ESC European Society of Cardiology

European Heart Journal - Digital Health (2024) **5**, 314–323
https://doi.org/10.1093/ehjdh/ztae024

ORIGINAL ARTICLE
Artificial intelligence (machine learning, deep learning)

Unlocking the potential of artificial intelligence in electrocardiogram biometrics: age-related changes, anomaly detection, and data authenticity in mobile health platforms

Kathryn E. Mangold [1], Rickey E. Carter[2], Konstantinos C. Siontis[1], Peter A. Noseworthy[1], Francisco Lopez-Jimenez [1], Samuel J. Asirvatham[1], Paul A. Friedman[1], and Zachi I. Attia [1,*]

[1]Department of Cardiology, Mayo Clinic, 200 1st Street SW, Rochester, MN 55905, USA; and [2]Department of Quantitative Health Sciences, Mayo Clinic, 4500 San Pablo Road, Jacksonville, FL 32224, USA

Received 12 September 2023; revised 5 March 2024; accepted 12 March 2024; online publish-ahead-of-print 23 April 2024

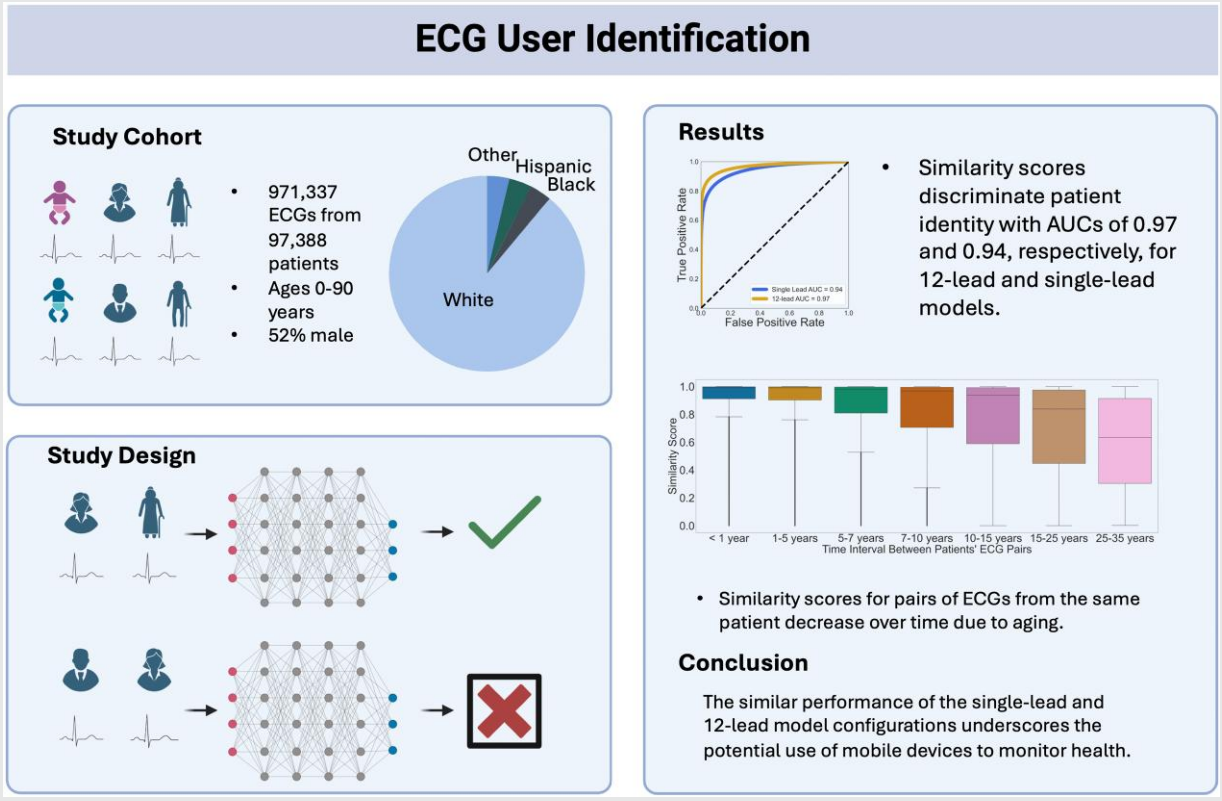
Aims Mobile devices such as smartphones and watches can now record single-lead electrocardiograms (ECGs), making wearables a potential screening tool for cardiac and wellness monitoring outside of healthcare settings. Because friends and family often share their smart phones and devices, confirmation that a sample is from a given patient is important before it is added to the electronic health record.

Methods and results We sought to determine whether the application of Siamese neural network would permit the diagnostic ECG sample to serve as both a medical test and biometric identifier. When using similarity scores to discriminate whether a pair of ECGs came from the same patient or different patients, inputs of single-lead and 12-lead medians produced an area under the curve of 0.94 and 0.97, respectively.

Conclusion The similar performance of the single-lead and 12-lead configurations underscores the potential use of mobile devices to monitor cardiac health.

* Corresponding author. Tel: +1 507 422 0699, Email: attia.itzhak@mayo.edu

Graphical Abstract



Keywords ECG • Siamese neural networks • Biometric • Aging

Introduction

Artificial intelligence (AI) algorithms based on the 12-lead electrocardiogram (ECG) can predict cardiac disease[1–4] and patient age and sex.[5] Smartwatches and smartphones able to record a single-lead ECG have emerged as powerful tools to track cardiac wellness outside of the clinic.[6,7] Twelve-lead algorithms have been adapted for use with single-lead inputs with success.[8,9] One study invited Mayo Clinic patients to transmit watch ECGs that were then incorporated into their electronic health record (EHR). Because smartwatches can be shared among family and friends, user identification is necessary to ascertain that ECGs belong to the expected patient. We hypothesized that given an individual's baseline ECG, a neural network could quantify the likelihood that subsequent recordings belong to the same individual.

Siamese neural networks are a deep learning architecture that excel in comparing multiple inputs.[10] The network receives paired inputs and sends the pair to twin encoding structures to extract features. These features are compared and then assigned a similarity score. Siamese networks have been used outside healthcare for user identification[11,12] and in healthcare for the detection and monitoring of Alzheimer's disease.[13,14] They also have been used in monitoring disease progression with various imaging modalities.[15]

Numerous previous studies have investigated ECG biometric feature extraction using a variety of methods from frequency transforms to convolutional neural networks.[16,17] Most applicable, however, are studies using Siamese networks trained on public ECG data sets[17–21] with at most 549 patients to classify individuals from their ECGs.

In this study, we sought to test the ability of a Siamese network to identify patients based on ECG biometrics within a large data set. To the best of our knowledge, this is the largest cohort of patients studied in an ECG biometric study. Additionally, given that an individual's ECG evolves with aging, we compared individuals' longitudinal ECG changes to assess the impact on identification and the potential role of the Siamese network approach for assessing biological age.

