Check for updates

Artificial intelligence-enhanced electrocardiography in cardiovascular disease management

Konstantinos C. Siontis, Peter A. Noseworthy, Zachi I. Attia and Paul A. Friedman ✉

Abstract | The application of artificial intelligence (AI) to the electrocardiogram (ECG), a ubiquitous and standardized test, is an example of the ongoing transformative effect of AI on cardiovascular medicine. Although the ECG has long offered valuable insights into cardiac and non-cardiac health and disease, its interpretation requires considerable human expertise. Advanced AI methods, such as deep-learning convolutional neural networks, have enabled rapid, human-like interpretation of the ECG, while signals and patterns largely unrecognizable to human interpreters can be detected by multilayer AI networks with precision, making the ECG a powerful, non-invasive biomarker. Large sets of digital ECGs linked to rich clinical data have been used to develop AI models for the detection of left ventricular dysfunction, silent (previously undocumented and asymptomatic) atrial fibrillation and hypertrophic cardiomyopathy, as well as the determination of a person's age, sex and race, among other phenotypes. The clinical and population-level implications of AI-based ECG phenotyping continue to emerge, particularly with the rapid rise in the availability of mobile and wearable ECG technologies. In this Review, we summarize the current and future state of the AI-enhanced ECG in the detection of cardiovascular disease in at-risk populations, discuss its implications for clinical decision-making in patients with cardiovascular disease and critically appraise potential limitations and unknowns.

Department of Cardiovascular Medicine, Mayo Clinic, Rochester, MN, USA.

✉e-mail: *friedman.paul@ mayo.edu*

https://doi.org/10.1038/ s41569-020-00503-2

The electrocardiogram (ECG) is a ubiquitous tool in clinical medicine that has been used by cardiologists and non-cardiologists for decades. The ECG is a low-cost, rapid and simple test that is available even in the most resource-scarce settings. The test provides a window into the physiological and structural condition of the heart but can also give valuable diagnostic clues for systemic conditions (such as electrolyte derangements and drug toxic effects). Although the acquisition of the ECG recording is well standardized and reproducible, the reproducibility of human interpretation of the ECG varies greatly according to levels of experience and expertise. In this setting, computer-generated interpretations have been used for several years. However, these interpretations are based on predefined rules and manual pattern or feature recognition algorithms that do not always capture the complexity and nuances of an ECG. However, artificial intelligence (AI) in the form of deep-learning convolutional neural networks (CNNs), which have been used mostly in computer vision, image processing and speech recognition, has now been adapted to analyse the routine 12-lead ECG. This development has resulted in fully automated AI models mimicking human-like interpretation of

the ECG, with potentially greater diagnostic fidelity and workflow efficiency than traditional rule-based computer interpretations.

Indeed, the ECG is an ideal substrate for deep-learning AI applications. The ECG is widely available and yields reproducible raw data that are easy to store and transfer in a digital format. In addition to fully automated interpretation of the ECG, rigorous research programmes using large databanks of ECG and clinical datasets, coupled with powerful computational capabilities, have been able to demonstrate the utility of the AI-enhanced ECG (referred to as the AI–ECG in the remainder of this Review) as a tool for the detection of ECG signatures and patterns that are unrecognizable by the human eye. These patterns can signify cardiac disease, such as left ventricular (LV) systolic dysfunction, silent atrial fibrillation (AF) and hypertrophic cardiomyopathy (HCM), but might also reflect systemic physiology, such as a person's age and sex or their serum potassium levels, as reviewed in detail herein. Among several potential clinical applications, a single 12-lead ECG might therefore allow the rapid phenotyping of an individual's cardiovascular health and help to guide targeted diagnostic testing in an efficient and potentially

Key points

- The feasibility and potential value of the application of advanced artificial intelligence methods, particularly deep-learning convolutional neural networks (CNNs), to the electrocardiogram (ECG) have been demonstrated.
- CNNs developed with the use of large numbers of digital ECGs linked to rich clinical datasets might be able to perform accurate and nuanced, human-like interpretation of ECGs.
- CNNs have also been developed to detect asymptomatic left ventricular dysfunction, silent atrial fibrillation, hypertrophic cardiomyopathy and an individual's age, sex and race on the basis of the ECG alone.
- CNNs to detect other cardiac conditions, such as aortic valve stenosis and amyloid heart disease, are in active development.
- These approaches might be applicable to the standard 12-lead ECG or to data obtained from single-lead or multilead mobile or wearable ECG technologies.
- Evidence on patient outcomes, as well as the challenges and potential limitations from the real-world implementation of the artificial intelligence-enhanced ECG, continues to emerge.

cost-effective manner. Some of the main ECG datasets that have been used in the development of AI–ECG applications are listed in TABLE 1.

This transformative progress has not occurred without potential limitations and challenges that require attention. Challenges with AI applications are not necessarily unique to the ECG and include the need for data quality control, external validity, data security and the demonstration of superior patient outcomes with the implementation of AI-enabled tools, such as the AI–ECG. In this Review, we focus on the promise, clinical capabilities, research opportunities, gaps and risk of the application of AI to the ECG for the diagnosis and management of cardiovascular disease. Finally, we discuss the numerous challenges to the development, validation and implementation of the AI–ECG in medicine.

Deep-learning methods applied to the ECG

Deep learning is a subfield of machine learning that uses neural networks with many layers (hence the term 'deep') to learn a function between a set of inputs and a set of outputs. The training is done by presenting to the network a set of input data with their corresponding output labels, and the model learns certain rules by applying and adjusting the network weights to minimize an error function until the model outputs are as close as possible to the actual data values. The strength of deep neural networks lies in using their ability to identify novel relationships in the data independent of features selected by a human. In the past, conventional statistical models, such as linear and logistic regression, and even neural networks with few layers (sometimes called 'shallow models') were developed on the basis of inputs using human-selected data features. For example, in ECG analysis, the inputs were morphological and temporal features, such as the QT, QRS and RR intervals or QRS and/or T-wave morphology, and the outputs would be the ECG rhythm, the serum potassium level or the LV ejection fraction (LVEF). As the inputs are features that were selected by humans, the model is limited to those features alone. In addition, any random or systematic error in calculating features will propagate to the output and will limit the accuracy of the model.

Conversely, in deep-learning networks, the representation of the input is learned by the network itself. The most common model for this type of representation learning is a subtype of neural networks called 'CNNs'. CNNs were originally designed to solve computer vision tasks such as image recognition because they use a set of convolutional filters to select features used to represent the input into the model[1]. These filters start with random weights and, as the network self-trains, the architecture optimizes both the features used as a representation of the data and the rules applied to this representation and the outputs. Therefore, a neural network can be thought of as having two sequential components: the feature extraction layers (typically convolutional filters) and the mathematical model (such as pooling layers), which take feature output as their input to perform analysis and create the ultimate output. The number and the shape of the convolutional filters are selected by the model architect and will affect the learned representation. In the ECG, one dimension is the spatial axis (each row in the input represents a time series from one of the leads) and the second dimension is the temporal axis (each column represents the voltage sampled at a specific time point across all leads) (FIG. 1a). Therefore, a convolution can be horizontal (combining information from more than one time point but only within one lead), vertical (combining information from many leads but at a single time point) or both horizontal and vertical (combining information on both axes). Using deep learning, the model can learn a representation of the input data that includes features that are relevant to the task that we are trying to accomplish without any human bias and without the need for human selection and engineering of features, which can be time-consuming, inaccurate and dependent on expertise and current physiological theories.

The agnostic approach in a neural network is an optimal representation, but this approach is also non-linear, and the learned associations between input and output data are unexplainable at present, making the model a black box — humans cannot understand how the network makes its decisions — which is one of the concerns raised regarding the clinical application of deep-learning CNNs[2]. Therefore, less agnostic machine-learning models, such as the more traditional logistic regression, reinforcement learning and random forest models, still hold promise and can help to inform research and clinical practice. For example, reinforcement learning is a field of AI providing the framework for training of a clinical decision model in which certain decisions (model input) under specific conditions are linked to long-term outcomes. The optimal reinforcement learning model is built through a continuous process of updating of inputs and outcomes (or rewards)[3]. In addition, CNNs are supervised models, which require data to have a label assigned during training. Conversely, unsupervised machine-learning methods, such as clustering models, do not require labelling of input data for their development, thereby minimizing the risk of random or systematic error introduced by a human.

In this Review, we also discuss natural language processing applications. The field of natural language processing integrates AI, computer science and linguistics

Table 1 | **Major contemporary ECG databases used to develop deep-learning AI–ECG applications**

Dataset	Location	Enrolment period	Number of patients	ECG type	Conditions, outcomes and scope of application	Refs
Telehealth Network of Minas Gerais	Brazil	2010–2018	1,676,384	12-lead	Automated ECG interpretation	9
Mayo Clinic	USA	1994–2017	449,380	12-lead	Automated ECG interpretation; left ventricular dysfunction, silent atrial fibrillation, hypertrophic cardiomyopathy, serum potassium level, and age, sex and race/ethnicity	22–24,32, 45,48,59,66
Geisinger	USA	1984–2019	253,397	12-lead	Overall survival	67
Huazhong University, Wuhan	China	2012–2019	71,520	12-lead	Automated ECG interpretation	11
iRhythm Technologies/ Stanford University	USA	2013–2017	53,549	Single-lead, ambulatory ECG monitoring	Classification of 12 rhythm types	8
University of California, San Francisco	USA	2010–2017	36,186 (ECGs)	12-lead	Left ventricular mass, left atrial volume, early diastolic mitral annulus velocity, pulmonary arterial hypertension, hypertrophic cardiomyopathy, amyloidosis and mitral valve prolapse	46
Health eHeart Study	Multinational	2016–2017	9,750	Single-lead, smartwatch-based	Passive detection of atrial fibrillation	14
China Physiological Signal Challenge 2018	China	2018	6,877	12-lead	Classification of 9 rhythm types	68
Cleveland Clinic	USA	2003–2012 and 2017–2018	946	12-lead	Response to cardiac resynchronization therapy	69

AI–ECG, artificial intelligence-enhanced electrocardiogram; ECG, electrocardiogram.

methodology to structure unstructured data (such as free text in clinical notes in electronic health records), thereby making these data analysable. These methods include rule-based recognition of word patterns, text vectorization and topic modelling.

