Improved Prediction of Atrial Fibrillation by Identifying ECG Patient Trajectories Over Time

Myrte Barthels[1,2,3], Henri Gruwez[1,4,5], Thomas De Cooman[3], Noella Pierlet[1], David Duncker[6], Pieter Vandervoort[1,2,4]

[1] Doctoral School of Medicine and Life Sciences, Hasselt University, Hasselt, Belgium
[2] Department Future Health, Hospital East-Limburg, Genk, Belgium
[3] Qompium NV, Hasselt, Belgium
[4] Department of Cardiology, Hospital East-Limburg, Genk, Belgium
[5] Department of Cardiovascular sciences, University of Leuven, Leuven, Belgium
[6] Department of Cardiology and Angiology, Hannover Medical School, Hannover, Germany

Abstract

Atrial fibrillation (AF) risk can be predicted from a single-point ECG one month before onset. This study investigates whether this AF risk score can be improved by combining the risk score of multiple consecutive ECGs over time. We compare the predictive performance of a single-point risk score with the average and maximum ECG risk scores within one year prior to AF onset as well as trajectories of one-year interval averaged and maximized scores up to five years before onset. K-means clustering was used to visualize trajectory patterns among patients. We included 13036 patients with an AF prevalence of 22.6%. The results showed a slight increase in performance for the one-year interval maximized approach, but no further improvement was observed using the trajectory approach. Trajectory clustering revealed substantial variability in AF onset patterns, suggesting diversification in the development of AF across patients.

1. Introduction

Atrial fibrillation (AF) is a cardiac arrythmia that causes irregular hearbeats with high risk for stroke and mortality when left untreated. At onset, AF is paroxysmal and often of asymptomatic nature causing AF to remain undiagnosed and untreated in 13% to 35% of people with AF [1]. The landmark study from Attia et al. demonstrated to detect AF from sinus rhythm (SR) ECGs up to 31 days before onset with an area under the receiver operator curve (AUROC) of 0.87, indicating early identification of AF risk even at the paroxysmal stage [2]. However, ECG recordings within 31 days prior to onset are often scarce, especially for asymptomatic patients who are unaware of their condition. More early risk prediction using routine care ECGs could provide a solution, as studies have demonstrated to predict AF risk one, two, and five years before onset with AUROCs of 0.85, 0.73, and 0.91, respectively. These findings highlight the presence of early signs of AF in the ECG. [3–5].

Despite these advances, the pathway to AF development is poorly understood, and the variability in symptoms and outcomes among patients suggests that multiple pathways to AF may exist. Traditional AF prediction models based on single-point assessments fail to capture the temporal context and exposure duration. Investigating the relationship between consecutive ECGs over time may provide deeper insights into AF patterns and enhance prediction accuracy. Rahman et al. investigated these longitudinal patterns within the CHARGE-AF score and its underlying risk factors [6]. They reported increased risk of incident AF for systolic blood pressure trajectories, however, no improvement was found in other risk factors and the composite CHARGE-AF score over using the single-point score. On the contrary, Melzi et al. reported that assessing longitudinal patterns of consecutive neural network prediction scores revealed differences depending on ECG follow-up frequency and AF diagnosis certainty [7].

In this study we hypothesized that an AF risk score derived from a single-point ECG can be improved by combining the risk score of multiple ECGs over time. Assessing patterns in longitudinal risk prediction could allow for early stratification of at-risk patients.

2. Materials and methods

2.1. Data

The dataset was obtained from a retrospective study at Ziekenhuis Oost-Limburg (ZOL, Genk, Belgium) and was

previously described by Gruwez et al. [8]. It contains all 10-second 12 lead ECGs along with diagnostic labels from patients older than 18 years obtained at ZOL between October 1, 2002 and August 31, 2021. AF risk scores were derived using a pretrained AF prediction model described by Gruwez et al. [8]. We defined a window-of-interest including all ECGs taken five years prior to a patient specific reference date. For patients with a history of AF rhythms (AF and atrial flutter), the reference date was the acquisition date of the first AF ECG. For patients with no AF rhythms, the reference date was set five years before the last measured SR ECG, to ensure the patient was likely free from AF episodes. To avoid data leakage, the same train, test and validation set were used as in the original AF prediction model. This study was reviewed and approved by the medical ethical committee of ZOL.

