Artificial Intelligence and Atrial Fibrillation

Ojasav Sehrawat[1], Anthony Kashou[1], and Peter Noseworthy[1]

[1]Mayo Clinic Department of Cardiovascular Medicine

December 29, 2021

Abstract

In the context of atrial fibrillation (AF), traditional clinical practices have thus far fallen short in several domains such as identifying patients at risk of incident AF or patients with concomitant undetected paroxysmal AF. Novel approaches leveraging artificial intelligence have the potential to provide new tools to deal with some of these old problems. In this review we focus on the roles of artificial intelligence-enabled ECG pertaining to AF, potential roles of deep learning (DL) models in the context of current knowledge gaps, as well as limitations of these models. One key area where DL models can translate to better patient outcomes is through automated ECG interpretation. Further, we overview some of the challenges facing AF screening and the harms and benefits of screening. In this context, a unique model was developed to detect underlying hidden AF from sinus rhythm and is discussed in detail with its potential uses. Knowledge gaps also remain regarding the best ways to monitor patients with ESUS and who would benefit most from oral anticoagulation. The AI-enabled AF model is one potential way to tackle this complex problem as it could be used to identify a subset of high-risk ESUS patients likely to benefit from empirical oral anticoagulation. Role of DL models assessing AF burden from long duration ECG data is also discussed as a way of guiding management. There is a trend towards the use of consumer-grade wristbands and watches to detect AF from photoplethysmography data. However, ECG currently remains the gold standard to detect arrythmias including AF. Lastly, role of adequate external validation of the models and clinical trials to study true performance is discussed.

Introduction

The ECG has an incredible untapped diagnostic and prognostic potential. This easy to use, omnipresent technology has been largely unchanged for a century but still remains critical to the everyday clinical workflow. It is non-invasive, provides rapid actionable insights, easy to perform and readily available at a low cost compared to advanced cardiac imaging modalities.[1] These attributes make it a go-to baseline evaluation test for many patients across the spectrum of healthcare settings.[1,2]

Currently, the ECG is used primarily as a diagnostic tool rather than a broad screening tool for conditions other than rhythm disorders.[3-5] This is largely because standard ECG interpretation using analogue or feature-based approaches lack the negative predictive value to exclude cardiac disease (such as myocardial or valvular heart disease) in ECGs that 'appear normal'. Furthermore, the accuracy of results depends considerably on the interpreter's competency level. However, this could soon change as technological advances are breathing new life into this century old modality.

This potential change has come forth due to recent computational advancements which have allowed a significant improvement in the machine learning algorithms alongside a vast availability of well-annotated digitalized ECG data.[6] Particularly important has been the application of a branch of machine learning known as deep learning (DL) to the ECG[7] which has allowed investigators to derive new insights from these voltage-time matrices. It is also now evident that largely undefined markers of health exist that, while unapparent to an expert cardiologist, can be recognized by deep neural network (DNN) models.[8] Information at the point-of-care can also be leveraged to facilitate 'cardiologist-level' ECG interpretation,[6,9] and even go beyond human capability to determine age and sex,[10] left ventricular dysfunction,[11] predicting hypertrophic cardiomyopathy[12] and atrial fibrillation (AF) from sinus rhythm[11] amongst others.

In the context of AF, traditional clinical practices have thus far fallen short in several domains including identifying patients at risk of incident AF or those with concomitant undetected paroxysmal AF. Innovative approaches leveraging AI have the potential to provide solutions to solve some of these old problems. In this review we focus only on the roles of AI-enabled ECG (AI-ECG) as it relates to AF, the potential role of DL models in the context of current knowledge gaps, as well as the current limitations of these models.

Artificial Neural Networks

Artificial neural networks (ANNs) are computational predictive analytical systems inspired by the human nervous system. These consist of a high number of computational nodes (called neurons) that spread across various layers. In simplest terms, it is comprised of three basic layers with distinct functions.

1. An input layer, in which data (usually multidimensional vector) is ingested and distributed to the hidden layer.
2. A hidden layer, which makes decisions to assign random weights within each node to determine if it detriments or improves the output (referred to as learning). When multiple hidden layers are stacked together to perform a complex pattern recognition task, it is referred to as DL. There are two fundamental learning approaches, supervised and unsupervised learning. Supervised learning is most often used for image-based pattern recognition tasks. In this form, for every training set (e.g., a single digital ECG), input vectors are associated to one or more pre-assigned labels [13]. In contrast unsupervised learning does not have pre-selected labels, but rather seeks to find clusters of salient features from the data itself.
3. An output layer where, in the case of supervised learning, a system output is compared to the preset label output and backpropagated to the previous layers with the goal of reducing classification error as much as possible by tuning the weights.

Convolutional neural networks (CNN) are similar to traditional ANN, but are particularly well suited for image recognition tasks, and thus reduce the parameters required to set up the model. The capability of these models to analyze subtle details from abstract data is remarkable and far superior to humans. More details on these models are described elsewhere.[14]

The AF Epidemic

AF has emerged as a major public health problem and global epidemic.[15,16] Not only is it the most common arrythmia but it is also responsible for more morbidity than any other rhythm disorder.[17,18] Data from the Framingham Heart Study revealed a lifetime risk of approximately 1 in 4 for men and women of age 40 years or older.[19] This signifies a substantial burden on healthcare costs with approximately 350,000 hospitalizations, 5 million office visits, 276,000 emergency visits and 234,000 outpatient visits attributable to AF annually within the United States translating to an estimated cost of treatment at 6.65 billion dollars.[18] The prevalence is further estimated to increase 2.5 to 3-fold which projects to 5.6 million affected Americans 2050.[20] Additionally, global AF burden is not well known and is most likely underestimated in many regions outside of North America and Europe.[21]

AF is vastly heterogenous with a myriad of causes, presentations and complications. There is a need to improve contemporary practices as the prevention and treatment outcomes in AF are sub-optimal.[22] AF is an independent predictor of all-cause mortality, hemorrhage, left ventricular dysfunction and thromboembolism.[23,24] A global collaborative effort to deal with the growing concern of AF is underway especially in the past 2 decades in areas of basic sciences, epidemiology, genetics along with clinical studies and now with the use of AI.