Fully automated interpretation of ECGs

One of the top priorities for the application of AI to the interpretation of ECGs is the creation of comprehensive, human-like interpretation capability. Since the advent of the digital ECG more than 60 years ago[4], ongoing effort has been towards rapid, high-quality and comprehensive computer-generated interpretation of the ECG. The problem seems tractable; after all, ECG interpretation is a fairly circumscribed application of pattern recognition to a finite dataset. Early programs for the interpretation of digital ECGs could easily recognize fiducial points, make discrete measurements and define common quantifiable abnormalities[5–7]. Modern technologies have moved beyond these rule-based approaches to recognize patterns in massive quantities of labelled ECG data[8,9].

Several groups have worked to create AI-driven algorithms, and some of these algorithms are already in limited clinical use[10]. Some studies have developed CNNs from large datasets of single-lead ECGs and then applied them to the 12-lead ECG. For instance, using 2 million labelled single-lead ECG traces collected in the Clinical Outcomes in Digital Electrocardiology study, one group used a CNN to identify six types of abnormalities on the 12-lead ECG[9]. This study demonstrated the feasibility of this approach, but widespread implementation or external validation in other 12-lead ECG datasets is

forthcoming. Another group conducted a similar study of the application of CNNs to single-lead ECGs and demonstrated that the CNN could outperform practising cardiologists for some diagnoses[8]. However, whether this approach will translate to clinically useful software for 12-lead ECG interpretation remains to be seen.

In an evaluation published in 2020, a CNN was developed for the multilabel diagnosis of 21 distinct heart rhythms based on the 12-lead ECG using a training and validation dataset of >80,000 ECGs from >70,000 patients[11]. The reference standard consisted of consensus labels by a committee of cardiologists. In a test dataset of 828 ECGs, the optimal network exactly matched the gold standard labels in 80% of the ECGs, significantly exceeding the performance of a single cardiologist interpreter. The model had a mean area under the curve (AUC) receiver operating characteristic score of 98%, sensitivity of 87% and specificity of 99%. Our group has worked with our internal dataset of >8 million ECGs performed for clinical indications (which have all been labelled by expert ECG readers and are linked to the respective electronic health record) to generate a comprehensive ECG-interpretation infrastructure. We demonstrated that a CNN can identify 66 discrete codes or diagnosis labels, with favourable diagnostic performance[12]. Lately, we have developed a novel method that uses a CNN to extract ECG features and a transformer network to translate ECG features into ECG codes and text strings[12]. This process creates a model output that more closely resembles that of a human ECG reader — presenting information in a similar order, with similar language — and also makes sense of associated

codes, thereby avoiding the presentation of contradictory or mutually exclusive interpretations that would not be presented by a human reader[12].

This technology will be particularly important as we increasingly rely on ECG data obtained through novel, consumer-facing applications, which are massively scalable. For instance, AI–ECG algorithms have been applied to single-lead ECG traces obtained through mobile, smartwatch-enabled recordings for the detection of AF[13]. This democratization of ECG technology will exponentially increase the volume of signals that

demand interpretation, which might quickly outstrip the capacity of human ECG readers. We anticipate that these models will be essential in facilitating tele-health technologies (that is, automatic, patient-facing or consumer-facing technologies) and could allow the creation of core laboratory facilities capable of ingesting and processing massive quantities of data. However, we caution that the signal quality obtained with these devices can be inconsistent, and AI–ECG might be less able than human expert over-readers to classify the heart rhythm using poor-quality tracings, as noted in
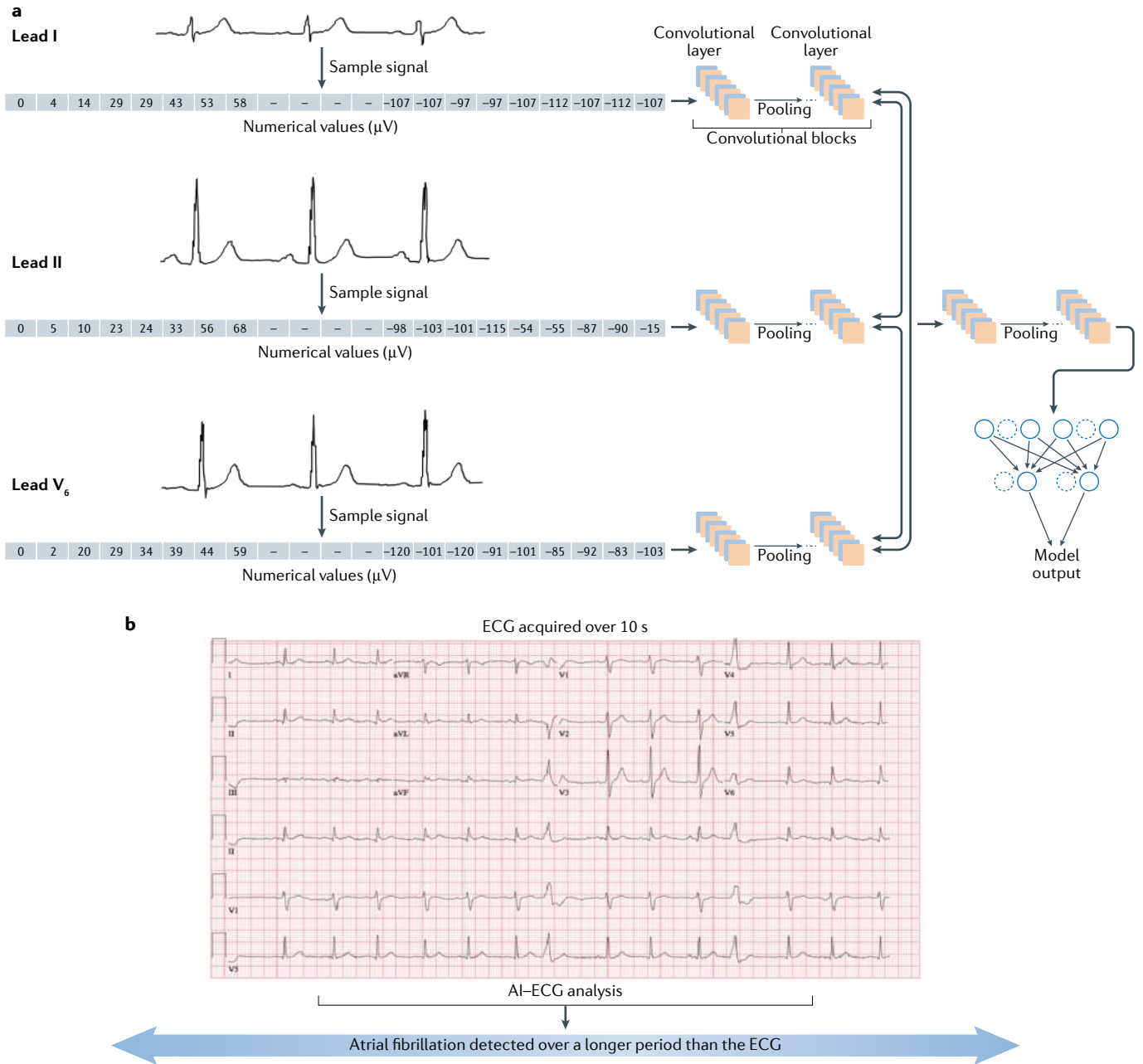

Fig. 1 | **Development of a convolutional neural network using the 12-lead ECG and application to detect silent atrial fibrillation. a** | The analogue electrocardiogram (ECG) signal is converted to a digital recording, resulting in a list of numerical values corresponding to the amplitude of the signal. (The numerical values depicted are arbitrary and shown for illustrative purposes only.) These numerical values are then convolved with the network weights within each lead and across leads, feeding sequential layers of convolutions until the final model output is reached. **b** | With the use of a trained, deep-learning artificial intelligence-enhanced ECG (AI–EGG) model, a one-off, standard, 12-lead, sinus-rhythm ECG can become a surrogate for prolonged rhythm monitoring for the detection of silent atrial fibrillation. Part **a** adapted with permission from REF.[70], Elsevier.

the aforementioned smartwatch study[13]. Similarly, in another study, a deep neural network developed using ECG recordings from smartwatches demonstrated good performance for passive detection of AF compared with the reference standard of AF diagnosed from 12-lead ECGs, but its performance was much less robust in reference to a self-reported history of persistent AF[14].

Nevertheless, although great progress has been made towards a comprehensive, human-like ECG-interpretation package, the realization remains on the horizon. Even in its most modern incarnation, the package lacks the accuracy needed for implementation without human oversight[15]. Additionally, computer-derived ECG interpretation has the potential to influence human over-readers and, if inaccurate, can serve as a source of bias or systematic error. This concern is particularly relevant if the algorithms are derived in populations that are distinct from those in which the algorithms are applied. This limitation underscores the need for a diverse derivation sample (ideally including a diverse patient population and varied means of data collection that reflect real-world practices), rigorous external validation studies and phased implementation with ongoing assessment of model performance and effectiveness. Quality control systems based on regular over-reads by expert interpreters of ECGs who are blinded to the output of the CNN will allow continuous calibration of the model.

The ECG as a deep phenotyping tool

Interpretation of an ECG by a trained cardiologist relies on established knowledge of what is normal or abnormal on the basis of more than a century of experience with assessing the ECG in patient care and based on our understanding of the electropathophysiology of various cardiac conditions. Despite the enormous potential to gain insights into cardiac health and disease from an expert interpretation of the ECG, the information gain is limited by the interpreter's finite ability to detect isolated characteristics or patterns fitting established rules. However, hidden in plain sight might be subtle signals and patterns that do not fit traditional knowledge and that are unrecognizable by the human eye. Harnessing the power of deep-learning AI techniques together with the availability of large ECG and clinical datasets, developing tools for systematic extraction of features of ECGs and their association with specific cardiac diagnoses has become feasible. Of course, some conditions are not reflected in the ECG, which even an AI–ECG cannot resolve — even if these technologies can see beyond an expert reader's capacity, they cannot see what is not there. In this section, we review the latest advances in the application of deep-learning AI techniques to the 12-lead ECG for the detection of asymptomatic cardiovascular disease that might not be readily apparent, even to expert eyes.

