Journal of Electrocardiology 88 (2025) 153850

Contents lists available at ScienceDirect

Journal of Electrocardiology

journal homepage: www.jecgonline.com

An ECG-based machine-learning approach for mortality risk assessment in a large European population

Martina Doneda [a,b], Ettore Lanzarone [c], Claudio Giberti [d], Cecilia Vernia [e], Andi Vjerdha [e], Federico Silipo [f], Paolo Giovanardi [g,h,*]

[a] Politecnico di Milano, Department of Electronics, Information and Bioengineering, Via Ponzio 34/5, 20133 Milan, Italy
[b] National Research Council, Institute for Applied Mathematics and Information Technologies, Via Alfonso Corti 12, 20133 Milan, Italy
[c] University of Bergamo, Department of Management, Information and Production Engineering, Via Einstein 2, 24044 Dalmine (BG), Italy
[d] University of Modena and Reggio Emilia, Department of Sciences and Methods for Engineering, Via G. Amendola 2, 42122 Reggio Emilia, Italy
[e] University of Modena and Reggio Emilia, Department of Physics, Informatics and Mathematics, Via Campi 213/b, 41125 Modena, Italy
[f] Health Authority and Services and Azienda Ospedaliero-Universitaria of Modena, Department of Clinical Engineering, Via del Pozzo 71, 41100 Modena, Italy
[g] Health Authority and Services of Modena, Department of Primary Care, Via del Pozzo 71, 41100 Modena, Italy
[h] Modena University Hospital, S. Agostino–Estense Hospital, Via Giardini 1355, 41126 Baggiovara, Modena, Italy

ARTICLE INFO

Keywords:
Age
Mortality
ECG
Heart rate
Corrected QT
Machine learning

ABSTRACT

Aims: Through a simple machine learning approach, we aimed to assess the risk of all-cause mortality after 5 years in a European population, based on electrocardiogram (ECG) parameters, age, and sex.
Methods: The study included patients between 40 and 90 years old who underwent ECG recording between January 2008 and October 2022 in the metropolitan area of Modena, Italy. Exclusion criteria established a patient cohort without severe ECG abnormalities, namely, tachyarrhythmias, bradyarrhythmias, Wolff-Parkinson-White syndrome, second- or third- degree AV block, bundle-branch blocks, more than three premature beats, poor signal quality, and presence of pacemakers and implantable cardioverter- defibrillators. Mortality was assessed using a set of logistic regression models, differentiated by age group, to which the Akaike Information Criterion was applied. Model fitting was evaluated using confusion matrix-related performance metrics, the area under the receiver operating characteristic (ROC) curve (AUC), and the predictive significance against the no-information rate (NIR).
Results: 53692 patients were enrolled, of whom 14353 (26.73 %) died within 5 years of ECG registration. The logistic regression model distinguished between those who died and those who survived based on the predicted mortality probability for all age groups, obtaining a significant difference between the predicted mortality and the NIR in 14 of the 55 age groups. Good accuracy and performance metrics were observed, resulting in an average AUC of 0.779.
Conclusions: The proposed model showed a good predictive performance in patients without severe ECG abnormalities. Therefore, this study highlights the potential of ECGs as prognostic rather than diagnostic tools.

Introduction

Cardiovascular (CV) diseases have emerged as the leading cause of death worldwide over the last half century [1]. In parallel, the use of risk assessment and prognostic tools has grown in clinical practice, and several clinical and instrumental scores have been proposed.

However, finding effective, low-cost, non-invasive, and readily deployable risk stratification methodologies remains a major challenge, despite the availability of basic disease screening equipment and the use of electronic health record systems [2].

Among the available CV diagnostic tests, the electrocardiogram (ECG) provides detailed information on the structure and electrical activity of the heart. The ongoing diffusion of digitized ECGs, improvements in ECG processing, and the creation of sizable ECG databases have

* Corresponding author at: Cardiology Service, Department of Primary Care, Health Authority and Services of Modena, Via del Pozzo 71, 41100 Modena, Italy.
E-mail addresses: martina.doneda@polimi.it, martina.doneda@mi.imati.cnr.it (M. Doneda), ettore.lanzarone@unibg.it (E. Lanzarone), claudio.giberti@unimore.it (C. Giberti), cecilia.vernia@unimore.it (C. Vernia), 276500@studenti.unimore.it (A. Vjerdha), f.silipo@ausl.mo.it (F. Silipo), p.giovanardi@ausl.mo.it (P. Giovanardi).

https://doi.org/10.1016/j.jelectrocard.2024.153850

Available online 29 November 2024

opened up a wide range of new opportunities for ECG-based diagnostics.

Many studies have found clinical relevance for several ECG parameters and advances in computational power have recently enabled a more widespread use of statistical tools, such as neural networks, thus further expanding the opportunities for ECG interpretation [3,4].

We argue that risk-scoring systems based on ECG parameters and developed with the use of statistically sound techniques and artificial intelligence (AI) tools could further improve the prognostic value of the ECG.

In this study, we exploit standard ECG parameters, age, and sex to assess the risk of all-cause mortality in an Italian population without severe ECG abnormalities, using a simple machine learning (ML) approach. This particular population makes the prediction task more challenging, due to the heterogeneity of individuals, which makes the proposed approach a useful tool to support mass screening.

To the best of our knowledge, this is the first European study aiming to create and validate such an ECG-based ML approach on a large population.

Methods

Population and data

The study was conducted according to the ethical standards of the Declaration of Helsinki (1975, revised 2013) and was approved by the local ethics committee of the *Area Vasta Emilia Nord*, Modena, Italy (protocol number 2605/2021, approval date September 21, 2021).

Due to the anonymous, retrospective, and observational nature of this study, informed consent could not be obtained from the enrolled patients. All data were anonymized by associating a numerical personal identification code unique to each patient.

Study population

The potential patient population coincided with the population of the metropolitan area of Modena, located in central-southern Emilia Romagna, northern Italy, with a total of 702635 inhabitants spread across 47 municipalities.[1] In this area, as in the whole of Italy, healthcare is provided mainly through the branches of the public National Healthcare System.

Patients with a digitized ECG stored in any facility in the metropolitan area of Modena (emergency departments, hospital units and services, and inpatient and outpatient clinics) from January 2008 to October 2022 were eligible for inclusion in the study.