2.2. Extrapolation from multiple ECGs

We defined two approaches to extrapolate the information of multiple ECGs over time. The first approach computed the average/maximum of the ECG risk scores within a fixed interval before AF onset [2]. Since our dataset is collected retrospective from routine care, patients have varying ECG frequencies and timings. We chose the optimal interval length n as the balance between including multiple ECGs, to obtain summary statistics, and minimizing time to onset, to ensure closeness of AF onset. We evaluated intervals of one month, six months, and one year. We discarded intervals shorter than the average inter-ECG follow-up time and selected the shortest interval with an ECG frequency closest to 1.

The second approach included temporal context by defining AF risk score trajectories. The five year window was segmented into intervals of length n, obtaining $\frac{5}{n}$ timesteps, represented by the average or maximum ECG score. To ensure adequate temporal spread, patients were included if they had ECGs in at least k different intervals, with k the average number of ECGs per patient within the five-year window. Gaps, intervals with no ECGs, were interpolated from the ECG scores within the surrounding intervals.

2.3. Model and evaluation metrics

The primary outcome of the study was to compare the last available single-point assessment of the AF risk score to the two multi-ECG approaches. For the trajectory approach, the final AF risk score was derived from a new prediction model trained on top of the trajectory vector. We modelled a fully connected (FC) neural network consisting of two hidden dense layers with 64 and 32 units, respectively, each using ReLU activation and a dropout rate of 0.5, and a final layer that produces a probability between 0 and 1 using the Sigmoid activation. The model was trained for 10 epochs using the Adam optimizer, an initial learning rate of 0.001 and a binary cross-entropy loss. AF prediction performance was compared using AUROC and the area under the precision recall curve (AUPRC). Two-sided 95% confidence intervals (CI) for the AUROC were estimated using the Sun and Xu optimization of the Delong method. The AUPRC was approximated as the average precision, and the CI was calculated using the bi-normal interval approximation.

In a secondary analysis, K-means clustering was employed to visualize different trajectory patterns among patients. The optimal number of clusters was determined using the Elbow, Silhouette, and Davies-Bouldin methods. For each cluster, the prevalence of AF, the average age and the gender prevalence was assessed to identify common patterns among patient groups.

3. Results

3.1. Data

We identified 45878 patients with at least one ECG five years prior to the reference date, of which 20% developed future AF. The average time in between follow up ECGs was 0.7 ± 0.7 years. This discarded both the one month and six month interval, and fixes n at one year. Patients had on average 0.7 ± 1.3 ECGs within the interval. The average number of ECGs within the five-year window-of-interest was 3.5 ± 3.6 ECGs, from which k was set as 3. After restraining the selected windows to at least k non-empty intervals, 13036 patients remained with an AF prevalence of 22.6%, indicating more longitudinal data is available for at-risk AF patients. Demographic details for the train, validation and test set are presented in Table 1.

Table 1. Dataset demographics. AF = atrial fibrillation.

Demographics	**Train**	**Validation**	**Test**
Patients	9130	1301	2605
Female (%)	40.0	40.0	40.2
Mean age	67.2	66.7	67.4
(years)	$\pm$12.0	$\pm$12.2	$\pm$12.2
AF patients (%)	22.1	23.9	23.5

3.2. Performance

Table 2 summarizes the performance metrics, comparing the single-point AF prediction to the one-year interval outcome and the full five-year trajectory for the average and maximum mode.

Table 2. Performance in terms of AUROC-score [95% CI] and AUPRC-score [95% CI]. AUROC = area under the receiver operating curve, AUPRC = area under the precision recall curve, CI = confidence interval.

Approach / **Mode**	**AUROC**	**AUPRC**
Single-point ECG		
Last	0.80 [0.78, 0.82]	0.58 [0.54, 0.62]
Last one-year segment		
Average	0.80 [0.78, 0.82]	0.59 [0.55, 0.63]
Max	0.81 [0.79, 0.83]	0.60 [0.56, 0.64]
Five-year trajectory		
Average	0.80 [0.78, 0.82]	0.59 [0.55, 0.63]
Max	0.81 [0.79, 0.83]	0.61 [0.57, 0.64]

No major improvements were observed in AUROC when comparing modes as well as the prediction approaches. The AUPRC showed a slight increase in performance using the maximum mode in both the last one-year segment and the trajectory. However, using the trajectory model did not improve prediction over the one-year segment outcome.

