GenECG: A synthetic image-based ECG dataset to augment artificial intelligence-enhanced algorithm development

Neil Bodagh[1], Kyaw Soe Tun[2], Adam Barton[3], Malihe Javidi[4], Darwon Rashid[4], Rachel Burns[4], Irum Kotadia[1], Magda Klis[1], Ali Gharaviri[4], Vinush Vigneswaran[4], Steven Niederer[1], Mark O'Neill[1], Miguel O Bernabeu*[4], Steven E Williams*[4].

[1] King's College London, London, United Kingdom

[2] Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom

[3] Neurolabs, Edinburgh, United Kingdom

[4] University of Edinburgh, Edinburgh, United Kingdom

*Equally contributing senior authors.

Corresponding author

Dr Neil Bodagh, St Thomas' Hospital Westminster Bridge Road, London, SE1 7EH, neil.bodagh@kcl.ac.uk, ORCiD: 0000-0001-5475-5298

Abstract

Artificial intelligence-enhanced electrocardiogram (AI-ECG) analysis has the potential to transform care of cardiovascular disease patients. Most algorithms rely on digitised signal data and are unable to analyse paper-based ECGs, which remain in use in numerous clinical settings. An image-based ECG dataset incorporating artefacts common to paper-based ECGs, which are typically scanned or photographed into electronic health records, could facilitate development of clinically useful image-based algorithms. This paper describes the creation of GenECG, a high-fidelity, synthetic image-based dataset containing 21,799 ECGs with artefacts encountered in routine care. Iterative clinical Turing tests confirmed the realism of the synthetic ECGs: expert observer accuracy of discrimination between real-world and synthetic ECGs fell from 63.9% (95% CI 58.0%-69.8%) to 53.3% (95% CI: 48.6%-58.1%) over three rounds of testing, indicating that observers could not distinguish between synthetic and real ECGs. GenECG is the first publicly available synthetic image-based ECG dataset to pass a clinical Turing test. The dataset will enable image-based AI-ECG algorithm development, ensuring the translation of AI-ECG research developments to the clinical workspace.

Introduction

The use of synthetic data in healthcare can facilitate the development of high-fidelity, fully anonymised patient datasets on a previously unachievable scale.[1] Synthetic patient data can be used as training data to develop artificial intelligence-enhanced (AI) algorithms offering the potential to revolutionise the scope and utility of AI within the healthcare setting. Several studies have already highlighted the potential benefits that AI algorithms may offer when applied to the electrocardiogram (ECG),[2] a low-cost, non-invasive, ubiquitous tool which provides important information about the electrical activity of the heart.[3] For example, AI-ECG can detect electrolyte imbalances,[4] identify left ventricular systolic dysfunction[5] and predict risks of paroxysmal arrhythmia[6] and all-cause mortality.[7] Cardiovascular diseases represent the leading causes of morbidity and mortality worldwide.[8] The ability of AI-ECG to facilitate automated ECG interpretation and detect patterns imperceptible to human observers[2] presents a significant opportunity to improve the care of patients with cardiovascular disease, ultimately contributing to a reduction in global cardiovascular disease burden.

However, despite the potential uses for AI-ECG, current algorithms are primarily limited to analyses of digitised ECG signals rather than ECG images. This is reflective of the composition of currently available public ECG datasets. Whilst multiple signal-based datasets exist,[9–11] the availability of image-based ECG data is limited. To our knowledge, only one publicly available image-based ECG dataset exists.[12] However, this dataset lacks the artefacts encountered with real-world paper-based ECGs, and it is substantially smaller than most digital ECG datasets. Nevertheless, numerous healthcare settings continue to rely on printed or scanned ECG images.[13,14] Given the ongoing, widespread use of paper-based ECGs, a disconnect exists between commonly available data types and the AI algorithms designed for their analysis. The creation of a dataset comprising ECG images could enable the development of image-based AI algorithms for use in scenarios where ECG signal data is unavailable. Such a dataset should capture the full range of diversity in paper-based ECGs incorporating artefacts common to clinical practice.

The aims of this study were to 1) create a large, publicly available image-based ECG dataset of labelled images from clinical ECG recordings; 2) modify the images to simulate the composition of a real-world image-based ECG dataset; 3) demonstrate the fidelity of the modified ECGs by testing the discrimination of synthetic from real-world images by healthcare professionals; and 4) evaluate the performance of pre-existing image-based AI

algorithms on synthetic images containing artefacts and compare their performance with artefact-free ECGs.

Results

ECG recreation

Synthetic ECG images were created for all 21,799 unique PTB-XL ECGs (Dataset A: ECG images without artefact) (Figure 1A, B). Methods to check ECG lead position confirmed that all ECG leads were plotted in the correct location according to the AHA/ACCF/HRS 'Recommendations for the Standardisation and Interpretation of the ECG' (Figure 1A and B).[15] There was a perfect correlation between measured and actual sine wave amplitude and frequency for each ECG lead.

Clinical validation: Turing tests

Image degradation techniques were applied to a randomly selected subset of synthetic ECGs for clinical validation via Turing tests. The results of the initial two rounds of Turing Tests indicated that healthcare professionals were able to distinguish real-world ECGs from synthetic images (Round one accuracy 63.9% (95% CI 58.0%-69.8%)), round two accuracy 59.8% (95% CI 55.9%-63.7%)) (Table 1). Qualitative feedback (summarised in Figure 2) was used to iteratively improve the fidelity of the ECG images (Figure 1C, D). In the third round of Turing tests, the Accuracy, True Recognition Rate and False Recognition Rate were 53.3% (95% CI 48.6%-58.1%), 53.0% (95% CI 48.7%-57.2%) and 53.7% (95% CI 47.4%-60.0%) respectively (Table 1). The Fleiss-Kappa score of 0.049 (95% Confidence interval 0.007-0.092) indicated a high degree of inter-observer variability. The area under the curve-receiver operating characteristic (AUC-ROC) curve score was 0.480 (95% CI 0.432-0.529) indicating that level of confidence was not an accurate predictor of an observer's ability to accurately distinguish a real-world ECG from a synthetic image (Figure 3). The response "Not at all confident" was the most frequently obtained level of confidence for both the real-world and synthetic ECG images. Further qualitative feedback was obtained with five observers commenting that it was difficult to distinguish real-world ECGs from synthetically created images.