Current Computerized ECG Interpretation Models

Computerized algorithms designed to interpret 12-lead standard (12 SL) ECGs have been commercially available since early 1980s.[25] Currently, computer generated ECG interpretations are widely used to provide 'first pass' interpretations which are then over-read by trained technicians and physicians. Most of these models use traditional statistical algorithms and require intensive feature extraction and engineering to compute. Misdiagnoses are not uncommon with this approach especially if the reliance on computer generated interpretation is high.[13,26] For instance, in a retrospective review computer interpreted diagnosis of AF was incorrect in 35% of the 1085 patients.[27]Furthermore, the physician ordering the ECG failed to correct the interpretation in 10% of these patients. This has significant downstream effects e.g., inappropriate management and unnecessary testing.[27] In one recent study, about 10% misinterpretation rates were reported for AF- 47% of which were not corrected by overreading primary care physicians.[28]Analysis from the SAFE (screening for atrial fibrillation in the elderly) trial, assessing diagnostic accuracy of general practitioners (from 49 practices) and traditional computer interpretations to diagnose AF revealed 20% missed AF cases and 8% false positives compared to reference standard cardiologists.[29] These studies highlight the limitations of current computer algorithms used to interpret ECGs and the devastating consequences on patient outcomes when there is an overreliance on them.

AI-Enabled ECG Interpretation Model:

There is much room for improvement in the current automated ECG interpretation.[30,31] This change could likely translate to better outcomes as incorrect labels are associated with incorrect physician overreads.[27,32,33] End-to-end DNN models to interpret ECG have recently shown great promise to replace the 'feature-based' computer algorithms currently used.

In one study, a DNN model was trained on a dataset of 91,232 ECGs to detect 12 rhythm classes from a single lead, patch based ambulatory monitor.[6] Results showed an AUC>0.91 and a superior model performance compared to separate annotations made by 6 cardiologists. AUC of the model for AF was as high as 0.96 but the small testing dataset of only 328 ECGs limits the reliability of results for any individual class [34]. For external validation, the model was trained and tested using 2017 Physionet Challenge data with a relatively larger testing dataset (performance for AF diagnosis- F1 score 0.84). However, the testing dataset was not randomly selected, and more rare diagnoses were purposefully included which makes the results less generalizable.[6]

In another proof-of-concept study, a DNN model was trained using over 2 million ECGs to detect 6 abnormalities from 12 SL ECG.[7] However, individual diagnostic accuracy for the 6 abnormalities selected is hard to comment on due to a smaller testing dataset used to assess several distinct diagnoses. For instance, testing dataset included only 13 AF cases.[7]

PhysioNet/computing in cardiology challenge 2017,[35]allowed an opportunity for external validation of algorithms from 75 teams which were trained using a common training dataset of 8,528 single lead ECGs. These then competed head-to-head on a hidden dataset to diagnose AF amongst the 4 labeled outputs (normal, AF, other, noisy). Winning algorithms (F1 score 0.83) varied from hand featured models using random forest, extreme gradient boosting to CNN and recurrent neural networks. However, the authors concluded that training set was not sufficient to allow an advantage for more complex algorithms that require enormous data for parameter and hyperparameter tuning. Furthermore, although training set was relatively larger (3,658 ECGs with 311 AF cases), only 27.3% of it was used to rank the algorithms.[35]

A preliminary study demonstrated that triaging patients in emergency care setting from interpretation of 12 SL ECG using a CNN.[36] The results were also compared to their conventional ECG algorithm currently in use. Output was mapped to 16 pre-specified groups to provide actionable information. Results showed slightly better performance than the traditional interpretation but was statistically significant. One of the limitations was again a small testing dataset which included few emergency readings (total 60).[36]

Our group has developed a DNN model to make a comprehensive 12-lead ECG interpretation using 2.5 million standard 12 SL-ECGs from 720,000 patients.[9,37] A 'transformer model' was also incorporated to translate the output into 66 discrete readable ECG diagnostic codes and make a multilabel prediction comparable to current computer automated programs. We previously showed the performance of this model according to all individual 66 codes included using a testing dataset of 499,917 ECGs.[37] The overall performance was an AUC of [?]0.98 for 62 of the 66 reported codes. Recently, the performance was evaluated head-to-head against the traditional ECG interpretation software currently in use at our institute and the final cardiologist over-read diagnosis.[9] Results showed an average ideal or acceptable diagnosis of 91.8% (AI-enabled interpretation) vs 86.6% (computer generated interpretation) vs 94% (final clinical diagnosis). In some ways these studies show the potential of DNN models to provide a level of ECG interpretation previously confined to the realms of field experts.

Challenges in Screening

There are many uncertainties regarding our current approaches to AF screening. Remaining questions include which subgroup of patients to screen, the best modality to use for screening and the subset of subclinical AF (SCAF) cases likely to benefit from oral anticoagulants (OAC). Two recently published large screening studies have produced somewhat contradictory results but provide important new insights about such questions.