Detection of LV systolic dysfunction. The systolic function of the left ventricle, traditionally quantified as the LVEF by echocardiography, is a key measure of cardiac function. A reduced LVEF defines a large subgroup of patients with heart failure, but a decline in LVEF can be asymptomatic for a long time before any symptoms

trigger evaluation. Indeed, up to 6% of people in the community might have asymptomatic LV dysfunction (LVEF <50%)[16]. A low LVEF has both prognostic and management implications[17]. Detection of a low LVEF should trigger a thorough evaluation for any reversible causes that should be addressed in a timely fashion to minimize the extent of permanent myocardial damage. The early initiation of optimal medical therapy can result in improvements in systolic LV function and quality of life, but can also reduce heart failure-related morbidity and mortality[18]. However, in the absence of symptoms, identifying these patients remains a challenge and, therefore, asymptomatic LV dysfunction might be underrecognized. Several approaches to screening patients for asymptomatic LV systolic dysfunction have been investigated, including risk factors, the standard 12-lead ECG, echocardiography and measuring the levels of circulating biomarkers[19–21]. However, none of these approaches has sufficient diagnostic accuracy or cost-effectiveness to justify routine implementation clinically.

The potential of the AI–ECG as a marker of asymptomatic LV dysfunction has been demonstrated. With the use of linked ECG and echocardiographic data from 44,959 patients at the Mayo Clinic (Rochester, MN, USA), a CNN was trained to identify patients with LV dysfunction, defined as LVEF of ≤35% by echocardiography, on the basis of the ECG alone[22]. The model was then tested on a completely independent set of 52,870 patients, and its AUC was 0.93 for the detection of LV dysfunction, with corresponding sensitivity, specificity and accuracy of 93.0%, 86.3% and 85.7%, respectively. In individuals in whom the CNN seemingly incorrectly detected LV dysfunction (apparent false positive tests), those with a positive AI screen were fourfold more likely to develop LV dysfunction over a mean follow-up of 3.4 years than those with a negative AI screen, suggesting the ability of the model to detect LV dysfunction even before a decline in LVEF measured by echocardiography. Akin to the paradigm of the ischaemic cascade in myocardial infarction and ischaemic heart disease, the preclinical cascade of cellular-level alterations (such as changes to calcium homeostasis) and mechanical function alterations (such as abnormal lusitropy and strain rate) might be reflected in the ECG and detectable by a trained deep-learning CNN AI–ECG model. This observation raises the possibility that this model could be used to identify patients with early or subclinical LV dysfunction or even those with normal ventricular function who are at risk of heart failure. In a subsequent prospective validation from our group, the AI–ECG algorithm was applied in 3,874 patients who underwent transthoracic echocardiography and an ECG recording within 30 days[23]. In these patients, the algorithm was able to detect an LVEF of ≤35% with 86.8% specificity, 82.5% sensitivity and 86.5% accuracy (AUC 0.918). This algorithm has been validated externally in patients presenting with dyspnoea to the emergency department[24]. A validation effort in a multicentre, international cohort is also under way. Of note, the algorithm has also demonstrated high accuracy for detecting low LVEF when applied to a single-lead ECG, thereby allowing its application with the use of smartphone-based or even stethoscope-based electrodes.

Although further rigorous clinical testing for patient outcomes is required, these data highlight the AI–ECG as a potential means to overcome the limitations of previously tested screening biomarkers for asymptomatic LV dysfunction. Other future applications of the AI–ECG algorithm for the detection of LV dysfunction might include the longitudinal monitoring of patients with established congestive heart failure who are receiving medical therapy, the prognostication of the risk of incident cardiomyopathy in patients receiving cardiotoxic chemotherapy in whom cardioprotective therapy might be instituted prophylactically or the longitudinal monitoring of patients with valvular heart disease in whom the development of LV dysfunction would signify an indication for surgical intervention. Additionally, the FDA issued an Emergency Use Authorization for the 12-lead AI–ECG algorithm to detect LV dysfunction in patients with coronavirus disease 2019 (COVID-19)[25]. Broad clinical application across other specialties might also be possible.

Detection of silent AF from a sinus-rhythm ECG. AF portends an increased risk of impaired quality of life, stroke and heart failure, and results in frequent visits to the emergency department and frequent inpatient admissions. Among patients with an embolic stroke of undetermined source (ESUS), previously called 'cryptogenic stroke', who undergo 30-day rhythm monitoring, about 15% are found to have previously undiagnosed paroxysmal AF[26]. In these patients, anticoagulation lowers the recurrence of stroke and might lower mortality, whereas in the absence of documented AF, anticoagulation offers no clinical benefit and increases the risk of bleeding[27,28]. However, the diagnosis of AF can be elusive because up to 20% of patients are completely asymptomatic, and another approximately one-third of patients have atypical symptoms[29]. Moreover, AF is only intermittent (or paroxysmal) in many patients. Despite extensive research on the topic, the value of screening individuals for AF remains a matter of debate, and the US Preventive Services Task Force states that the data are currently insufficient to recommend routine AF screening in general populations[30].

In the recent Apple Heart Study[31], the largest pragmatic evaluation of AF screening in a general population using a smartwatch-enabled photoplethysmography technology, 0.52% of participants received notifications of possible AF over an average of >3 months of monitoring. In approximately one-third of these individuals, AF was later confirmed by week-long patch ECG monitoring. This finding suggests that, although mass screening of unselected populations is feasible with current technologies, the yield of this approach is low and the clinical effect is uncertain. Simple and highly accurate approaches to the detection of asymptomatic paroxysmal AF would be important for the selection of patients for early institution of oral anticoagulation for the prevention of AF-related morbidity and death.

To assess the likelihood of silent AF, a group of investigators from the Mayo Clinic developed a CNN to predict AF on the basis of a standard 12-lead ECG obtained during sinus rhythm[32]. The algorithm was developed using nearly half a million digitally stored ECGs from 126,526 patients and was validated and tested in separate internal datasets. The model applied convolutions on a temporal axis and across multiple leads to extract morphological and temporal features during the training and validation processes (FIG. 1a). Patients with at least one ECG showing AF within 31 days after the sinus-rhythm ECG were classified as being positive for AF. In the testing dataset, the algorithm demonstrated an AUC of 0.87, sensitivity of 79.0%, specificity of 79.5% and an accuracy of 79.4% in detecting patients with documentation of AF using only information from the sinus-rhythm ECG[32]. Therefore, the algorithm can detect nearly concomitant, unrecognized AF, rather than predict the long-term risk of AF. Conceptually, this AI tool converts a routine 10-s, 12-lead ECG into the equivalent of a prolonged rhythm-monitoring tool (FIG. 1b), although the duration of 'monitoring' and its yield require validation. Additionally, this tool can be applied retroactively to digitally stored ECGs from patients with a previous ESUS. This algorithm might facilitate targeted AF surveillance (such as using an ambulatory rhythm-monitoring patch or implantable loop recorder) in subsets of high-risk patients. This work is preliminary, but we are currently assessing the performance of this algorithm in identifying patients who might benefit from prospective AF screening or monitoring (with Holter or extended monitoring) and, ultimately, various stroke-prevention strategies. We also note that other groups have derived similar AF risk-prognostication tools that examine other electrophysiological parameters, such as signal-averaged ECG-derived P-wave analysis[33].

ECGs are ubiquitously performed for a variety of screening, diagnostic and monitoring purposes, thereby providing ample opportunities for the application of this algorithm. The ultimate clinical utility of this approach will be determined by the observed positive and negative predictive values of the algorithm when applied to a given population and by the cost and downstream consequences, particularly for patient outcomes, related to follow-up diagnostic testing and therapies.

Unlike the traditional risk-prediction models that comprise predefined variables, the CNN described above is agnostic, because we do not know what ECG features the CNN is 'seeing' and which factors drive its performance. The performance of the algorithm is likely to be based on a combination of ECG signatures that are known risk factors for AF (such as LV hypertrophy, P-wave amplitude, atrial ectopy and heart rate variability) as well as others that are currently unknown or are not obvious to the human eye, in combination, in a non-linear manner[34]. The ECG is also likely to contain information that correlates with known clinical risk factors.

AI-enabled ECG and rhythm tools in AF care. In addition to screening individuals for silent AF, CNNs can also be developed from ECGs or other rhythm-monitoring data (including those derived from permanently implanted cardiac devices) for the stratification of stroke risk and the refinement of decision-making

about oral anticoagulant use. In an analysis using data from implanted cardiac devices in >3,000 patients with AF (including 71 patients with stroke), three different supervised machine-learning models of AF burden signatures were developed to predict the risk of stroke (random forest, CNN and L1 regularized logistic regression)[35]. In the testing cohort, the random forest model had an AUC of 0.66, the CNN model had an AUC of 0.60 and the L1 regularized logistic regression model had an AUC of 0.56. By contrast, the CHA_2DS_2–VASc score, the most widely used stroke-prediction scheme in current practice[36], had an AUC of 0.52 for stroke prediction. However, the highest AUC (0.63) was achieved when the CHA_2DS_2–VASc score was combined with the random forest and CNN models[35], indicating the prognostic strength of approaches that combine AI-enriched models with traditional clinical tools. The performance of this model is still quite modest. The integration of additional information from the clinical history, imaging tests and circulating biomarkers might further improve risk stratification but this task is beyond current AI capabilities. For example, in an unsupervised cluster analysis of approximately 10,000 patients with AF in the ORBIT-AF registry, including patient-specific clinical data, medications, and laboratory, ECG and imaging data, four clinically relevant phenotypes of AF were identified, each with distinct associations with clinical outcomes (low comorbidity, behavioural comorbidity, device implantation and atherosclerotic comorbidity clusters)[37]. However, although this finding offers a proof of concept, the clinical utility of these clusters has not yet been demonstrated. The hope is that phenotype-specific treatment strategies will lead to superior patient outcomes, but testing is required.