Methods

Data set characteristics

The data set consisted of 971 337 ECGs from 97 388 distinct, randomly selected patients who had provided research authorization and whose data were in the MUSE data management system. The ECG signal processing included standard baseline artefact removal and noise reduction prior to storage in MUSE. The patients were 53% male with a minimum and maximum age of 0 and 90 years, respectively, at the time of ECG recording. The race and ethnicity of the patients was 90% non-Hispanic White, 4% 'Other', 3% Hispanic or Latino, and 3% Black. The average minimum age was 41.5 ± 25.2 years, and the average maximum age was 52.0 ± 29.5 years. There was an average of 10.0 ± 12.7 ECGs on record per patient with an average interval of 10.4 ± 9.8 years between the first and last recording. Digital 10-s ECGs were collected between May 1987 and January 2023 at Mayo Clinic at a

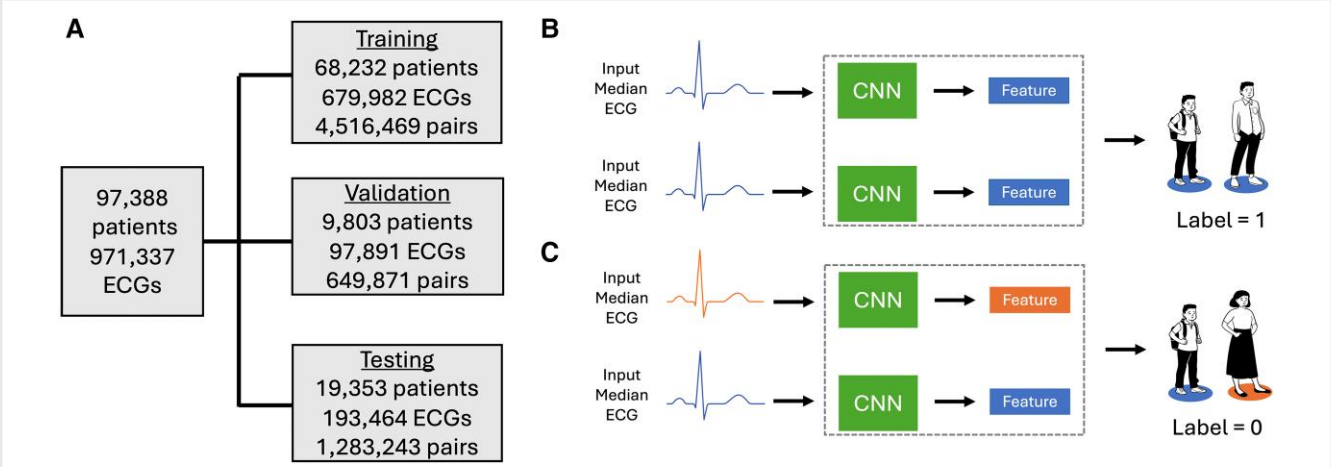


Figure 1 Overview of the study. (*A*) Splitting of the original data set by patient in a ratio of 7:1:2. (*B*) Given a pair of two input electrocardiograms from the same patient labelled one, the Siamese network is trained to extract features from both inputs simultaneously, compare these features, and finally generate a high similarity score. (*C*) Conversely, given a pair of two electrocardiograms originating from two different patients (labelled zero), the network is trained to extract features that when compared produce a low similarity score.

sampling frequency of 500 or 250 Hz and were resampled to 500 Hz on Marquette ECG Machine. Median beats were created by the MUSE system (GE Healthcare) and represent the ECG morphology.

Siamese model training

The data set was first grouped by patient and then split into training, validation, and testing sets with a ratio of 7:1:2, respectively (*Figure 1A*). Data splitting was performed by patient identification number, and all ECGs from a given patient were in the same data set to avoid data leakage. *Table 1* displays age, sex, race/ethnicity, and cardiac rhythm abnormalities for the training, validation, and testing sets. Initial grouping by patient allowed the network to be tested on unique patients not seen during training or validation. Electrocardiograms in the training, validation, and testing sets were then paired within the patient set so that each pair consisted of two distinct ECGs: both from the same patient or one each from different patients. Labels for distinct ECGs from the same patient were designated as one while labels of zero (*Figure 1B*) were reserved for distinct ECGs stemming from different patients (*Figure 1C*). Given memory constraints, only the first 100 pairs within a patient were included (label one). To keep the label balance approximately equal, the number of desired pairs including a specific patient paired with another (label zero) was proportional to the number of his or her ECG pairs (label one). When forming ECG pairs between distinct patients, because age and sex are known to affect the ECG,[5] patients were only randomly paired with other patients in the same age (by decade) and sex category. Pairing distinct patient ECGs only within the same age decade and sex category encourages the model to focus on learning the differences between patient biometrics rather than just the differences in age and sex. In total, 53% of ECG pairs were within a patient's history while 47% were pairs of ECGs from different patients. There were 4 516 469 (training), 649 871 (validation), and 1 283 243 (testing) pairs. Most of the testing ECG pairs were in sinus rhythm (76%), followed by one ECG in sinus rhythm and one ECG in atrial fibrillation (sinus/atrial fibrillation, 11%) and sinus/other (8%). There were variations in the ECG pair rhythm breakdowns when stratifying by same and distinct patient pairs across training, validation, and testing data sets (see Supplementary material online, *Figure S1*).

Electrocardiogram pairs were first fed into the same (Siamese) residual neural network portion to extract features. To test whether beat-to-beat variation affected classification of each pair of ECGs, four Siamese neural networks were created that each have a pair of ECGs as inputs (*Figure 2A*). The first two networks had inputs of 1.2-s medians for a single lead (lead I, matrix of 600 × 1) and 12 leads (matrix of 600 × 12). The other two networks had inputs of 2.5 s of an ECG for a single lead (lead I, matrix of 1250 × 1) and 12 leads (matrix of 1250 × 12). All these Siamese neural networks were implemented in TensorFlow 2.4.x (Google) and Python 3.9.x. Each feature extraction branch of the twin networks (*Figure 2A*) consisted of convolutional layers, batch normalization, residual connections, leaky RELU, and max pooling layers with filter numbers, kernel sizes, and pooling sizes listed in *Figure 2B–D*. The extracted features of size 256 were then compared with an L1 norm (implemented with the tf.keras.layer.Lambda layer) followed by a fully connected layer of size 128, dropout of 0.5, leaky RELU, and a single dense output with sigmoid activation (*Figure 2A*). This final network output represents the probability that the two ECGs are from the same individual. The model was trained with the Adam optimizer with binary cross entropy as the loss function and a learning rate of 0.001 with a batch size of 512. The area under the curve (AUC) under the receiver operating characteristic (ROC) curve was monitored for early stopping to avoid overfitting. Accuracy metrics were calculated based on the optimal threshold, which we defined as the closest point to (0, 1) on the validation ROC curve.[22]

Testing of the model

The model was first tested with patients not seen during training or validation. Positive and negative pairs from these unseen patients were generated in the same way as during training and validation. The pairs were then used in a simulated scenario of user identification in which ECGs acquired over a span of years were analysed. Because the ECG is affected by the natural aging process[23–25] like other biometrics such as fingerprint scans or iris imaging,[26,27] the similarity scores may drop over time. In other words, two ECG records from a patient acquired years apart could have lower similarity score than two records collected closer in time. Because of this expected drop in similarity score, the model may not recognize a patient who submits a new ECG to the medical record after a long-time interval if a strict acceptance threshold is used. Rather than simply using the first ECG as a baseline, we tested systematically larger subsequent sets of ECG records to create an evolving baseline. This way the model could successfully verify patients who have ECG records spanning decades, even using high thresholds for identity. An evolving baseline mirrors a classification[26–28] process in which each new record is systematically compared to a specific subset of previous ECGs designated as the baseline: A new ECG record is compared to each member of the baseline, and the final similarity score is an average of these similarity scores. By averaging the scores across the members of the evolving baseline, the model can better approximate the patient's evolving ECG baseline for successful user verification.