The main CV risk factors and their prevalence in the resident population were as follows: diabetes, 5.7 %; systemic arterial hypertension, 23.4 %; dyslipidemia, 38.8 %; tobacco smoking, 17.4 %. The prevalence of CV diseases and their main comorbidities in the resident population were as follows: CV diseases, 7.6 %; cerebrovascular diseases, 1.5 %; chronic obstructive pulmonary diseases 7.5 %; dementia, 3.1 %; cancer, 4.9 %; chronic kidney disease, 1.2 %.

Electrocardiography

ECGs were recorded at rest in the supine position using a standard 12-lead tracing at 25 *mm/s* speed and 10 *mm/mV* amplitude, with a sampling rate of at least 500 *samples/second*, and were archived into a MUSE® centralized electronic archive (GE Marquette Medical System, Milwaukee, WI, USA). Automated analyses were performed using a digitized, computer-assisted multi-channel program (GE 12SL ECG Analysis), a healthcare system that uses validated algorithms.

ECG diagnoses were then supervised and confirmed by experienced cardiologists to complete the quality control.

The following ECG parameters were automatically measured from the ECGs stored in the MUSE® electronic archive, to serve as predictors [5,6]:

- Heart rate (HR), expressed in beats per minute (*BPM*);
- QRS axis and T wave axis, expressed in degrees (*deg*), derived from the limb leads;
- Frontal QRS-T angle, expressed in *deg*, calculated as the absolute difference between the QRS and T wave axis [7];
- P wave duration, expressed in milliseconds (*msec*), obtained by averaging the wave duration over the 12 ECG leads;
- T wave duration, expressed in *msec*, obtained by averaging the wave duration over the 12 ECG leads;
- QRS duration, expressed in *msec*, obtained by averaging the complex duration over the 12 ECG leads;
- PR interval, expressed in *msec*, obtained by averaging the interval duration over the 12 ECG leads;
- Corrected QT (QTc) interval, expressed in *msec*, calculated using Bazett's, Fridericia's, and Framingham methods and obtained by averaging the interval duration over the 12 leads and
- ST-segment upward (ST elevation) or downward (ST depression) displacement, expressed in millivolts (*mV*), automatically estimated in the V5 lead at the J-point (the point where the QRS complex ends and joins the ST segment) and at the mid-point (the point located halfway between the J-point and the beginning of the T wave).

Other collected data

Patient's age in years, collected on the day of ECG registration, and sex as a binary variable, equal to 1 for females and 0 for males, were also recorded and included as predictors.

Follow-up

The follow-up period was 5 years. All-cause mortality was evaluated for each patient within 5 years after the date of the (first) ECG recording (i.e., the mortality variable was equal to 1 if the patient died within 5 years or equal to 0 if they survived for at least 5 years).

All-cause mortality and emigrations from the metropolitan area of Modena were anonymously assessed using the electronic medical records of the Health Authority and Services of Modena, Italy.

Inclusion and exclusion criteria

Patients younger than 40 years and older than 90 years on the date of ECG recording were excluded, as were patients who emigrated during the follow-up period. Furthermore, only patients with a follow-up period of at least 5 years or who died within 5 years of the (first) ECG recording were included in the study.

ECGs were discarded if they were classified as incomplete or if they presented technical problems, such as poor signal quality, waveform recognition errors, or electrode interchanges. The presence of a pacemaker or an implantable cardioverter-defibrillator was an exclusion criterion too. ECGs were also excluded in the presence of atrial fibrillation or atrial flutter, supraventricular tachycardia, Wolff-Parkinson-White syndrome, second- or third-degree AV block, complete or incomplete left or right bundle-branch block, or had more than 3 premature atrial or ventricular beats.

In the case of patients with more than one ECG stored in the centralized dataset, only the earliest recorded one was used for this study. No information was available on the clinical reason(s) the ECGs were performed for.

Statistical analyses

We considered the all-cause mortality rate after 5 years as the dependent output variable. We linked this rate to the predictors using a binomial logistic regression model, the output of which is the probability $\rho \in [0,1]$, that the event of death will occur.

The formulation of a logistic model is as follows:

[1] Official website of the *Provincia di Modena* (www.provincia.modena.it), accessed October 1, 2022.

$$\rho = \frac{exp(\alpha_0 + \sum_{i=1}^{N} \alpha_i X_i)}{1 + exp(\alpha_0 + \sum_{i=1}^{N} \alpha_i X_i)} \quad (1)$$

where X_i ($i = 1, \ldots, N$) denotes the predictors, α_i the coefficients to be estimated, and α_0 the intercept of the model.

Different age groups were analyzed, considering all lower bounds between 40 and 85 years and upper bounds between 45 and 90 years, with 5-year steps on both bounds, for a total of 55 different age groups. This differentiation was made because of the impact of age on mortality, as assessed by preliminary analyses and previous works [8]. The use of restricted age groups also allowed us to confirm that the performance of the logistic regression was not inferior when studying large age groups. This sub-division was also functional in assessing the effect of age-dependent ECG predictors.

We also applied the Akaike information criterion (AIC) to the model fitted with the data of each age group to identify the most important predictors and discard the others, with the ultimate aim of avoiding overfitting [9].

For each of the 55 age groups, we divided the data into two parts: 80 % were used to train the model, while 20 % were used to test it. Within the training set, we used a 10-fold cross-validation to train our model, ensuring that each entry was used as a validation sample only once. Finally, we stratified the ρ values according to the corresponding observed mortality data. This allowed us:

1. to analyze whether the logistic model was able to distinguish the two populations based on ρ. Operatively, a Wilcoxon rank-sum test was applied to verify whether the two distributions of ρ values were different. This can be done on both the training set and the test set.
2. to convert the predicted probability ρ into a binary output. Indeed, we computed a threshold k as the mean between the medians of the two stratified distributions. This threshold was applied to classify the patients into those at risk, with $\rho > k$, or not. Because k is part of the model, this could only be done on the training set. Medians were chosen over means to reduce the impact of outliers. In the presence of class imbalance, as in our case, the value of the threshold k may differ considerably from 0.5 [10].

All analyses were performed using the R open-source statistical software, version 4.2.1 (R-project.org).