In the 'clinical outcomes in systematic screening for atrial fibrillation' (STROKESTOP)[38] trial, population aged 75 to 76 years without an earlier diagnosis of AF were screened using intermittent ECG recordings over 2 weeks yielded 3% detection of new AF cases with 90% patients eventually put on OAC.[39] AF detection rate was only 0.5% from the first ECG used to screen.[39] This highlights the low yield of using a single ECG for screening even in elderly population because of a paroxysmal nature of the arrhythmia. With further continuation of this study (median follow up 6.9 years), screening resulted in a slight benefit in outcomes (i.e., stroke, emboli and bleeding) compared to standard of care.[38] These results suggest that screening is safe and beneficial in elderly population.[38]

In contrast, the 'implantable loop recorder detection of atrial fibrillation to prevent stroke' (LOOP)[40] trial screened patients aged 70-90 years, with at least one additional stroke risk factor.[40] In the intervention group, implantable loop recorders were used for prolonged monitoring to detect AF and started on OAC if AF lasted more than 6 minutes. Results showed 3 times increase in AF detection and OAC initiation however, it did not result in better outcomes (stroke and emboli prevention) compared to the control group. Different results from this study have been attributed to detection of very short episodes of AF with loop recorder which might not benefit from anticoagulation.[40]

The best approach to population wide screening also remains unresolved. Systematic screening using (pulse palpation alongside an ECG) has not been shown to be superior to opportunistic screening (pulse palpation followed by ECG if the former is positive) in various clinical trials.[29,41,42]

Benefits and Harms of Screening

There is 30% first year mortality rate amongst AF patients who have stroke and another 30% are permanently disabled.[43]Therefore, one would expect that there is merit in detecting AF cases early in the course at the stage of paroxysmal and subclinical stage before considerable remodeling has occurred to allow for a higher likelihood of spontaneous conversion, early OAC to prevent strokes, and overall better outcomes for patients.[44] Meta-analyses data consistently show lower risk of primary composite outcomes of stroke and emboli in AF patients treated with warfarin and direct acting OAC.[45-47]

Implementation of a risk modification plan (such as weight reduction, decreasing alcohol intake, treatment of obstructive sleep apnea) in patients identified to be at a future high risk of developing AF could also alter the overall outcomes.[48] Especially as considerable research efforts to identify risk factors of AF are transpiring, an effective screening plan could eventually become crucial.

Lastly, another important consideration is harm from screening which is not extensively studied.[49] Potential harms include misinterpretation of ECGs, with false positive results leading to unnecessary testing and treatment with OAC. Additionally, with increasing screening modalities being employed, anticoagulation rates are bound to increase and thus, there is a need to be mindful of increased bleeding risk.[49]

Ideally, the screening modality needs to be cheap, widely available, non-invasive and non-cumbersome for the patient. Stratifying patients who are at higher risk is likely beneficial to increase diagnostic yield of any modality, to be cost effective and reduce unnecessary testing and distress to patients. Thus, having a tool which can make prediction of incident AF using only a 12 SL ECG has potential to affect patient outcomes.

AI-ECG AF Model

To tackle the complex problem of screening for AF, especially pertaining to paroxysmal and SCAF, our group developed a unique model to predict the likelihood of a person having underlying hidden AF from a sinus rhythm 'apparently normal' ECG without any additional information.[11] The rationale for this study was that mechanical remodeling in the form of myocyte hypertrophy, fibrosis and chamber enlargement might lead to ECG changes yet undefined and too subtle to be studied effectively by human potential. For instance, evidence of interatrial block (Bayes syndrome) which is seldom reported on ECGs correlates to both risk of incident AF and stroke.[50,51]

Over 20 years of ECG data from 180,922 patients and 649,931 normal sinus rhythm ECGs were analyzed. Patients were randomly assigned into 3 groups- 70% for training dataset, 10% for internal validation (optimization and selecting hyperparameters) and 20% testing set (previously unexposed ECGs). Within each dataset there were 2 groups- patients with at least one ECG confirmed AF/atrial flutter diagnosis and patients with no AF rhythms recorded. A 31-day window period preceding the AF recorded ECG was taken in the disease group and all sinus rhythm ECGs were included from the control group. This short 31-day period prior to AF diagnosis was taken to include ECGs with the maximum potential markers associated with AF and left atrial remodeling.

To optimize performance, eight independent leads (leads I, II, and V1–6) were selected because any linear function of the leads could be learned by the models (8×5000 matrix). The model was then tested on a dataset of 130,802 sinus rhythm ECGs (3051 verified AF cases) (Table 1). Model performed well with an AUC of 0·87 when a single ECG was used with no additional information and an AUC of 0.90 when multiple ECGs were used.

Potential Applications of AI-ECG AF Model

AI-ECG model thresholds can be adjusted to be more specific for patients with low pretest probability such as healthy population. This could help make a more cost-effective strategy to rule in patients for further testing (Figure 1).

To rule out patients with AF, a higher sensitivity is needed such as in patients with cryptogenic stroke. Patients with a higher likelihood of AF could then undergo more extensive monitoring (eg., beyond the recommended 30-day monitoring period) (Figure 1).[52]

As part of a risk score to predict 5- or 10-year probability of incident AF, which may have utility in prevention trials and screening programs.[53] Previously this has been attempted with various risk scores, such as the Framingham AF and CHARGE-AF risk scores.[54,55] Although not developed for this purpose, the usefulness of the model as an independent predictor of future AF has been externally validated in a population-based Mayo Clinic study of Ageing.[56] In the study, both CHARGE-AF score and AI-ECG AF model had similar performance (C statistic of 0.69 for both). Combining the two resulted in a slight increase in overall performance. Participants with an AI-ECG AF model output of >0.5 had a cumulative incidence of 21.5% at 2 years and 52.2% at 10 years.[56]

In another recent study, a methodology similar to AI-ECG AF model was used to create an ECG-based algorithm for AF prediction. This model was trained using sinus rhythm ECGs at baseline and a window of interest for ECG was at least 1-3 years before AF diagnosis (compared to within a 30-day window in the Mayo Clinic model). They evaluated the performance of this model using UK Biobank data and showed an overall comparable performance to the CHARGE-AF risk score. There was a modest improvement in model performance when the AI-ECG was added to CHARGE-AF. These studies demonstrate that the ECG holds value in predicting AF (not just detecting concomitant AF) and that the clinical features captured by CHARGE-AF likely explain a lot of the predictive poster of the AI-ECG model.[57]

As a tool to guide management, for instance, identification of a high-risk subset of patients with ESUS may help determine who may benefit from empirical anticoagulation. This utility is currently planned as a follow up study to the BEAGLE trial.[34]

It may have utility to guide management in difficult cases when AF is highly suspected but has eluded diagnosis (case report).[58]

Cryptogenic Stroke

Cryptogenic strokes are defined as symptomatic cerebral infarcts for which no probable cause is identified after thorough standard evaluation.[59] About one-third of transient ischemic attacks [60] and ischemic strokes are of undetermined etiology (cryptogenic).[60] These numbers have decreased over time from as high as 40% in 1970's[61] to as low as 10-15% today in advanced centers with extensive testing modalities.[29,62] This highlights the importance of better ways to investigate patients in order to initiate an appropriate and timely secondary prevention strategy.