A fully automated, electronic health record-embedded platform powered with AI–ECG capabilities and other advanced machine-learning methods, including natural language processing, could be trained to collect data from the ECG, AF patterns and other diagnostic tests or even clinical notes in order to continuously assess the risk of stroke. When a high risk of stroke is detected, the clinician is alerted and the timely initiation of anticoagulation could prevent a potentially devastating adverse clinical event. Similarly, real-time modelling of the risk of stroke could be realized on the basis of information collected via wearable ECG technologies or consumer-facing, smartphone-based ECG technologies. Ultimately, developing an easy-to-implement tool that provides clinically actionable data that are fully vetted and validated will be required to allow interventions for stroke prevention.

Detection of HCM. HCM is infrequent in the general population, with an estimated prevalence of 1 in 200 to 1 in 500 individuals[38,39]. However, HCM is one of the leading causes of sudden cardiac death among adolescents and young adults. HCM is also associated with substantial morbidity in all age groups[40]. Over the past 15 years, interest has focused on the screening of at-risk populations for HCM. This interest is often rekindled by highly publicized sudden deaths of athletes and other young adults, events that are devastating

and potentially preventable if a diagnosis of HCM had been established.

In most cases, a diagnosis of HCM can be established with echocardiography combined with the clinical history, but the widespread use of echocardiography for the detection of HCM in otherwise asymptomatic individuals is impractical. Therefore, alternative modalities, such as the ECG, have been considered as a means for screening. More than 90% of patients with HCM have electrocardiographic abnormalities[41], but these abnormalities are non-specific and can be indistinguishable from LV hypertrophy. Generally, ECG screening has relied on manual or automated detection of particular features, such as LV hypertrophy, left axis deviation, prominent Q waves and T-wave inversions. However, these approaches have insufficient diagnostic performance to justify routine ECG screening[42]. Moreover, several sets of ECG criteria have been proposed to distinguish between HCM and athletic heart adaptation, but their diagnostic performance has been inconsistent when external validations have been attempted[43,44]. The nature of a deep-learning AI approach might offer the advantage of an agnostic and unbiased approach to the ECG-based detection of HCM that does not rely on traditional criteria for LV hypertrophy.

With use of the ECGs of 2,500 patients with a validated diagnosis of HCM and >50,000 age-matched and sex-matched control individuals without HCM, an AI–ECG CNN was trained and validated to diagnose HCM on the basis of the ECG alone[45]. In an independent testing cohort of 612 patients with HCM and 12,788 control individuals, the AUC of the CNN was 0.96 (95% CI 0.95–0.96) with sensitivity of 87% and specificity of 90%. The performance of the model was robust in subgroups of patients meeting the ECG criteria for LV hypertrophy and among those with normal ECGs[45]. Importantly, performance was even better in younger patients (aged <40 years) but declined with increasing age. Furthermore, the performance of the model did not seem to be affected by the sarcomeric mutation status of the patient, given that the model-derived probabilities for a diagnosis of HCM were a median of 97% and 96% in patients with HCM who either had or did not have confirmed variants in sarcomere-encoding genes, respectively[45]. The algorithm developed had equally favourable performance when implemented on the basis of a single lead (rather than all 12 leads of the ECG), meaning that this algorithm could be applied as a screening test on a large scale and across various resource settings. FIGURE 2 shows an example of a woman aged 21 years with massive septal hypertrophy who underwent surgical septal myectomy[45]. Despite only modest abnormalities on her ECG before myectomy, the AI–ECG algorithm indicated a probability of HCM of 72.6%, whereas after myectomy, the AI–ECG algorithm indicated a probability of HCM of only 2.5%, despite more obvious and striking ECG abnormalities.

Another group of investigators used a large, 12-lead ECG dataset to train machine-learning models for the detection of HCM together with other elements of cardiac structure (LV mass, left atrial volume and early diastolic mitral annulus velocity) and disease (pulmonary

arterial hypertension, cardiac amyloidosis and mitral valve prolapse)[46]. Although a different model architecture was applied from that in the aforementioned study, including a novel combination of CNNs and hidden Markov models, the performance of the model for the detection of HCM was also quite favourable, with an AUC of 0.91. Of note, the researchers also reported good performance for the detection of pulmonary arterial hypertension (AUC 0.94), cardiac amyloidosis (AUC 0.86) and mitral valve prolapse (AUC 0.77).

The favourable diagnostic performance of these models suggests that HCM screening based on fully automated AI–ECG algorithms might be feasible in the future. External validations in other populations with greater racial diversity, as well as in athletes and in adolescents, will be crucial in the evaluation of the AI–ECG algorithm as a future screening tool for HCM in individuals at risk of adverse outcomes, particularly sudden cardiac death. Direct comparisons with other possible screening methods and cost-effectiveness and other practical implementation issues also need to be evaluated.

Detection of hyperkalaemia. Numerous studies have shown that either hyperkalaemia or hypokalaemia is associated with increased mortality, and evidence suggests that the mortality associated with hyperkalaemia might be linked to underdosing of evidence-based therapies[47]. Our group has evaluated the performance of an AI–ECG CNN for the detection of hyperkalaemia in patients with chronic kidney disease[48,49]. In the latest large-scale evaluation, the model was trained to detect serum potassium levels of ≥5.5 mmol/l using >1.5 million ECGs from nearly 450,000 patients who underwent contemporaneous assessment of serum potassium levels. This level of potassium was chosen because this threshold was thought to be clinically actionable. At this cut-off point, the model demonstrated 90% sensitivity and 89% sensitivity in a multicentre, external validation cohort[49]. This algorithm could be applied to detect clinically silent but clinically significant hyperkalaemia without a blood draw and could facilitate remote patient care, including diuretic dosing, timing of haemodialysis or adjustment of medications such as angiotensin-converting enzyme inhibitors or angiotensin-receptor blockers in the setting of heart failure or chronic kidney disease.

Antiarrhythmic drug management. Dofetilide and sotalol are commonly used for the treatment of AF. Their antiarrhythmic effect is exerted on the myocardium by prolonging the duration of the repolarization phase, meaning that QT prolongation is an anticipated effect of these drugs. Owing to the ensuing risk of substantial QT prolongation and potentially fatal ventricular proarrhythmia, patients require close monitoring with a continuous ECG in the hospital setting when these drugs are used, particularly for dofetilide. In addition, with the long-term use of these medications, the QT interval should be intermittently assessed because dose adjustments might be necessary in cases of substantial QT prolongation, concomitant medications with QT-prolonging effects and fluctuations in renal function (both sotalol and dofetilide are primarily metabolized through the kidneys). Using serial 12-lead ECGs and linked information on plasma dofetilide concentrations in 42 patients who were treated with dofetilide or placebo in a crossover randomized clinical trial, a deep-learning algorithm predicted plasma dofetilide concentrations with good correlation ($r = 0.85$)[50]. By comparison, a linear model of the corrected QT interval correlated with dofetilide concentrations with a coefficient of 0.64 (REF.[50]). This finding suggests that the QT interval might not accurately reflect the plasma dofetilide concentration in some patients and so might underestimate or overestimate the proarrhythmic risk. Machine-learning approaches, including supervised, unsupervised and reinforcement learning, have also been used to determine the optimal dosing regimen during dofetilide treatment[51].

In the future, patients treated with dofetilide, sotalol or other antiarrhythmic medications might avoid the need for hospitalization for drug loading or office visits for routine surveillance ECGs by monitoring their own ECG using smartphone-based tools powered with AI capabilities to determine the plasma concentrations of the drug or the risk of drug-related toxic effects. The development of these AI algorithms applied to a single-lead ECG is still in progress.

Wearable and mobile ECG technologies

AI algorithms can be applied to wearable technologies, enabling rapid, point-of-care diagnoses for patients and consumers. Although many algorithms have been derived using 12-lead ECG data, some studies have

a Before myectomy
AI–ECG probability of HCM 72.6%

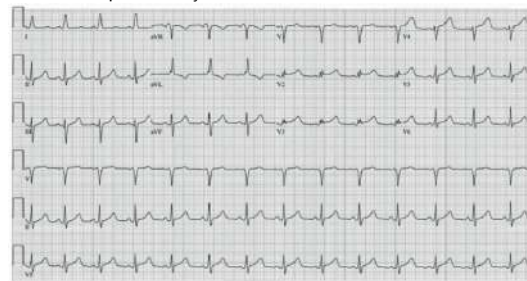**b After myectomy**
AI–ECG probability of HCM 2.5%