Table 1 Electrocardiogram data set demographics and rhythms overview

	Training (n = 679 982 ECGs)	Validation (n = 97 891 ECGs)	Testing (n = 193 464 ECGs)
Age	59.5 ± 22.4	59.6 ± 22.3	59.4 ± 22.35
% male	52.9%	52.9%	52.4%
Race and ethnicity			
Non-Hispanic White	609 786 (90%)	88 062 (90%)	172 867 (89%)
Other	27 701 (4%)	4081 (4%)	7528 (4%)
Hispanic or Latino	22 423 (3%)	3076 (3%)	6967 (4%)
Black	20 072 (3%)	2672 (3%)	6102 (3%)
Rhythms			
Infarction present	94 105 (14%)	12 749 (13%)	26 582 (13%)
Left bundle branch abnormality	20 817 (3%)	2591 (3%)	5844 (3%)
Low voltage	25 133 (4%)	3439 (4%)	7199 (4%)
Long QT	43 376 (6%)	6387 (7%)	12 682 (7%)
Right bundle branch abnormality	62 112 (9%)	8558 (9%)	17 344 (9%)
Left ventricular hypertrophy	23 805 (4%)	3058 (3%)	6759 (3%)
Atrial flutter/fibrillation	63 057 (10%)	9520 (10%)	17 920 (9%)
First-degree AV block	72 470 (11%)	10 741 (11%)	20 946 (11%)
Second-degree AV block	12 008 (2%)	1770 (2%)	3382 (2%)

Statistical considerations

Reported AUCs are not customary because patient ECG pairing introduces clustering that violates the independence assumption of the Mann–Whitney *U* statistic.[29] Area under the curves are reported without confidence intervals because measures of uncertainty are not likely appropriate for this paired data set. Differences in distributions were deemed statistically significant with the Kruskal–Wallis test. Otherwise, measures are reported as means and standard deviation.

Results

Single-lead and 12-lead median Siamese network performance

The 97 388 patients and 971 337 ECGs were split into training, validation, and testing patient wise so that a patient's ECGs are restricted to one set as described in the Methods. Similarity scores discriminated whether pairs of ECGs came from the same patient or different patients with an AUC of 0.94 and 0.97, respectively, for the single-lead and 12-lead models with median ECGs (*Figure 3A*). Similarly, the AUC under the precision recall curve was 0.96 and 0.98 for the single-lead and 12-lead median models.

When applying the optimal validation threshold to the testing data set, the single-lead model had a sensitivity of 0.86, specificity of 0.89, a positive predictive value of 0.90, a negative predictive value of 0.85, and a F1 score of 0.88. The 12-lead model had a sensitivity of 0.91, specificity of 0.92, a positive predictive value of 0.93, a negative predictive value of 0.89, and a F1 score of 0.91.

Short rhythm vs. median electrocardiogram Siamese network performance

While the previous models used the median beat, we next affirmed that this strategy would also work with the first 2.5 s of each of the 12 leads. The ROC AUCs for the single-lead and 12-lead 2.5-s models were 0.92 and 0.95, respectively.

Median model performance by rhythm

When stratifying the median single-lead results by rhythm, pairs of ECGs with both records in sinus rhythm, atrial fibrillation/flutter, or 'other' rhythms had testing AUCs of 0.96, 0.92, and 0.88, respectively. Pairs with one record in sinus rhythm and one record in atrial fibrillation/flutter or 'other rhythm' produced AUCs of 0.90 and 0.86, respectively. Pairs with one record in 'other rhythm' and one record in atrial fibrillation/flutter showed further reduced performance with testing AUC of 0.84 (*Figure 3B*).

Median model performance by sex, age decade, and race/ethnicity

When the stratifying the testing median single-lead results by sex, male and female ECG pairs had consistent performance with an AUC of 0.94. When stratifying by patient age in decades, pairs of ECGs where patients were age 0–10 showed a reduction in performance with an AUC of 0.88 compared to AUCs of least 0.94 in all other decades (*Figure 3C*). Model performance was consistent across different race and ethnicity groups with AUCs of 0.95 for Hispanic and Black while non-Hispanic White and Other groups demonstrated an AUC of 0.94.

Median model performance for patients with frequent electrocardiograms

Supplementary material online, *Figure S2A*, illustrates the network performance for 1975 testing patients who had an average of more than three ECGs a year in the Mayo system with an AUC of 0.92 and 0.95, respectively, for the single-lead and 12-lead configuration median models (see Supplementary material online, *Figure S2A*). Electrocardiogram pair discrimination stratified by rhythm ranged from the best for pairs of sinus rhythms (AUC of 0.94) to the worst for pairs with one record in sinus rhythm and the other record in 'other' rhythm (AUC 0.83; see Supplementary material online, *Figure S2B*). Electrocardiogram pairs of children under 10 years also had the worst classification performance with an AUC of 0.85