AF prevalence as high as 24.6% has been noted in patients presenting with first time ischemic strokes and was especially high amongst elderly females.[63,64] In CRYSTAL AF (Cryptogenic Stroke and Underlying AF) trial, AF detection was compared between insertable cardiac monitors [65] and conventional follow-up in patients with cryptogenic stroke or TIA.[65] Results showed detection rate of AF at 8.9% vs 1.4% at 6 months and 12.4% vs 2% at 12 months for ICM vs conventional follow-up respectively. This proves that many cases go undetected after the first thromboembolic event and foreshadow a recurrence which could have been prevented. In the same trial, mean time in AF was only 4.3 minutes a day and about 74% patients were asymptomatic, highlighting the increasingly difficult diagnosis of paroxysmal AF in these patients.[65] Since prolonged monitoring is inconvenient and expensive, the authors called for further studies to determine which risk factors could better delineate which patients would derive the most clinical benefit from extensive monitoring.

Current recommendation for patients with cryptogenic stroke is that a prolonged rhythm monitoring of about 30 days within 6 months of index event is reasonable.[66] Furthermore, even though a single 1-hour episode of AF during 2 years of monitoring doubles the risk of stroke, the treatment benefit of anticoagulants vs antiplatelet agents is not clearly defined amongst low burden paroxysmal AF patients.

A large subgroup of cryptogenic strokes (80-90%), in which the cause is almost always embolic (superficial or deep large infarcts) is termed embolic stroke of undetermined significance (ESUS).[59]Low burden paroxysmal AF forms an important underlying cause for these patients. However, empiric anticoagulation is not recommended as bleeding risk seems to outweigh the benefits.[67,68]

Therefore, knowledge gaps remain regarding the best way to monitor patients with ESUS and who would benefit most from oral anticoagulation. The AI-ECG AF model is one potential way to tackle this complex problem as it could be used to identify a subset of high-risk ESUS patients most likely to benefit from empirical oral anticoagulation. In a retrospective study of stroke patients, we found a strong association of probability output >0.2 as noted by the AI-ECG AF model and detection of AF by ambulatory cardiac monitoring OR of 5.47 (95% CI 1.51-22.51; P = 0.01).[52] However, we were limited by the detection rate of AF amongst the ESUS patients due to a shorter average monitoring time. A clinical trial 'batch enrollment for AI-guided intervention to lower neurologic events in unrecognized AF' (BEAGLE) is currently underway to assess the use of this model.[34] Participants are selected using natural language processing tools if they are at high risk for incident AF as identified using the algorithm and are eligible to receive anticoagulation if AF is detected. Enrolled patients are then mailed a device to continuously monitor for up to 30 days. This is a completely off-site trial and will allow vetting and validation of the model in a real-world scenario.[34]

AF Burden

Currently, AF is classified as paroxysmal, persistent or permanent, but treatment generally relies on a binary classification (i.e., presence or absence of AF to determine the decision for anticoagulation).[69] A quantitative approach to define the burden of AF has been suggested. This factors in the number of episodes or the proportion of time a patient spends in AF during a particular monitoring period.[70] AF burden has been noted to have a predictive value for AF-related strokes and thus, could guide management.[70,71] However, even with wearable and implantable technologies allowing longer monitoring durations, burden of AF is difficult to determine. Also, the minimum AF duration required for anticoagulation is not well defined.[72]

Another scenario in which AF burden becomes important is SCAF. Higher AF burden along with the traditional risk factors (CHA_2DS_2-VASc score) are more likely to benefit from oral anticoagulants than low to medium burden of AF and without risk factors.[73] A general rule of thumb for anticoagulation, regardless of indication, is the need to maintain efficacy while minimizing the bleeding risk.[74]However, the exact cut-offs are not well defined and leads to considerable variation in management depending on the provider.[75] With advancements in wearable and intracardiac monitoring techniques as well as their wider availability,[76] detection of SCAF is also increasing and the knowledge gaps need to be addressed more clearly. Two highly anticipated trials Apixaban for the Reduction of Thrombo-Embolism in Patients With Device-Detected Sub-Clinical Atrial Fibrillation[77] and Non-vitamin K Antagonist Oral Anticoagulants in Patients With Atrial High Rate Episodes[78] will likely help inform management in these patients.

A preliminary study using DNN model has been able to assess AF burden from 24-hour Holter monitor data.[79] Although this proves feasibility of developing the prototype, translation of such models into clinical use would require validation and clearer guidelines as to the importance of AF burden and duration for guiding anticoagulation therapy.

Deep Learning Approach to AF Detection Using Photoplethysmography Data

Photoplethysmography (PPG) data is a pulse pressure waveform detected using a light source and a photodetector.[80] This technology has been successfully used previously to detect oxygen saturation, heart rate and respiratory rate. On PPG data, AF is manifested as variable pulse-to-pulse intervals and morphologies.[80] Several studies have tested DL algorithms on PPG data to detect AF, although many of these included very few patients.[81,82]

There is also a trend towards the use of consumer-grade wristbands and watches to detect AF. In Huawei heart study, of the 187,912 individuals who used smart devices for PPG analysis, only 0.23% received 'suspected AF' notification.[83] Out of those undergoing an effective follow up, 87% were confirmed as having AF. These results show an impressive positive predictive value of 91.6%.[83]

In Apple Heart Study, Apple smartwatches were used to detect AF using a proprietary PPG based algorithm.[84] Of the 419,297 patients enrolled over an 8-month period, only 2161 (0.52%) received a notification for irregular pulse. After a telemedicine visit, ECG patches were mailed to the patients to be worn for up to 7 days. Among 450 patients who returned patches with analyzable data, only 34% were diagnosed with AF, even though the algorithm was designed to maximize specificity. Furthermore, sensitivity and thus usefulness as a screening tool could not be assessed from this study.[84]

The Fitbit Heart Study- another large prospective study applying a similar methodology is currently underway.[85] These studies highlight the barriers to mass screening for AF in healthy populations particularly given the low benefit to cost ratio.

