

ORIGINAL ARTICLE



Age and Sex Estimation Using Artificial Intelligence From Standard 12-Lead ECGs

BACKGROUND: Sex and age have long been known to affect the ECG. Several biologic variables and anatomic factors may contribute to sex and age-related differences on the ECG. We hypothesized that a convolutional neural network (CNN) could be trained through a process called deep learning to predict a person's age and self-reported sex using only 12-lead ECG signals. We further hypothesized that discrepancies between CNN-predicted age and chronological age may serve as a physiological measure of health.

METHODS: We trained CNNs using 10-second samples of 12-lead ECG signals from 499 727 patients to predict sex and age. The networks were tested on a separate cohort of 275 056 patients. Subsequently, 100 randomly selected patients with multiple ECGs over the course of decades were identified to assess within-individual accuracy of CNN age estimation.

RESULTS: Of 275 056 patients tested, 52% were males and mean age was 58.6 ± 16.2 years. For sex classification, the model obtained 90.4% classification accuracy with an area under the curve of 0.97 in the independent test data. Age was estimated as a continuous variable with an average error of 6.9 ± 5.6 years (R -squared = 0.7). Among 100 patients with multiple ECGs over the course of at least 2 decades of life, most patients (51%) had an average error between real age and CNN-predicted age of <7 years. Major factors seen among patients with a CNN-predicted age that exceeded chronologic age by >7 years included: low ejection fraction, hypertension, and coronary disease ($P < 0.01$). In the 27% of patients where correlation was >0.8 between CNN-predicted and chronologic age, no incident events occurred over follow-up (33 ± 12 years).

CONCLUSIONS: Applying artificial intelligence to the ECG allows prediction of patient sex and estimation of age. The ability of an artificial intelligence algorithm to determine physiological age, with further validation, may serve as a measure of overall health.

VISUAL OVERVIEW: A [visual overview](#) is available for this article.



Zachi I. Attia, MSc
Paul A. Friedman, MD
Peter A. Noseworthy, MD
Francisco Lopez-Jimenez, MD, MSc
Dorothy J. Ladewig, BS
Gaurav Satam, MS, MBA
Patricia A. Pellikka, MD
Thomas M. Munger, MD
Samuel J. Asirvatham, MD
Christopher G. Scott, MS
Rickey E. Carter, PhD
Suraj Kapa, MD

Key Words: artificial intelligence
■ coronary disease
■ electrocardiography ■ hypertension
■ neural network

© 2019 The Authors. *Circulation: Arrhythmia and Electrophysiology* is published on behalf of the American Heart Association, Inc., by Wolters Kluwer Health, Inc. This is an open access article under the terms of the [Creative Commons Attribution Non-Commercial-NoDerivs](#) License, which permits use, distribution, and reproduction in any medium, provided that the original work is properly cited, the use is noncommercial, and no modifications or adaptations are made.

<https://www.ahajournals.org/journal/circep>

WHAT IS KNOWN?

- The 12-lead ECG may change in an individual over time due to normal aging, and males and females may exhibit differences in their ECGs.
- Artificial intelligence techniques, including convolutional neural networks, may strongly correlate features of an ECG with specific phenotypic findings (such as low ejection fraction) not otherwise identifiable through standard algorithms or by humans.

WHAT THE STUDY ADDS?

- Using convolutional neural networks, a trained computer system is able to identify whether an individual is male or female from a 12-lead ECG with an area under the curve of 0.97.
- A trained neural network can determine an individual's age from a 12-lead ECG alone within 7 years of their actual age.
- When the convolutional neural network-predicted age exceeds a patient's actual age by at least 7 years, there is a higher incidence of cardiovascular comorbidities, potentially suggesting that the convolutional neural network-predicted age from 12-lead ECGs may correlate with physiological health.

The ECG was first invented in 1901 by Willem Einthoven. Over the ensuing century, it has become a mainstay for risk stratification, disease identification, and cardiovascular management.^{1,2} In the current age of machine learning and artificial intelligence (AI), it may be possible to identify novel uses of the ECG. Recent studies suggest that using advanced computing approaches, the ECG may be used for serum potassium level estimation,³ screening for asymptomatic left ventricular dysfunction,⁴ and detecting cocaine usage.⁵ In the nonmedical space, recent data has also suggested that the ECG can serve as a sort of biometric sensor similar to a retinal scan or a fingerprint as it contains subject-specific features.^{6,7}

It is theoretically possible that diagnostic tests such as the ECG, which carry the ability to discriminate between individuals, might be able to more deeply assess aspects of an individual, such as predicting age and sex. The value of such approaches lies in further validating the role of the ECG in recognizing not just a disease but in revealing more details about an individual that may factor into overall health. Additionally, discrepancies between chronological age and the age determined by processing the raw electrical signals generated by cardiac tissue may serve as a marker for occult cardiac or noncardiac disease, which may impact long-term outcomes.

While age and sex are known to affect the ECG,⁸⁻¹⁰ prior work to see if the ECG could predict age has focused on Bayesian or other analytic techniques depen-

dent on specific feature extraction as the foundation for the analysis, and also largely included healthy patients.¹⁰ We sought to determine whether the application of AI algorithms to a large ECG patient data set would be capable of predicting age and sex reported by patients, independent of additional clinical data. We further sought to determine whether discrepancies between ECG age and chronological age might be a marker of physiological health.

METHODS

The data, analytic methods, and study materials that support the findings of this study are available from the corresponding author on reasonable request.

Data Sources and Study Population

We identified 774 783 adult patients (18 years or older) with at least one digital, standard 10 second 12-lead ECG acquired in the supine position between January 1994 and February 2017. Data was collected from the Mayo Clinic digital vault, with institutional review board approval. For patients with multiple ECGs, only the earliest ECG was selected.

ECGs were acquired at a sampling rate of 500 Hz using a GE-Marquette ECG machine (Marquette, WI) and stored using the MUSE data management system.