In the secondary analysis, the optimal number of clusters determined by the Elbow, Silhouette, and Davies-Bouldin methods were 4, 4, and 5, respectively. Therefore, we opted to use 4 clusters for the final analysis. Figure 1 displays the mean trajectory for each of the 4 clusters. Each point represents the average outcome of the trajectory input (maximum mode) at that timestamp across all cluster members, with CIs around the mean depicted by the colored band. The trajectories for all 4 clusters are parallel, showing a decrease in cluster membership (from 63.7% to 4.7%) and an increase in AF prevalence (from 11.4% to 74.7%) and mean age (from 64.2 $\pm$ 12.1 to 76.2 $\pm$ 9.1) with increasing cluster number (Table 3). Within each cluster, AF patients showed a steeper trajectory slope in the last one-year segment compared to non-AF patients.

Figure 2 represents the results with 10 clusters. As compared to the 4 cluster model, more variability and overlapping trajectories were observed.

Table 3. Cluster distribution. AF = atrial fibrillation, mem. = members, prev. = prevalance.

Cluster	**0**	**1**	**2**	**3**
Members (%)	63.7	22.1	9.4	4.7
Female (%)	40.3	38.1	39.6	49.6
Mean age	64.2	72.0	74.8	76.2
(years)	$\pm$12.1	$\pm$10.3	$\pm$9.5	$\pm$9.1
AF mem. (%)	31.0	31.3	22.7	15.0
AF prev. (%)	11.4	33.3	56.7	74.8

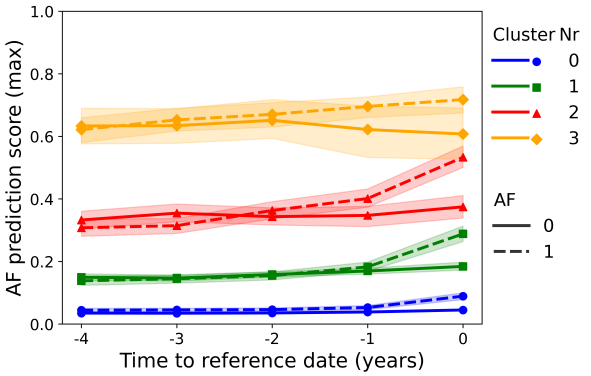

Figure 1. K-Means results showing 4 trajectory clusters. Continuous lines represent the non-AF patient group within the cluster while dashed lines represent the AF patient group. AF = atrial fibrillation.

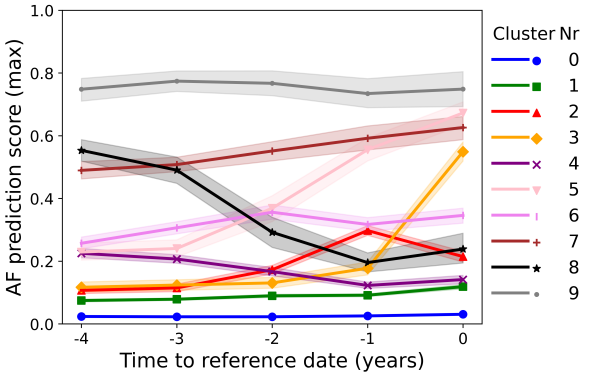

Figure 2. K-means results showing 10 trajectory clusters. AF = atrial fibrillation.

4. Discussion and conclusion

This retrospective study evaluated whether a single-point ECG AF risk score could be improved by using multiple consecutive ECGs. Performance results showed a modest improvement when using the one-year interval maximum AF risk score. No further improvements were found using five-year trajectories. However, the data selection procedure, tailored to extract multi-ECG trajectories, revealed that at-risk AF patients have in general more ECG measurements in time.

The secondary analysis using K-means clustering identified four distinct trajectory patterns based on the maximized AF risk score. As demonstrated in Figure 1, the clusters show a parallel progression in AF risk, which validates why trajectories do not provide additional information over a single-point assessment. As AF risk increases with age, the gradual increase, seen over all four clusters, might merely reflect the aging over time.

Distinction between AF and non-AF patients within trajectory clusters is expressed only in the last one-year interval of each cluster, confirming the inferior results for the trajectory approach in the primary analysis. This differentiation is most pronounced in cluster 3, however, the limited number of cluster members and the large CI interval hamper generalizabilty to all AF patients. Moreover, more than 60% of AF patients remain within the lower clusters (0 and 1), were the prevalence of non-AF patients is much higher. Expanding to more clusters as shown in Figure 2 results in a high diversity of patterns in the lower clusters. This suggests that while the current single-point AF prediction model serves as a general predictor, it may not capture the nuanced development of AF in all patients. This aligns with the study of Baqal et al., indicating that an ECG-based AF prediction model struggled to predict AF after a cardiac procedure [9], highlighting the need for caution in applying these models across different patient populations.