Some key limitations of using PPG data include high false positive rate due to detection of other arrythmias mimicking AF, low signal to noise ratio and motion artifacts.[80] One study showed that almost 40% of the signal was unusable.[86] Due to these limitations, the 12-SL ECG remains the gold standard to detect arrythmias including AF.

Scalability

In primary care settings and emergency units, there is often a lack of trained specialists to interpret complex

ECGs and make rapid diagnosis.[87] Even the combined accuracy of practitioners in these settings with current computer interpretations for AF diagnosis remains insufficient.[29] This need is even greater in less developed countries which contribute about 75% to the overall cardiovascular mortality.[7] As discussed previously, automated AI-ECG interpretation could enable expert level diagnoses and streamline clinical workflow in these settings.

A network once trained can be fine-tuned to widespread applications and smart devices. For instance, we have shown the application of an algorithm trained using 12 lead ECG to detect serum potassium levels and applied it to a single lead ECG recorded using a smartphone.[88]

To bring these tools to the point-of-care we have incorporated AI-ECG tools into the electronic health record as an 'AI-ECG Dashboard'.[8] With the click of a button, clinicians can have all the ECGs available for that patient analyzed with AI-ECG probability outputs for various diagnoses. A major advantage of incorporating AI-ECG algorithms into practice is their ability to keep learning indefinitely as more information is added. This essentially creates a self-improving healthcare system.[36]

The application of DNN models to long duration signals like 24-hour Holter monitor data to detect paroxysmal AF has shown good preliminary results.[79] This has broad applications for development of a new generation of real-time analytical tools to detect AF from the growing number of ambulatory monitoring devices.

Limitations

After the development of an AI-ECG model, rigorous external validation in diverse populations and clinical trials to demonstrate superior clinical outcomes to the standard of care is needed. To date, most models have not undergone rigorous evaluation and thus their results remain confined to their own datasets. Furthermore, due to the several knowledge gaps surrounding AF, it becomes challenging to prove immediate benefits in clinical outcomes.

A limitation to widespread applicability of these models is the need to be tailored to the target application. Although, cross domain utility of a pre-trained model to perform the same function has shown promising results. For instance, in a study a DNN model was trained to detect paroxysmal AF from 24-hour Holter ECG and then fine tuned to a completely different input data of smartwatch PPG signal with superior results (AUC of 0.97).[79] However, as noted by authors, this study was limited by a lack of gold standard labels to study true performance.[79]

Several other limitations surrounding AI such as data security, perpetuating bias, physician acceptance and regulatory considerations are also important to be addressed and have been discussed elsewhere in more detail.[8,89]

Conclusion

Algorithms using AI to interpret ECGs in various new ways have been developed. While still much work needs to be done, these technologies have shown enormous potential in a short span of time. With further advancements and continuous research, these novel ways of interpretation may well become part of everyday clinical workflow. However, rigorous testing is key with development of new models. Testing datasets need to be large to show accurate results, especially when several different outputs are being considered. Finally, the models need to undergo appropriate external validation and clinical trials to note generalizable benefits in diverse populations when compared to the standard of care. This may also help alleviate some of the current hesitance in clinicians from acceptance of AI tools.

Figures/Tables

Table 1

Summary (non-exhaustive) list of various DL models, with performance discussed in context of AF for this article.

(AF: Atrial Fibrillation; AUC: Area under the receiver operating characteristic curve; DNN: Deep neural network; F1: F1 score; S12L-ECG: 12 lead standard ECG; SN: Sensitivity; SP: Specificity)

Figure 1

Overview of performance, validation, potential uses and ongoing work with AI-ECG AF model from Mayo Clinic.

(AF: Atrial Fibrillation; AI: Artificial Intelligence; AUC: Area under the receiver operating characteristic curve; ESUS: Embolic stroke of undetermined significance; RCT: Randomized controlled trial)

REFERENCES

1. Schlant RC, Adolph RJ, DiMarco JP, et al. Guidelines for electrocardiography. A report of the American College of Cardiology/American Heart Association Task Force on Assessment of Diagnostic and Therapeutic Cardiovascular Procedures (Committee on Electrocardiography). *Circulation.* 1992;85(3):1221-1228.

2. Hornick J, Costantini O. The Electrocardiogram: Still a Useful Tool in the Primary Care Office. *Med Clin North Am.*2019;103(5):775-784.

3. Maron BJ, Friedman RA, Kligfield P, et al. Assessment of the 12-lead ECG as a screening test for detection of cardiovascular disease in healthy general populations of young people (12-25 Years of Age): a scientific statement from the American Heart Association and the American College of Cardiology. *Circulation.*2014;130(15):1303-1334.

4. Garvey JL, Zegre-Hemsey J, Gregg R, Studnek JR. Electrocardiographic diagnosis of ST segment elevation myocardial infarction: An evaluation of three automated interpretation algorithms. *J Electrocardiol.*2016;49(5):728-732.

5. Kudenchuk PJ, Maynard C, Cobb LA, et al. Utility of the prehospital electrocardiogram in diagnosing acute coronary syndromes: the Myocardial Infarction Triage and Intervention (MITI) Project. *J Am Coll Cardiol.* 1998;32(1):17-27.

6. Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med.* 2019;25(1):65-69.

7. Ribeiro AH, Ribeiro MH, Paixão GMM, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat Commun.*2020;11(1):1760.

8. Siontis KC, Noseworthy PA, Attia ZI, Friedman PA. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nat Rev Cardiol.* 2021;18(7):465-478.