Overview of AI Model Development

We implemented a convolutional neural network (CNN)¹¹ using the Keras Framework with a Tensorflow (Google, Mountain View, CA) backend and Python. While CNNs are mainly applied to images, we adjusted the network architecture to have spatial and temporal feature extraction layers. The network operates by adjusting the weights of the convolutional filters during training to extract meaningful and relevant features in an unsupervised way. Both the sex and age network was built using stacked blocks of convolutional, max pooling, and batch normalization¹²; each block was followed by a nonlinear activation function. After the first group of blocks extracted temporal features, another spatial block was used to fuse data from all leads, and then the extracted features were used in a fully connected network. The difference between the age and sex networks was in the output layer. The sex detection network had 2 outputs (male and female) and was activated with a Softmax output, while the age network had a single output (age) as a continuous number, without a following nonlinear function (linear activation). Although we used the same data for training both the sex and age algorithms, the networks were trained separately, but with the same data split for training, validation, and testing sets. Of note, the only data inputted for training were the raw digital 12-lead ECG signal and the associated age and self-reported sex of each individual. Figure 1 summarizes the network architecture for each network. Hyperparameters (batch size, initial learning rate, number of neurons in the fully connected layers, and number of convolutional layers) were changed during training to get the optimal model based on the validation set. Initial learning rates tested were in the range 3e-3 to 1e-5, and batch size was in the range of 16 to 128. Selected hyperparameters were

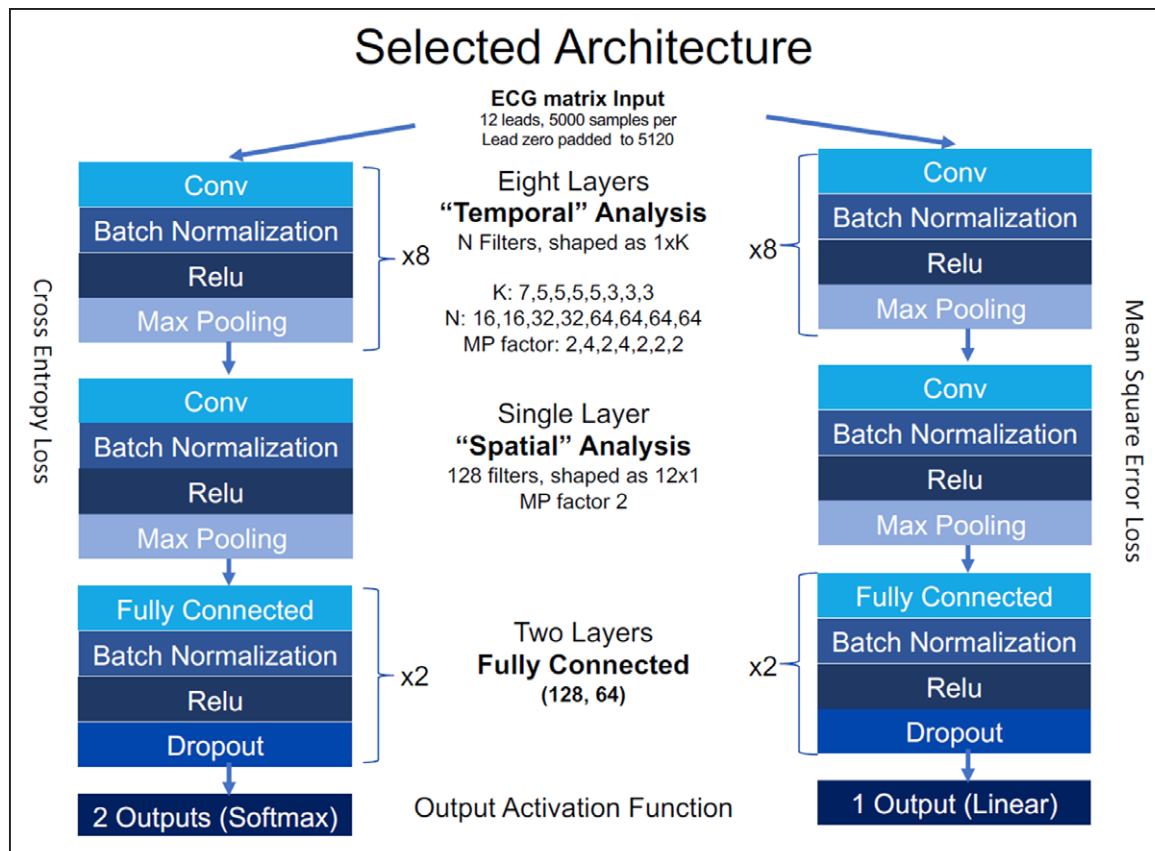


Figure 1. Construction of neural network architecture.

Shown is the construction of the neural network architecture for the age and sex networks. Further details are discussed under the methods section.

learning rate of $3e-4$ and batch size of 64 for the age network and 16 for the sex network.

Data Sets for Developing and Testing the Network

ECGs have some biometric features that might be learnt in the process of training the network. To avoid cross-contamination of the training and testing data sets, only a single ECG per patient was used. After an initial split into the development (64% of the population) and holdout (36% of the population) data sets, the development data set was further divided into training (75% of the development set) and internal validation data (25%). The training and validation sets were mutually exclusive with respect to patient identification. For training both networks, ECGs were fed to the network, and the network weights were optimized using the Adam optimizer¹³ with categorical cross-entropy as the loss function for the sex network and mean squared error for the age network. The internal validation data set was used to select the optimal model and hyper parameters and to find the optimal threshold for sex classification.

Models Evaluation and Statistical Methods

Separate models for sex and age were tested using the same holdout data set. The sex network is a binary classifier (male versus female), so the model output is the probability that

the ECG was obtained from a male—denoted as P (convolutional CNN-predicted sex). The output P is a continuous number in the range 0 to 1, and in cases of inconclusive ECG–sex discrimination, P will be closer to 0.5. Due to the nonlinear nature of the network and the uneven number of male and female ECGs used for training, a trivial threshold of 0.5, meaning $P > 0.5$ implies male, may not be the optimal threshold to maximize classification accuracy for males and females. To address this, a receiver operating characteristic curve was created using P as the linear discriminator to estimate the area under the curve (AUC) and accuracy at various candidate thresholds. An optimal threshold was selected using the internal validation data set, and uncertainty in classification was evaluated using the output probability P to analyze if the model could evaluate its own classification results by rejecting samples with low certainty. To evaluate whether specific portions of the ECG were critical in arriving at a specific sex estimation, separate neural networks were trained on 35 000 independent individuals using discrete portions of the ECG (P wave, QRS complex, etc), and similarly a receiver operating characteristic curve was generated to determine accuracy in prediction of sex. For the age estimation network, the network output was an estimate for age based on the ECG as an input (CNN-predicted age). The R square metric and the mean average error metric were used to evaluate the network. In the training stage, the networks were fed the whole training set, and in each iteration (epoch) the model was evaluated on the validation set, a software callback that saved only the best model based

on the validation performance was used. The training was stopped if, for 5 consecutive epochs, the validation performance did not improve (to avoid training in the overfitting stage). The age prediction optimal model was trained for 20 epochs while the sex detection model achieved optimal results after 12.

