

Fig. 3 | Framework for Al-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 ECGs8. 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 Watch¹⁴. Newer iterations of the Apple Watch now allow users to confirm the presence of AF using electrophysiological signals obtained through a single bipolar vector14.

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 outcomes⁵⁵. 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 trial⁵⁶. The EAGLE trial⁵⁷ 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

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 ago⁴, 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 abnormalities⁵⁻⁷. 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 use¹⁰. 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°. 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 diagnoses⁸. 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 patients¹¹. 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 performance¹². 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 strings¹². 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 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 CHA₂DS₂-VASc score, the most widely used stroke-prediction scheme in current practice36, had an AUC of 0.52 for stroke prediction. However, the highest AUC (0.63) was achieved when the CHA₂DS₂-VASc score was combined with the random forest and CNN models35, 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 phenotypespecific 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 groups⁴⁰. 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 abnormalities41, 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 screening⁴². 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 alone⁴⁵. 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 ECGs⁴⁵. 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, respectively⁴⁵. 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 myectomy45. 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 therapies⁴⁷. 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 cohort49. 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.

a Before myectomy AI–ECG probability of HCM 72.6%



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 OT 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 treatment⁵¹.

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

b After myectomy AI–ECG probability of HCM 2.5%



Fig. 2 | **The Al–ECG to detect HCM.** Use of an artificial intelligence-enhanced electrocardiogram (Al–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.

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 trial⁵⁷ 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 trial⁵⁸ 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 HCM⁴⁵. 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 $^{59}. \ \mbox{FIGURE}$ 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 analysis60. 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.

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

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