9. Anthony HKaSKMaAJDaW-YKaZIAaRECaPAFa. An artificial intelligence–enabled ECG algorithm for comprehensive ECG interpretation: Can it pass the 'Turing test'? *Cardiovascular Digital Health Journal.* 2021;2(3):164-170.

10. Attia ZI, Friedman PA, Noseworthy PA, et al. Age and Sex Estimation Using Artificial Intelligence From Standard 12-Lead ECGs. *Circ Arrhythm Electrophysiol.* 2019;12(9):e007284.

11. 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.* 2019;394(10201):861-867.

12. Ko WY, Siontis KC, Attia ZI, et al. Detection of Hypertrophic Cardiomyopathy Using a Convolutional Neural Network-Enabled Electrocardiogram. *J Am Coll Cardiol.* 2020;75(7):722-733.

13. Guglin ME, Thatai D. Common errors in computer electrocardiogram interpretation. *Int J Cardiol.* 2006;106(2):232-237.

14. O'Shea K, Nash R. An introduction to convolutional neural networks. *arXiv preprint arXiv:151108458.* 2015.

15. Patel NJ, Deshmukh A, Pant S, et al. Contemporary trends of hospitalization for atrial fibrillation in the United States, 2000 through 2010: implications for healthcare planning. *Circulation.*2014;129(23):2371-2379.

16. Chugh SS, Havmoeller R, Narayanan K, et al. Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. *Circulation.* 2014;129(8):837-847.

17. Samuel M, Brophy JM. Challenges in Assessing the Incidence of Atrial Fibrillation Hospitalizations. *Can J Cardiol.*2019;35(10):1291-1293.

18. Coyne KS, Paramore C, Grandy S, Mercader M, Reynolds M, Zimetbaum P. Assessing the direct costs of treating nonvalvular atrial fibrillation in the United States. *Value Health.* 2006;9(5):348-356.

19. Lloyd-Jones DM, Wang TJ, Leip EP, et al. Lifetime risk for development of atrial fibrillation: the Framingham Heart Study. *Circulation.* 2004;110(9):1042-1046.

20. Go AS, Hylek EM, Phillips KA, et al. Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study. *JAMA.* 2001;285(18):2370-2375.

21. Lip GYH, Brechin CM, Lane DA. The global burden of atrial fibrillation and stroke: a systematic review of the epidemiology of atrial fibrillation in regions outside North America and Europe. *Chest.* 2012;142(6):1489-1498.

22. Morin DP, Bernard ML, Madias C, Rogers PA, Thihalolipavan S, Estes NA. The State of the Art: Atrial Fibrillation Epidemiology, Prevention, and Treatment. *Mayo Clin Proc.* 2016;91(12):1778-1810.

23. Yuan S, Larsson SC. No association between coffee consumption and risk of atrial fibrillation: A Mendelian randomization study. Nutr Metab Cardiovasc Dis. 2019 Nov;29(11):1185-1188. doi: 10.1016/j.numecd.2019.07.015. Epub 2019 Jul 27. PMID: 31558414. In.

24. Stewart S, Hart CL, Hole DJ, McMurray JJ. A population-based study of the long-term risks associated with atrial fibrillation: 20-year follow-up of the Renfrew/Paisley study. *Am J Med.*2002;113(5):359-364.

25. Hongo RH, Goldschlager N. Overreliance on computerized algorithms to interpret electrocardiograms. *Am J Med.* 2004;117(9):706-708.

26. Shah AP, Rubin SA. Errors in the computerized electrocardiogram interpretation of cardiac rhythm. *J Electrocardiol.*2007;40(5):385-390.

27. Bogun F, Anh D, Kalahasty G, et al. Misdiagnosis of atrial fibrillation and its clinical consequences. *Am J Med.*2004;117(9):636-642.

28. Lindow T, Kron J, Thulesius H, Ljungström E, Pahlm O. Erroneous computer-based interpretations of atrial fibrillation and atrial flutter in a Swedish primary health care setting. *Scand J Prim Health Care.* 2019;37(4):426-433.

29. Mant J, Fitzmaurice DA, Hobbs FD, et al. Accuracy of diagnosing atrial fibrillation on electrocardiogram by primary care practitioners and interpretative diagnostic software: analysis of data from screening for atrial fibrillation in the elderly (SAFE) trial. *BMJ.*2007;335(7616):380.

30. Schläpfer J, Wellens HJ. Computer-Interpreted Electrocardiograms: Benefits and Limitations. *J Am Coll Cardiol.*2017;70(9):1183-1192.

31. Madias JE. Computerized interpretation of electrocardiograms: Taking stock and implementing new knowledge. *J Electrocardiol.*2018;51(3):413-415.

32. Novotny T, Bond R, Andrsova I, et al. The role of computerized diagnostic proposals in the interpretation of the 12-lead electrocardiogram by cardiology and non-cardiology fellows. *Int J Med Inform.* 2017;101:85-92.

33. Anh D, Krishnan S, Bogun F. Accuracy of electrocardiogram interpretation by cardiologists in the setting of incorrect computer analysis. *J Electrocardiol.* 2006;39(3):343-345.

34. Yao X, Attia ZI, Behnken EM, et al. Batch enrollment for an artificial intelligence-guided intervention to lower neurologic events in patients with undiagnosed atrial fibrillation: rationale and design of a digital clinical trial. *Am Heart J.* 2021;239:73-79.

35. Clifford GD, Liu C, Moody B, et al. AF Classification from a Short Single Lead ECG Recording: the PhysioNet/Computing in Cardiology Challenge 2017. *Comput Cardiol (2010).* 2017;44.

36. Smith SW, Walsh B, Grauer K, et al. A deep neural network learning algorithm outperforms a conventional algorithm for emergency department electrocardiogram interpretation. *J Electrocardiol.*2019;52:88-95.

37. Anthony HKaW-YKaZIAaMSCaPAFaPAN. A comprehensive artificial intelligence–enabled electrocardiogram interpretation program.*Cardiovascular Digital Health Journal.* 2020;1(2):62-70.