Estimation of CNN-Predicted Age Over Time and Impact of Clinical Variables

Within the data set, an algorithm to extract those patients with multiple ECGs recorded over a minimum 2 decades of life was created, and 100 patients were randomly selected. The choice of 100 patients was (1) to permit feasibility of deep chart review and (2) given this was meant to be an exploratory analysis of how this algorithm might perform in a random limited subset of patients. All ECGs at every chronologic year of life were evaluated using the trained network for age, and CNN-predicted age was plotted against the chronologic age at which that ECG was obtained. If multiple ECGs were obtained at any year of life, the median CNN-predicted age at that year of life was used. After this, clinical histories for patients were extracted via chart review. Clinical comorbidities, including prior history of myocardial infarction, low ejection fraction, coronary disease, hypertension, diabetes mellitus, atrial arrhythmias, and prior history of cardiac surgery were included. In addition, incident events over the follow-up period including hospitalization for hemodynamic shock, incident cancer, incident comorbidities that were first diagnosed over the follow-up period, new diagnoses of coronary disease, heart failure, or thromboembolic events, and cardiac transplantation were recorded. The R-squared correlation values between CNN-predicted and chronologic age were then calculated per patient and effect of prior comorbidities, and incident events on the R-squared or differences between CNN-predicted and chronologic age were determined. To compare impact of preexisting comorbidities on deviation of CNN-predicted age from chronologic age over time, the mean deviation of CNN-predicted age from chronologic age was calculated over the course of the follow-up period. The frequency of specific comorbidities among patients whose CNN-predicted exceeded the chronologic age by greater than 7 years, was within 7 years of the chronologic age, or was less than the chronologic age by 7 years was recorded. χ^2 analysis was used to determine whether frequency of comorbidities was significantly different between groups. $P < 0.01$ was considered significant.

RESULTS

Study Population

A total of 774783 patients with ECG were evaluated. The first ECG was used for developing and testing the model. Patients' mean age was 58.6 ± 16.2 , and 52% were male. In the development set, 399750 unique patients' ECGs were in the training set, and 99977 in the internal validation set. The remaining 275056 ECGs were in the holdout testing set.

Sex Classification

All the ECGs in the internal validation set and in the holdout data set were tested using the network. The algorithm output for each ECG included a probability number (P) between 0 and 1. A lower number implied high probability of being female while a number closer to 1 implied high probability of being male. The AUC of the holdout data set was 0.968 versus an AUC of 0.973 on the internal validation data set (Figure 2A). Using the Youden index calculated on the internal validation set ($Y = 0.517$) the overall accuracy was 90.4% (95% CI, [90.3%–90.5%]) 90.6% (95% CI, [90.4%–90.7%]) for females and 90.3% (95% CI, [90.2%–90.5%]) for males.

After rejecting the 10% ECGs with intermediate P values ($0.31 < P < 0.69$) which suggested an inability to differentiate male versus female for the given ECG (seen in about 10% of all ECGs), the accuracy of the classification improved to 94.3%. When applying the same technique wherein P values between 0.17 and 0.83 were excluded (accounting for 20% of all ECGs), accuracy improved even further to 96.8%. These findings suggest the classifier certainty can be derived from the output probability.

When testing the classifier in a younger population (< 45 years of age), the overall accuracy for sex identification was 93% versus 89% in patients older than 55. In addition, a separate independent cohort of 35000 individuals was used to test whether specific ECG features (P wave area in lead I, QRS duration, P wave time to peak in lead I, QTc interval, and T wave area in lead V4) could be used separately to determine sex. The receiver operating characteristic curves are summarized in Figure 2B for each portion of the P-QRS-T complex, with QRS duration carrying the highest AUC (0.683) but none individually exceeding 0.7.

Age Estimation

As the output of the age estimation network was a continuous variable, the statistics of the absolute error were calculated together with the overall correlation and the explained variance (R squared). For the holdout data set, the mean absolute error was 6.9 ± 5.6 years and R squared of 0.7 ($r = 0.837$, $P < 0.0001$). A scatter plot with chronologic age versus CNN-predicted age is presented in Figure 3A. For detection of age ≥ 40 , the AUC was 0.94 with a sensitivity of 87.8%, specificity of 86.8%, and accuracy of 87%. For the multi-group classification to the age groups of 18 to 25, 25 to 50, 50 to 75, and 75 and above the overall accuracy was 71.6% (95% CI [71.5%–71.9%]; Figure 3B).

When supplying the network the results of the sex network and the reported sex of the patient, for the 10% of patient with an inaccurate sex estimation, the R squared for age estimation decreased to 0.62

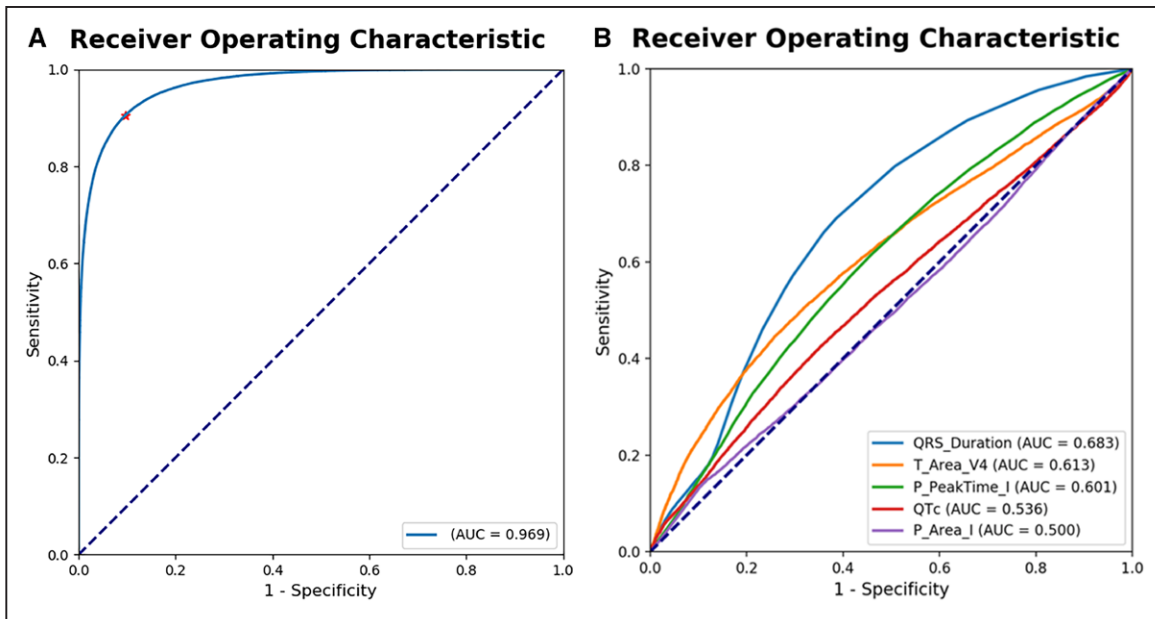


Figure 2. Receiver operating characteristic (ROC) of sex classification. **A**, Shown is the ROC curve for sex classification in the validation set. Overall area under the curve (AUC) was 0.97. **B**, Shows a separate network derived from specific ECG features—QRS duration, area of the T wave under V4, the time to peak of the P wave in lead I, the QTc interval, and the area under the P wave in lead I.