Performance metrics

The goodness-of-fit of the binarized output, based on the threshold k, was assessed using the confusion matrix, which reports the totals of *true positive* (TP), *true negative* (TN), *false positive* (FP), and *false negative* (FN) predictions over all subjects in the test set. These values were used to compute a set of performance metrics, reported in the Supplementary Materials. Performance was also assessed employing the receiver operating characteristic (ROC) curve, which evaluates the behavior of a binary classification model at various classification thresholds by plotting *Sensitivity* against 1-*Specificity* for different threshold values. Based on the ROC, we also computed the area under the curve (AUC) as a performance metric.

Finally, we conducted a hypothesis test (one-tailed binomial exact test) to assess whether the accuracy achieved by our model was higher than the non- information rate (NIR), i.e., the prevalence of the largest class [11].

A p-value less than 0.05 on the test was used to determine the significance of our predictions.

Results

A total of 375207 ECGs were stored in the MUSE® centralized electronic archive of the National Health System facilities of Modena from January 2008 to October 2022.

Because some patients had more than one ECG stored in the dataset, and only the first one for each patient was included, 117630 ECGs were discarded. Patients with multiple ECGs recorded in different facilities were detected by tracking an individual identification code in the MUSE archive. Then, an additional 13466 ECGs were excluded because they were incorrect or had poor signal quality.

A total of 69023 subjects were excluded because their age was outside the age range for enrollment, and 80710 were excluded because they did not complete the follow-up period or emigrated.

Abnormal ECGs were manually flagged as such in the system by physicians. Based on this information 40686 patients were excluded due to the following conditions: 2803 had a pacemaker or a cardioverter defibrillator, 15930 suffered from atrial fibrillation or atrial flutter, 5301 from supraventricular or ventricular tachycardia or numerous premature beats, 803 from second- or third-degree atrioventricular block, and 15849 from complete or incomplete left or right bundle-branch block.

Based on the above-listed criteria, 53692 patients were enrolled; 39339 of them (73.27 %) survived for 5 years, while 14353 (26.73 %) died within 5 years of the date of ECG registration. The enrolled patients included 26883 males (50.07 %) and 26809 females (49.93 %). The enrolled patients' mean age was 64.49 years (Fig. 1).

A total of 12919 subjects (corresponding to 24.06 % of the sample) were enrolled from the emergency department. The remaining patients (40773 subjects, corresponding to 75.94 % of the sample) were enrolled from hospital units and services (including internal medicine, cardiology, surgery, orthopedics, and gynecology) and from inpatient and outpatient clinics (such as anaesthesiology for preoperative risk assessment, cardiology, diabetology, and geriatrics).

The patients enrolled included healthy subjects and patients with acute conditions or chronic diseases, provided they had no severe ECG abnormalities. In the enrolled population, the prevalence of CV risk factors, CV diseases, and comorbidities aligned with those in the resident population (see Section 2.1.1) but were not distinguishable at the individual level due to the retrospective and anonymous nature of the study.

Of the subjects included in the study, 19602 presented one or more of the following non-severe ECG abnormalities: QRS-T angle >90 *deg*, found in 2720 records; PR interval > 200 *msec*, found in 3611 records; QTc > 460 *msec*, found in 8663 records; QRS axis < −15 *deg*, found in 10110 records and QRS axis > +90 *deg*, found 841 records.

5-years all-cause mortality rates for the 55 age groups are reported in Table 1.

Preliminary analyses suggested to not consider two of the three QTc

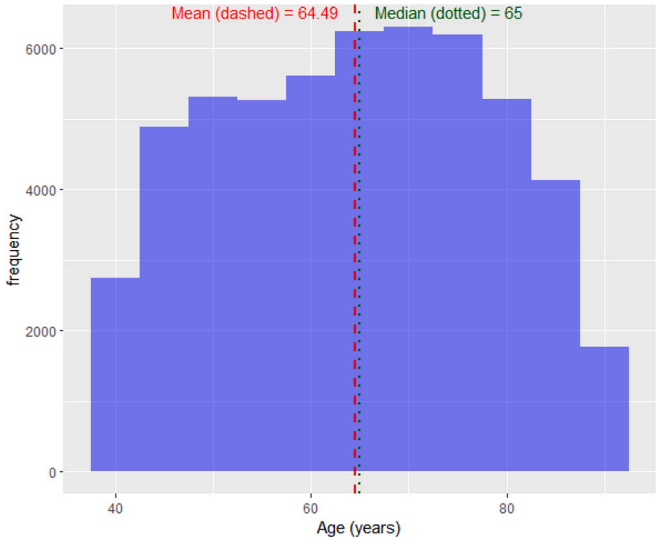

Fig. 1. Age distribution of the enrolled population.

Table 1
All-cause mortality at 5 years stratified over the age groups.

		Age upper bound									
		90	85	80	75	70	65	60	55	50	45
Age lower bound	**85**	82.03 %									
	80	72.30 %	65.72 %								
	75	60.25 %	53.62 %	44.06 %							
	70	51.36 %	45.44 %	37.59 %	31.33 %						
	65	43.98 %	38.64 %	31.79 %	26.04 %	20.60 %					
	60	38.96 %	34.10 %	27.99 %	22.81 %	18.18 %	15.26 %				
	55	35.17 %	30.71 %	25.16 %	20.45 %	16.36 %	13.78 %	11.69 %			
	50	31.88 %	27.72 %	22.61 %	18.24 %	14.46 %	11.96 %	9.90 %	8.07 %		
	45	29.02 %	25.13 %	20.36 %	16.25 %	12.69 %	10.26 %	8.22 %	6.39 %	4.81 %	
	40	26.73 %	23.05 %	18.56 %	14.67 %	11.32 %	8.98 %	7.04 %	5.34 %	3.94 %	2.93 %

methods and one of the two points used for ST-segment evaluation from the pool of predictors because of their strong collinearities. Therefore, because Bazett's formula for QTc and ST fluctuation at the midpoint of the V5 derivation revealed greater predictive power, the other variables were not considered. Regarding QTc, we acknowledge that there is no consensus regarding the best correction method; however, despite its limitations, Bazett's formula revealed a greater prognostic value in some works [12,13], in particular in Giovanardi et al. [14], who considered the same study population as this work.