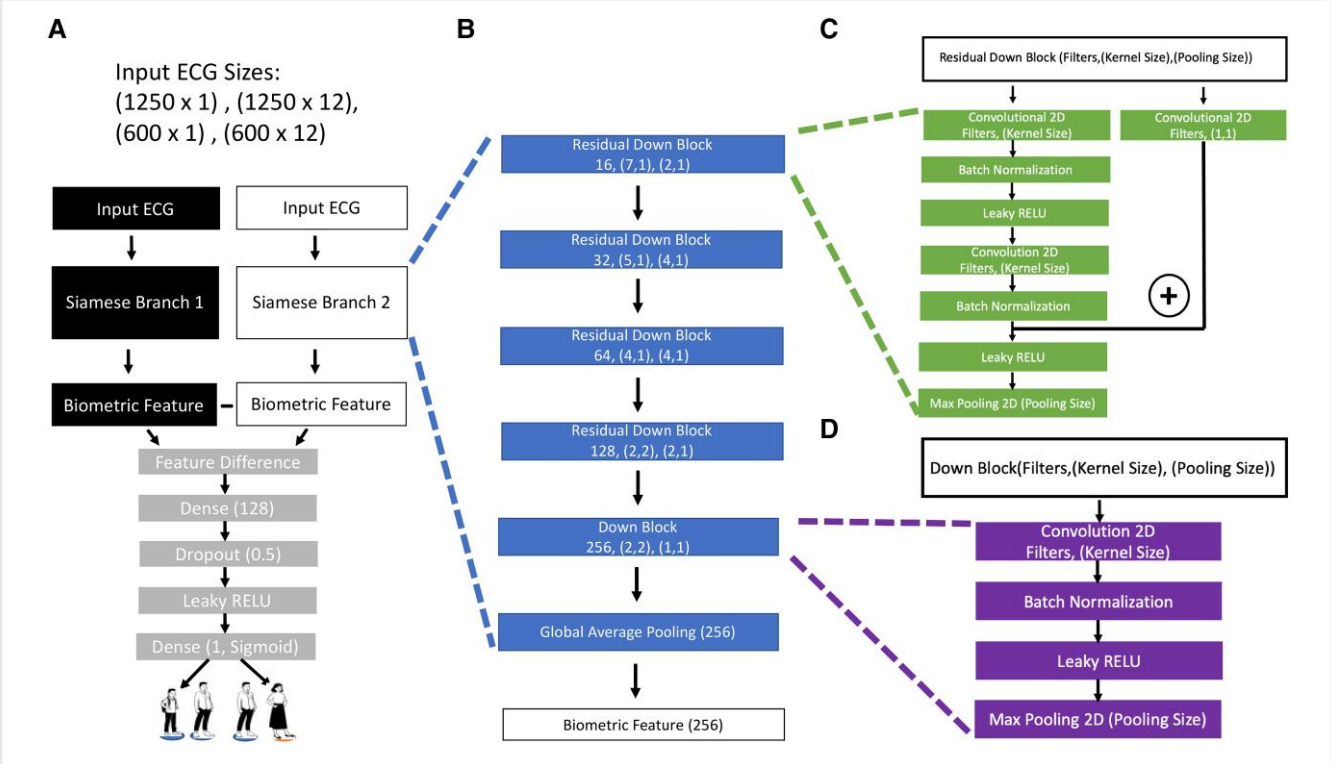


Figure 2 Architecture of the Siamese neural network. (*A*) The Siamese backbone with the (*B*) feature extraction. The feature extraction layers include (*C*) residual down blocks and (*D*) down blocks.

compared to AUCs of least 0.9 in the other decade categories (see Supplementary material online, *Figure S2C*).

Median model performance with diseased electrocardiograms

To assess the model's performance to ascertain same patient identity with various diseased ECGs as in *Table 1*, *Figure 4* displays the distribution of similarity scores for pairs where both ECGs are in the specified condition and where only one ECG is afflicted. Similarity probability scores were higher for pairs where both ECGs were afflicted compared to where only one ECG is afflicted, where model accuracy was at least 88% accurate. Model performance remained above 80% accuracy for one ECG of a same patient pair afflicted among low voltage, myocardial infarction, and first-degree atrioventricular block. The model was between 70 and 80% accurate for same patient pairs with one afflicted ECG with atrial fibrillation/flutter, right bundle branch abnormality, second-degree atrioventricular block, left ventricular hypertrophy, and long QT. Model performance was lowest at 61% accurate for patient pairs with one afflicted ECG in left bundle branch abnormality.

Use of the Siamese network for user identification over time

Because normal biological aging is known to affect the ECG[30,31] and can be detected by deep learning,[5] we plot the decrease in the similarity scores for ECG pairs within testing patient sets (both ECGs from the same patient) with increasing time intervals between acquisition times (*Figure 5*). In other words, even though the ECGs belong to the same patient, aging and disease change the ECG, and at some point, the patient may stop 'looking' like themselves (see Supplementary material online, *Figure S3*). Across all 16 511 testing patients with multiple ECGs on record, ECG pairs with a large difference in acquisition times show lower similarity scores. Electrocardiogram pair time difference groups across 1–35 years showed a significant change in similarity score distributions with medians dropping from 0.99 to 0.63 ($P < 0.001$).

Because of this natural decrease in similarity scores over time, a strategy to allow for accurate self-identification when ECG records are submitted over a large time interval is needed. *Figure 6* illustrates the use of an evolving baseline, a type of classification identification where new records are systematically compared to increasing numbers of prior ECGs.[26–28] We introduce this new type of classification identification to improve self-identification when ECG records are added over time. The evolving baseline allows for the continued use of higher similarity threshold, a necessity when adding to a medical record, when performing user identification.

Discussion

Our work had three main findings. First, we found that an ECG signal can be classified as coming from the same or a different individual when compared to a baseline sample using a Siamese network. This finding may be important as wearable devices are shared and because these signals are imported into the EHR. The model could be used in conjunction with standard safeguarding measures to verify patient identity where diverse patient identification scenarios, including different sex pairings, would be regularly encountered, before adding to the record. Second, we found that a single-lead ECG (lead I) had a practical performance near to that of the 12-lead ECG. Similarly, the median beat for practical purposes performed near to the short rhythm