($r=0.78$, $P<0.00001$) even though the networks were trained separately.

Change in ECG-Estimated Age Over Time

For the 100 patients with longitudinal data, Figure 4 shows the numbers of patients with specific R-squared values between CNN-predicted age and chronologic age over the course of a minimum 20 year follow-up time span. The plurality of patients ($N=24$) had an R-squared value of 0.8 to 0.9 with 4 patients having an R-squared 0.9 to 1.0. Among patients with an R-squared value of 0.9 to 1.0, all had no significant medical comorbidities and were otherwise healthy, obtaining ECGs only for general physical examinations, executive physical examinations, or

for routine noncardiac orthopedic or minor surgical procedures (Figure 5A presents one patient example of this). These patients all had minimal difference between ECG-estimated age and chronological age (± 7 years). Several patients ($N=17$) had R-squared values of ≤ 0.3 . Among these patients, multiple incident events including new-onset low ejection fraction, incident myocardial infarction, new-onset hypertension and diabetes mellitus, cardiac transplantation, and major events such as hemodynamic shock occurred over the follow-up period, resulting in major deviations from chronological age during the year in which the events occurred, (Figure 5B presents one patient example of this). In most of these patients, with stabilization from the incident clinical insult, CNN-predicted age again approximated chronological age. (Figure 5B)

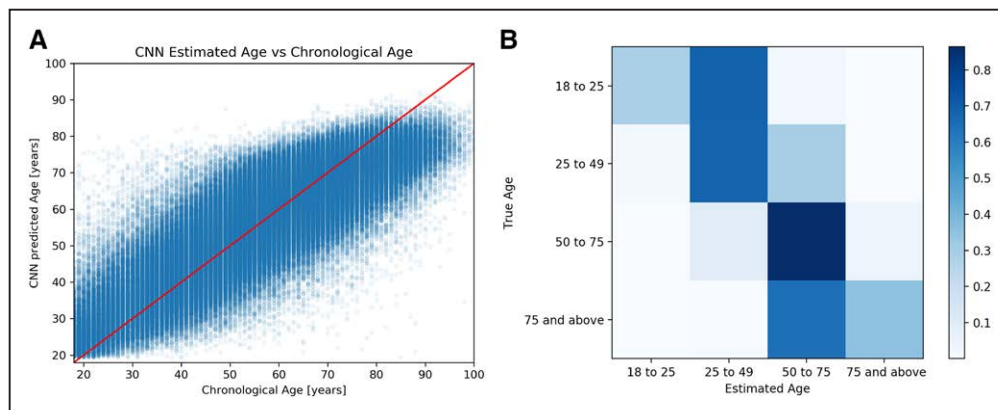


Figure 3. Convolutional neural network (CNN)-predicted age vs reported age. **A**, Shown is the estimated CNN-predicted ECG age (blue) vs the reported chronological age (in years; red—identity line). The model R squared was 0.7 with a Pearson correlation of $r=0.837$. **B**, Demonstrates a multi-group classification to the age (in years) groups of 18 to 25, 25 to 50, 50 to 75, and 75 and accuracy of CNN-predicted age (x axis—estimated age) vs the actual age (y axis) in terms of the percentage of patients with a specific actual age who had a specific corresponding CNN-predicted age within a similar range (eg, a patient from 18–25 y of age having a CNN-predicted age from 18–25 y).

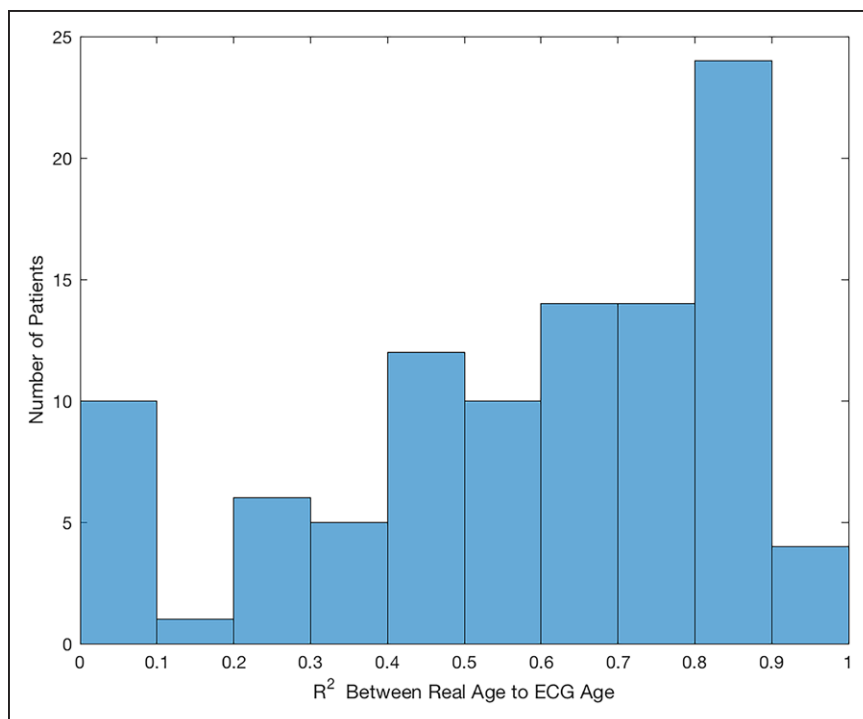


Figure 4. R-squared distribution for correlation between convolutional neural network (CNN)-predicted age and chronologic age. Shown is the distribution of the number of patients with specific R-squared values correlating CNN-predicted age (listed as ECG age in x axis) with chronologic age. Higher R squared values suggest a closer linear correlation between ages while low R squared values suggest poor correlation over time.

