Supplementary Material

Online Text S1. ECGs in SR coming from both FA and non-FA patients

Groups contruction

After cleaning the database, two groups of ECGs in SR were established. The first group included ECGs from patients who did not develop AF along their clinical history. To select the ECGs included in this group, patients presenting an automatic interpretation of “Sinus Rhythm” in all their ECGs were identified and selected, and their previous to the last ECG was included in this group. The group was composed of n=31,867 ECGs and was called SR-SR group. The second group was composed of ECGs in SR from patients who later developed AF. To select the ECGs included in this group, patients whose ECGs were labelled with “Atrial Fibrillation” were selected, their first ECG with AF was identified and then the previous ECG in SR was selected and included in this group. Through this process, a group of n=2,628 ECGs in SR was obtained, corresponding to patients who eventually developed AF. The group was called SR-AF. From the original 48,022 patients, 13473 patients did not meet the ECG selection rules and therefore their ECGs were excluded from both groups. Figure 2 shows the selection process in two representative patients.

Exclusion and inclusion rules

Several exclusion criteria were used to further clean both groups datasets: (i) patients with age <18 years (n = 0 in the SR-AF group and n = 87 in the SR-SR group) or with unknown sex (n = 4 in the SR-AF group and n = 62 in the SR-SR group), (ii) patients with extrasystoles (n = 328 in the SR-AF group and n = 1,504 in the SR-SR group), with atrial to ventricular ratio > 2 or < (-2) (n = 442 in the SR-AF group and n = 2,071 in the SR-SR group), with an average number of P waves per QRS complex ≠ 1 (n = 87 in the SR-AF group and n = 309 in the SR-SR group), and with a number of QRS complexes in the rhythm group higher than the average number of P waves per QRS complex (n = 263 in the SR-AF group and n = 1,007 in the SR-SR group) (some patients presented these conditions simultaneously). These last two exclusion criteria were incorporated to avoid atrial oversensing secondary to artifacts and to increase the specificity of sinus rhythm diagnosis.

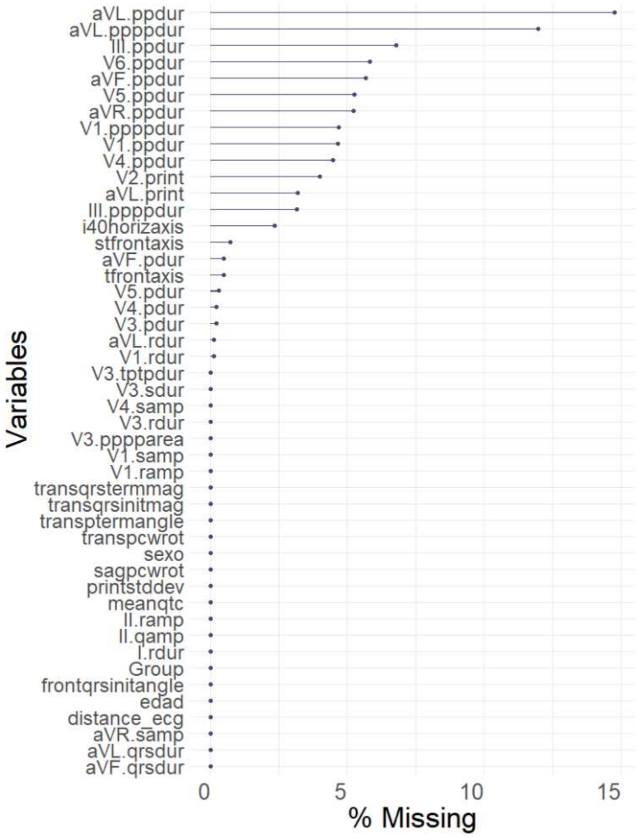
Lastly, we selected patients, in both groups, with a time interval between the previous-to-last and the last ECG (both ECGs in SR in the SR-SR group, and in SR and AF in the SR-AF group) between 1 week and 2 years (n = 1,296 in the SR-AF group and n = 18,906 in the SR-SR group).

Quality control by cardiologist’s assessment

Finally, a quality control assessment of software automatic interpretation was conducted by visually interpreting all the SR-AF ECGs. This resulted in 207 ECGs removed from the SR-AF group due to incorrect classification as AF of non-AF ECGs. The same process was performed for a 5% of the SR-SR ECGs, removing only 2 ECGs. The revision was performed by two experienced cardiologists (AC and AV) and eventual discrepancies were solved by a third cardiologist (LJJB). The final working cohort comprised 523 and 16,009 ECGs in the SR-AF and SR-SR groups, respectively. The whole process of patient selection is shown in Figure 2.