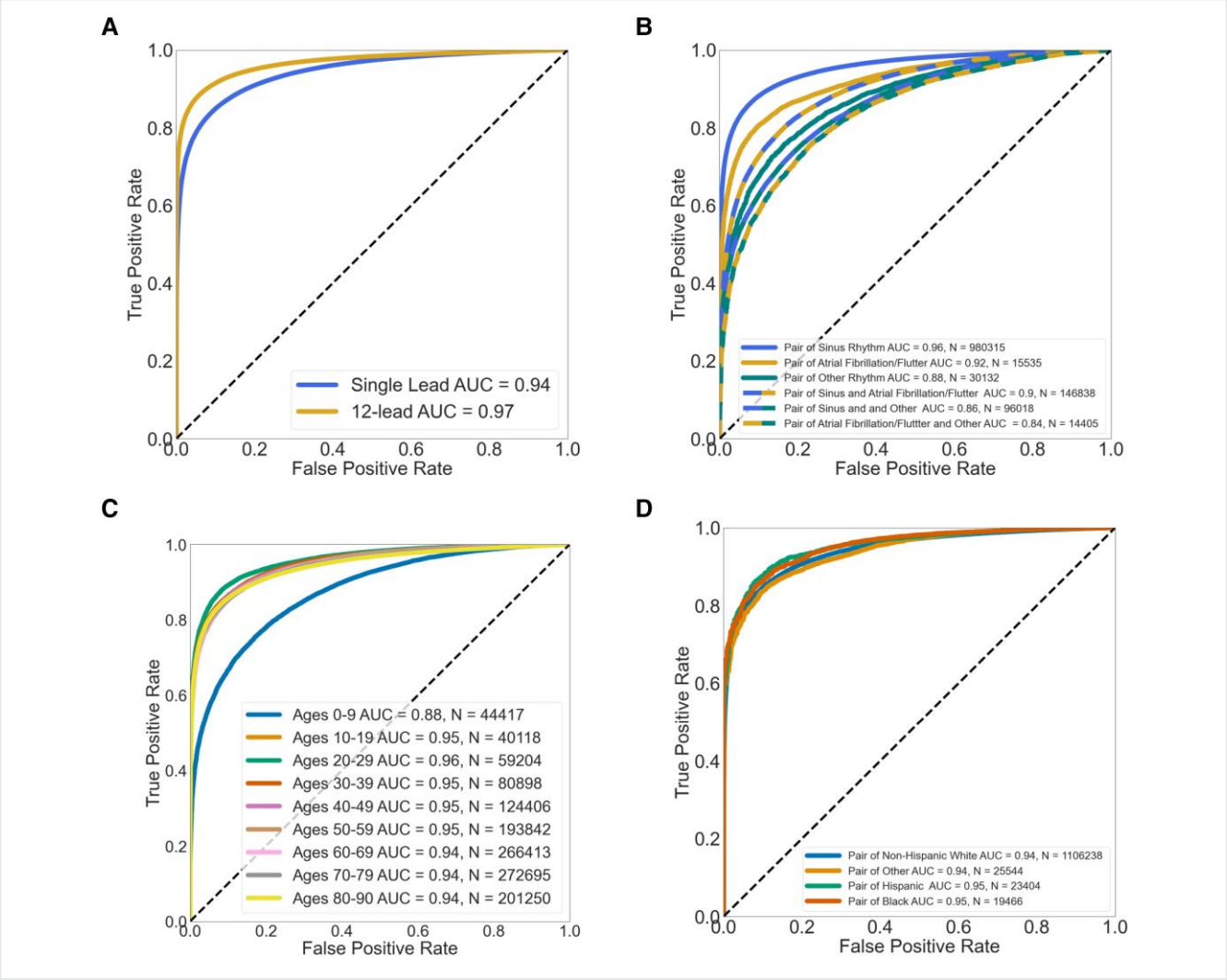


Figure 3 (*A*) Receiver operating characteristic for the testing data set using the median single-lead and 12-lead electrocardiogram Siamese networks. (*B*) Median single-lead receiver operating characteristics stratified by rhythms of the pairs. (*C*) Median single-lead receiver operating characteristics stratified by the decade of the electrocardiogram pair. (*D*) Median single-lead receiver operating characteristics stratified by race of the electrocardiogram pair.

ECG. Taken together, these results could make this approach useful for signals from mobile and wearable devices. However, further research will be needed to understand the AUC performance drops between single-lead and 12-lead ECGs as well as median and short rhythm signals. Median beat processing helps to denoise and normalize a signal, which may be important with signals acquired from nonmedical environments. Since mobile recordings are acquired by patients, and patients might share their devices with family members or friends, identity ascertainment in noisy, uncontrolled environments becomes important. Third, we found that baseline ECG signals change over time, and that a novel form of baseline averaging improves identity performance, particularly when there is a long-time interval between signal acquisitions.

Siamese neural networks are an ideal deep learning to architecture for long-term monitoring of patient data. Once a baseline record is established for the patient, the Siamese network can determine whether the next record added is sufficiently like the baseline. Because the data set is split on the patient level, there was no overlap of patient data between the training and testing sets. Splitting by patient assures the model learns how to discriminate between individuals generally so that the model may successfully be used on unseen patients. Importantly, failure of self-identification may prove a powerful marker for an important cardiac pathophysiologic event as quantified in *Figure 4* and prompt the user to seek medical evaluation.

The ECG is affected by the natural aging process, so deep learning similarity scores will drop over time, even in the absence of intercurrent illness.[5,32,33] Other biometric detectors such as fingerprint scans or iris imaging[26,27] are also affected by this aging process.[23–25] Comparing the drop in similarity scores observed for ECG pairs within the patient sets, grouped by time interval between the acquisition times of the pair, highlights the natural aging process (*Figure 5*). It invites many practical questions on how the baseline ECG used for Siamese network pairing will need to be changed over time. In this study, we review ~35 years of digitally acquired records while intervals as short as 5–7 years introduce a change in the similarity score distribution ($P < 0.001$). This observation then suggests how often the baseline, or as referenced in the biometrics literature as 'template update,' should be updated at a patient level.[25,34]