In patients with CNN-predicted age having a deviation from chronologic age of more than 7 years, there was a higher incidence of preexisting comorbidities including prior diagnoses of myocardial infarction, low ejection fraction, coronary disease, hypertension, and atrial fibrillation. ($P < 0.01$; Table). In one patient of interest, CNN-predicted age exceeded chronologic age for most of the follow-up period, but after subsequent cardiac transplant, CNN-predicted age decreased but was still greater than the age of the donor. After a period of weight loss with improvement in other comorbidities including hypertension and diabetes mellitus, the

patient's CNN-predicted age further decreased closer to the donor age and, in fact, to less than the patient's chronologic age. (Figure 5C)

DISCUSSION

In the age of smartphones, wearables and the internet of things, access to and acquisition of ECGs is easier and more cost-effective than ever.^{14,15} In addition to the demonstration of proof of concept that ECG is a biomarker, this work demonstrates that the CNN-predicted age appears to reflect changes in health sta-

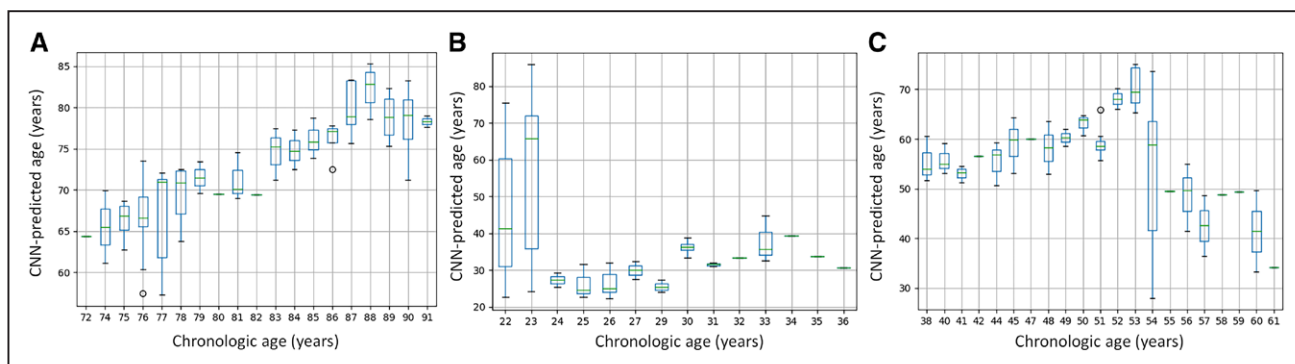


Figure 5. Correlation between convolutional neural network (CNN) predicted age and chronologic age in different patients. **A**, Shown is the correlation of CNN-predicted age (in years) with chronologic age (in years) at every year of life in an otherwise healthy patient. Note the linear association of age. Also note that the CNN-predicted age is younger than the chronological age. **B**, Shows correlation between CNN-predicted age and chronologic age in a woman with pulmonary hypertension with incident major hospitalizations. She was chronically managed with beta-blockers but developed supraventricular tachycardia and acute pulmonary embolus at age 22. In the second year, she developed acute cardiogenic and hemodynamic shock due to endocarditis and acute right ventricular failure. Thereafter, she was well controlled on medications for her pulmonary hypertension without further major incident events. With stabilization of her medical condition her CNN-predicted age dropped and matched her chronological age after age 23. **C**, shows an example of a patient with multiple incident myocardial infarctions, low ejection fraction, and eventual cardiac transplant. Before age 54, he had numerous myocardial infarctions, and CNN-predicted age significantly surpassed actual age. At age 54, he underwent a heart transplant from a donor who died at the age of 16, with subsequent decrease in CNN-predicted age to 50 y. He remained well for several years with subsequent loss of weight to normal weight and cessation of blood pressure and diabetes mellitus medications at the age of 60, with further drop in CNN-predicted age to below chronological age.

Table. Distribution of Preexisting Comorbidities Among 100 Patients With ECGs Over Decades

Comorbidity*	Overall (N=100)	CNN-Predicted Age Exceeds Chronologic Age by >7 y (N=49)	CNN-Predicted Within +7 to -7 of Chronologic Age (N=40)	CNN-Predicted Age Is Less Than Chronologic Age by >7 y (N=11)	P Value†
No comorbidities	18 (18%)	0	11 (28%)	7 (64%)	<0.01
Hypertension	56 (56%)	36 (73%)	16 (40%)	4 (36%)	<0.01
Diabetes mellitus	33 (33%)	16 (33%)	17 (43%)	0 (0%)	0.03
Low ejection fraction (EF <50%)	20 (20%)	19 (39%)	1 (3%)	0 (0%)	<0.01
Myocardial infarction	16 (16%)	15 (31%)	1 (3%)	0 (0%)	<0.01
Atrial fibrillation	27 (27%)	22 (45%)	5 (13%)	0 (0%)	<0.01
Coronary artery disease	47 (47%)	41 (84%)	6 (15%)	0 (0%)	<0.01
Any cardiac surgery	25 (25%)	25 (51%)	0 (0%)	0 (0%)	<0.01

*A given patient could be included in multiple different comorbidity categories. In other words, a given patient may have both atrial fibrillation and diabetes mellitus.

† χ^2 analysis used to determine significance of difference in disease incidence between groups.

tus. This may suggest that physiological age is distinct from chronological age, and may have useful clinical applications.

Accuracy of Age and Sex Estimation

In a chart review of a random selection of patients, several clinical factors associated with mortality including myocardial infarction, low ejection fraction, and coronary artery disease, appeared to correlate with deviations between CNN-predicted age and chronological age. This may account for some differences between CNN-predicted and chronological age, and supports the hypothesis that CNN-predicted age might prove a useful marker for physiological age. The higher incidence of these diseases with older age might also explain the corresponding differences in sex estimation as sex identification was more accurate in younger than older patients.

In fact, the lower accuracy of sex in older individuals may be hypothesized to be due to changes in hormone levels that may result in greater difficulty for the neural network to differentiate men from women. It is well recognized that circulating sex hormones may impact ECG characteristics.^{16–18} However, we could not account for hormonal status or gender in our population (only reported sex as male or female). Thus, there may have been factors in individual patients (eg, hormone therapy, gender reassignment, incident diseases, etc) that may have resulted in inaccuracy of sex or age assignment by the neural network. Further work is needed to better understand the nature of these outliers.

Certainty Values and the Black Box of Neural Networks

We have demonstrated that by applying advanced AI techniques to the ECG, it is feasible to accurately

extract sex and age in most individuals. Furthermore, the system is able to provide the user with a certainty value about its categorization of an individual to a specific sex. The ability of such a system to provide a certainty value is critical in future applications to make it less of a black box. Limitations to such AI approaches include current inability to simply explain the features of the ECG that result in a given result (such as assignment of a specific age or sex). Figure 6A through 6D show example ECGs from patients with their actual age/sex, their CNN-predicted age and sex, and their comorbid conditions (if any). The reasons for the limitation of explainability includes that the AI is not relying on a singular measure or a simple scoring metric, but rather a complex set of factors that include varying levels of nonlinear interaction.