Fig. 2 | **The AI–ECG to detect HCM.** Use of an artificial intelligence-enhanced electrocardiogram (AI–ECG) model to detect obstructive hypertrophic cardiomyopathy (HCM) in a woman aged 21 years before (part **a**) and after (part **b**) septal myectomy. Adapted with permission from REF.[45], Elsevier.
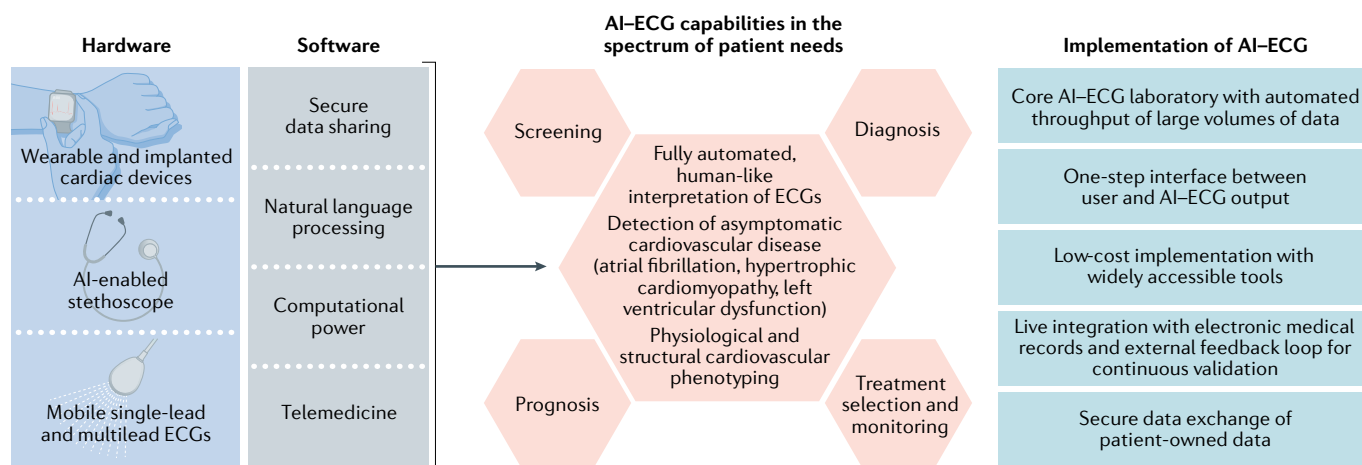
Fig. 3 | **Framework for AI–ECG applications in clinical practice.** Current, versatile electrocardiogram (ECG)-recording technologies (wearable and implantable devices, smartwatches and e-stethoscopes) coupled with the ability to store, transfer, process and analyse large amounts of digital data are increasingly allowing the deployment of artificial intelligence (AI)-powered tools in the clinical arena, addressing the spectrum of patient needs. The science of AI-enhanced ECG (AI–ECG) implementation, including the interface between patients and the AI–ECG output, integration of AI–ECG tools with electronic health records, patient privacy, and cost and reimbursement implications, is in its infancy and continues to evolve.

demonstrated favourable performance even when algorithms are deployed on single-lead ECGs[8]. The performance of AI–ECG algorithms for the detection of HCM or the determination of serum potassium levels when applied to single-lead ECGs has been shown not to be significantly different from the performance when applied to 12-lead ECGs[45,52]. Additionally, signals other than the ECG can also be analysed using AI approaches. For instance, a deep neural network has been developed to detect AF passively from photoplethysmography signals obtained from the Apple Watch[14]. Newer iterations of the Apple Watch now allow users to confirm the presence of AF using electrophysiological signals obtained through a single bipolar vector[14].

The COVID-19 pandemic has highlighted the need for rapid, point-of-care diagnostic testing. For instance, early interest in the use of hydroxychloroquine or azithromycin for the treatment of the infection resulted in a marked increase in the use of these medications. Given the potential for these medications to prolong myocardial repolarization and increase the risk of dangerous ventricular arrhythmias, the FDA issued an Emergency Use Authorization for the use of mobile devices to record and monitor the QT interval in patients taking these medications. The FDA also issued an Emergency Use Authorization for the detection of low LVEF as a potential complication of COVID-19 using the AI–ECG algorithm integrated in the digital Eko stethoscope (Eko Devices)[22,25]. Furthermore, an international consortium is currently evaluating the ECG as a potential means to diagnose COVID-19, cardiac involvement or the risk of cardiac deterioration, given the known ECG changes and cardiac involvement in patients with COVID-19 (REFS[53,54]). Although results are not yet available, these types of investigation emphasize the potential power of digitally delivered AI technologies for timely deployment at the point of care and large-scale implementation.

Implementation of AI–ECG
In contrast to data obtained through the clinical history, medical record review or imaging tests, the ease and consistency with which ECG data can be obtained and analysed for the development and implementation of AI models are likely to accelerate the uptake of the AI–ECG in clinical applications, with ensuing increases in workflow efficiency. The demonstrated capabilities of the AI–ECG described in the previous sections have the potential to influence the spectrum of patient care, including screening, diagnosis, prognostication, and personalized treatment selection and monitoring (FIG. 3). The preliminary data on the performance of the AI–ECG algorithms are clearly promising, but these technologies will be meaningful only inasmuch as they improve our clinical practice and patient outcomes[55]. To this end, several AI technologies are currently being tested in various clinical applications.

The algorithm to identify LV dysfunction using the ECG is currently being evaluated in a large-scale, pragmatic, cluster randomized clinical trial[56]. The EAGLE trial[57] randomly assigned >100 clinical teams (or clusters) either to have access to the new AI screening tool results or to usual care at nearly 50 primary care practices (which will encompass >400 clinicians and >24,000 patients) in the Mayo Clinic Health System. Eligible patients include adults who undergo ECG for any reason and in whom low LVEF has not been previously diagnosed. The primary outcome is the detection of low LVEF (<50%), as determined by standard echocardiography. The objective of this study is twofold: to evaluate the real-world efficacy of the algorithm in identifying patients with asymptomatic or previously unrecognized LV dysfunction in primary care practices and to understand how information derived from AI algorithms is interpreted and acted on by clinicians — how do humans and machines interact? This study will

validate (or refute) the utility of this approach and will help us to understand potential barriers and opportunities for the implementation of AI in clinical practice. Regardless of its results, the EAGLE trial[57] will be an important study because it will be the prototype study for the implementation of AI-enabled tools.

Similarly, we are developing a protocol to assess the algorithm to identify concomitant silent AF or the risk of near-term AF using a 12-lead ECG obtained during normal sinus rhythm. The BEAGLE trial[58] will seek to evaluate the utility of this AI algorithm for targeted AF screening in patients who would have at least a moderate risk of stroke if they had AF. Subsequent follow-up studies will test the role of empirical anticoagulation in selected patients without known documented AF who have a high probability of AF according to the AI–ECG. In the first pilot phase, patients at various levels of risk of AF on the basis of their AI–ECG will be fitted with cardiac event monitors to provide surveillance for incident AF. We anticipate that the AI–ECG output will help us to identify patients who are at risk of AF and that the AI–ECG might increase the diagnostic yield of screening.

Another application of this algorithm might be in guiding treatment decisions for patients with ESUS. We postulate that these patients might benefit from intensified screening or even empirical anticoagulation on the basis of a high probability of AF, as indicated by the AI–ECG algorithm. Several studies have shown no benefit of empirical anticoagulation in patients with ESUS[27,28], but the AI–ECG might help to identify a subset of patients with ESUS in whom recurrent strokes can be prevented. At this point, this concept is speculative, but we intend to pursue this hypothesis by examining existing datasets and possibly in a prospective clinical trial.

We have demonstrated the potential utility of the AI–ECG for the diagnosis of HCM[45]. Screening individuals

for HCM has attracted considerable controversy, driven mostly by the poor diagnostic performance of the ECG and the downstream consequences of false positive or equivocal findings. External validation of this algorithm in completely independent cohorts of patients with HCM is currently in progress. Evaluations of this algorithm in various populations, including family members of patients with HCM, patients with undifferentiated syncope, athletes or even unselected patients through retrospective medical record review, are also being planned. These studies are likely to demonstrate the utility and limits of the AI–ECG for screening individuals for rare conditions and will help us to understand how to apply appropriate thresholds based on Bayesian concepts of pretest risk and downstream effects of a test, including costs.

In addition to straightforward diagnostic tests, the AI–ECG might help to refine the clinical workflow. One such application might be the initiation and monitoring of antiarrhythmic medications that carry a risk of cardiotoxicity or proarrhythmia and require close monitoring for QT prolongation and new ventricular arrhythmias in order to guide initial and subsequent dosing[50]. Other potential future applications of the AI–ECG in various settings with direct effects on cardiac clinical care are listed in BOX 1.

At the Mayo Clinic, we have developed an internal website (AI–ECG Dashboard) where the medical record number of any patient of interest can be entered for all their ECGs to be retrospectively analysed, with probabilities reported for the presence of LV systolic dysfunction, silent AF and HCM, and their AI–ECG age and sex prediction also reported[59]. FIGURE 4a depicts the AI–ECG Dashboard output for the example of a patient with an ESUS who had a high AI–ECG-derived probability of AF on an ECG that preceded the thromboembolic event by 12 years. At 5 years after the initial stroke, this patient had a recurrent stroke and was clinically documented to have AF shortly thereafter, which was 17 years after the first ECG had indicated an elevated risk of AF by AI analysis[60]. FIGURE 4b depicts an example of a different patient with a history of cardiomyopathy who underwent heart transplantation in 2005, at which time the probability of low LVEF dropped precipitously and remained low until 2020, when the patient experienced graft rejection with LV dysfunction. At that point, the AI–ECG reported a high probability of low LVEF, correlating accurately with the clinical syndrome. The AI–ECG Dashboard has now been integrated into the electronic medical record, and clinicians can rapidly access the results of the AI analysis on all of a patient's available ECGs. Other tools, such as those for AI–ECG detection of valvular heart disease, cardiac amyloidosis and pulmonary arterial hypertension, have been developed and are undergoing testing before their addition to the dashboard.