Assessment on pre-existing image-based algorithms

Using the same architecture as an image-based model developed by Bridge *et al.,*[16] we trained a synthetic model with artefact-free images from an open access real-world dataset and then fine-tuned it on synthetic images. Initially, the synthetic model achieved an AUC score of 0.956 (95% CI 0.936-0.977) on an open access image-based dataset[12](Figure

4A). The initial model was less able to distinguish normal from abnormal on 43 synthetic images (AUC score 0.592, 95% CI 0.421-0.763) (Figure 4B). Fine-tuning of the model led to a substantial improvement in image classification (AUC score 0.945, 95% CI 0.876-1.000) (Figure 4C). The fine-tuned model subsequently achieved an AUC score of 0.896 (95% CI 0.864-0.928) on 400 real-world ECG images, down from 0.956 (95% CI 0.936-0.977), demonstrating that the model did not over-fit throughout the fine-tuning process (Figure 4D).

We assessed the performance of 75 abnormal ECGs both with and without artefact on the ECG-Dx© image-based algorithm.[13] The algorithm was able to correctly identify abnormal diagnoses for 51/75 (68%) artefact-free images compared with 29/75 (39%) images containing artefact (Figure 5).

Dataset creation and release

Upon completion of the 'Turing Tests', the artefact generation algorithm was deemed capable of recreating life-like ECGs from the PTB-XL database. A dataset of 21,799 images containing PTB-XL ECGs with the incorporation of artefacts common to photographed images was created (Dataset B: ECG images which appear photographed) (Figure 1C, D).

Discussion

In the present study, we have created a synthetic, image-based ECG dataset – GenECG – from a publicly available signal-based dataset. Clinical Turing tests confirmed the fidelity of the ECG images containing artefact with iterative development resulting in the creation of synthetic images that were indistinguishable from real-world ECGs. Pre-existing image-based AI algorithms exhibited good performance levels on synthetic images without artefacts, but poor performance on synthetic images with artefacts. Importantly, the accuracy of these algorithms could be improved through fine-tuning without overfitting on synthetic ECG images with artefacts. These findings highlight the potential for synthetic ECG data to augment clinically useful image-based AI-ECG algorithm development. GenECG will facilitate algorithm development on a diverse array of images containing artefacts and ensure real-world utility of image-based AI-ECG analysis. This presents a significant opportunity to capitalise on the benefits of synthetic patient data in healthcare and ensure the translation of AI-ECG analysis from the research setting to the clinical workspace.

The field of AI-ECG has grown considerably in recent years with signal-based algorithms now demonstrating diagnostic capabilities comparable to experienced clinicians.[17,18] Studies have also shown that AI-ECG may be able to detect patterns unrecognisable by the human eye thereby facilitating precise screening and prediction for disease, and the phenotyping of patients with pre-existing disease.[19] Whilst promising, the requirement for signal data to be inputted into these algorithms presents a barrier to use in a range of real-world clinical areas where signal data may not be readily available.[14] Such areas include hospital settings, where paper-based ECGs continue to be used, and non-hospital settings such as remote healthcare areas or pre-hospital settings. Several groups have attempted to develop ECG digitisation tools to address the issue.[20–22] Such tools have been designed to derive signal data from paper-based ECGs. Unfortunately, these tools are limited by the requirements for manual intervention to ensure correct ECG lead identification by the user, are unable to process large volumes of paper-based ECGs and often require users to individually input single ECGs one at a time.

Even if a reliable automatable ECG digitisation method was available, there is an absence of a universal format for the storage and exchange of digitised ECG data,[23] and a lack of interoperability amongst the array of ECG storage formats that do exist.[24] These factors further limit the integration of re-digitised ECG signal data into signal-based AI-ECG models. In contrast, paper-based ECGs have a standardised lead format which is widely recognised by several major international cardiac societies.[15] The GenECG dataset was created according to these specifications thereby overcoming issues pertaining to digitised ECG data format heterogeneity and obviating the requirement to digitise paper-based ECGs.

Whilst some image-based AI-ECG algorithms have been developed,[13,16] there is a relative absence of large ECG image-based datasets. Real-world image-based datasets have previously been created,[12] but are limited in comparison to signal-based datasets in their size and scope,[9–11] and lack the artefacts encountered with ECGs when they are encountered in clinical practice. In the present study, we have been able to utilise an existing signal-based dataset to create a vast image-based ECG dataset which is indistinguishable from real-world ECGs. In creating the ECG images from the PTB-XL database, GenECG benefits from the PTB-XL dataset’s broad scope and variety.[9]

To date, AI-ECG algorithms based on signal data have required ECG signal denoising to improve their accuracy and diagnostic yield.[25] In the present study, we have demonstrated the addition of simulated noise to ECGs images. In clinical practice,

artefacts are frequent and can result from patient-related issues such as movement, or non-patient related issues such as interference by devices.[26] Paper-based ECGs are typically photographed or scanned, and subsequently uploaded to electronic health records and this process also introduces artefacts. The incorporation of these artefacts into synthetic image-based ECG datasets is necessary to ensure that such repositories represent real-world clinical datasets as closely as possible. Within our study, the series of 'clinical Turing tests' ensured that the dataset that we have created is life-like with generated images incorporating the multitude of artefacts typically encountered with real-world ECGs. To our knowledge, this is the first study to examine the realism of synthetic ECG images using Turing testing.

In the third round of clinical Turing tests, the Accuracy, True Recognition Rate and False Recognition Rates were all in the region of 50% indicating an inability of observers to distinguish real-world ECGs from synthetically created images. Furthermore, the AUC-ROC score of 0.480 provided evidence to suggest that user certainty was unable to influence healthcare professionals' ability to distinguish real-world ECGs from synthetic images. The series of clinical Turing tests that were performed provided a robust validation method to confirm the realism of the synthetic ECG dataset. Whilst Turing tests have been used to validate other forms of synthetic medical imaging data,[27–30] our iterative methodology is novel as it integrates healthcare professionals' feedback to enhance the realism of the synthetic images. The methodology described in this study presents a framework that can be used by other disciplines to generate large, life-like synthetic patient datasets reducing the requirement to prospectively create new patient datasets.