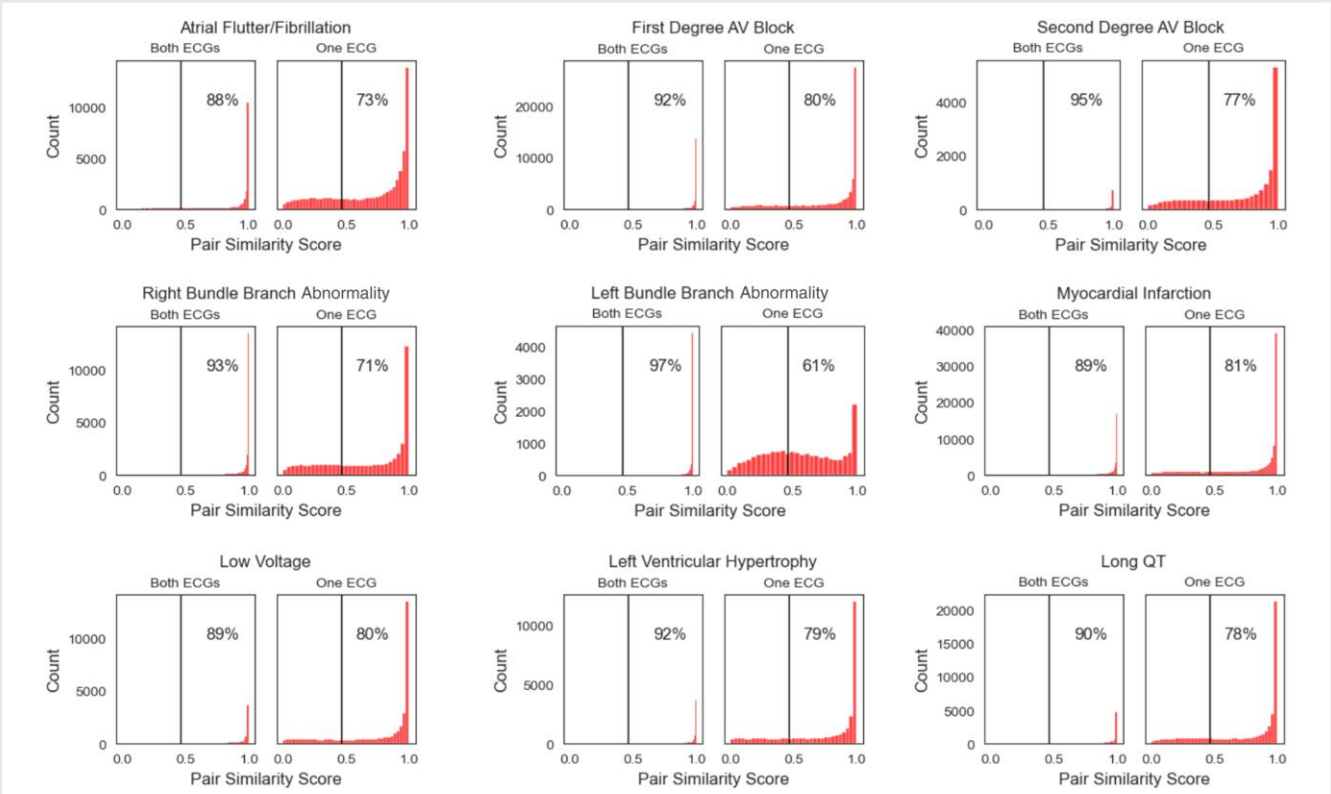


Figure 4 Distribution of same patient pair electrocardiogram similarity scores where either both electrocardiograms of the pair are in the rhythm/disease or only one electrocardiogram of the pair is afflicted. Vertical lines indicate the optimal validation threshold while percentages indicate the correct same patient pair classification based on this threshold.

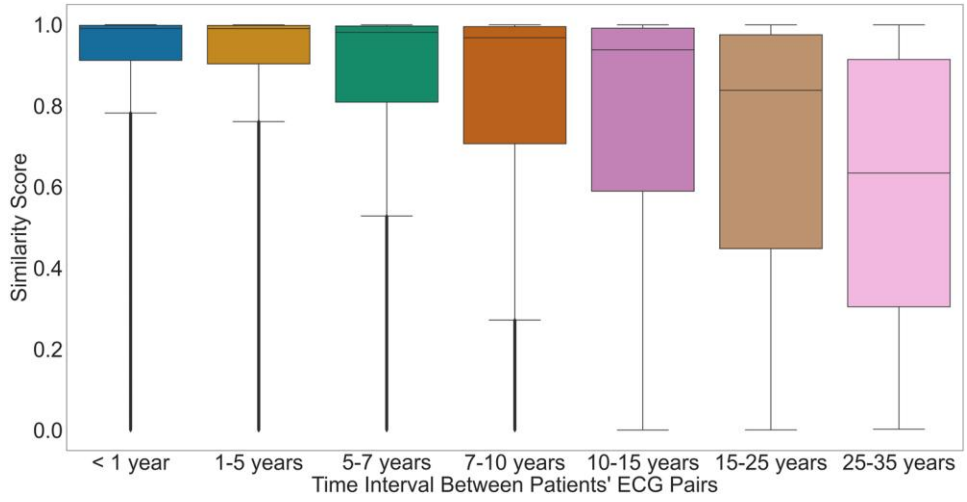


Figure 5 A total of 16 511 testing within patients' electrocardiogram set pairs (both electrocardiograms are from the same patient) with progressively larger differences in acquisition times show significantly different distributions in similarity scores with lower medians ($P < 0.001$).

The optimal time for recapturing a baseline remains unknown and can be affected by the patient's overall health as acute disease can significantly affect the ECG. However, the rate of change can also be an opportunity to use this network to assess acute changes: for example, if a patient's similarity score drops dramatically over a short period of time or whether it recovers in response to medical therapies and interventions. As shown in

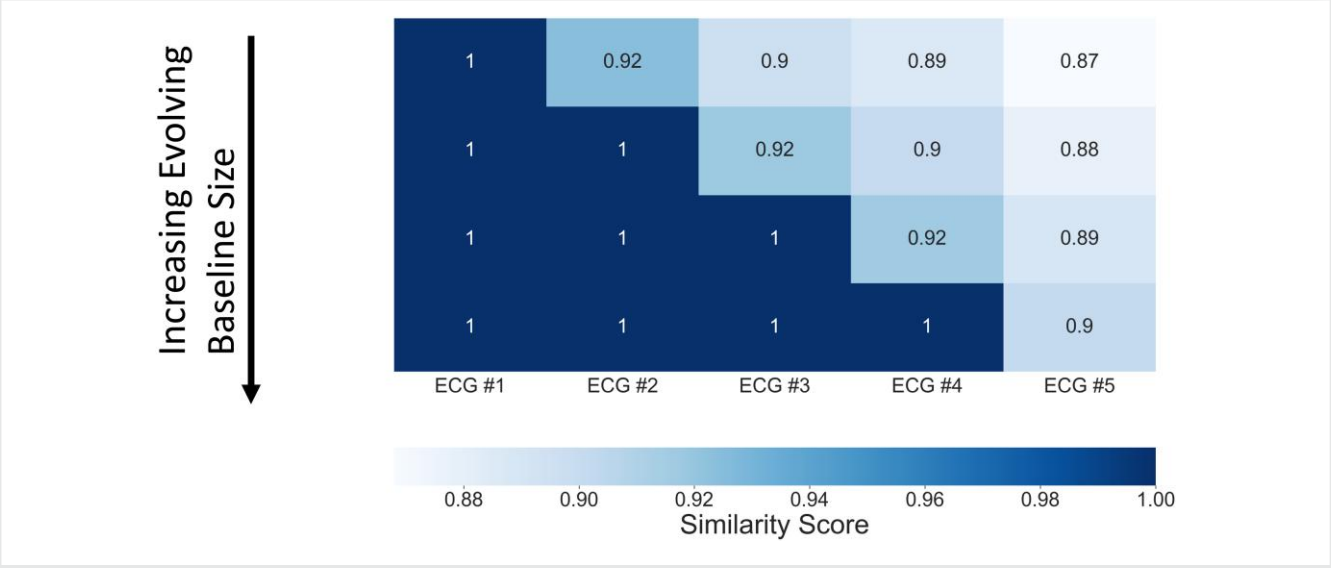


Figure 6 Updating the baseline with future records ('evolving') improves similarity scores. Here, there are 747 patients, each with five electrocardiograms total on record, where the difference between each electrocardiogram is between 0 and 2000 days. In the first row, the first electrocardiogram on record is the sole baseline like what is commonly done with other biometrics. By the time the fifth electrocardiogram record is added to the system, the similarity score has dropped to 0.87. However, the subsequent rows illustrate if this baseline is expanded by each consecutive record (row 2: electrocardiogram #1, electrocardiogram #2; row 3: electrocardiogram #1, electrocardiogram #2, electrocardiogram #3), and then the drop in the similarity score is less, only down to 0.90 when using the first four electrocardiograms as a baseline, compared to 0.87 when only using electrocardiogram #1 as a baseline.

our previous work on ECG estimated age,[5] an individual does not necessarily age chronologically given the various life factors.[32,33,35–40] We suggest developing an 'evolving baseline' based on the ECGs seen so far (*Figure 6*) that indicate whether the latest record should be added to the patient's repository. Intriguingly, the frequency of when the system signals a lower similarity score between the current baseline ECG(s) could be in of itself a metric of aging and a key part of personalized medicine.