Online Text S2. Data Imputation

Data imputation is a technique aimed to overcome the issue of missing values in a data record. Several ECGs measures presented missing values, as can be shown in the plot below, ordered by the percentage of missing values:



Most of the variables presenting more than 1% of missing values were related with P wave duration, “ppdur” which is the duration of the negative part of the P wave and “ppppdur” which is the total duration (positive and negative parts) of the P wave. Also, the PR-interval, “print”, in leads V2 and aVL presented more than 1% of missing values.

A valid procedure to replace missing data by imputed values implies that one is certain about the underlying distribution of both the missing and non-missing data and, in the case of replacing the missing data with a subset of the non-missing ones, both types of data obey the same distribution. Because we do not know the exact mechanism by which the ECG software assign a NA instead of a calculated value, we cannot use the others calculated values to replace (randomly, or by a mean, etc.) the missing ones. If, for example, the ECG software replace extreme small values, because it is unable to calculate them by NAs, replacing these values by using the non-NAs data will certainly introduce a bias in the calculation. If noise is also considered, this will distort even more the underlying distribution of the missing data.

Nonetheless, we have imputed data in those measures with NAs, in such a way that no variable was excluded a priori. After repeating the univariate selection process in the same fashion as was done with the non-imputed data, 46 variables were selected for being used in the multivariate model, as shown below:

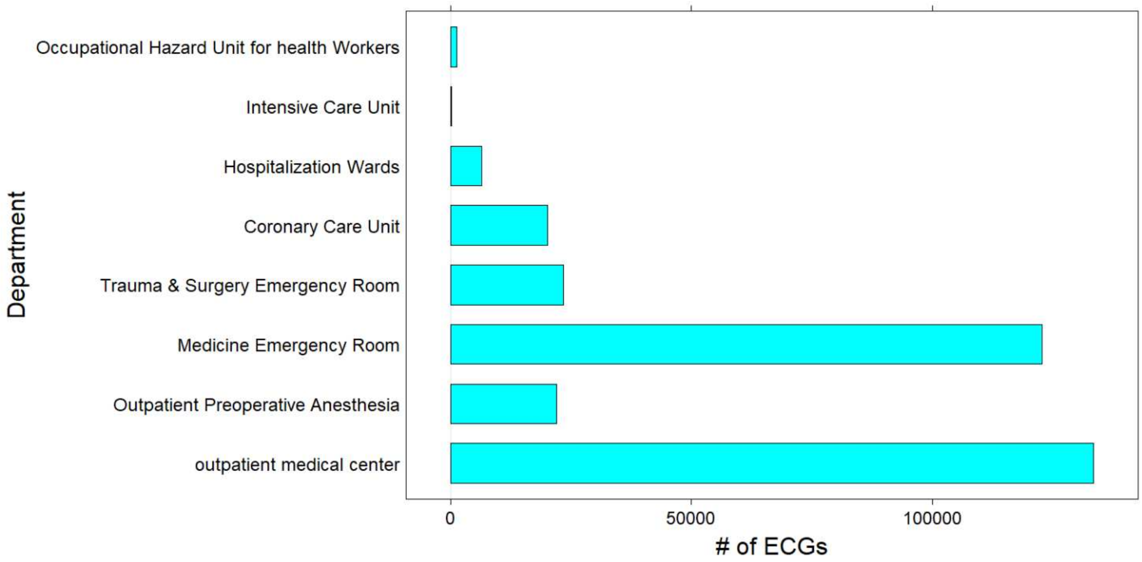
```
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)        1.112e+01  8.440e-01  13.173  < 2e-16 ***
age               -4.398e-02  4.706e-03  -9.345  < 2e-16 ***
sex               -4.099e-02  1.171e-01  -0.350 0.726286
distance_ecg      -3.930e-01  1.000e-01  -3.928 8.58e-05 ***
aVF.pdur           1.576e-03  1.935e-03   0.814 0.415454
V3.pdur            2.275e-03  1.873e-03   1.215 0.224556
V4.pdur           -2.864e-03  2.043e-03  -1.402 0.160911
V5.pdur            4.604e-03  1.911e-03   2.409 0.016000 *
III.ppdur         -4.125e-03  1.873e-03  -2.203 0.027621 *
aVR.ppdur          2.981e-03  2.910e-03   1.024 0.305682
aVL.ppdur         -1.636e-03  2.148e-03  -0.762 0.446168
aVF.ppdur         -1.526e-03  2.184e-03  -0.699 0.484789
V1.ppdur          -3.659e-03  1.573e-03  -2.326 0.020036 *
V4.ppdur          -6.961e-03  2.071e-03  -3.362 0.000775 ***
V5.ppdur           2.930e-03  2.606e-03   1.124 0.260849
V6.ppdur          -2.677e-03  2.819e-03  -0.950 0.342358
III.ppppdur       -3.581e-03  2.152e-03  -1.664 0.096091 .
aVL.ppppdur       -6.009e-03  2.516e-03  -2.388 0.016937 *
V1.ppppdur        -7.188e-03  2.176e-03  -3.304 0.000954 ***
V3.pppparea        8.954e-03  7.263e-03   1.233 0.217634
II.qamp           -2.685e-04  9.195e-04  -0.292 0.770291
II.ramp           -5.276e-04  1.699e-04  -3.106 0.001898 **
V1.ramp            4.371e-04  4.652e-04   0.940 0.347379
I.rdur             3.042e-03  3.077e-03   0.989 0.322788
aVL.rdur           6.927e-04  2.848e-03   0.243 0.807853
V1.rdur            8.609e-05  4.239e-03   0.020 0.983799
V3.rdur            1.028e-03  3.618e-03   0.284 0.776210
aVR.samp          -1.586e-04  1.740e-04  -0.911 0.362093
V1.samp           -3.741e-04  1.288e-04  -2.903 0.003691 **
V4.samp            2.262e-04  1.505e-04   1.503 0.132747
V3.sdur           -1.104e-03  2.832e-03  -0.390 0.696601
aVL.qrsdur        -1.364e-03  3.913e-03  -0.349 0.727358
aVF.qrsdur        -1.795e-03  4.053e-03  -0.443 0.657865
V3.tptpdur        -2.414e-03  9.239e-04  -2.613 0.008982 **
aVL.print          1.702e-03  1.823e-03   0.934 0.350509
V2.print          -2.427e-03  1.749e-03  -1.387 0.165308
printstddev       -1.038e-02  5.617e-03  -1.849 0.064492 .
i40horizaxis      -9.059e-04  1.839e-03  -0.493 0.622361
stfrontaxis       -1.087e-03  7.777e-04  -1.398 0.161997
tfrontaxis        -1.017e-03  1.202e-03  -0.846 0.397330
meanqtc           -4.902e-03  1.591e-03  -3.080 0.002069 **
transpcwrot       -4.625e-03  1.331e-03  -3.474 0.000513 ***
transptermangle    3.837e-05  4.117e-04   0.093 0.925742
transqrsinitmag    1.137e-04  3.078e-04   0.369 0.711796
transqrstermmag   -2.963e-04  3.126e-04  -0.948 0.343276
frontqrsinitangle -5.350e-04  5.859e-04  -0.913 0.361166
sagpcwrot          1.896e-03  1.274e-03   1.488 0.136780
```

In this model, five variables III.ppdur, V1.ppdur, V4.ppdur, aVL.ppppdur, and V1.ppppdur were new as compare to the original results, because they were initially excluded due to the presence of high percentage of NAs. Moreover, all of them are related with the negative part of the P wave. Taking into account these facts and the arguments exposed above, we have decided not imputing data and exclude those variables presenting more than 1% of missing values.

Online Text S3. Lasso Regression

(Intercept)	6.5886402675
age	-0.0330613273
sex	.
distance_ecg	.
aVF.pdur	.
V3.pdur	.
V4.pdur	.
V5.pdur	.
aVR.ppdur	.
aVL.ppdur	.
V5.ppdur	-0.0018101138
V6.ppdur	-0.0033574133
III.ppppdur	-0.0023406085
aVL.ppppdur	.
V1.ppppdur	-0.0048949965
V3.pppparea	0.0031968552
V2.qamp	.
V1.ramp	.
I.rdur	.
aVL.rdur	.
V1.rdur	.
V3.rdur	.
V6.rdur	.
aVL.samp	.
I.vat	.
aVL.vat	.
V1.tptpdur	.
V3.tptpdur	.
V5.tptpdur	.
V3.tpdur	.
V5.tpdur	.
aVL.print	.
printstddev	.
i40horizaxis	.
stfrontaxis	.
sthorizaxis	.
meanqtc	-0.0008322995
transpcwrot	-0.0022047081
transptermangle	.
transqrsinitmag	.
transqrstermmag	.
frontpcwrot	.
frontqrscwrot	.
frontqrsinitangle	.
sagpcwrot	.
sagpinitangle	.
high.low	.