Potential challenges and solutions

The AI–ECG technologies offer great promise, but it is important to acknowledge several potential challenges. Given that models are often derived from high-quality databases with meticulously obtained ECGs and

---

Box 1 | Potential future applications of the AI–ECG

- Atrial fibrillation
  - Beyond the CHA₂DS₂–VASc score — electrocardiogram (ECG) markers of stroke risk integrated with electronic health record-based clinical information, imaging and biomarkers
  - Rhythm-pattern recognition (implantable and wearable cardiovascular electronic devices) to predict atrial fibrillation events and artificial intelligence-enhanced ECG (AI–EGG)-guided 'pill-in-pocket' (oral anticoagulation or class Ic antiarrhythmic drugs) in patients with paroxysmal atrial fibrillation
  - Smartphone-based surveillance of the QT interval for patients in whom treatment with sotalol, dofetilide or other medications that can affect repolarization is initiated or for patients longitudinally treated with sotalol, dofetilide or other medications that can affect repolarization
- Sudden cardiac death: identifying which patients should receive an implantable cardioverter–defibrillator and integration with clinical and imaging markers
  - Beyond the left ventricular ejection fraction in patients with ischaemic or non-ischaemic cardiomyopathy
  - Risk stratification in patients with hypertrophic cardiomyopathy or inherited arrhythmia syndromes
- Predicting the exacerbation of congestive heart failure on the basis of continuous ECG ambulatory data to prevent emergency department visits and hospitalizations
- Severity and staging of heart failure, valvular heart disease and pulmonary arterial hypertension to guide clinical decision-making, prognosis and monitoring
- Angina and stress ECG analysis — identifying high-risk patients who will benefit from invasive coronary evaluation and possible intervention

---

well-phenotyped patients, their application to ECGs obtained in routine clinical practice in real-world settings might be poor. Similarly, although the models might perform well in one population, they require rigorous evaluation for external validity in diverse populations. The aforementioned AI–ECG model for low LVEF has been validated in racially diverse cohorts, but validation data for the other AI–ECG model algorithms are pending. Although models might perform well in terms of their individual performance characteristics, this performance does not always translate into meaningful and actionable clinical information. For instance, screening tests for very rare conditions might be limited by low positive predictive value when applied to populations with low pretest probability of the disease. Although an algorithm might seem to predict a disease state well, if this information does not add to other readily available data (such as age, sex and comorbidities), the algorithm will add very little to clinical risk stratification. Many routinely used screening and diagnostic tests often do not produce consistent improvements in downstream patient outcomes, thereby offering little incremental value to clinical care[61]. Similar assessments of the effects of routinely integrating the AI–ECG into clinical practice and its implications for patient outcomes and costs will be important. Clearly, the delivery of AI-related applications in the clinical environment generates a new set of previously unrecognized challenges.
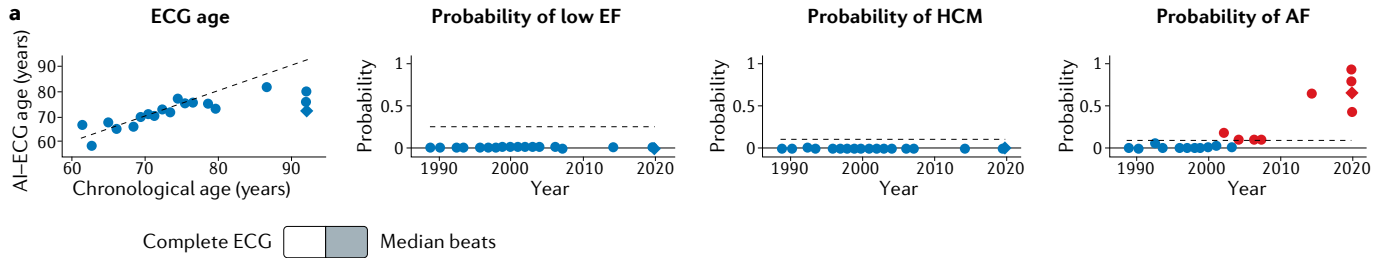
As with most other AI-enabled tools, the development of AI–ECG models requires large datasets for training, validation and testing. In some cases, multicentre collaborations might be necessary to assemble the sample sizes required for the development of high-fidelity models. Such collaboration is particularly pertinent when the condition of interest is rare or when an urgent clinical need dictates the rapid development of a model, such as the AI–ECG tools for the diagnosis and risk stratification of COVID-19 during the current pandemic. In this process and during external validation of any AI–ECG model, large amounts of patient data are exchanged between research teams worldwide, generating concerns for the security and protection of sensitive patient information that might be susceptible to cyberattacks or other threats[62]. In the current environment, the use of traditional encryption methods might not be sufficient to alleviate these concerns. Among other possible novel data-protection solutions, blockchain technology might allow the secure and traceable sharing of patient data between investigators and institutions for the development, validation and clinical implementation of AI tools by generating a decentralized marketplace of securely stored patient data specifically intended to be used in AI applications[63,64]. Of note, even a miniscule artificial perturbation of the input data (such as a single pixel in an image or an ECG) that is unrecognizable by the human eye might lead an otherwise well-trained CNN to misclassify the data and generate false output. CNNs are clearly vulnerable to adversarial perturbations of input data, and shielding against these vulnerabilities will be important for their future widespread implementation[65].

We must also acknowledge that a risk of perpetuating bias exists with AI. Algorithms that are derived using current clinical practice patterns and outcomes will reflect the disparity present and pose a risk of reinforcing and perpetuating these inequities. To address this risk, investigators and consumers of these algorithms must consider how the data are collected (in what population and of what quality), how the models perform, how the model should best be applied and how the findings fit within existing knowledge.
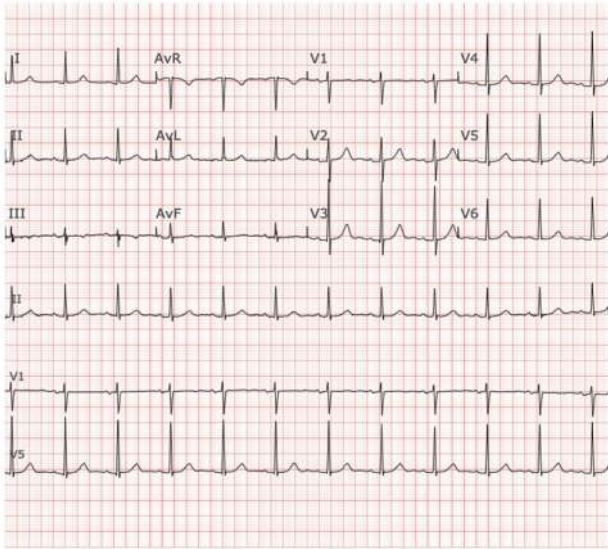
Clinical and technical challenges also exist to the application of the AI–ECG to patient care. AI analysis relies on standardized digital ECG acquisition that is not universal in current clinical environments, even in high-resource settings. Poor-quality data can prevent the effective development and application of AI–ECG technologies. Current AI tools cannot analyse ECGs stored as images (such TIFF or PDF file types), which limits generalizability in some clinical environments. Furthermore, the infrastructure to integrate AI–ECG results into electronic health records and make them available at the point of clinical care is not widely available.

An evidence base for the real-world effects of AI–ECG analysis on clinical care and ultimately on important patient outcomes remains to be established. Outcomes to be considered are the uptake by clinicians, effect on clinical decision-making (how does the result of AI analysis affect downstream testing and therapies?) and ultimately whether AI–ECG analysis improves the provider and patient experience and patient-specific clinical outcomes. Similarly, how does the AI–ECG algorithm complement existing clinical or laboratory markers of disease? For example, can we use the AI–ECG algorithm to detect LV systolic dysfunction in addition to using measurements of plasma levels of N-terminal prohormone of B-type natriuretic peptide and what does it add? Randomized clinical trials of the pragmatic implementation of the AI–ECG, such as the EAGLE[56,57] and BEAGLE[58] trials discussed previously, are now being conducted and will provide unique insights. Rigorous testing of AI–ECG tools in randomized trials is crucial for demonstrating their value, as with any other form of clinical intervention.

Lastly, the regulatory aspects of incorporating AI–ECG-derived diagnoses for direct clinical care are now starting to be formulated. The barrier to approval by regulatory bodies might be quite different from that for devices or medications and might vary according to the ability of a clinician to 'over-read' the AI–ECG findings. For algorithms that streamline workflow but perform a task that is usually done by humans (for example, rhythm determination), the algorithm might be approved as a tool to aid clinician workflow. However, if the models perform an analysis that cannot be done by expert clinicians (for example, determination of 'ECG age' or the risk of future AF), additional approval complexities might exist. Similarly, if algorithms are meant to be applied for patients and consumers at the point of care (without clinician oversight), a higher bar for performance of the model might be required. A legal framework to support AI-based clinical decision-making has not yet been established.
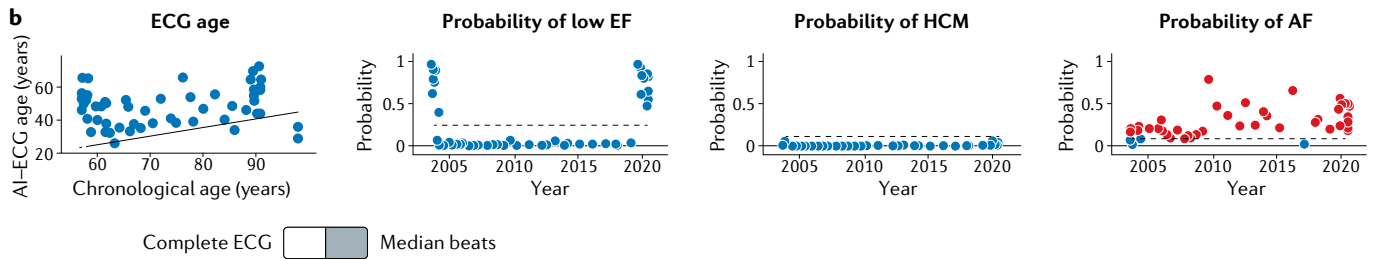
a

ECG age | Probability of low EF | Probability of HCM | Probability of AF

Complete ECG ▭ Median beats

Normal sinus rhythm. Normal ECG

Results (screenshot from AI–ECG Dashboard)