This issue of neural network explainability faces the additional issue of understanding the clinical features driving the decision making of the algorithm. In this study, we did not seek to specifically compare the benefits of a CNN approach for predicting age and sex from the ECG to other approaches to machine learning, such as principal component analysis, Bayesian approaches, or other high-level statistical methodologies. The neural network approach is subject to inherent limitations, including the reliance on a big data approach, which is characterized by its own flaws related to data verification, the as-of-yet inability to oversee the system's reasoning given the limitations in achieving immediate explainability, and limited research on how best to apply it to medicine. However, neural networks may offer a higher level of accuracy than other statistical methods as it may operate independently of the biases of the investigator, only dependent on the data that is input. The reason for this is that a neural network learns through repeated exposure over time to associations between inputs. Of course, one key limitation is when consid-

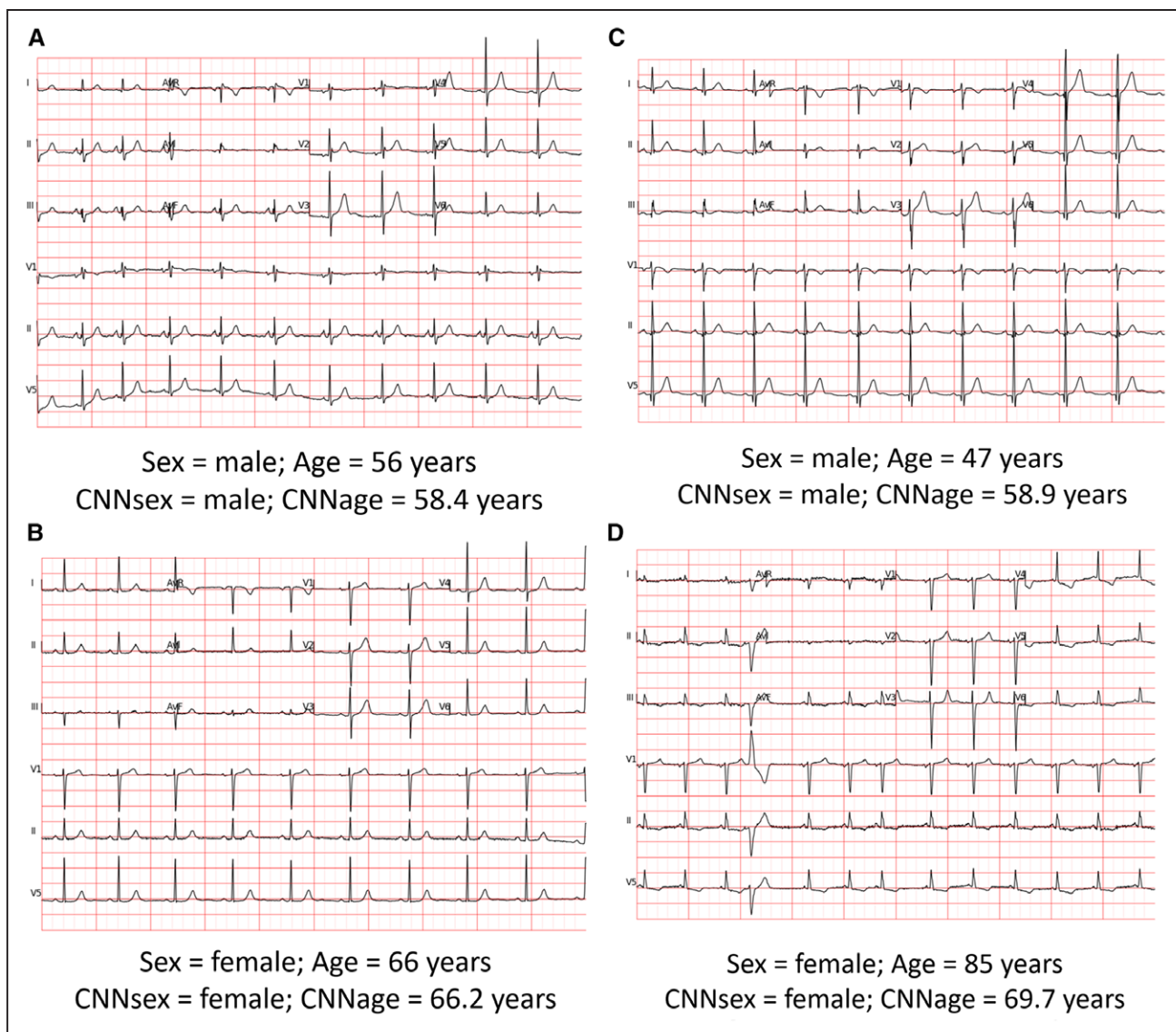


Figure 6. Example electrocardiograms from patients comparing actual age and sex against CNN-predicted age and sex.

Shown are 4 examples of patient ECGs. Patient (A) is a moderately healthy man without significant cardiovascular disease whose CNN-predicted age and sex were similar to actual. Similarly, patient (C) is a moderately healthy woman without significant cardiovascular disease whose CNN-predicted age and sex were similar to actual. Patient (B) is a man with history of myocardial infarction, hypertension, and diabetes mellitus whose CNN-predicted age was >10 y above his actual age. Finally, patient (D) is a very healthy 85 y old woman with no significant diseases who underwent a routine stress test as part of a general check-up.

ering the nature of the data inputted and resultant generalizability to real-world cohorts. For example, all of our cases were, effectively, patients, whether for general wellness evaluations or for evaluation of active cardiac conditions. Our population (both those used to train the algorithm and those in whom a deeper dive was performed) was randomly selected, but we cannot specifically speak to the percentage that had total absence of any cardiovascular disease. Thus, future work on (1) whether algorithms derived from such a cohort are relevant when studied in an ostensibly healthy population and (2) whether other demographic differences may exist in our population (eg, ethnicity, age distribution, etc) that further impact generalizability will need to be done.

Future Research and Potential Implications

Sex Identification

It is well recognized that hormonal concentrations can impact electrocardiographic findings. Prior work has suggested that sex might be identifiable using electrocardiography. Features that contribute to sex identification may include QT interval characteristics, ST segment, and others.^{9,16,18} The value of demonstrating that the ECG can accurately identify sex has implications for future research.¹⁵ It is possible that the hormonal shifts may drive changes on ECG. Furthermore, these findings may suggest that the ECG can be used in large population-based studies to reconfirm manual data entry regarding sex. These studies also question the relevance of includ-

ing both sex and ECG in risk prediction models. Prior work suggested that the addition of age and sex to the ECG did not result in improved prediction of a low ejection fraction.² This may be due to the fact that elements of the ECG are already accounting for both these factors. Finally, further work is needed to understand the potential relevance when there exists discordance between the CNN-predicted sex and actual sex. This last point is probably the most interesting as the question of why such an algorithm is wrong when it is generally correct may occur due to any number of factors (eg, exogenous hormone intake [whether illicit or prescribed], transgender status or efficacy of appropriate therapy in achieving gender transition).