Online Figure S1. Distribution of patients by services and units



Online Table S1. Variable's definition

ECG variable	Definition
frontqrsinitangle	Frontal QRS complex initial angle (degrees)
meanqtc	QT rate correction by Bazett's formula (QTc = QT / √RR)
pdur	Duration of positive part of P wave (milliseconds)
pppparea	Area of positive and negative part of P wave (Ashman units; 40 ms x 0.1 mV))
printstddev	Standard deviation of the PR interval in the rhythm group (milliseconds)
qamp	Q wave amplitude (millivolts)
qrsdur	Duration of QRS complex (milliseconds)
ramp	R wave amplitude (millivolts)
rdur	Duration of positive part of R wave (milliseconds)
samp	S wave amplitude (millivolts)
sdur	Duration of S wave (milliseconds)
stfrontaxis	Mean ST wave frontal axis (degrees)
tfrontaxis	Mean T wave frontal axis (degrees)
sagpcwrot	Sagital P wave clockwise rotation (100 to -100)
samp	S wave amplitude (millivolts)
tptpdur	Duration of positive and negative part of T wave (milliseconds)
transpcwrot	Transverse P wave clockwise rotation (100 to -100)
transptermangle	Transverse P wave terminal angle (degrees)
transqrsinitmag	Transverse QRS complex initial vector magnitude (millivolts)
transqrstermmag	Transverse QRS complex terminal vector magnitude (millivolts)

Online Table S2.

```
                             Estimate Std. Error z value Pr(>|z|)    
(Intercept)                  -6.35787    0.49007 -12.973  < 2e-16 ***
age(60,65]                    0.73662    0.26935   2.735 0.006242 ** 
age(65,70]                    1.16286    0.23802   4.886 1.03e-06 ***
age(70,75]                    0.94937    0.24438   3.885 0.000102 ***
age(75,80]                    1.18380    0.22688   5.218 1.81e-07 ***
age(80,85]                    1.66033    0.20893   7.947 1.91e-15 ***
age(85,90]                    1.75675    0.22136   7.936 2.09e-15 ***
age(90,95]                    1.98339    0.26501   7.484 7.20e-14 ***
age(95,100]                   2.21916    0.48421   4.583 4.58e-06 ***
age(100,103]                  2.64218    1.16183   2.274 0.022957 *  
aVF.pdur[0,90]                0.21135    0.11863   1.782 0.074815 .  
aVF.pdur(200,279]             1.68737    0.85100   1.983 0.047390 *  
V3.pdur[0,75]                 0.11322    0.12226   0.926 0.354413    
V4.pdur[0,95]                 0.30871    0.12732   2.425 0.015320 *  
V4.pdur(150,233]              0.74000    0.22749   3.253 0.001142 ** 
V5.pdur[0,90]                 0.29089    0.12149   2.394 0.016648 *  
V5.pdur(175,268]             -0.09605    0.52859  -0.182 0.855805    
V3.pppparea[-88,5]            0.16043    0.12175   1.318 0.187601    
II.qamp(-168,0]               0.27638    0.34362   0.804 0.421213    
II.ramp(1.35e+03,4.04e+03]   -0.01356    0.23796  -0.057 0.954554    
V1.ramp[-206,183]             0.12896    0.16195   0.796 0.425877    
V1.ramp(600,3.3e+03]         -0.44850    0.42662  -1.051 0.293122    
I.rdur(50,181]                0.15306    0.12803   1.196 0.231880    
aVL.rdur(50,187]             -0.05657    0.13024  -0.434 0.664039    
V1.rdur[0,25]                 0.17116    0.13298   1.287 0.198057    
V1.rdur(50,169]               0.27609    0.32816   0.841 0.400172    
V3.rdur[0,40]                -0.08310    0.12034  -0.691 0.489839    
V3.rdur(75,184]              -0.36450    0.34851  -1.046 0.295610    
aVR.samp[-2.29e+03,-1e+03]   -0.18334    0.28487  -0.644 0.519837    
aVR.samp(-250,0]             -0.03578    0.14785  -0.242 0.808775    
V1.samp(-1.13e+03,0]          0.31674    0.20379   1.554 0.120114    
V4.samp(-868,0]              -0.17337    0.15092  -1.149 0.250657    
V3.sdur[0,20]                 0.07222    0.19260   0.375 0.707669    
V3.sdur(50,148]               0.18871    0.14374   1.313 0.189233    
aVL.qrsdur[0,70]             -0.08904    0.15635  -0.570 0.569009    
aVL.qrsdur(100,208]          -0.14680    0.19897  -0.738 0.460627    
aVF.qrsdur[0,70]             -0.00456    0.16640  -0.027 0.978139    
aVF.qrsdur(100,208]           0.14629    0.19362   0.756 0.449919    
V3.tptpdur[0,150]            -0.08817    0.20521  -0.430 0.667437    
V3.tptpdur(250,588]           0.46503    0.13502   3.444 0.000573 ***
printstddev(7.5,44]           0.07690    0.11121   0.691 0.489287    
stfrontaxis[-90,0]           -0.06521    0.20275  -0.322 0.747731    
stfrontaxis(70,269]           0.14044    0.13655   1.028 0.303721    
tfrontaxis[-90,0]             0.25039    0.20553   1.218 0.223128    
tfrontaxis(70,269]           -0.11847    0.14469  -0.819 0.412901    
meanqtc(450,708]              0.45221    0.12120   3.731 0.000191 ***
transpcwrot(-50,100]          0.48787    0.13392   3.643 0.000270 ***
transptermangle[0,300]        0.26509    0.12249   2.164 0.030445 *  
transqrsinitmag(300,2.32e+03] -0.04478   0.13450  -0.333 0.739180    
transqrstermmag(300,2.24e+03]  0.24637   0.12149   2.028 0.042566 *  
frontqrsinitangle[0,150]     -0.10430    0.17934  -0.582 0.560827    
frontqrsinitangle(250,360]    0.02972    0.25210   0.118 0.906157    
sagpcwrot[-100,50]            0.12177    0.13433   0.906 0.364682    
```