ECG Date	Main Rhythm	Heart Rate	QT/QTc	Real Age	ECG Age	P of Male (%)	P of Low EF (%)	P of AF (%)	P of HCM (%)
XX/XX/2019	Atrial flutter	116	326/453	92.2	72.0	17.42%	0.59%	65.63%	0.01%
XX/XX/2019	Sinus tachycardia	109	356/477	92.2	79.7	16.35%	0.93%	43.52%	0.01%
XX/XX/2019	Normal sinus rhythm	79	426/488	92.1	79.7	7.77%	0.39%	92.94%	0.03%
XX/XX/2019	Sinus rhythm	92	398/492	92.1	75.7	7.55%	1.87%	78.84%	0.22%
XX/XX/2014	Normal sinus rhythm	69	434/465	86.7	81.5	1.04%	1.11%	65.15%	0.10%
XX/XX/2007	Normal sinus rhythm	63	456/466	79.7	72.9	0.57%	0.75%	10.56%	0.18%
XX/XX/2006	Normal sinus rhythm	69	436/463	78.6	74.9	2.94%	1.77%	10.28%	0.08%
XX/XX/2004	Normal sinus rhythm	63	448/454	76.6	75.3	2.99%	0.63%	10.71%	0.18%
XX/XX/2003	Normal sinus rhythm	68	436/460	75.5	75.0	0.13%	0.57%	1.75%	0.35%
XX/XX/2002	Normal sinus rhythm	65	436/449	74.5	76.9	2.39%	0.59%	18.00%	0.04%
XX/XX/2000	Normal sinus rhythm	74	404/445	73.4	71.5	0.59%	1.05%	2.61%	0.10%
XX/XX/1999	Normal sinus rhythm	85	388/461	72.4	72.5	1.66%	0.53%	1.48%	0.01%
XX/XX/1998	Normal sinus rhythm	71	420/456	71.3	70.1	1.91%	0.70%	0.42%	0.10%
XX/XX/1997	Normal sinus rhythm	85	388/461	70.4	70.9	0.39%	0.55%	0.45%	0.10%
XX/XX/1996	Normal sinus rhythm	68	416/442	69.4	69.6	0.46%	1.21%	0.24%	0.15%
XX/XX/1995	Normal sinus rhythm	68	444/472	68.4	65.8	0.99%	0.71%	0.50%	0.17%

b

ECG age | Probability of low EF | Probability of HCM | Probability of AF

Complete ECG ▭ Median beats

Probable sinus tachycardia with first-degree AV block
Premature ventricular complexes. Right-axis deviation.
Incomplete right bundle branch block. Low anterior forces.
Non-specific ST and T-wave abnormalities.
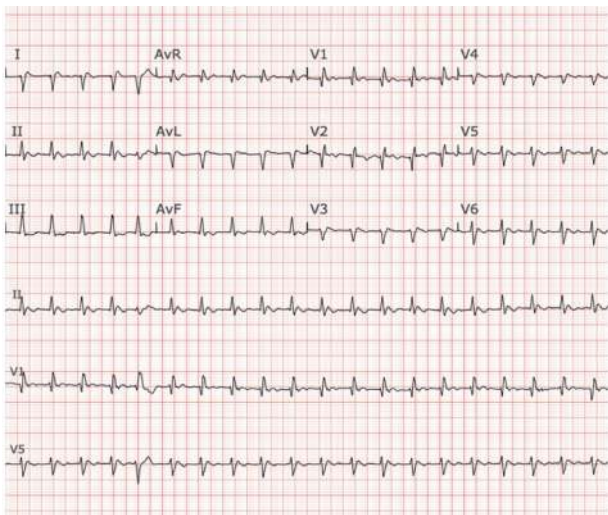
Results (screenshot from AI–ECG Dashboard)

ECG Date	Main Rhythm	Heart Rate	QT/QTc	Real Age	ECG Age	P of Male (%)	P of Low EF (%)	P of AF (%)	P of HCM (%)
XX/XX/2020	Atrial tachycardia	95	430/537	40.4	58.8	22.89%	55.78%	18.65%	1.25%
XX/XX/2020	Junctional tachycardia	127	212/307	40.4	43.4	7.90%	85.06%	21.91%	0.48%
XX/XX/2020	Junctional tachycardia	128	310/452	40.4	59.4	42.88%	83.57%	36.26%	2.81%
XX/XX/2020	Sinus tachycardia	104	388/510	40.4	62.7	23.73%	66.18%	27.97%	0.08%
XX/XX/2020	Sinus tachycardia	107	368/491	40.4	57.3	38.91%	87.47%	47.73%	0.03%
XX/XX/2020	Atrial tachycardia	118	320/448	40.3	71.1	13.76%	48.80%	51.07%	0.47%
XX/XX/2020	Atrial tachycardia	115	350/484	40.1	43.1	20.40%	83.08%	51.82%	1.89%
XX/XX/2019	Supraventricular tachycardia	140	302/459	39.9	51.1	42.99%	86.01%	24.97%	8.47%
XX/XX/2019	Left posterior fascicular block	142	292/449	39.8	53.7	20.46%	96.27%	57.62%	0.03%
XX/XX/2019	Sinus tachycardia	106	366/483	39.8	67.9	18.69%	95.79%	44.59%	0.12%
XX/XX/2019	Sinus tachycardia	116	332/461	39.8	57.5	62.02%	62.84%	52.21%	1.37%
XX/XX/2019	Sinus tachycardia	121	322/457	39.6	63.4	29.47%	99.72%	42.74%	0.81%
XX/XX/2019	Sinus tachycardia	112	358/488	39.1	45.4	22.38%	3.48%	20.73%	0.01%
XX/XX/2018	Sinus tachycardia	108	336/448	38.0	33.1	60.99%	0.41%	32.23%	0.03%
XX/XX/2017	Normal sinus rhythm	93	348/432	37.9	47.3	18.37%	1.81%	29.62%	0.00%
XX/XX/2017	Normal sinus rhythm	94	354/442	37.1	39.6	66.81%	2.85%	3.15%	0.00%
XX/XX/2016	Normal sinus rhythm	90	346/423	36.1	54.2	11.23%	2.37%	66.66%	0.01%
XX/XX/2015	Normal sinus rhythm	99	352/451	35.1	46.0	11.45%	2.21%	22.40%	0.01%
XX/XX/2014	Sinus rhythm	97	310/393	34.1	38.4	16.19%	3.03%	37.29%	0.02%
XX/XX/2013	Sinus tachycardia	105	406/533	33.8	52.7	3.54%	1.61%	42.22%	0.03%
XX/XX/2013	Accelerated Junctional rhythm	104	362/476	33.1	64.9	20.88%	6.29%	24.44%	0.02%
XX/XX/2012	Normal sinus rhythm	90	352/430	32.5	37.7	25.20%	1.49%	52.82%	0.00%
XX/XX/2012	Normal sinus rhythm	91	354/435	32.1	39.3	7.32%	0.30%	25.22%	0.00%
...	...	...	...	...	...	...	...	...	...
XX/XX/2003	Sinus tachycardia	113	292/398	23.5	52.3	87.18%	99.87%	16.13%	0.06%

◀ Fig. 4 | **AI–ECG Dashboard linked from within the electronic health record for point-of-care application. a** | A patient with embolic stroke of undetermined source had an increased probability of silent atrial fibrillation (AF; red data points) that predated the clinical documentation of atrial flutter. **b** | A patient with a history of heart transplantation in 2005 who experienced graft rejection with left ventricular systolic dysfunction in 2020. At that point, the artificial intelligence-enhanced electrocardiogram (AI–ECG) reported a high probability of low ejection fraction (EF), correlating with the graft rejection. AV, atrioventricular; HCM, hypertrophic cardiomyopathy.

Conclusions

The true potential of deep-learning AI applied to the ubiquitous 12-lead ECG is starting to be realized. The utility of the AI–ECG is being demonstrated as a tool for comprehensive human-like interpretation of the ECG, but also as a powerful tool for phenotyping of cardiac health and disease that can be applied at the point of care. The implementation of the AI–ECG is still in its infancy, but a continuously growing clinical investigation agenda will determine the added value of these AI tools, their optimal deployment in the clinical arena and their multifaceted and so-far largely unpredictable implications. As with any medical tool, the AI–ECG must be vetted, validated and verified, and clinicians must be trained to use it properly, but when integrated into medical practice, the AI–ECG holds the promise to transform clinical care.