Because of the natural aging process, we wish to illustrate how the ECG features from a single recording or in combination, as in the case when using the 'evolving baseline,' could be harnessed to populate a self-submission repository of patient ECG records over a lifetime. This algorithm could be used with other biometrics, such as voice,[41,42] to insure accuracy of medical data collected in noisy, nonmedical environments. It is interesting to highlight that between 5 and 7 years, the distribution of similarity scores significantly changes (*Figure 5*; $P < 0.001$) while the average error in our previously published ECG age estimate model was 6.9 ± 5.6 years. Perhaps these results suggest a resolution limit of the AI-ECG age-based estimates.

While this study immediately addresses a practical problem of user identification for single-lead ECG records in the EHR, changes in similarity scores due to aging expand its scope to additional biological marker of aging. This measure may be a frontline measure in the clinic for a relatively easy to acquire and cheap picture of the patient wellness, especially as wearables with mobile form factors continue to increase in popularity.

Limitations

Our work is best understood in the context of its limitations. We acknowledge that many Mayo Clinic patients with many ECGs on record are old or sick. Thus, the model may not recognize the individual differences between healthy, young people. Indeed, the average chronological age at the first ECG is 42 ± 25 years and at the last ECG is 52 ± 30 years, which reflects this aging population. However, given the rise in the use of smartwatches, we hope increasing numbers of patients will participate in recording ECGs regardless of age or disease history and the model estimates will improve.

All recordings used for this study were done in the clinic, and we only mimicked the use of a single lead by selecting lead I of the 12-lead ECG. Future studies will be needed to understand the model's sensitivity to noise and artefacts. The same brands of ECG recordings and digital signal processing were used throughout this study: more validation studies will be needed to investigate pairing records acquired from different institutions. Smart device recordings may contain more noisy artefacts compared to clinic recordings along with differing signal processing specifications. Smart device recordings also may not be recorded in the supine position, which may even further impede the model's performance. Forthcoming formal validation studies using data collected from a smart device of various brands will be vital to test the model generalizability. While the single-lead and 12-lead models have promising sensitivity and specificity of at least 86% based on the optimal validation threshold, external validation studies on diverse data sets will be critical to minimize patient misclassifications when implemented in clinical practice.

The data set was heavily biased towards sinus rhythms (see Supplementary material online, *Figure S1*), and the performance of the mixed rhythm pairs was lower compared to pairs in the same rhythm (*Figure 3B*). Because smartwatches may be used to diagnose new arrhythmias, future studies could explore data sets with a higher weighting towards mixed rhythm pairs to help improve user identification. This higher weighting may also be of use to improve the performance of single ECG afflicted pairs with left bundle branch abnormality, for example. The performance drops seen in the paediatric cases, indicative of rapidly changing cardiac electrophysiology[43] and thus ECG biometric features, suggest the need for future studies to focus on paediatric performance.

From a statistical standpoint, the use of estimated AUCs is not customary because patient ECG pairing introduces clustering that violates the independence assumption of the Mann–Whitney *U* statistic.[29] Measures of uncertainty in the AUCs are not likely appropriate, so more appropriate measures should the subject of future research.

Conclusions

Here, we show that the ECG permits self-identification, which could help limit errors in medical signals acquired in nonmedical environments. We introduce a novel form of baseline adjustment for expected age-related changes in ECG, while unexpected rates of change may be key to implementing true personalized medicine by prompting medical evaluation.

Supplementary material

Supplementary material is available at *European Heart Journal – Digital Health*.

Acknowledgements

The graphical abstract and Figure 1 were created with BioRender.com.

Funding

This publication was made possible through the support of the Ted and Loretta Rogers Cardiovascular Career Development Award Honoring Hugh C. Smith MD.

Conflict of interest: R.E.C., K.C.S., P.A.N., F.L.-J., S.J.A., P.A.F., Z.I.A., and Mayo Clinic have licensed AI-ECG models to Anumana and might benefit from its commercialization.

Data availability

All requests for raw and analysed data and related materials, excluding programming code, will be reviewed by the Mayo Clinic legal department and Mayo Clinic Ventures to verify whether the request is subject to any intellectual property or confidentiality obligations. Requests for patient-related data not included in the paper will not be considered. Any data and materials that can be shared will be released via a material transfer agreement.

