicting patient-level new-onset atrial fibrillation from population-based nationwide electronic health records: protocol of FIND-AF for developing a precision medicine prediction model using artificial intelligence. *BMJ Open* 2021;**11**:e052887.
38. Hagan MT, Demuth HB, Beale M. *Neural Network Design*. PWS Publishing Co.; 1997.
39. Hinton G. Deep learning—a technology with the potential to transform health care. *JAMA* 2018;**320**:1101–1102.
40. Li Y, Rao S, Solares JRA, Hassaine A, Ramakrishnan R, Canoy D *et al.* BEHRT: transformer for electronic health records. *Sci Rep* 2020;**10**:1–12.
41. Solares JRA, Raimondi FED, Zhu Y, Rahimian F, Canoy D, Tran J *et al.* Deep learning for electronic health records: a comparative review of multiple deep neural architectures. *J Biomed Inform* 2020;**101**:103337.
42. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN *et al.* Attention is all you need. *Adv Neural Inf Process Syst* 2017;5998–6008.
43. Deo RC. Machine learning in medicine. *Circulation* 2015;**132**:1920–1930.
44. Buhrmester V, Münch D, Arens M. Analysis of explainers of black box deep neural networks for computer vision: a survey. *Mach Learn Knowl Extr* 2021;**3**:966–989.
45. Ooka T, Johno H, Nakamoto K, Yoda Y, Yokomichi H, Yamagata Z. Random forest approach for determining risk prediction and predictive factors of type 2 diabetes: large-scale health check-up data in Japan. *BMJ Nutr Prev Health* 2021;**4**:140–148.
46. Lowres N, Olivier J, Chao T-F, Chen S-A, Chen Y, Diederichsen A *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* 2019;**16**:e1002903.
47. van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC Med Res Method* 2014;**14**:1–13.
48. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T *et al.* Large-scale assessment of a smartwatch to identify atrial fibrillation. *N Engl J Med* 2019;**381**:1909–1917.