Published online 1 February 2021

1. Krizhevsky, A., Sustskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **25**, 1106–1114 (2012).
2. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **1**, 206–215 (2019).
3. Gottesman, O. et al. Guidelines for reinforcement learning in healthcare. *Nat. Med.* **25**, 16–18 (2019).
4. Pipberger, H. V., Freis, E. D., Taback, L. & Mason, H. L. Preparation of electrocardiographic data for analysis by digital electronic computer. *Circulation* **21**, 413–418 (1960).
5. Caceres, C. A. & Rikli, A. E. The digital computer as an aid in the diagnosis of cardiovascular disease. *Trans. NY Acad. Sci.* **23**, 240–245 (1961).
6. Caceres, C. A. et al. Computer extraction of electrocardiographic parameters. *Circulation* **25**, 356–362 (1962).
7. Rikli, A. E. et al. Computer analysis of electrocardiographic measurements. *Circulation* **24**, 643–649 (1961).
8. Hannun, A. Y. et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat. Med.* **25**, 65–69 (2019).
9. Ribeiro, A. H. et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat. Commun.* **11**, 1760 (2020).
10. Smith, S. W. et al. A deep neural network learning algorithm outperforms a conventional algorithm for emergency department electrocardiogram interpretation. *J. Electrocardiol.* **52**, 88–95 (2019).
11. Zhu, H. et al. Automatic multilabel electrocardiogram diagnosis of heart rhythm or conduction abnormalities with deep learning: a cohort study. *Lancet Digit. Health* **2**, E348–E357 (2020).
12. Kashou, A. H. et al. A comprehensive artificial intelligence–enabled electrocardiogram interpretation program. *Cardiovasc. Digit. Health J.* **1**, 62–70 (2020).
13. Bumgarner, J. M. et al. Smartwatch algorithm for automated detection of atrial fibrillation. *J. Am. Coll. Cardiol.* **71**, 2381–2388 (2018).
14. Tison, G. H. et al. Passive detection of atrial fibrillation using a commercially available smartwatch. *JAMA Cardiol.* **3**, 409–416 (2018).
15. Schlapfer, J. & Wellens, H. J. Computer-interpreted electrocardiograms: benefits and limitations. *J. Am. Coll. Cardiol.* **70**, 1183–1192 (2017).
16. Redfield, M. M. et al. Burden of systolic and diastolic ventricular dysfunction in the community: appreciating the scope of the heart failure epidemic. *JAMA* **289**, 194–202 (2003).
17. Wang, T. J. et al. Natural history of asymptomatic left ventricular systolic dysfunction in the community. *Circulation* **108**, 977–982 (2003).
18. Yancy, C. W. et al. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Circulation* **128**, e240–e327 (2013).
19. Vasan, R. S. et al. Plasma natriuretic peptides for community screening for left ventricular hypertrophy and systolic dysfunction: the Framingham Heart Study. *JAMA* **288**, 1252–1259 (2002).
20. Gruca, T. S., Pyo, T. H. & Nelson, G. C. Providing cardiology care in rural areas through visiting consultant clinics. *J. Am. Heart Assoc.* **5**, e002909 (2016).
21. Costello-Boerrigter, L. C. et al. Amino-terminal pro-B-type natriuretic peptide and B-type natriuretic peptide in the general community: determinants and detection of left ventricular dysfunction. *J. Am. Coll. Cardiol.* **47**, 345–353 (2006).
22. Attia, Z. I. et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat. Med.* **25**, 70–74 (2019).
23. Attia, Z. I. et al. Prospective validation of a deep learning electrocardiogram algorithm for the detection of left ventricular systolic dysfunction. *J. Cardiovasc. Electrophysiol.* **30**, 668–674 (2019).
24. Adedinsewo, D. et al. An artificial intelligence-enabled ECG algorithm to identify patients with left ventricular systolic dysfunction presenting to the emergency department with dyspnea. *Circ. Arrhythm. Electrophysiol.* **13**, e008437 (2020).
25. FDA. *Emergency use of the ELECT during the COVID-19 pandemic* https://www.fda.gov/media/137930/download (2020).
26. Gladstone, D. J. et al. Atrial fibrillation in patients with cryptogenic stroke. *N. Engl. J. Med.* **370**, 2467–2477 (2014).
27. Hart, R. G. et al. Rivaroxaban for stroke prevention after embolic stroke of undetermined source. *N. Engl. J. Med.* **378**, 2191–2201 (2018).
28. Diener, H. C. et al. Dabigatran for prevention of stroke after embolic stroke of undetermined source. *N. Engl. J. Med.* **380**, 1906–1917 (2019).
29. Siontis, K. C. et al. Typical, atypical, and asymptomatic presentations of new-onset atrial fibrillation in the community: characteristics and prognostic implications. *Heart Rhythm.* **13**, 1418–1424 (2016).
30. US Preventive Services Task Force. Screening for atrial fibrillation with electrocardiography: US Preventive Services Task Force recommendation statement. *JAMA* **320**, 478–484 (2018).
31. Perez, M. V. et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. *N. Engl. J. Med.* **381**, 1909–1917 (2019).
32. Attia, Z. I. 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* **394**, 861–867 (2019).
33. Palano, F. et al. Assessing atrial fibrillation substrates by P wave analysis: a comprehensive review. *High Blood Press. Cardiovasc. Prev.* **27**, 341–347 (2020).
34. Dewland, T. A. et al. Atrial ectopy as a predictor of incident atrial fibrillation: a cohort study. *Ann. Intern. Med.* **159**, 721–728 (2013).
35. Han, L. et al. Atrial fibrillation burden signature and near-term prediction of stroke: a machine learning analysis. *Circ. Cardiovasc. Qual. Outcomes* **12**, e005595 (2019).
36. Lip, G. Y., Nieuwlaat, R., Pisters, R., Lane, D. A. & Crijns, H. J. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the Euro Heart Survey on atrial fibrillation. *Chest* **137**, 263–272 (2010).
37. Inohara, T. et al. Association of atrial fibrillation clinical phenotypes with treatment patterns and outcomes: a multicenter registry study. *JAMA Cardiol.* **3**, 54–63 (2018).
38. Semsarian, C., Ingles, J., Maron, M. S. & Maron, B. J. New perspectives on the prevalence of hypertrophic cardiomyopathy. *J. Am. Coll. Cardiol.* **65**, 1249–1254 (2015).
39. Maron, B. J. et al. Prevalence of hypertrophic cardiomyopathy in a general population of young adults. Echocardiographic analysis of 4111 subjects in the CARDIA study. *Circulation* **92**, 785–789 (1995).
40. Maron, B. J., Haas, T. S., Murphy, C. J., Ahluwalia, A. & Rutten-Ramos, S. Incidence and causes of sudden death in U.S. college athletes. *J. Am. Coll. Cardiol.* **63**, 1636–1643 (2014).
41. McLeod, C. J. et al. Outcome of patients with hypertrophic cardiomyopathy and a normal electrocardiogram. *J. Am. Coll. Cardiol.* **54**, 229–233 (2009).
42. Maron, B. J. et al. Assessment of the 12-lead electrocardiogram 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. *J. Am. Coll. Cardiol.* **64**, 1479–1514 (2014).
43. Corrado, D. et al. Recommendations for interpretation of 12-lead electrocardiogram in the athlete. *Eur. Heart J.* **31**, 243–259 (2010).
44. Uberoi, A. et al. Interpretation of the electrocardiogram of young athletes. *Circulation* **124**, 746–757 (2011).
45. Ko, W. Y. et al. Detection of hypertrophic cardiomyopathy using a convolutional neural network-enabled electrocardiogram. *J. Am. Coll. Cardiol.* **75**, 722–733 (2020).
46. Tison, G. H., Zhang, J., Delling, F. N. & Deo, R. C. Automated and interpretable patient ECG profiles for disease detection, tracking, and discovery. *Circ. Cardiovasc. Qual. Outcomes* **12**, e005289 (2019).
47. Ferreira, J. P. et al. Abnormalities of potassium in heart failure: JACC state-of-the-art review. *J. Am. Coll. Cardiol.* **75**, 2836–2850 (2020).
48. Galloway, C. D. et al. Development and validation of a deep-learning model to screen for hyperkalemia from the electrocardiogram. *JAMA Cardiol.* **4**, 428–436 (2019).
49. Attia, Z. I. et al. Novel bloodless potassium determination using a signal-processed single-Lead ECG. *J. Am. Heart Assoc.* **5**, e002746 (2016).
50. Attia, Z. I. et al. Noninvasive assessment of dofetilide plasma concentration using a deep learning (neural network) analysis of the surface electrocardiogram: a proof of concept study. *PLoS ONE* **13**, e0201059 (2018).
51. Levy, A. E. et al. Applications of machine learning in decision analysis for dose management for dofetilide. *PLoS ONE* **14**, e0227324 (2019).
52. Yasin, O. Z. et al. Noninvasive blood potassium measurement using signal-processed, single-lead ECG acquired from a handheld smartphone. *J. Electrocardiol.* **50**, 620–625 (2017).
53. Shi, S. et al. Association of cardiac injury with mortality in hospitalized patients with COVID-19 in Wuhan, China. *JAMA Cardiol.* **5**, 802–810 (2020).
54. Bangalore, S. et al. ST-segment elevation in patients with Covid-19 — a case series. *N. Engl. J. Med.* **382**, 2478–2480 (2020).
55. Vollmer, S. et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *BMJ* **368**, l6927 (2020).

56. Yao, X. et al. ECG AI-Guided Screening for Low Ejection Fraction (EAGLE): rationale and design of a pragmatic cluster randomized trial. *Am. Heart J.* **219**, 31–36 (2020).

57. US National Library of Medicine. *ClinicalTrials.gov* https://clinicaltrials.gov/ct2/show/NCT04000087 (2020).

58. US National Library of Medicine. *ClinicalTrials.gov* https://clinicaltrials.gov/ct2/show/NCT04208971 (2020).

59. Attia, Z. I. et al. Age and sex estimation using artificial intelligence from standard 12-lead ECGs. *Circ. Arrhythm. Electrophysiol.* **12**, e007284 (2019).

60. Kashou, A. H. et al. Recurrent cryptogenic stroke: a potential role for an artificial intelligence-enabled electrocardiogram? *Heart Rhythm. Case Rep.* **6**, 202–205 (2020).

61. Siontis, K. C., Siontis, G. C., Contopoulos-Ioannidis, D. G. & Ioannidis, J. P. Diagnostic tests often fail to lead to changes in patient outcomes. *J. Clin. Epidemiol.* **67**, 612–621 (2014).

62. Price, W. N. 2nd & Cohen, I. G. Privacy in the age of medical big data. *Nat. Med.* **25**, 37–43 (2019).

63. Krittanawong, C. et al. Integrating blockchain technology with artificial intelligence for cardiovascular medicine. *Nat. Rev. Cardiol.* **17**, 1–3 (2020).

64. Kuo, T. T., Gabriel, R. A., Cidambi, K. R. & Ohno-Machado, L. EXpectation Propagation LOgistic REgRession on permissioned blockCHAIN (ExplorerChain): decentralized online healthcare/genomics predictive model learning. *J. Am. Med. Inform. Assoc.* **27**, 747–756 (2020).

65. Su, J., Vargas, D. V. & Kouichi, S. One pixel attack for fooling deep neural networks. *arXiv* https://arxiv.org/abs/1710.08864 (2017).

66. Noseworthy, P. A. et al. Assessing and mitigating bias in medical artificial intelligence: the effects of race and ethnicity on a deep learning model for ECG analysis. *Circ. Arrhythm. Electrophysiol.* **13**, e007988 (2020).

67. Raghunath, S. et al. Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network. *Nat. Med.* **26**, 886–891 (2020).

68. Chen, T. M., Huang, C. H., Shih, E. S. C., Hu, Y. F. & Hwang, M. J. Detection and classification of cardiac arrhythmias by a challenge-best deep learning neural network model. *iScience* **23**, 100886 (2020).

69. Feeny, A. K. et al. Machine learning of 12-lead QRS waveforms to identify cardiac resynchronization therapy patients with differential outcomes. *Circ. Arrhythm. Electrophysiol.* **13**, e008210 (2020).

70. Lopez-Jimenez, F. et al. Artificial intelligence in cardiology: present and future. *Mayo Clin. Proc.* **95**, 1015–1039 (2020).