The performance of pre-existing image-based algorithms on our GenECG data indicates poor generalisability of existing algorithms to real-world settings. In training and fine-tuning the model developed by Bridge *et al.*[16], we have demonstrated that utilising synthetic ECG data may overcome the limited efficacy of pre-existing image-based algorithms on real-world ECG images. Importantly, the performance of the fine-tuned model on real-world images has also demonstrated that over-fitting can be avoided. Our synthetic ECG images provide a large data repository which can be used to facilitate the development of AI-ECG algorithms on images containing artefacts. This will ensure the translation of AI-ECG analysis from the research setting to the clinical workplace.

Limitations

A limitation of our study is that ECGs were reconstructed from clinically recorded signal data. Whilst the clinical Turing tests confirmed the fidelity of the images created, whether

fully synthetic ECGs images can be created is not yet known. A further limitation is that the ECGs created are of a single layout type. Whilst several major cardiac societies have advocated a standardised simultaneous lead format for paper-based ECGs,[15] ECG layouts can vary in clinical practice. It is therefore crucial to ensure that future image-based datasets capture this variability such that image-based artificial intelligence-enhanced algorithms can be applied to ECGs regardless of the ECG layout encountered.

In the present study, we have shown how a pre-existing ECG signal-based dataset can be utilised to synthetically create a high-fidelity ECG image-based dataset. Our diverse dataset will enable the development of image-based AI-ECG algorithms offering the potential to bridge the gap that exists between AI-ECG research and current clinical practice. On a wider scale, our study presents a framework which other disciplines can utilise to synthetically generate large, life-like patient datasets. The use of synthetic patient data to augment AI algorithm development presents an exciting opportunity to enhance the use of AI in the healthcare setting.

Methods

PTB-XL Dataset

Input ECG signals were provided by the PTB-XL database which contains signal data representing 21,799 clinical ECGs from 18,869 patients, stored in Wave Form Database Format (WFDB).[9] The PTB-XL ECGs are configured as 12 channel binary files with a resolution of 1μv / LSB at 500Hz (each sample is 0.002 sec). Annotated by two cardiologists, there are a total of 71 different ECG statements observed throughout the dataset. The statements cover form, rhythm, and diagnostic labels in a machine-readable form. The diagnostic labels are organised into five superclasses and 24 subclasses as described in [9].

ECG recreation

For each PTB-XL ECG, an ECG image was created according to recommendations outlined in the ‘AHA/ACCF/HRS Recommendations for the Standardization and Interpretation of the Electrocardiogram’ document,[15] comprising a continuous ten second recording divided into three rows and four columns consisting of 2.5 seconds of data for each lead where column one represents leads I, II and II; column two represents aVR, aVL and aVF; column three represents V1, V2 and V3 and column four represents V4, V5 and V6. An additional rhythm strip containing 10 seconds of data (lead II) was included for rhythm analysis.

The Blender (Blender Foundation, Amsterdam, the Netherlands) software platform was used to create synthetic ECG images using custom code (developed by AB). ECG images were recreated by sampling 2.5s epochs for each lead, positioned according to the AHA/ACCF/HRS recommendations[15] and with lead markers, lead labels and calibration scales added. Resulting waveform traces were superimposed onto a paper grid (with a resolution set to 150Hz (25mm/s) horizontally and 10.0mm/mV vertically) leading to the generation of a single waveform layout for each PTB-XL ECG. The resolution of the waveform image was set to 5 pixels/mm with a final image output size of 1397 x 1029 pixels for a 10 second trace.

To validate the accuracy of initial ECG images created from signal data, ECG files representing sine waves of known amplitude (0mV) and frequency (1.25Hz), one per ECG lead, were created using MATLAB and written to file using the WFDB Toolbox for MATLAB/Octave.[31] A total of 12 test ECGs were created, consisting of ECGs with the sine wave at a single ECG lead location with all other leads set to a constant electrical potential of 0mV. These files were converted into ECG images using the same code used for ECG recreation. All validation ECGs were manually inspected to confirm the correct location of the ECG leads. For each lead of each simulated ECG containing a sine wave, the amplitude and cycle length (frequency) were measured by an observer (NB) blinded to the actual amplitude and frequency.

Creation of synthetic ECG images

To add realistic artefacts to ECG images (i.e. to make it appear as though ECG images had been photographed), ECG images were passed to a second render which placed each ECG image trace on a 3D model comprising a paper sheet positioned in a synthetically developed workspace. In total, 352 unique geometric variations were created from eight paper sheet variations, eleven workspaces and four synthetic workspace orientations. The Blender platform's bpy module was used to create an automated Python script for ECG image generation. For each ECG, a mesh and synthetic workspace were randomly selected, and the location and rotation of the ECG paper sheet, camera, and light sources were randomly adjusted. Varying degrees of stucci noise were subsequently applied to enhance the realism of the rendered images.

Clinical Turing Tests

A series of visual Turing tests were designed using previously published healthcare research utilising image-based Turing testing[27–30] and conducted to assess the fidelity of synthetic ECG images via an online survey (Qualtrics, Provo, UT). In all rounds of

Turing tests, healthcare professionals were provided with a series of 60 images comprising 30 synthetically created ECGs and 30 photographs of real-world ECGs. ECG images were redacted in areas where text may appear. Images were displayed one-by-one to participants and shown in uniform order. Participants were asked to select whether they thought the images were real or synthetic and, in the second and third rounds, to rate their confidence using a five-point Likert scale (Figure 6). At the end of each survey, healthcare professionals were asked to provide qualitative feedback through a series of open questions. Feedback was summarised and used to iteratively improve the dataset's fidelity. All readers decided whether each image was real or synthetic without any time limit and no prior knowledge regarding the number of real or synthetic images. To avoid bias, healthcare professionals were only allowed to complete one round of clinical Turing tests.