Table 2 shows the values of the predictors included in the regression model, distinguishing between males and females. Continuous variables are reported as medians together with 2nd and 98th percentiles, as none of the variables resulted to be normally distributed (normality assessed by means of Q-Q plots). After applying the AIC in all age groups, we assessed the goodness-of-fit of the logistic regression models in terms of the *p*-values of the Wilcoxon rank- sum test, as described at the end of Section 2.2. The test was employed to verify whether the distributions of ρ values, stratified according to the observed mortality, were statistically different. All tests performed returned *p*-values of less than 0.001, considering the predictions made in both the training and test sets. These results implied that the logistic regression model was always capable of distinguishing the two populations (deceased and survivors) based on the predicted ρ. The values obtained for threshold k in each age group are reported in Table 3(a), while the boxplots of the stratified distributions can be found in the Supplementary Materials.

The comparison of model performance with respect to the NIR on the test set is reported in Table 3(b), in terms of the p-values of the hypothesis test. Results showed a significant difference in 14 of the 55 age groups, mainly because there were too few patients in the less represented age groups, which led to non- significant differences.

Therefore, we report the performance metrics of the models only for the 14 significant age groups (Table 4). Interestingly, in all age groups, the values of accuracy and balanced accuracy do not differ substantially (average absolute difference equal to 0.65 %), further supporting good performance of the regression in these age groups, despite the imbalance between those who died and those who survived after 5 years.

Fig. 2 reports the ROCs for these age groups. The resulting average AUC over these groups was equal to 0.779. Details about the individual ROCs with their AUCs are reported in the Supplementary Materials.

Table 5 reports the results of model fitting, averaged over the 14 significant age groups. The *Weight* column reports the average value of the coefficient across the age groups multiplied by the average value of the predictor over the entire population. This approximates the impact of each predictor on the calculation of ρ to distinguish between predictors that tend to drive predictions and those that impact them only marginally, as sorted by the *Ranking* column. Columns *Global results* give general information on the predictors across the age groups. The column *Appearance tally* reports how many times the predictor was not discarded by the AIC in the 14 age groups analyzed, while *Effect direction* summarizes the effect of the predictor intuitively. The symbol "+" is used to denote that a predictor contributed positively to the value of ρ, increasing its value, and vice versa for symbol "-".

Discussion

This work presents a simple ML model to assess the risk of long-term mortality, built using standard ECG parameters, age, and sex. The model was trained on a large Italian population over a period of 15 years, and with the target of assessing the 5-year all-cause mortality, which is a clinically meaningful and well-defined endpoint.

We believe that this model has the potential for application in a variety of settings: in primary care, to evaluate the mortality risk; in surgical units, to assess the risk of surgical procedures; in hospital medical units and emergency departments, for the management of CV and other diseases.

The clinical scores currently in use may present limitations, and they rely heavily on CV risk factors, comorbidities, and blood analyses, which are available only for a subset of patients. These scores generally do not use ECG parameters, with the notable exception of TIMI and GRACE, which include an evaluation of the ST segment [15].

We purposely built the model using standard ECG parameters, age, and sex (and not clinical data, laboratory tests, or waveform analysis)

Table 2
Values of the predictors included in the regression for all patients included in the study and stratified by sex (median values with 2nd and 98th percentiles in brackets).

Predictor	All patients Median [2 %, 98 %]		Males Median [2 %, 98 %]		Females Median [2 %, 98 %]	
Age *[years]*	65	[41, 89]	65	[41, 88]	66	[41, 89]
HR *[bpm]*	71	[49, 110]	70	[48, 109]	73	[51, 110]
QRS axis *[deg]*	18	[−57, 87]	14	[−60, 88]	21	[−51, 87]
T wave axis *[deg]*	41	[−21, 140]	39	[−23, 137]	42	[−18, 142]
Frontal QRS-T angle *[deg]*	24	[1, 150]	26	[1, 152]	23	[1, 149]
P wave duration *[msec]*	102	[44, 130]	104	[44, 132]	100	[44, 128]
T wave duration *[msec]*	196	[35, 254]	192	[54, 248]	202	[1, 256]
QRS duration *[msec]*	88	[68, 124]	92	[72, 130]	84	[66, 116]
PR interval *[msec]*	156	[114, 230]	160	[116, 238]	152	[112,220]
QTc *[msec]*	433	[382, 507]	428	[379, 505]	437	[388, 507]
ST elevation/depression at midpoint *[mV]*	1.9	[−6.9, 11.7]	2.9	[−6.9, 13.1]	0.9	[−6.4, 7.8]

Table 3
k threshold values as determined in the training set (a) and p-values of the comparisons of the logistic regression with respect to the NIR (b) for the different age groups.

(a)

		Age upper bound									
		90	**85**	**80**	**75**	**70**	**65**	**60**	**55**	**50**	**45**
Age lower bound	**85**	0.817									
	80	0.710	0.653								
	75	0.590	0.531	0.429							
	70	0.510	0.452	0.367	0.306						
	65	0.452	0.391	0.316	0.258	0.198					
	60	0.408	0.351	0.283	0.228	0.180	0.152				
	55	0.374	0.323	0.259	0.208	0.163	0.140	0.119			
	50	0.353	0.302	0.235	0.191	0.148	0.123	0.102	0.078		
	45	0.335	0.284	0.224	0.173	0.134	0.107	0.084	0.063	0.047	
	40	0.320	0.269	0.210	0.165	0.123	0.100	0.074	0.056	0.039	0.031

(b)

		Age upper bound									
		90	**85**	**80**	**75**	**70**	**65**	**60**	**55**	**50**	**45**
Age lower bound	**85**	1.000									
	80	1.000	1.000								
	75	<0.001	<0.001	<0.001							
	70	<0.001	<0.001	0.254	0.999						
	65	<0.001	<0.001	0.999	1.000	1.000					
	60	<0.001	<0.001	1.000	1.000	1.000	1.000				
	55	<0.001	<0.001	1.000	1.000	1.000	1.000	1.000			
	50	<0.001	0.224	1.000	1.000	1.000	1.000	1.000	1.000		
	45	<0.001	0.328	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	40	<0.001	0.799	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 4
Performance metrics in the age ranges with a significant difference with respect to the NIR.