Several limitations should be noted. First, non-AF patients appearing in high score trajectories may be undetected due to the paroxysmal and often asymptomatic nature of AF. Additionally, the use of non-controlled retrospective data necessitated summarization techniques concealing frequency and timing details of ECGs. Lastly, results indicate that the last one-year interval allows for improved performance, however, further exploration within this interval was not possible due to the low ECG density.

Future research should focus on time-aware models that can handle varying input lengths. Additionally, work is needed to understand the underlying physical process related to AF onset. Personalized risk scores that account for individual variability and relative ECG changes could increase prediction accuracy. It remains to be investigated if the ECG is rich enough to reflect these subtle changes.

In conclusion, combining multiple consecutive ECG AF risk scores within one year prior to AF onset offers a modest enhancement in prediction performance over a single-point assessment. Results suggest that AF development should be actively monitored within the last year and indicate high diversity toward AF onset among patients.

Acknowledgments

This study is part of Limburg Clinical Research Center, supported by Hasselt University, Ziekenhuis Oost-Limburg and Jessa Hospital.

References

[1] Diederichsen SZ, Haugan KJ, Brandes A, Graff C, Krieger D, Kronborga C, Holst AG, Nielsen JB, Køber L, Højberg S, Svendsen JH. Incidence and predictors of atrial fibrillation episodes as detected by implantable loop recorder in patients at risk: From the loop study. Am Heart J Jan. 2020;219:117–127.

[2] Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, Carter RE, Yao X, Rabinstein AA, Erickson BJ, Kapa S, Friedman PA. An artificial intelligence-enabled ecg algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. The Lancet Sep. 2019;394(10201):861–867.

[3] Raghunath S, Pfeifer JM, Ulloa-Cerna AE, Nemani A, Carbonati T, Jing L, vanMaanen DP, Hartzel DN, Ruhl JA, Lagerman DB, Rocha DB, Stoudt NJ, Schneider G, Johnson KW, Zimmerman N, Leader JB, Kirchner HL, Griessenauer CJ, Hafez A, CW CWG, Fornwalt BK, Haggerty CM. Deep neural networks can predict new-onset atrial fibrillation from the 12-lead ecg and help identify those at risk of atrial fibrillation-related stroke. Circ Mar. 2021;143(13):1287–1298.

[4] Christopoulos G, Graff-Radford J, Lopez CL, Yao X, Attia Z, Rabinstein AA, Petersen RC, Knopman DS, Mielke M, Kremers W, Vemuri P, Siontis KC, Friedman PA, Noseworthy PA. Artificial intelligence-electrocardiography to predict incident atrial fibrillation: A population-based study. Circ Arrhythm Electrophysiol Dec. 2020;13(12):e009355.

[5] Biton S, Gendelman S, Ribeiro AH, Miana G, Moreira C, Ribeiro ALP, Behar JA. Atrial fibrillation risk prediction from the 12-lead electrocardiogram using digital biomarkers and deep representation learning. Eur Heart J Dig Health Dec. 2021;2(4):576–585.

[6] Rahman F, Yin X, Larson MG, Ellinor PT, Lubitz SA, Vasan RS, McManus DD, Magnani JW, Benjamin EJ. Trajectories of risk factors and risk of new-onset atrial fibrillation in the framingham heart study. Hypertension Jul. 2016;68(3):597–605.

[7] Melzi P, Vera-Rodriguez R, Tolosana R, Sanz-Garcia A, Cecconi A, Ortega GJ, Jimenez-Borreguero LJ. Prediction of atrial fibrillation from sinus-rhythm electrocardiograms based on deep neural networks: Analysis of time intervals and longitudinal study. IRBM Dec. 2023;44(6):100811.

[8] Gruwez H, Barthels M, Haemers P, Verbrugge F, Dhont S, Meekers E, Wouters F, Nuyens D, Pison L, Vandervoort P, Pierlet N. Detecting paroxysmal atrial fibrillation from an electrocardiogram in sinus rhythm: External validation of the ai approach. JACC Clin Electrophysiol Aug. 2023; 9(8.3):1771–1782.

[9] Baqal O, Habib EA, Hasabo EA, Galasso F, Barry T, Arsanjani R, Sweeney JP, Noseworthy P, Fortuin FD. Artificial intelligence-enabled electrocardiogram (ai-ecg) does not predict atrial fibrillation following patent foramen ovale closure. Int J Cardiol Heart Vasc Feb. 2024;51:101361.

Address for correspondence:

Myrte Barthels
Corda Campus, Kempische Steenweg 303/27, 3500 Hasselt
myrte.barthels@uhasselt.be