38. 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.* 2021;398(10310):1498-1506.

39. 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(25):2176-2184.

40. Svendsen JH, Diederichsen SZ, Højberg S, et al. Implantable loop recorder detection of atrial fibrillation to prevent stroke (The LOOP Study): a randomised controlled trial. *Lancet.*2021;398(10310):1507-1516.

41. Fitzmaurice DA, McCahon D, Baker J, et al. Is screening for AF worthwhile? Stroke risk in a screened population from the SAFE study.*Fam Pract.* 2014;31(3):298-302.

42. Swancutt D, Hobbs R, Fitzmaurice D, et al. A randomised controlled trial and cost effectiveness study of systematic screening (targeted and total population screening) versus routine practice for the detection of atrial fibrillation in the over 65s: (SAFE) [ISRCTN19633732].*BMC Cardiovasc Disord.* 2004;4:12.

43. Menke J, Lüthje L, Kastrup A, Larsen J. Thromboembolism in atrial fibrillation. *Am J Cardiol.* 2010;105(4):502-510.

44. Danias PG, Caulfield TA, Weigner MJ, Silverman DI, Manning WJ. Likelihood of spontaneous conversion of atrial fibrillation to sinus rhythm. *J Am Coll Cardiol.* 1998;31(3):588-592.

45. Risk factors for stroke and efficacy of antithrombotic therapy in atrial fibrillation. Analysis of pooled data from five randomized controlled trials. *Arch Intern Med.* 1994;154(13):1449-1457.

46. van Walraven C, Hart RG, Connolly S, et al. Effect of age on stroke prevention therapy in patients with atrial fibrillation: the atrial fibrillation investigators. *Stroke.* 2009;40(4):1410-1416.

47. Tereshchenko LG, Henrikson CA, Cigarroa J, Steinberg JS. Comparative Effectiveness of Interventions for Stroke Prevention in Atrial Fibrillation: A Network Meta-Analysis. *J Am Heart Assoc.*2016;5(5).

48. Romero JR, Morris J, Pikula A. Stroke prevention: modifying risk factors. *Ther Adv Cardiovasc Dis.* 2008;2(4):287-303.

49. Jonas DE, Kahwati LC, Yun JDY, Middleton JC, Coker-Schwimmer M, Asher GN. Screening for Atrial Fibrillation With Electrocardiography: Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA.* 2018;320(5):485-498.

50. Martínez-Sellés M, Massó-van Roessel A, Álvarez-García J, et al. Interatrial block and atrial arrhythmias in centenarians: Prevalence, associations, and clinical implications. *Heart Rhythm.*2016;13(3):645-651.

51. Arboix A, Martí L, Dorison S, Sánchez MJ. Bayés syndrome and acute cardioembolic ischemic stroke. *World J Clin Cases.*2017;5(3):93-101.

52. Rabinstein AA, Yost MD, Faust L, et al. Artificial Intelligence-Enabled ECG to Identify Silent Atrial Fibrillation in Embolic Stroke of Unknown Source. *J Stroke Cerebrovasc Dis.*2021;30(9):105998.

53. Alonso A, Roetker NS, Soliman EZ, Chen LY, Greenland P, Heckbert SR. Prediction of Atrial Fibrillation in a Racially Diverse Cohort: The Multi-Ethnic Study of Atherosclerosis (MESA). *J Am Heart Assoc.*2016;5(2).

54. Schnabel RB, Sullivan LM, Levy D, et al. Development of a risk score for atrial fibrillation (Framingham Heart Study): a community-based cohort study. *Lancet.* 2009;373(9665):739-745.

55. Alonso A, Krijthe BP, Aspelund T, et al. Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the CHARGE-AF consortium. *J Am Heart Assoc.*2013;2(2):e000102.

56. Christopoulos G, Graff-Radford J, Lopez CL, et al. Artificial Intelligence-Electrocardiography to Predict Incident Atrial Fibrillation: A Population-Based Study. *Circ Arrhythm Electrophysiol.* 2020;13(12):e009355.

57. Khurshid S, Friedman S, Reeder C, et al. Electrocardiogram-based Deep Learning and Clinical Risk Factors to Predict Atrial Fibrillation. *Circulation.* 2021.

58. Kashou AH, Rabinstein AA, Attia IZ, et al. Recurrent cryptogenic stroke: A potential role for an artificial intelligence-enabled electrocardiogram? *HeartRhythm Case Rep.* 2020;6(4):202-205.

59. Saver JL. CLINICAL PRACTICE. Cryptogenic Stroke. *N Engl J Med.* 2016;374(21):2065-2074.

60. Li L, Yiin GS, Geraghty OC, et al. Incidence, outcome, risk factors, and long-term prognosis of cryptogenic transient ischaemic attack and ischaemic stroke: a population-based study. *Lancet Neurol.*2015;14(9):903-913.

61. Sacco RL, Ellenberg JH, Mohr JP, et al. Infarcts of undetermined cause: the NINCDS Stroke Data Bank. *Ann Neurol.*1989;25(4):382-390.

62. Wolf ME, Grittner U, Böttcher T, et al. Phenotypic ASCO Characterisation of Young Patients with Ischemic Stroke in the Prospective Multicentre Observational sifap1 Study. *Cerebrovasc Dis.* 2015;40(3-4):129-135.

63. Marini C, De Santis F, Sacco S, et al. Contribution of atrial fibrillation to incidence and outcome of ischemic stroke: results from a population-based study. *Stroke.* 2005;36(6):1115-1119.

64. Paciaroni M, Agnelli G, Caso V, et al. Atrial fibrillation in patients with first-ever stroke: frequency, antithrombotic treatment before the event and effect on clinical outcome. *J Thromb Haemost.* 2005;3(6):1218-1223.

65. Sanna T, Diener HC, Passman RS, et al. Cryptogenic stroke and underlying atrial fibrillation. *N Engl J Med.*2014;370(26):2478-2486.

66. Kernan WN, Ovbiagele B, Black HR, et al. Guidelines for the prevention of stroke in patients with stroke and transient ischemic attack: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke.*2014;45(7):2160-2236.