Assessment on pre-existing image-based algorithms

To examine the performance of currently available image-based algorithms on the GenECG dataset, synthetically created images were inputted into two image-based AI-ECG algorithms.[13,16]

Bridge *et al.,* have previously developed an image-based algorithm capable of distinguishing 'normal' ECGs from 'abnormal' ECGs, and this has demonstrated good performance on scanned ECG printouts.[16] The dataset used to train the model was unavailable. We therefore trained the model with 1682 artefact-free images from an open access real-world, image-based dataset (split into train (n=1082), validation (n=200) and test datasets (n=400)).[12] The trained model was applied to evaluate its performance on 43 synthetic ECG images containing image degradation techniques. Following the initial results, the model had low efficiency on our synthetic images since it was not exposed to images that resembled our images during training. Therefore, the model was fine-tuned on 215 synthetic images (split into train (n=150), validation (n=22) and test (n=43) images), while the weights of the model trained on the open access dataset were used as initial weights for this new model. To assess the generalisation power of the synthetic-model and to ensure that the model did not over-fit throughout the fine-tuning process, we evaluate the trained synthetic-model over the open access test dataset (n=400). AUC-ROC analysis was performed to evaluate the performance of the models.

ECG Dx© (https://www.cards-lab.org/ecgdx) is a publicly available automated diagnostic algorithm capable of detecting six diagnoses (atrial fibrillation, sinus tachycardia, sinus bradycardia, left bundle branch block, right bundle branch block and first-degree

atrioventricular block). We selected 75 abnormal ECGs from the PTB-XL dataset. Images with and without image degradation techniques were inputted into the web-based platform. The corresponding classifiers were compared with labels assigned from the PTB-XL dataset.

Statistical Analysis

For each round of Turing Tests, we measured the Accuracy (overall proportion of ECGs correctly identified as 'real-world' or 'synthetic'), True Recognition Rate (proportion of real-world ECGs identified correctly) and False Recognition Rate (proportion of synthetic ECGs identified correctly) using adapted terminology from previous Turing tests used in healthcare research.[28,30] The Fleiss-Kappa score was calculated to evaluate the degree of inter-observer agreement. For the second and third rounds of clinical Turing tests, confidence Likert scale scores were converted to a signed ordinal scale for AUC-ROC score analysis. The data were analysed using SPSS version 29 (IBM Corp., Amonk, NY).

Data availability statement

The ECG images described in the study were created from the PTB-XL database.[9] The ECG images will be used for a British Heart Foundation Data Science Centre open challenge (https://bhfdatasciencecentre.org/areas/unstructured-data/imaging-open-challenge). Following this challenge, Dataset A and B will be made publicly available via a Creative Commons license.

Code Availability statement

The code used to facilitate batch image generation will be made publicly available via a Creative Commons license.

Figure 1. Synthetic ECG images recreated from the PTB-XL dataset. Panels A and B represent artefact-free images. Panels C and D represent the same images following the application of image degradation techniques to make it appear as though the images have been photographed. Panels A and Panels C have been recreated from 00074_hr_1R.dat. Panels B and D have been recreated from 00067_hr_1R.dat.

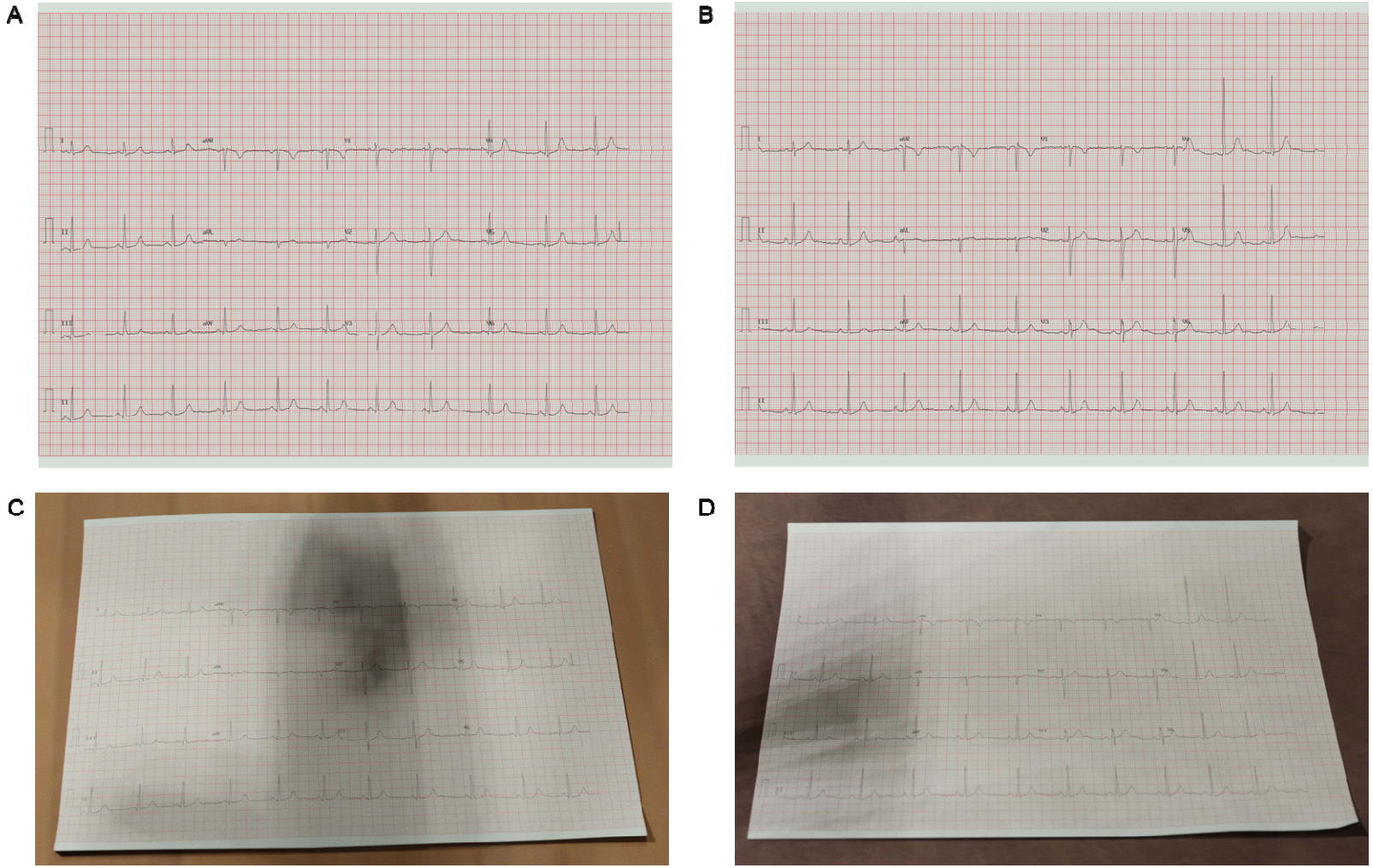


Figure 2. Example of a synthetic ECG used in the initial Turing test with a summary of the qualitative feedback provided by healthcare professionals.