Age group	Sensitivity	Specificity	Accuracy	Balanced accuracy	PPV	NPV	Detection rate	Detection prevalence	Cohen's κ	F1-score
75–90	69.04 %	70.43 %	69.60 %	69.73 %	77.69 %	60.40 %	41.33 %	53.20 %	38.42 %	73.11 %
75–85	64.72 %	66.23 %	65.42 %	65.48 %	68.79 %	62.01 %	34.62 %	50.33 %	30.81 %	66.69 %
75–80	62.64 %	65.59 %	64.28 %	64.11 %	59.39 %	68.61 %	27.91 %	46.99 %	28.08 %	60.97 %
70–90	68.20 %	73.57 %	70.80 %	70.89 %	73.27 %	68.54 %	35.13 %	47.95 %	41.68 %	70.64 %
70–85	64.96 %	68.63 %	66.98 %	66.79 %	62.92 %	70.50 %	29.26 %	46.50 %	33.49 %	63.93 %
65–90	69.75 %	75.02 %	72.80 %	72.38 %	66.97 %	77.35 %	29.35 %	43.82 %	44.51 %	68.33 %
65–85	67.28 %	71.21 %	69.72 %	69.25 %	58.91 %	78.02 %	25.58 %	43.42 %	37.47 %	62.82 %
60–90	68.85 %	75.81 %	73.10 %	72.33 %	64.50 %	79.23 %	26.82 %	41.59 %	44.13 %	66.60 %
60–85	68.07 %	72.76 %	71.16 %	70.41 %	56.32 %	81.53 %	23.17 %	41.14 %	38.87 %	61.64 %
55–90	70.58 %	75.52 %	73.76 %	73.05 %	61.45 %	82.28 %	25.13 %	40.89 %	44.61 %	65.70 %
55–85	69.60 %	73.60 %	72.37 %	71.60 %	53.83 %	84.55 %	21.34 %	39.65 %	39.94 %	60.71 %
50–90	71.27 %	77.28 %	75.37 %	74.28 %	59.41 %	85.22 %	22.67 %	38.16 %	46.10 %	64.80 %
45–90	71.36 %	78.80 %	76.66 %	75.08 %	57.61 %	87.21 %	20.52 %	35.62 %	46.83 %	63.75 %
40–90	71.96 %	80.47 %	78.19 %	76.21 %	57.34 %	88.72 %	19.24 %	33.55 %	48.50 %	63.82 %

because these parameters are accurate, quick and easy to measure, and have established clinical relevance [16,17,18]. Moreover, as in some previously published studies [19,20], we were unable to accurately characterize the enrolled population from a clinical perspective because of the retrospective nature of this work such as the previous studies cited above. In addition, ECG parameters and intervals have precise reference values. Some of them show linear behaviors with respect to aging [8] and are continuous variables, as opposed to clinical data, which are often coarsely expressed in binary form.

Many ECG parameters have demonstrated clinical and prognostic value due to the influence of CV risk factors and aging on heart conduction tissue [21,22,23]. The contemporary use of multiple ECG parameters in risk models improves their clinical relevance, and the use of AI could provide further added value. Indeed, ECG represents an ideal substrate for AI models, as it is standardized, widely available, reproducible, and easy to transfer in a digital format [3,24].

Previous studies have already explored the prognostic importance of ECG with AI and ML techniques, albeit with some key differences with respect to our work. Sun et al. [25] recently developed an ECG-based ML model to predict short- and long-term mortality in a large cohort of Canadian hospitalized patients, combining ECG parameters, ECG traces, and laboratory tests. Similarly, Raghunath et al. [19] used a large regional electronic record from Geisinger, US, to train a deep neural network to predict 1-year mortality from voltage- time traces. Additionally, Tsai et al. [26] trained a deep learning model to predict 1-year all-cause mortality and major CV events (MACES) in a Chinese hospital population. Finally, Hughes et al. [20] developed a long-term mortality risk estimator using yet another deep learning model trained on a large ECG dataset from three American universities. Other authors applied ECG-based ML approaches to predict mortality or the appearance of MACES in specific risk categories, e.g., in patients with COVID-19 infection, in patients hospitalized in intensive care units, or in patients with acute myocardial infarction or pulmonary hypertension [27,28,29,30].

The above-mentioned studies included all types of ECG and heart rhythm abnormalities and therefore observed high mortality rates. In

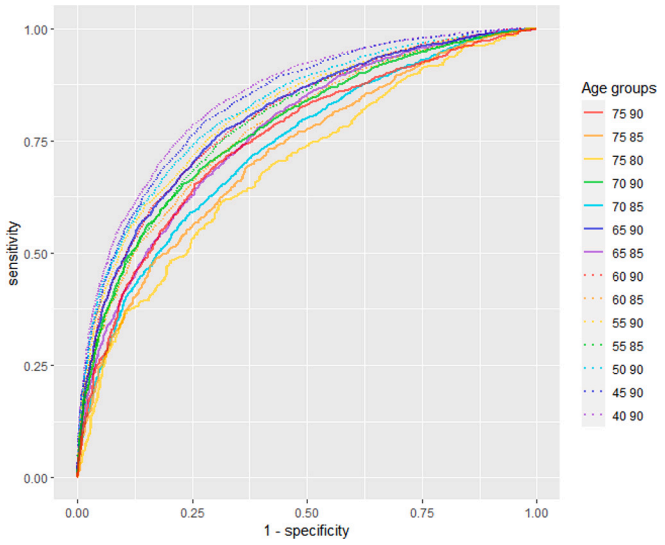

Fig. 2. ROC curves for the age ranges with a significant difference with respect to the NIR.

Table 5
Summary of model fitting over the 14 age groups with a significant difference with respect to the NIR.

Predictor	Weight Value	Ranking	Global results	
			Appearance tally	Effect direction
(Intercept)	(13.917)	1	(14)	
Age *[years]*	7.911	2	14	+
QTc *[msec]*	4.423	3	14	+
HR *[bpm]*	1.287	4	14	+
P wave duration *[msec]*	0.766	5	14	−
QRS duration *[msec]*	0.725	6	14	−
Female sex	0.559	7	14	−
PR interval *[msec]*	0.428	8	14	−
T wave duration *[msec]*	0.371	9	14	−
Frontal QRS-T angle *[deg]*	0.183	10	14	+
ST elevation/ depression at midpoint *[mV]*	0.140	11	1	−
QRS axis *[deg]*	0.058	12	14	+
T wave axis *[deg]*	0.049	13	9	+