67. Diener HC, Sacco RL, Easton JD, et al. Dabigatran for Prevention of Stroke after Embolic Stroke of Undetermined Source. *N Engl J Med.*2019;380(20):1906-1917.

68. Hart RG, Connolly SJ, Mundl H. Rivaroxaban for Stroke Prevention after Embolic Stroke of Undetermined Source. *N Engl J Med.*2018;379(10):987.

69. January CT, Wann LS, Alpert JS, et al. 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task

Force on practice guidelines and the Heart Rhythm Society. *Circulation.*2014;130(23):e199-267.

70. Chen LY, Chung MK, Allen LA, et al. Atrial Fibrillation Burden: Moving Beyond Atrial Fibrillation as a Binary Entity: A Scientific Statement From the American Heart Association. *Circulation.*2018;137(20):e623-e644.

71. Swiryn S, Orlov MV, Benditt DG, et al. Clinical Implications of Brief Device-Detected Atrial Tachyarrhythmias in a Cardiac Rhythm Management Device Population: Results from the Registry of Atrial Tachycardia and Atrial Fibrillation Episodes. *Circulation.*2016;134(16):1130-1140.

72. Rankin AJ, Tran RT, Abdul-Rahim AH, Rankin AC, Lees KR. Clinically important atrial arrhythmia and stroke risk: a UK-wide online survey among stroke physicians and cardiologists. *QJM: An International Journal of Medicine.* 2014;107(11):895-902.

73. Freedman B, Boriani G, Glotzer TV, Healey JS, Kirchhof P, Potpara TS. Management of atrial high-rate episodes detected by cardiac implanted electronic devices. *Nat Rev Cardiol.*2017;14(12):701-714.

74. Kumbhani DJ, Cannon CP, Beavers CJ, et al. 2020 ACC Expert Consensus Decision Pathway for Anticoagulant and Antiplatelet Therapy in Patients With Atrial Fibrillation or Venous Thromboembolism Undergoing Percutaneous Coronary Intervention or With Atherosclerotic Cardiovascular Disease: A Report of the American College of Cardiology Solution Set Oversight Committee. *J Am Coll Cardiol.*2021;77(5):629-658.

75. Perino AC, Fan J, Askari M, et al. Practice Variation in Anticoagulation Prescription and Outcomes After Device-Detected Atrial Fibrillation. *Circulation.* 2019;139(22):2502-2512.

76. Noseworthy PA, Kaufman ES, Chen LY, et al. Subclinical and Device-Detected Atrial Fibrillation: Pondering the Knowledge Gap: A Scientific Statement From the American Heart Association.*Circulation.* 2019;140(25):e944-e963.

77. Lopes RD, Alings M, Connolly SJ, et al. Rationale and design of the Apixaban for the Reduction of Thrombo-Embolism in Patients With Device-Detected Sub-Clinical Atrial Fibrillation (ARTESiA) trial.*Am Heart J.* 2017;189:137-145.

78. Kirchhof P, Blank BF, Calvert M, et al. Probing oral anticoagulation in patients with atrial high rate episodes: Rationale and design of the Non-vitamin K antagonist Oral anticoagulants in patients with Atrial High rate episodes (NOAH-AFNET 6) trial. *Am Heart J.*2017;190:12-18.

79. Shashikumar SPaSAJaCGDaNS. Detection of Paroxysmal Atrial Fibrillation Using Attention-Based Bidirectional Recurrent Neural Networks. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; 2018.

80. Pereira T, Tran N, Gadhoumi K, et al. Photoplethysmography based atrial fibrillation detection: a review. *NPJ Digit Med.* 2020;3:3.

81. Tison GH, Sanchez JM, Ballinger B, et al. Passive Detection of Atrial Fibrillation Using a Commercially Available Smartwatch.*JAMA Cardiol.* 2018;3(5):409-416.

82. Poh MZ, Poh YC, Chan PH, et al. Diagnostic assessment of a deep learning system for detecting atrial fibrillation in pulse waveforms.*Heart.* 2018;104(23):1921-1928.

83. Guo Y, Wang H, Zhang H, et al. Mobile Photoplethysmographic Technology to Detect Atrial Fibrillation. *J Am Coll Cardiol.*2019;74(19):2365-2375.

84. Perez MV, Mahaffey KW, Hedlin H, et al. Large-Scale Assessment of a Smartwatch to Identify Atrial Fibrillation. *N Engl J Med.*2019;381(20):1909-1917.

85. Lubitz SA, Faranesh AZ, Atlas SJ, et al. Rationale and design of a large population study to validate software for the assessment of atrial fibrillation from data acquired by a consumer tracker or smartwatch:

The Fitbit heart study. *Am Heart J.* 2021;238:16-26.

86. Bonomi AG, Schipper F, Eerikäinen LM, et al. Atrial Fibrillation Detection Using a Novel Cardiac Ambulatory Monitor Based on Photo-Plethysmography at the Wrist. *J Am Heart Assoc.*2018;7(15):e009351.

87. Veronese G, Germini F, Ingrassia S, et al. Emergency physician accuracy in interpreting electrocardiograms with potential ST-segment elevation myocardial infarction: Is it enough? *Acute Card Care.*2016;18(1):7-10.

88. Yasin OZ, Attia Z, Dillon JJ, et al. Noninvasive blood potassium measurement using signal-processed, single-lead ecg acquired from a handheld smartphone. *J Electrocardiol.* 2017;50(5):620-625.

89. Kashou AH, May AM, Noseworthy PA. Artificial Intelligence-Enabled ECG: a Modern Lens on an Old Technology. *Curr Cardiol Rep.*2020;22(8):57.

Hosted file

`IMAGE 1.pdf` available at https://authorea.com/users/453112/articles/551038-artificial-intelligence-and-atrial-fibrillation

Hosted file

`Table 1.docx` available at https://authorea.com/users/453112/articles/551038-artificial-intelligence-and-atrial-fibrillation