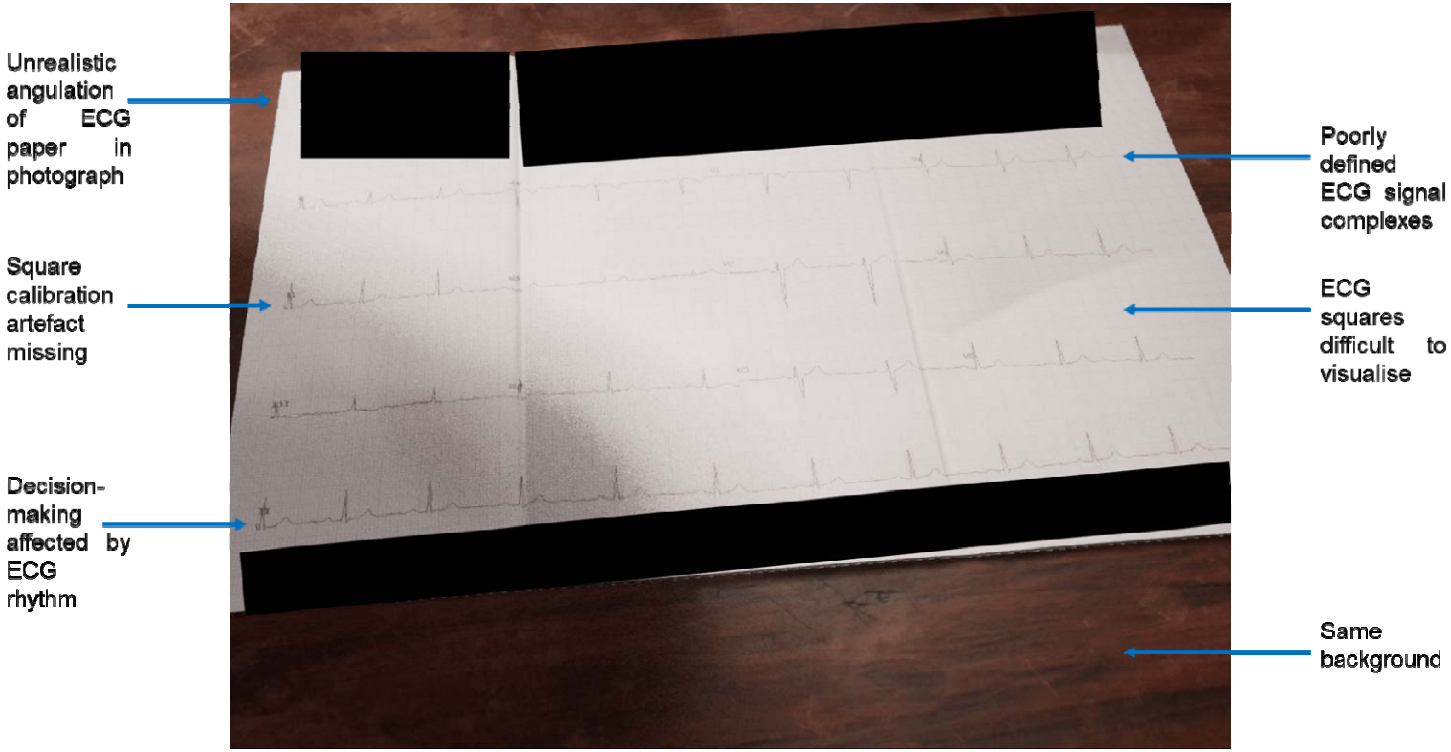
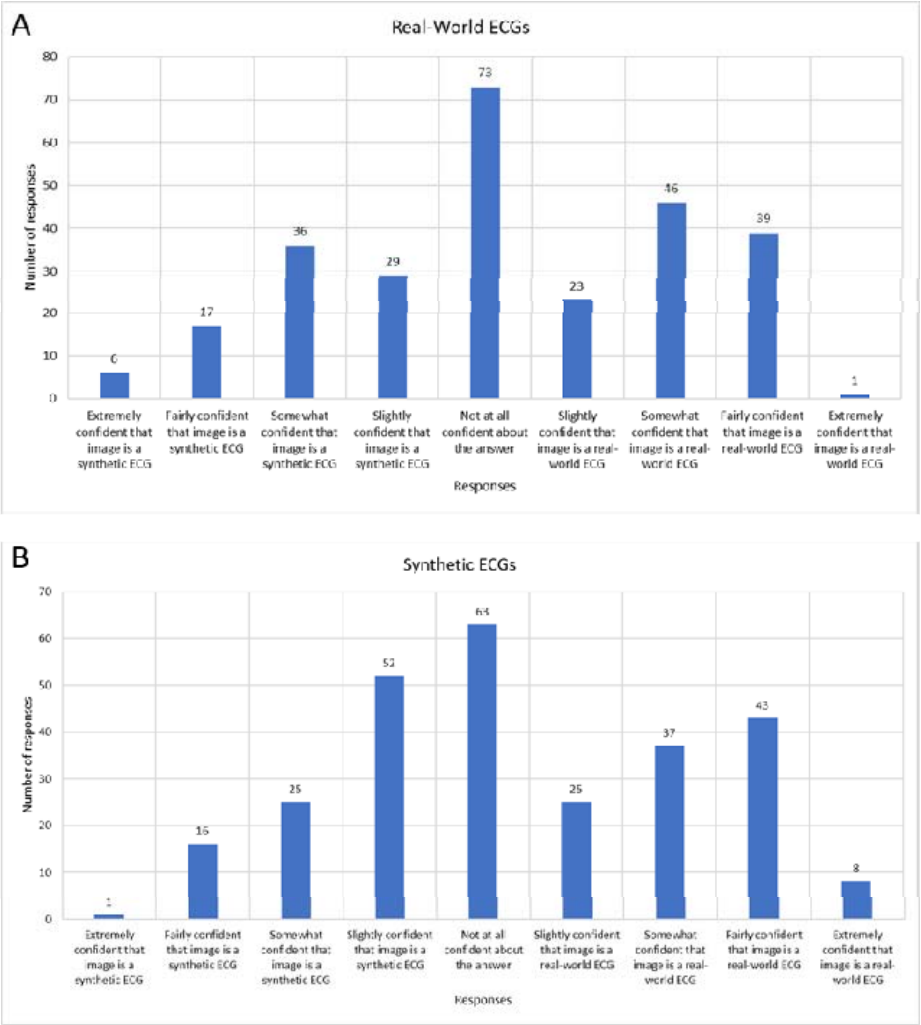