Columns *weight* report the average value of the coefficient across the age groups multiplied by the average value of the predictor over the entire population, specifying the ranking of the predictors based on the computed values. Columns *global results* give general information on the predictors: column *appearance tally* reports how many times the predictor was not discarded by the AIC, while for *effect direction* the symbol "+" denotes that the predictor contributes positively to the value of ρ, increasing its value, and vice versa for symbol "-".

contrast, we believe that our work presents the following unique contributions:

1. To the best of our knowledge, this is the first such study conducted in a European population. AI models may perform differently on different populations and ethnicities [31,32], and previously published studies on this topic were conducted on American or Asian populations.
2. Our study enrolled subjects without severe ECG abnormalities or arrhythmias, whose ECGs are often overlooked by physicians because they do not contain relevant clinical findings per se. The enrolled population represents a wide spectrum of patients for whom the estimation of the individual risk is crucial and remains a major challenge. The perspective with which this study was developed included risk estimation in primary prevention and in subjects with low CV risk. In comparison, previous studies enrolled subjects with a complete spectrum of ECG abnormalities, including tachyarrhythmias, bradyarrhythmias, and bundle branch blocks; only Raghunath et al. made a specific analysis of a subgroup of patients with normal ECGs [25,19,26].
3. Our model was intentionally built using only common ECG parameters, age, and sex. Notably, previous studies showed the superiority of ML models constructed through waveform analysis [25,19]. Despite using rather simple parameters, we obtained only slightly lower AUC values than in previous studies (average AUC over the age groups equal to 0.779). Accuracy, sensitivity, and specificity values were all comparable with those reported in the literature. Moreover, the average value of the obtained F1- scores was 65.23 %, a moderate value aligned with those obtained in other published works dealing with the prediction of the same endpoint. For example, our F1-score is comparable to that of Sun et al. [25] and higher than that of Hughes et al. [20]. Similarly, our average positive predictive value (PPV) was higher than the one reported by Hughes et al. [20] (average PPV of 11.12 % over the various tested datasets) and comparable with the one reported by Sun et al. [25] (63.66 %) in all age groups, supporting the claim that our model is able to predict mortality, as almost two thirds of its positive predictions were correct. Regarding negative predictive value (NPV), its average value of 76.73 % confirmed the ability of our model to predict true negatives, determining which patients are not at risk. Comparing the values of detection rate and detection prevalence with raw data on 5-year mortality (Table 1), it is possible to observe that they both showed monotonicity with respect to the latter; in particular, detection prevalence was only −2.78 % lower than the mortality rate on average. In any case, the model produced a number of false positives, with an average difference between detection rate and mortality prevalence of 12.99 %. Finally, Cohen's κ values obtained showed moderate agreement, according to Landis & Koch's guidelines [33].
4. We followed the enrolled subjects over a 5-year follow-up, whereas previous studies have focused mainly on a shorter follow-up (e.g., 30-day or 1-year). Our longer follow-up period was linked to the lower mortality rates that we observed, given the exclusion of subjects with severe ECG abnormalities. The main advantage of this approach is that, using a longer follow-up, we can assess risk farther back in time, potentially providing opportunities for preventive interventions.
5. Previous studies have shown that clinical scores may have a lower predictive value in middle-aged and elderly adults [34,35], despite the increased prevalence of age related CV diseases. In contrast, our model performed better by including middle-aged and elderly patients.
6. Finally, our logistic regression model is simple and quick to use. It can also be easily retrained, as its coefficients can be updated after further observations. Therefore, it could play a useful role in mass screening. Another reason we chose such a simple ML model is that it is fully explainable and easily applicable, as the relationship between predictors and outcomes was of great interest in this study.

Several ECG parameters are known to play a clinical role in the monitoring of CV and non-CV diseases, but previous studies did not clearly reveal which predictors may be the most significant. In a cohort of 4615 elderly subjects, Lu et al. [36] observed that left ventricular hypertrophy, QTc, and PR interval were the ECG features mostly correlated with all-cause mortality, CV death, and unexplained death and developed a prognostic score based on the number of ECG abnormalities counted in each patient. Hirota et al. [37] simultaneously evaluated 438 ECG parameters, observing different correlations with all-cause and CV death, showing that the most important predictors were related to the QRS complex and the ST segment.

Our study showed that the ECG parameters considered have a

different prognostic importance, based on a global ranking of the predictors that we constructed regardless of the direction of the effect (Table 5). In particular, age, QTc, and HR had the greater prognostic impact with respect to the others. Notably, these parameters are well-known predictors of mortality [18,38,39].

We emphasize that if a predictor increased the value of ρ, that was not necessarily correlated with mortality (and vice versa). This is especially true for predictors with low fitted coefficients. Higher-order interactions could be present among predictors, and the effect direction of a particular predictor could also be the result of these effects, as a mathematical artifact to improve model fitting by counterbalancing linear and non-linear intra-predictor interactions. In our results, effect directions that did not match the literature were observed for P wave duration, QRS duration, PR interval, and QRS axis. It is worth noting that these predictors contributed only marginally to the estimation of ρ (see Table 5), but their presence contributed to the good fitting of the model, although their sign did not match the literature. To verify this, we retrained the model for the different age groups by not considering these predictors, and the performance of the alternative models decreased, albeit only slightly. Performance decreased (still marginally) when removing these predictors from the model without retraining. This shows that the fitted coefficients helped the models to make better estimates.

We also conducted analyses using only age and sex (without ECG parameters), obtaining consistently worse performance metrics and confirming the importance of ECG in predicting all-cause 5-year mortality. This justifies our methodological decision to perform the analyses in age-ranges and strengthens our assertion about the importance of considering ECG records that would otherwise have been overlooked as not clinically relevant. Biological age is an obvious predictor of outcome, but heart age, computed via ECG signals, and the gap between the two are receiving increasing recognition as interesting and helpful predictors of outcome, as reported by Lima et al. [40].

Finally, we believe that the appearance of acute diseases and the evidence of severe ECG abnormalities should be diagnosed, investigated, and treated regardless of the use of ECG-based prognostic tools. However, risk assessment remains a great challenge, and an ECG without severe abnormalities can give false reassurance, especially in asymptomatic or unconscious patients. The results of this study could prove useful in this field, as a decision aid tool to improve clinical evaluations, reduce costs, and optimize resource use.

Of course, although the ECG is increasingly becoming a prognostic tool rather than a diagnostic test, these innovative technological opportunities should still be considered as complementary to good clinical practice.