Online Table S3.

threshold	0	1	2	3	4
specificity	0,00	37,19	71,57	96,35	99,97
sensitivity	100,00	91,53	66,95	13,56	1,69
accuracy	2,89	38,76	71,44	93,96	97,14
npv	NA	99,33	98,65	97,40	97,16
ppv	2,89	4,15	6,54	9,94	66,67

Online Table S4.

Model for the scenario that exclude patients < 65 years

```
                   Estimate Std. Error z value Pr(>|z|)    
age              -0.0568609  0.0041870 -13.580  < 2e-16 ***
distance_ecg     -0.2320681  0.0908730  -2.554  0.01066 *  
aVF.pdur          0.0028670  0.0015820   1.812  0.06995 .  
V3.pdur           0.0026431  0.0016209   1.631  0.10296    
V1.ramp           0.0006640  0.0002901   2.289  0.02208 *  
I.vat            -0.0098615  0.0036461  -2.705  0.00684 ** 
printstddev      -0.0228687  0.0050007  -4.573 4.81e-06 ***
qrsfrontaxis     -0.0022906  0.0008459  -2.708  0.00677 ** 
stfrontaxis      -0.0010408  0.0006500  -1.601  0.10932    
transpcwrot      -0.0036024  0.0011767  -3.061  0.00220 ** 
transqrsinitmag   0.0004570  0.0003281   1.393  0.16363    
transqrstermmag  -0.0006946  0.0002507  -2.771  0.00560 ** 
frontpmaxangle   -0.0030867  0.0012047  -2.562  0.01040 *  
frontqrsinitmag  -0.0010162  0.0003486  -2.915  0.00356 ** 
sagpcwrot         0.0018273  0.0011487   1.591  0.11166    
```

Model for the scenario that randomly selected patients from SR-SR to match SR-AF

```
             Estimate Std. Error z value Pr(>|z|)    
age          0.009762   0.005489   1.778 0.075335 .  
aVL.pdur    -0.006502   0.002009  -3.237 0.001208 ** 
III.ppdur   -0.005624   0.001948  -2.887 0.003890 ** 
V1.ppppdur  -0.011170   0.002805  -3.982 6.83e-05 ***
V3.pppparea  0.022771   0.009807   2.322 0.020234 *  
meanqtint.1 -0.004724   0.001709  -2.764 0.005711 ** 
transpcwrot -0.005871   0.001350  -4.349 1.37e-05 ***
```

Model considering only age

```
             Estimate Std. Error z value Pr(>|z|)
age         -0.06066    0.00428  -14.17   <2e-16 ***
```