Figure 3. Confidence levels in ECG classification for Turing Test round three. Panels (A) and (B) represent the number of responses for each option for (A) real-world ECGs and (B) synthetically created ECGs. Panel C shows an Area under the curve-receiver operating characteristic (AUC-ROC) curve examining the impact of confidence level on ability to correctly identify an ECG image as 'real-world' or 'synthetically created.'



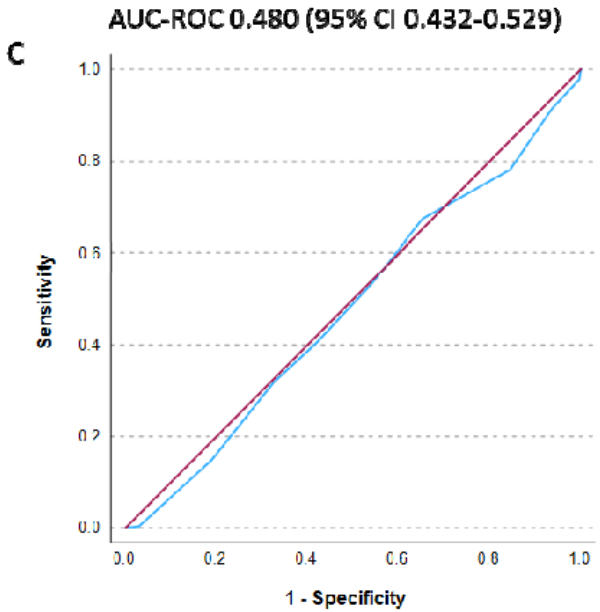


Figure 4. Area under the curve-receiver operator characteristic score curves for ECG images on the algorithm initially developed by Bridge et al., [16]. (A) represents the performance of the model on 400 test images obtained from an artefact-free real-world image-based dataset [12], (B) demonstrates the performance of the trained model on 43 test synthetic ECG images containing artefact, (C) demonstrates the performance of the trained model on the same synthetic ECG images following fine-tuning, (D) shows the performance of the fine-tuned model on the 400 test images from the artefact-free real-world image-based dataset.

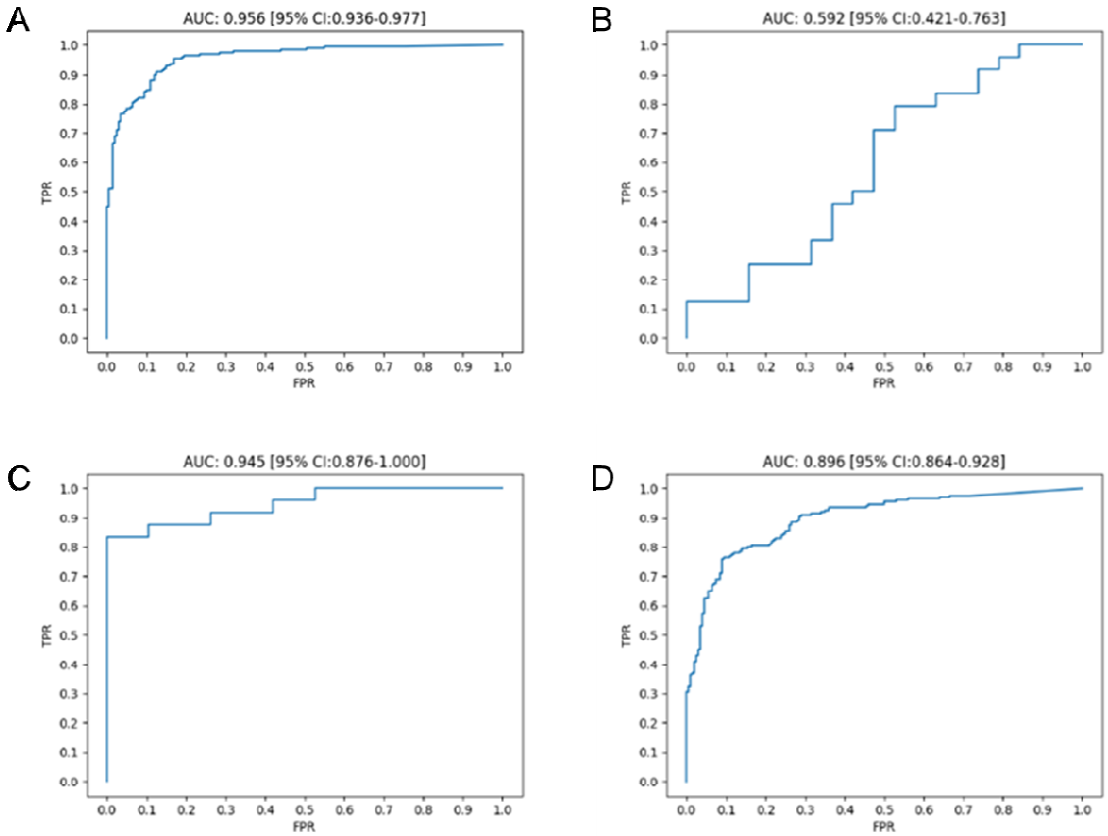


Figure 5. Performance of ECG Dx© image-based algorithm on abnormal images obtained from the PTB-XL dataset. Diagnoses along the Y axis represent PTB-XL labelled diagnoses.