Limitations

Because of its anonymous nature, this work presents some clinical and methodological limitations. It was not possible to determine, at the individual level, the exact presence of CV risk factors, CV diseases, and comorbidities. Similarly, the prevalence of patients using QTc modifying drugs, cardiotoxic drugs, and HR modifying drugs in our study population was unknown. For the same reason, we could not precisely estimate the prevalence of MACES or distinguish between CV mortality and other causes of death. Another limitation is also one of the innovative points of the study, namely it being carried out in a European population, thereby determining an inherent bias.

Another limitation is the inclusion only of patients with a follow-up period of at least 5 years or who died within 5 years; excluding patients with an observation period shorter than 5 years unless they died may have over-represented mortality. However, as we were not interested in describing the population, this was done on purpose because of the low mortality rate in the data and in light of the patient-specific approach considered in the study.

Finally, external validation testing was not possible, but we performed an internal cross-validation with a 10-fold strategy on a segregated 20 % of the enrolled population.

Conclusions

This is the first European study using a simple ML approach to assess long- term all-cause mortality risk, trained on a large Italian population. Despite the limitations, we claim that the results are interesting and useful for clinical practice because the proposed ML tool is easily interpretable, reproducible, and applicable. The sample size is large and significant, and the endpoint is relevant and accurate. The ML tool was intentionally developed using a simple mathematical description (logistic regression) and includes only common ECG parameters, together with age and sex. Further studies can be conducted, but our results underline the potential prognostic role of the proposed ECG- based ML approach, especially in subjects who do not show severe or clinically manifest ECG abnormalities. Indeed, this is a large and heterogeneous group of patients in whom estimating mortality risk remains a major challenge.

Funding

Cecilia Vernia acknowledges the financial support of the European Union - Next Generation EU (Grant PRIN 2022B5LF52). The other authors did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Martina Doneda: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Ettore Lanzarone:** Conceptualization, Methodology, Software, Formal analysis, Writing – review & editing. **Claudio Giberti:** Conceptualization, Formal analysis, Writing – review & editing. **Cecilia Vernia:** Validation, Formal analysis, Data curation, Writing – review & editing. **Andi Vjerdha:** Validation, Data curation, Writing – review & editing. **Federico Silipo:** Investigation, Data curation, Writing – review & editing, Supervision. **Paolo Giovanardi:** Conceptualization, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Supervision, Visualization, Project administration.

Declaration of generative AI and AI-assisted technologies in the writing process

The authors declare that the work has been written without the aid of AI-assisted technologies.

Declaration of competing interest

The authors declare no conflict of interest.

Data availability

The data used in this article can be shared upon request to the corresponding author with permission from the Health and Services Authority of Modena.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jelectrocard.2024.153850.

References

[1] World Health Organisation. The top 10 causes of death. www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death; 2020 (Accessed October 1st, 2022).

[2] Patel MS, Volpp KG, Asch DA. Nudge units to improve the delivery of health care. N Engl J Med 2018;378(3):214. https://doi.org/10.1056/NEJMp1712984.

[3] Ambale-Venkatesh B, Yang X, Wu CO, et al. Cardiovascular event prediction by machine learning: the multi-ethnic study of atherosclerosis. Circ Res 2017;121(9):1092–101. https://doi.org/10.1161/CIRCRESAHA.117.311312.

[4] Mincholé A, Camps J, Lyon A, Rodríguez B. Machine learning in the electrocardiogram. J Electrocardiol 2019;57:S61–4. https://doi.org/10.1016/j.jelectrocard.2019.08.008.

[5] Rijnbeek PR, Van Herpen G, Bots ML, et al. Normal values of the electrocardiogram for ages 16–90 years. J Electrocardiol 2014;47(6):914–21. https://doi.org/10.1016/j.jelectrocard.2014.07.022.

[6] van der Ende MY, Siland JE, Snieder H, van der Harst P, Rienstra M. Population-based values and abnormalities of the electrocardiogram in the general Dutch population: the LifeLines cohort study. Clin Cardiol 2017;40(10):865–72. https://doi.org/10.1002/clc.22737.

[7] Chua KC, Teodorescu C, Reinier K, et al. Wide QRS-T angle on the 12-Lead ECG as a predictor of sudden death beyond the LV ejection fraction. J Cardiovasc Electrophysiol 2016;27(7):833–9. https://doi.org/10.1111/jce.12989.

[8] Giovanardi P, Vernia C, Tincani E, Giberti C, Silipo F, Fabbo A. Combined effects of age and comorbidities on electrocardiographic parameters in a large non-selected population. J Clin Med 2022;11(13):3737. https://doi.org/10.3390/jcm11133737.

[9] Akaike H. A new look at the statistical model identification. IEEE Trans Automat Contr 1974;19(6):716–23. https://doi.org/10.1109/TAC.1974.1100705.

[10] Lanzarone E, Finotello A, Pane B, et al. Prediction model of isolated iliac and abdominal aneurysms. Eur J Clin Invest 2021;51(7):e13517. https://doi.org/10.1111/eci.13517.

[11] Kubát M, Matwin S. Addressing the curse of imbalanced training sets: one-sided selection. In: Proceedings of the 14th International Conference on Machine Learning. 97; 1997. p. 179–86.

[12] Noseworthy PA, Peloso GM, Hwang SJ, et al. QT interval and long-term mortality risk in the Framingham heart study. Ann Noninvasive Electrocardiol 2012;17(4):340–8. https://doi.org/10.1111/j.1542- 474X.2012.00535.x.

[13] Yazdanpanah MH, Naghizadeh MM, Sayyadipoor S, Farjam M. The best QT correction formula in a non-hospitalized population: the Fasa PERSIAN cohort study. BMC Cardiovasc Disord 2022;22(1):52. https://doi.org/10.1186/s12872-022-02502-2.

[14] Giovanardi P, Vernia C, Roversi S, et al. Prognostic value and limits of heart rate and QT—corrected in a large population. Hearts 2024;5(2):225–35. https://doi.org/10.3390/hearts5020015.

[15] Yanqiao L, Shen L, Yutong M, Linghong S, Ben H. Comparison of GRACE and TIMI risk scores in the prediction of in-hospital and long-term outcomes among east Asian non-ST-elevation myocardial infarction patients. BMC Cardiovasc Disord 2022;22(1):4. https://doi.org/10.1186/s12872-021-02311-z.