Age Identification

While we generally define age as the number of years passed since birth, patients with the same biologic age may have different health statuses, resulting in differences in physiological age. Whether the age predicted by the CNN is more characteristic of physiological age (ie, accounting for a patient who may appear older than

stated age) than biologic age and thus predict individuals more likely to have shorter life expectancy is unclear. Our review of a small but randomly identified cohort of patients supports the concept of a correlation between incident and chronic diseases and deviations between CNN-predicted age and chronologic age. However, this cohort reflected a subset of patients with multiple ECGs over decades, and thus many may have had an indication for serial ECGs, reflecting a sicker cohort. Thus, further research is needed to understand why the CNN over or underestimates a patient's age in certain clinical situations. The algorithm output, the ECG age may indicate a risk for or protection against future morbidity and mortality. For example, if a patient's biologic age is 60 but their ECG age predicts that they are 70, it may suggest underlying disease and potential risk. Future research into whether CNN-predicted age may predict short and long-term patient outcomes ought to be performed, especially given our study was focused on predicting age at a point in time a specific ECG was obtained, and not purposed to predict future aging. Finally, identifying objective factors that differentiate

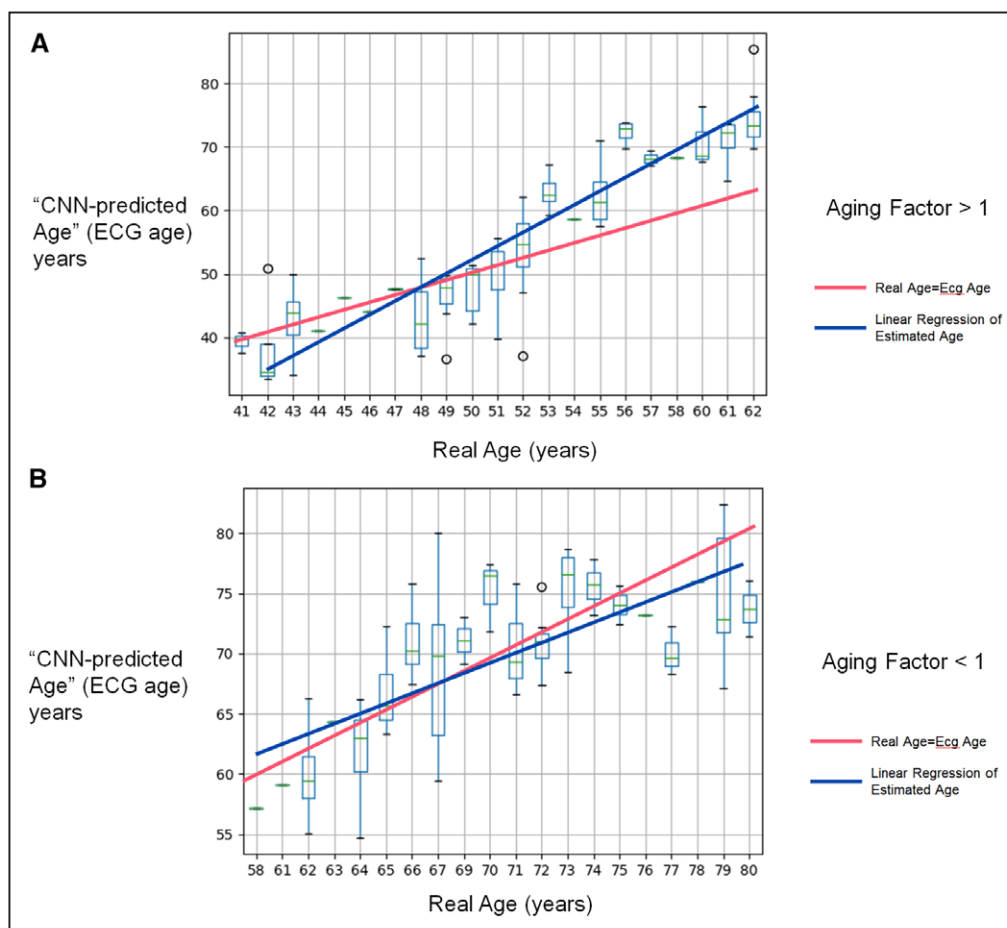


Figure 7. Potential use of linear regression modeling of ECG-estimated age progression as a biomarker. Shown are examples from 2 patients—one whose ECG-estimated age increased faster than chronologic age (**A**) and one whose ECG-estimated age increased slower than chronologic age (**B**). In these 2 patients' cases, patient (**A**) had an incident MI around age 50 and eventually developed a low ejection fraction. Patient (**B**) had hospitalizations in his late 60s for infectious and cancer issues but had no significant cardiovascular disease later in life along with resolution of his infectious and cancer diagnoses. The patient in (**A**) died at the age of 65 due to cardiovascular causes while the patient in (**B**) is still alive at the age of 85.

individuals' physiological health beyond chronologic age may also allow for an objective means to quantify individual health status. However, further outcomes research will be necessary to validate these concepts.

Given the small population in whom ECGs were acquired over decades, and given that our population had a preexisting indication for multiple ECGs and thus may be sicker than the general population (eg, the average age of our cohort was 58 years), implications of how people age (ie, temporal progression of their ECG-estimated age relative to their chronologic age) was beyond the scope of this study. However, it could be hypothesized that the CNN-predicted age may not just reflect preexisting comorbidities or the impact of incident events, but that temporal progression of CNN-predicted age relative to chronologic age per individual ECGs obtained over time could serve as a predictive biomarker for future events. Thus, future population-based research focusing on the potential for risk stratification based on the relative progression of CNN-predicted age compared with chronologic age is also necessary. (Figure 7)

Limitations

One key limitation to these results is that all individuals included were patients, and thus an ECG was obtained for some clinical indication. Whether these results are similarly accurate among an ostensibly healthy population is unknown and revalidation in such a cohort will be critical. The population as a whole was randomly selected and thus not precharacterized according to presence or absence of specific diseases or other factors thus making assumptions regarding their disease status or presence or absence of comorbidities difficult to assume since they were not rigorously characterized for this specific purpose. Furthermore, we are unable to indicate which features of the ECG contributed to the final network output and how the final output was reached, due to the very nature of neural networks. Finally, we could not consider other clinical characteristics (eg, we could only consider patients' self-reported sex) in the data set and thus understanding why certain outliers were in fact outliers is not possible with the current data set. This does reflect one critical limitation of our study as we could only consider sex as reported in the ECG system, where male/female is based on self-reported sex of the individual, and further details were not readily extractable from the health record without manual abstraction.