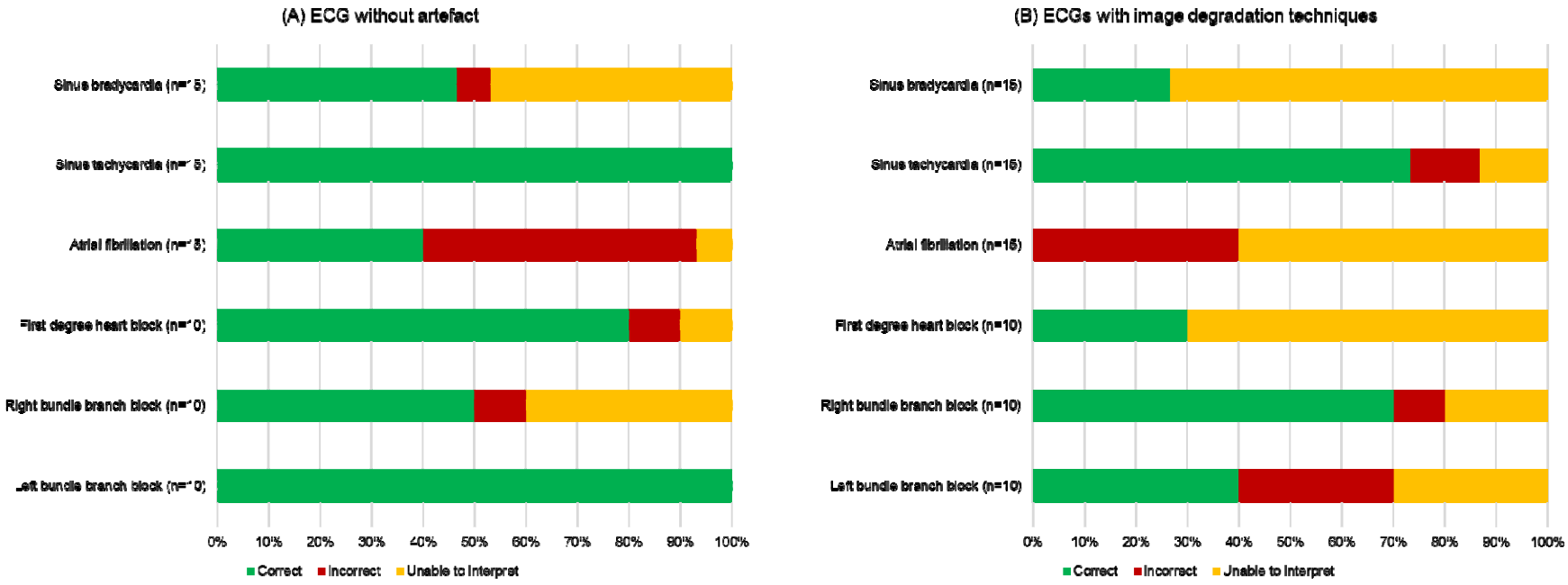


Figure 6. Clinical Turing Tests. ECG images were individually displayed on pages and observers asked to select whether they thought each ECG image was a real-world ECG or synthetically created. For the second round and third rounds, observers were asked to indicate their levels of confidence in their answers using a five-point Likert scale. Figures (A) and (B) demonstrate two separate questions showing examples of (A) a real world-ECG and (B) a synthetically created ECG. Images were redacted in areas where text may appear.

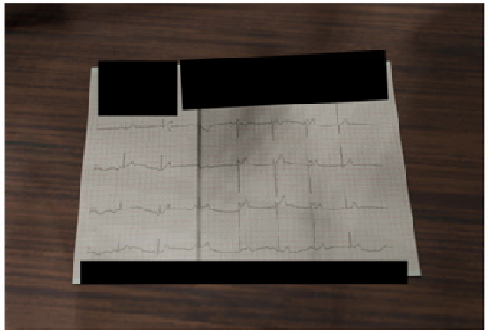
A Please state whether you think this ECG is a real-world ECG or synthetically created

- Real-world ECG
- Synthetically created ECG

Please indicate your level of confidence in the answer that you have provided for this ECG

- Not at all confident
- Slightly confident
- Somewhat confident
- Fairly confident
- Extremely confident

B Please state whether you think this ECG is a real-world ECG or synthetically created



- Real-world ECG
- Synthetically created ECG

Please indicate your level of confidence in the answer that you have provided for this ECG

- Not at all confident
- Slightly confident
- Somewhat confident
- Fairly confident
- Extremely confident

Table 1. Summarised results of iterative Turing tests. CI = Confidence interval. AUC-ROC = Area under the curve-receiver operating characteristic.

	Round One	Round Two	Round Three
Number of participants	9	8	9
Job role breakdown	2 Consultants 5 Registrars 1 Senior House Officer 1 Nurse	1 Consultant 3 Registrars 2 Senior House Officers 1 Physiologist 1 Nurse	2 Consultants 2 Registrars 3 Senior House Officers 1 Physiologist 1 Advanced Clinical Practitioner
Total number of electrocardiograms correctly identified	345/540	287/480	288/540
Accuracy	63.9% (95% CI 58.0%-69.8%)	59.8% (95% CI 55.9%-63.7%)	53.3% (95% CI 48.6%-58.1%)
Real world electrocardiograms correctly identified	179/270	148/240	143/270
True Recognition Rate	66.3% (95% CI 60.4%-72.2%)	61.7% (95% CI 55.3%-68.1%)	53.0% (95% CI 48.7%-57.2%)
Synthetic electrocardiograms correctly identified	166/270	139/240	145/270
False Recognition Rate	61.5% (95% CI 54.7%-68.3%)	57.9% (95% CI 51.4%-64.5%)	53.7% (95% CI 47.4%-60.0%)
Fleiss-Kappa Score	0.045 (95% CI 0.003-0.087)	0.003 (95% CI -0.044-0.051)	0.049 (95% CI 0.007-0.092)
AUC-ROC score	Not performed	0.565 (95% CI 0.514-0.616)	0.480 (95% CI 0.432-0.529)

Author Contributions

NB, MOB and SEW conceived the concept. NB, MJ, MOB and SEW wrote the paper. MK, AG, SN and MON provided intellectual input into the draft of the manuscript. AB, DR, and RB created the synthetic ECG images. NB, IK, and KT designed and conducted the Turing tests. NB, KT, MJ, and MOB examined the performance of the synthetic ECG images on pre-existing image-based algorithms. NB and VV collected and organised the dataset. MOB and SEW are equally contributing senior authors. All authors read, edited, and approved the final version.