[16] Aro AL, Anttonen O, Kerola T, et al. Prognostic significance of prolonged PR interval in the general population. Eur Heart J 2014;35(2):123–9. https://doi.org/10.1093/eurheartj/eht176.

[17] Whitbeck MG, Charnigo RJ, Shah J, et al. QRS duration predicts death and hospitalization among patients with atrial fibrillation irrespective of heart failure: evidence from the AFFIRM study. Europace 2014;16(6):803–11. https://doi.org/10.1093/europace/eut335.

[18] Nielsen JB, Graff C, Rasmussen PV, et al. Risk prediction of cardiovascular death based on the QTc interval: evaluating age and gender differences in a large primary care population. Eur Heart J 2014;35(20):1335–44. https://doi.org/10.1093/eurheartj/ehu081.

[19] Raghunath S, Ulloa Cerna AE, Jing L, et al. Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network. Nat Med 2020;26(6):886–91. https://doi.org/10.1038/s41591-020-0870-z.

[20] Hughes JW, Tooley J, Torres Soto J, et al. A deep learning-based electrocardiogram risk score for long term cardiovascular death and disease. NPJ Digital Med 2023;6(1):169. https://doi.org/10.1038/s41746-023- 00916-6.

[21] Aro AL, Huikuri HV, Tikkanen JT, et al. QRS-T angle as a predictor of sudden cardiac death in a middle-aged general population. Europace 2012;14(6):872–6. https://doi.org/10.1093/europace/eur393.

[22] Rautaharju PM, Zhang Z, Gregg RE, et al. Normal standards for computer-ECG programs for prognostically and diagnostically important ECG variables derived from a large ethnically diverse female cohort: the Women's Health Initiative (WHI). J Electrocardiol 2013;46(6):707–16. https://doi.org/10.1016/j.jelectrocard.2013.05.136.

[23] Jørgensen PG, Jensen JS, Marott JL, Jensen GB, Appleyard M, Mogelvang R. Electrocardiographic changes improve risk prediction in asymptomatic persons age 65 years or above without cardiovascular disease. J Am Coll Cardiol 2014;64(9):898–906. https://doi.org/10.1016/j.jacc.2014.05.050.

[24] Siontis KC, Noseworthy PA, Attia ZI, Friedman PA. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. Nat Rev Cardiol 2021;18(7):465–78. https://doi.org/10.1038/s41569-020-00503-2.

[25] Sun W, Kalmady SV, Sepehrvand N, et al. Towards artificial intelligence-based learning health system for population-level mortality prediction using electrocardiograms. NPJ Digital Med 2023;6(1):21. https://doi.org/10.1038/s41746-023-00765-3.

[26] Tsai DJ, Lou YS, Lin CS, et al. Mortality risk prediction of the electrocardiogram as an informative indicator of cardiovascular diseases. Digital Health 2023;9. https://doi.org/10.1177/20552076231187247. 20552076231187247.

[27] van de Leur R, Bleijendaal H, Taha K, et al. Electrocardiogram- based mortality prediction in patients with COVID-19 using machine learning. Netherlands Heart J 2022;30(6):312–8. https://doi.org/10.1007/s12471-022-01670-2.

[28] Jentzer JC, Kashou AH, Lopez-Jimenez F, et al. Mortality risk stratification using artificial intelligence-augmented electrocardiogram in cardiac intensive care unit patients. Eur Heart J Acute Cardiovasc Care 2021;10(5):532–41. https://doi.org/10.1093/ehjacc/zuaa021.

[29] Hayiroglu MI, Lakhani I, Tse G, Inar T, Inier G, Tekkesin AI. In- hospital prognostic value of electrocardiographic parameters other than ST-segment changes in acute myocardial infarction: literature review and future perspectives. Heart Lung Circ 2020;29(11):1603–12. https://doi.org/10.1016/j.hlc.2020.04.011.

[30] Liu CM, Shih ES, Chen JY, et al. Artificial intelligence-enabled electrocardiogram improves the diagnosis and prediction of mortality in patients with pulmonary hypertension. JACC 2022;2(3 part 1):258–70. https://doi.org/10.1016/j.jacasi.2022.02.008.

[31] Vyas DA, Eisenstein LG, Jones DS. Hidden in plain sight—reconsidering the use of race correction in clinical algorithms. N Engl J Med 2020;383(9):874–82. https://doi.org/10.1056/NEJMms2004740.

[32] Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019;366(6464):447–53. https://doi.org/10.1126/science.aax2342.

[33] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159–74.

[34] van Bussel EF, Richard E, Busschers WB, et al. A cardiovascular risk prediction model for older people: development and validation in a primary care population. J Clin Hyper 2019;21(8):1145–52. https://doi.org/10.1111/jch.13617.

[35] Cooney MT, Selmer R, Lindman A, et al. Cardiovascular risk estimation in older persons: SCORE OP. Eur J Prev Cardiol 2016;23(10):1093–103. https://doi.org/10.1177/2047487315588390.

[36] Lu TP, Chattopadhyay A, Lu KC, et al. Develop and apply electrocardiography-based risk score to identify community-based elderly individuals at high-risk of mortality. Front Cardiovasc Med 2021;8:738061. https://doi.org/10.3389/fcvm.2021.738061.

[37] Hirota N, Suzuki S, Arita T, et al. Relationship between resting 12- lead electrocardiogram and all-cause death in patients without structural heart disease: Shinken database analysis. BMC Cardiovasc Disord 2021;21(1):1–8. https://doi.org/10.1186/s12872-021-01864-3.

[38] Oikarinen L, Nieminen MS, Viitasalo M, et al. QRS duration and QT interval predict mortality in hypertensive patients with left ventricular hypertrophy: the losartan intervention for endpoint reduction in hypertension study. Hypertension 2004;43(5):1029–34. https://doi.org/10.1161/01.HYP.0000125230.46080.c6.

[39] Ristow B, Doubell A, Derman W, Heine M. Change in resting heart rate and risk for all-cause mortality. Eur J Prev Cardiol 2022;29(7):e249–54. https://doi.org/10.1093/eurjpc/zwab231.

[40] Lima EM, Ribeiro AH, Paixao GM, et al. Deep neural network-estimated electrocardiographic age as a mortality predictor. Nat Commun 2021;12(1):5117. https://doi.org/10.1038/s41467-021-25351-7.