References

1. Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G, *et al.* Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. *Nat Med* 2019;**25**:70–74.
2. Ko W-Y, Siontis KC, Attia ZI, Carter RE, Kapa S, Ommen SR, *et al.* Detection of hypertrophic cardiomyopathy using a convolutional neural network-enabled electrocardiogram. *J Am Coll Cardiol* 2020;**75**:722–733.
3. Grogan M, Lopez-Jimenez F, Cohen-Shelly M, Dispenzieri A, Attia ZI, Abou Ezzedine OF, *et al.* Artificial intelligence–enhanced electrocardiogram for the early detection of cardiac amyloidosis. *Mayo Clin Proc* 2021;**96**:2768–2778.
4. Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, *et al.* An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet* 2019;**394**:861–867.
5. Attia ZI, Friedman PA, Noseworthy PA, Lopez-Jimenez F, Ladewig DJ, Satam G, *et al.* Age and sex estimation using artificial intelligence from standard 12-lead ECGs. *Circ Arrhythm Electrophysiol* 2019;**12**:e007284.
6. Bayoumy K, Gaber M, Elshafeey A, Mhaimeed O, Dineen EH, Marvel FA, *et al.* Smart wearable devices in cardiovascular care: where we are and how to move forward. *Nat Rev Cardiol* 2021;**18**:581–599.
7. Strik M, Ploux S, Ramirez FD, Abu-Alrub S, Jaïs P, Haïssaguerre M, *et al.* Smartwatch-based detection of cardiac arrhythmias: beyond the differentiation between sinus rhythm and atrial fibrillation. *Heart Rhythm* 2021;**18**:1524–1532.
8. Attia ZI, Harmon DM, Dugan J, Manka L, Lopez-Jimenez F, Lerman A, *et al.* Prospective evaluation of smartwatch-enabled detection of left ventricular dysfunction. *Nat Med* 2022;**28**:2497–2503.
9. Bachtiger P, Petri CF, Scott FE, Park SR, Kelshiker MA, Sahemey HK, *et al.* Point-of-care screening for heart failure with reduced ejection fraction using artificial intelligence during ECG-enabled stethoscope examination in London, UK: a prospective, observational, multicentre study. *Lancet Digit Health* 2022;**4**:e117–e125.
10. Chicco D. Siamese neural networks: an overview. *ANN* 2021;**2190**:73–94.
11. Bromley J, Guyon I, LeCun Y, Säckinger E, Shah R. Signature verification using a" siamese" time delay neural network. *Int J Pattern Recogn Artif Intell* 1993;**7**:669–688.
12. Zhang C, Liu W, Ma H, Fu H. Siamese neural network based gait recognition for human identification. In: 2016 IEEE International Conference on Acoustics, Speech and Signal processing (ICASSP); 2016 Mar 20. IEEE. p.2832–2836.
13. Bhagwat N, Viviano JD, Voineskos AN, Chakravarty MM. Alzheimer's disease neuroimaging, I. Modeling and prediction of clinical symptom trajectories in Alzheimer's disease using longitudinal data. *PLoS Comput Biol* 2018;**14**:e1006376.
14. Mehmood A, Maqsood M, Bashir M, Shuyuan Y. A deep Siamese convolution neural network for multi-class classification of Alzheimer disease. *Brain Sci* 2020;**10**:84.
15. Li MD, Chang K, Bearce B, Chang CY, Huang AJ, Campbell JP, *et al.* Siamese neural networks for continuous disease severity evaluation and change detection in medical imaging. *npj Digital Medicine* 2020;**3**:48.
16. Pinto JR, Cardoso JS, Lourenço A. Evolution, current challenges, and future possibilities in ECG biometrics. *IEEE Access* 2018;**6**:34746–34776.
17. Prakash AJ, Patro KK, Samantray S, Pławiak P, Hammad M. A deep learning technique for biometric authentication using ECG beat template matching. *Information* 2023;**14**:65.
18. Ivanciu L, Ivanciu I-A, Farago P, Roman M, Hintea S. An ECG-based authentication system using Siamese neural networks. *J Med Biol Eng* 2021;**41**:558–570.
19. Ibtehaz N, Chowdhury ME, Khandakar A, Kiranyaz S, Rahman MS, Tahir A, *et al.* EDITH: ECG biometrics aided by deep learning for reliable individual authentication. *IEEE Trans Emerg Top Comput Intell* 2022;**6**:928–940.
20. Behrouzi P, Shirkani B, Hazratifard M. Using ECG signals in Siamese networks for authentication in digital healthcare systems. *J ISSN* 2022;**3**:1367–1373.
21. Hazratifard M, Agrawal V, Gebali F, Elmiligi H, Mamun M. Ensemble Siamese network (ESN) using ECG signals for human authentication in smart healthcare system. *Sensors* 2023;**23**:4727.
22. Unal I. Defining an optimal cut-point value in ROC analysis: an alternative approach. *Comput Math Methods Med* 2017;**3762651**.
23. Fenker SP, Bowyer KW. Fenker SP, Bowyer KW. Analysis of template aging in iris biometrics. In: 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops; 2012 Jun 16. IEEE. p.45–51.
24. Galbally J, Haraksim R, Beslay L. A study of age and ageing in fingerprint biometrics. *IEEE Trans Inf Forensics Secur* 2019;**14**:1351–1365.
25. Pisani PH, Mhenni A, Giot R, Cherrier E, Poh N, Ferreira de Carvalho ACPL, *et al.* Adaptive biometric systems: review and perspectives. *ACM Comput Surv (CSUR)* 2019;**52**:1–38.
26. Jain AK, Kumar A. In: Mordini E, Tzovaras D (eds.), *Second Generation Biometrics: the Ethical, Legal and Social Context*. Netherlands: Springer; 2012. p49–79.
27. Jain AK, Kumar A. Biometrics of next generation: an overview. *Second Gener Biom* 2010; **12**:2–3.
28. Gorodnichy DO. Evolution and evaluation of biometric systems. In: 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications; 2009 Jul 8. IEEE. p.1–8.
29. Mason SJ, Graham NE. Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: statistical significance and interpretation. *Q J R Meteorol Soc* 2002;**128**:2145–2166.
30. Malik M, Hnatkova K, Kowalski D, Keirns JJ, Gelderen EMV. QT/RR curvatures in healthy subjects: sex differences and covariates. *Am J Physiol Heart Circ Physiol* 2013;**305**: H1798–H1806.
31. Simonson E. The effect of age on the electrocardiogram. *Am J Cardiol* 1972;**29**: 64–73.
32. Ladejobi AO, Medina-Inojosa JR, Shelly Cohen M, Attia ZI, Scott CG, LeBrasseur NK, *et al.* The 12-lead electrocardiogram as a biomarker of biological age. *Eur Heart J Digit Health* 2021;**2**:379–389.
33. Lima EM, Ribeiro AH, Paixão GMM, Ribeiro MH, Pinto-Filho MM, Gomes PR, *et al.* Deep neural network-estimated electrocardiographic age as a mortality predictor. *Nat Commun* 2021;**12**:5117.
34. Akhtar Z, Ahmed A, Erdem CE, Foresti GL. Biometric template update under facial aging. In: 2014 IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM); 2014 Dec 9. IEEE. p.9–15.
35. Diez Benavente E, Jimenez-Lopez F, Attia ZI, Malyutina S, Kudryavtsev A, Ryabikov A, *et al.* Studying accelerated cardiovascular ageing in Russian adults through a novel deep-learning ECG biomarker [version 1; peer review: 1 approved with reservations]. *Wellcome Open Res* 2021;**6**:12.

36. Libiseller-Egger J, Phelan JE, Attia ZI, Benavente ED, Campino S, Friedman PA, *et al.* Deep learning-derived cardiovascular age shares a genetic basis with other cardiac phenotypes. *Sci Rep* 2022;**12**:22625.
37. Toya T, Ahmad A, Attia Z, Cohen-Shelly M, Ozcan I, Noseworthy PA, *et al.* Vascular aging detected by peripheral endothelial dysfunction is associated with ECG-derived physiological aging. *J Am Heart Assoc* 2021;**10**:e018656.
38. Shelly S, Lopez-Jimenez F, Chacin-Suarez A, Cohen-Shelly M, Medina-Inojosa JR, Kapa S, *et al.* Accelerated aging in LMNA mutations detected by artificial intelligence ECG–derived age. *Mayo Clin Proc* 2023;**98**:522–532.
39. Meenakshi-Siddharthan DV, Livia C, Peterson TE, Stalboerger P, Attia ZI, Clavell AL, *et al.* Artificial intelligence–derived electrocardiogram assessment of cardiac age and molecular markers of senescence in heart failure. *Mayo Clin Proc* 2023;**98**:372–385.
40. Hirota N, Suzuki S, Motogi J, Nakai H, Matsuzawa W, Takayanagi T, *et al.* Cardiovascular events and artificial intelligence-predicted age using 12-lead electrocardiograms. *IJC Heart Vasculature* 2023;**44**:101172.
41. Krawczyk S, Jain AK. Securing electronic medical records using biometric authentication. In International Conference on Audio-and Video-Based Biometric Person Authentication; 2005 Jul 20. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 1110–1119.
42. Leonard D, Pons AP, Asfour SS. Realization of a universal patient identifier for electronic medical records through biometric technology. *IEEE Trans Inf Technol Biomed* 2008;**13**:494–500.
43. Salameh S, Ogueri V, Posnack NG. Adapting to a new environment: postnatal maturation of the human cardiomyocyte. *J Physiol (Lond)* 2023;**601**:2593–2619.