Competing Interests

No competing interests to declare.

Acknowledgements

The study was funded by a University of Edinburgh Wellcome Trust iTPA award. The authors acknowledge the support of the British Heart Foundation Centre for Research Excellence Award III (RE/18/5/34216). The authors acknowledge the support of the British Heart Foundation (RG/20/4/34803). SEW is supported by the British Heart Foundation (FS/20/26/34952). IK is supported by the British Heart Foundation (FS/CRTF/21/24166).

References

1. Gonzalesid A, Guruswamy G, Smith SR. Synthetic data in health care: A narrative review. *PLOS Digital Health* 2023;2:e0000082.

2. Attia ZI, Harmon DM, Behr ER *et al.* Application of artificial intelligence to the electrocardiogram. *Eur Heart J* 2021;42:4717–30.

3. AlGhatrif M, Lindsay J. A brief review: history to understand fundamentals of electrocardiography. *J Community Hosp Intern Med Perspect* 2012;2:14383.

4. Kwon J myoung, Jung MS, Kim KH *et al.* Artificial intelligence for detecting electrolyte imbalance using electrocardiography. *Ann Noninvasive Electrocardiol* 2021;26, DOI: 10.1111/ANEC.12839.

5. Attia ZI, Kapa S, Lopez-Jimenez F *et al.* Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. *Nat Med* 2019;25:70–4.

6. Attia ZI, Noseworthy PA, Lopez-Jimenez F *et al.* An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet* 2019;394:861–7.

7. Raghunath S, Ulloa Cerna AE, Jing L *et al.* Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network. *Nature Medicine 2020 26:6* 2020;26:886–91.

8. Roth GA, Mensah GA, Johnson CO *et al.* Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019: Update From the GBD 2019 Study. *J Am Coll Cardiol* 2020;76:2982–3021.

9. Wagner P, Strodthoff N, Bousseljot RD *et al.* PTB-XL, a large publicly available electrocardiography dataset. *Sci Data* 2020;7:1–15.

10. Moody GB, Mark RG. The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine* 2001;20:45–50.

11. Taddei A, Distante G, Emdin M *et al.* The European ST-T database: standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography. *Eur Heart J* 1992;13:1164–72.

12. Khan AH, Hussain M, Malik MK. ECG Images dataset of Cardiac and COVID-19 Patients. *Data Brief* 2021;34, DOI: 10.1016/J.DIB.2021.106762.

13. Sangha V, Mortazavi BJ, Haimovich AD *et al.* Automated multilabel diagnosis on electrocardiographic images and signals. *Nat Commun* 2022;13, DOI: 10.1038/S41467-022-29153-3.

14. Bodagh N, Ali O, Kotadia I *et al.* Feasibility of artificial intelligence-enhanced electrocardiogram (AI-ECG) analysis in the current clinical environment: An online survey. *EP Europace* 2023;25, DOI: 10.1093/EUROPACE/EUAD122.533.

15. AHA/ACCF/HRS Scientific Statement: Recommendations for the Standardization and Interpretation of the Electrocardiogram: Part I: The Electrocardiogram and Its Technology | Heart Rhythm Society.

16. Bridge J, Fu L, Lin W *et al.* Artificial intelligence to detect abnormal heart rhythm from scanned electrocardiogram tracings. *J Arrhythm* 2022;38:425–31.

17. Ribeiro AH, Ribeiro MH, Paixão GMM *et al.* Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat Commun* 2020;11, DOI: 10.1038/S41467-020-15432-4.

18. Hannun AY, Rajpurkar P, Haghpanahi M *et al.* Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med* 2019;25:65–9.

19. Sau A, Ng FS. The emerging role of artificial intelligence enabled electrocardiograms in healthcare. *BMJ Medicine* 2023;2:e000193.

20. Ravichandran L, Harless C, Shah AJ *et al.* Novel Tool for Complete Digitization of Paper Electrocardiography Data. *IEEE J Transl Eng Health Med* 2013;1:1800107–1800107.

21. Baydoun M, Safatly L, Hassan OKA *et al.* High Precision Digitization of Paper-Based ECG Records: A Step Toward Machine Learning. *IEEE J Transl Eng Health Med* 2019;7, DOI: 10.1109/JTEHM.2019.2949784.

22. Wu H, Patel KHK, Li X *et al.* A fully-automated paper ECG digitisation algorithm using deep learning. *Scientific Reports 2022 12:1* 2022;12:1–12.

23. Badilini F, Young B, Brown B *et al.* Archiving and exchange of digital ECGs: A review of existing data formats. *J Electrocardiol* 2018, DOI: 10.1016/j.jelectrocard.2018.07.028.

24. Bond RR, Finlay DD, Nugent CD *et al.* A review of ECG storage formats. *Int J Med Inform* 2011;80:681–97.

25. Chatterjee S, Thakur RS, Yadav RN *et al.* Review of noise removal techniques in ECG signals. *IET Signal Processing* 2020;14:569–90.

26. Pérez-Riera AR, Barbosa-Barros R, Daminello-Raimundo R *et al.* Main artifacts in electrocardiography. *Ann Noninvasive Electrocardiol* 2018;23:23.

27. Shin K, Lee JS, Lee JY *et al.* An Image Turing Test on Realistic Gastroscopy Images Generated by Using the Progressive Growing of Generative Adversarial Networks. *J Digit Imaging* 2023;36:1760–9.

28. Veturi YA, Woof W, Lazebnik T *et al.* SynthEye: Investigating the Impact of Synthetic Data on Artificial Intelligence-assisted Gene Diagnosis of Inherited Retinal Disease. *Ophthalmology Science* 2023;3:100258.

29. Jang M, Bae H jin, Kim M *et al.* Image Turing test and its applications on synthetic chest radiographs by using the progressive growing generative adversarial network. *Sci Rep* 2023;13, DOI: 10.1038/S41598-023-28175-1.

30. Chuquicusma MJM, Hussein S, Burt J *et al.* How to Fool Radiologists with Generative Adversarial Networks? A Visual Turing Test for Lung Cancer Diagnosis. *Proceedings - International Symposium on Biomedical Imaging* 2017;2018-April:240–4.

31. Waveform Database Software Package (WFDB) for MATLAB and Octave v0.10.0.