Conclusions

Applying AI to the ECG allows robust detection of patient sex and estimation of age. The ability of an AI algorithm to determine physiological age, with further validation, has the potential to serve as a rapid mea-

sure of overall health status, though will require further validation in well-characterized, prospectively followed cohorts.

ARTICLE INFORMATION

Received February 8, 2019; accepted July 3, 2019.

Guest Editor for this article was Andrew E. Epstein, MD, FAHA, FACC, FHRS. The Data Supplement is available at <https://www.ahajournals.org/doi/suppl/10.1161/CIRCEP.119.007284>.

Correspondence

Suraj Kapa, MD, Department of Cardiovascular Medicine, Mayo Clinic College of Medicine, 200 First St SW, Rochester, MN 55905. Email kapa.suraj@mayo.edu

Affiliations

Department of Cardiovascular Medicine (Z.I.A., P.A.F., P.A.N., F.L.-J., P.A.P., T.M.M., S.J.A., S.K.), Department of Business Development (D.J.L., G.S.), and Department of Health Sciences Research (C.G.S.), Mayo Clinic College of Medicine, Rochester, MN. Health Sciences Research, Mayo Clinic College of Medicine, Jacksonville, FL (R.E.C.).

Sources of Funding

This study was funded using institutional funds at Mayo Clinic for data collection and statistical analyses.

Disclosures

None.

REFERENCES

- Holmvang L, Lüscher MS, Clemmensen P, Thygesen K, Grande P. Very early risk stratification using combined ECG and biochemical assessment in patients with unstable coronary artery disease (A thrombin inhibition in myocardial ischemia [TRIM] substudy). The TRIM Study Group. *Circulation*. 1998;98:2004–2009. doi: 10.1161/01.cir.98.19.2004
- Voss A, Dietz R, Fiehring H, Kleiner HJ, Kurths J, Sapanin P, Vossing HJ, Witt A. High resolution ECG, heart rate variability and nonlinear dynamics: tools for high risk stratification. *Proceedings of Computers in Cardiology Conference*. 1993;261–264.
- Attia ZI, DeSimone CV, Dillon JJ, Sapir Y, Somers VK, Dugan JL, Bruce CJ, Ackerman MJ, Asirvatham SJ, Striemer BL, Bukartky J, Scott CG, Bennet KE, Ladewig DJ, Gilles EJ, Sadot D, Geva AB, Friedman PA. Novel bloodless potassium determination using a signal-processed single-lead ECG. *J Am Heart Assoc*. 2016;5:e002746.
- Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G, Pellikka PA, Enriquez-Sarano M, Noseworthy PA, Munger TM, Asirvatham SJ, Scott CG, Carter RE, Friedman PA. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med*. 2019;25:70–74. doi: 10.1038/s41591-018-0240-2
- Hossain SM, Ali AA, Rahman M, Ertin E, Epstein D, Kennedy A, Preston K, Umbricht A, Chen Y, Kumar S. Identifying drug (cocaine) intake events from acute physiological response in the presence of free-living physical activity. *IPSN*. 2014;2014:71–82.
- Lourenço A, Silva H, Fred A. Unveiling the biometric potential of finger-based ECG signals. *Comput Intell Neurosci*. 2011;2011:1–8. doi: 10.1155/2011/720971
- Pinto JR, Cardoso JS, Lourenço A, Carreiras C. Towards a continuous biometric system based on ECG signals acquired on the steering wheel. *Sensors (Basel)*. 2017;17:E2228.
- Malik M, Hnatkova K, Kowalski D, Keirns JJ, van Gelderen EM. QT/RR curvatures in healthy subjects: sex differences and covariates. *Am J Physiol Heart Circ Physiol*. 2013;305:H1798–H1806. doi: 10.1152/ajpheart.00577.2013
- Salama G, Bett GC. Sex differences in the mechanisms underlying long QT syndrome. *Am J Physiol Heart Circ Physiol*. 2014;307:H640–H648. doi: 10.1152/ajpheart.00864.2013

10. Ball RL, Feiveson AH, Schlegel TT, Starc V, Dabney AR. Predicting "heart age" using electrocardiography. *J Pers Med*. 2014;4:65–78. doi: 10.3390/jpm4010065
11. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogueira I, Yao J, Mollura D, Summers RM. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging*. 2016;35:1285–1298. doi: 10.1109/TMI.2016.2528162
12. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. 2015;arXiv:1502.03167.
13. Kingma DP, Ba J. Adam: a method for stochastic optimization. 2014;arXiv:1412.6980.
14. Haberman ZC, Jahn RT, Bose R, Tun H, Shinbane JS, Doshi RN, Chang PM, Saxon LA. Wireless smartphone ECG enables large-scale screening in diverse populations. *J Cardiovasc Electrophysiol*. 2015;26:520–526. doi: 10.1111/jce.12634
15. Kvedar JC, Fogel AL. mHealth advances clinical research, bit by bit. *Nat Biotechnol*. 2017;35:337–339. doi: 10.1038/nbt.3851
16. Kurokawa J, Kodama M, Clancy CE, Furukawa T. Sex hormonal regulation of cardiac ion channels in drug-induced QT syndromes. *Pharmacol Ther*. 2016;168:23–28. doi: 10.1016/j.pharmthera.2016.09.004
17. van Noord C, Dörr M, Sturkenboom MC, Straus SM, Reffelmann T, Felix SB, Hofman A, Kors JA, Haring R, de Jong FH, Nauck M, Uitterlinden AG, Wallaschofski H, Witteman JC, Völzke H, Stricker BH. The association of serum testosterone levels and ventricular repolarization. *Eur J Epidemiol*. 2010;25:21–28. doi: 10.1007/s10654-009-9406-z
18. Zhang Y, Ouyang P, Post WS, Dalal D, Vaidya D, Blasco-Colmenares E, Soliman EZ, Tomaselli GF, Guallar E. Sex-steroid hormones and electrocardiographic QT-interval duration: findings from the third national health and nutrition examination survey and the Multi-Ethnic Study of Atherosclerosis. *Am J Epidemiol*. 2011;174:403–411. doi: 10.1093/aje/kwr172