ECG-Based Long-Term Prediction of Atrial Fibrillation

Livia Colucci[1], Mashroor Khan[2], Leif Sörnmo[1], Emma Svennberg[2], Martin Stridh[1]

[1] Department of Biomedical Engineering, Lund University, Lund, Sweden
[2] Department of Medicine Huddinge, Karolinska University Hospital, Stockholm, Sweden

Abstract

Identification of patients at risk to develop atrial fibrillation (AF) can enable more frequent follow-up and lead to earlier detection of AF. The aim of this project was to perform long-term AF prediction based on a novel set of ECG-derived features describing the characteristics of supraventricular arrhythmias present before new onset AF.

In total, 12 199 patients aged 75/76 performed repeated single-lead ECG measurements for 30-s during a period of 2–4 weeks. Arrhythmia detection was performed following quality control to ensure that only reliable detections were included. For each detected arrhythmic episode, a set of features were extracted including degree of prematurity, burden of supraventricular ectopic beats, and number of RR intervals between arrhythmic events. Following feature extraction, a 1D-convolutional neural network was employed to predict the long-term risk of developing AF.

On average, the trained prediction models led to an AUC of 0.60 for the test set. At the end of the observation period, the risk stratification curves for the model showed 94% and 88% probabilities of not developing AF for the low- and high-risk groups, respectively. These results correspond to a weighted F1 score of 0.72 for the test set.

The results show that ECG-derived features characterising supraventricular arrhythmias occurring before AF contribute to improved risk stratification for a future AF diagnosis.

1. Introduction

Atrial fibrillation (AF) screening is a useful tool in clinical practice as it allows identification of undiagnosed subjects, leading to earlier treatment with anticoagulation therapy, and thus a lower risk of complications [1]. Such screening may be used not only to detect AF, but also to predict the risk of patients developing AF in the short- or long-term future. Several studies have investigated this possibility, leading to short- and long-term prediction models. The aim of the former is mainly to detect paroxymal AF, hence the shorter prediction horizon usually falls within 3 months, while the longer prediction spans over a longer time frame, usually from 6 months and onwards. To develop long-term prediction models with the ability to stratify patients into different risk groups, both clinical parameters [2] and information derived from the ECG [3–5] have been employed. When based on the ECG, these models have been trained either on the raw ECG signal [3, 4], or on its features [5]. Using the former option, Raghunath et al. employed a deep convolutional neural network (CNN) and a Cox proportional hazard model to predict AF after one year based on 12-lead, 2.5-s and 10-s ECGs, and achieved a 0.83 area under the curve (AUC) [3]. Khurshid et al. also employed 12-lead, 10-s ECGs and a CNN, which was combined with a discrete-time survival model predicting the 5-year probability of being AF-free. In the internal validation set, this model obtained an AUC of 0.82 [4].

Several clinical studies have investigated the prognostic implications of supraventricular arrhythmias (SVAs) in the development of AF. Specifically, the presence of frequent supraventricular ecopic beats (SVEBs) was shown to be of significance for AF prediction [6–10]. Hygrell et al. investigated the prognostic implications of different degrees of complexity in the pattern of SVEB in relation to AF development [8]. In this work, the risk increased gradually from patients with isolated SVEBs (hazard ratio 2.1) or SVEBs in bigemini/trigemini (hazard ratio of 1.9/2.4) to non-sustained supraventricular tachycardias (SVTs) (hazard ratio ranging from 1.1 to 4.0 depending on the degree of regularity and presence of p-waves). The latter finding was supported by Johnson et al., who showed that the characteristics of non-sustained SVTs, namely their irregularity and the absence of P waves, were predictive of AF [9].

To our knowledge, no long-term AF prediction models have been trained based on the presence and characteristics of SVAs of different levels of complexity, episode duration and burden. Hence, the aim of this study was to develop a long-term AF risk prediction model based on a novel feature set describing the characteristics and burdens of various SVAs detected in screening ECGs.

2. Methods

In this work, data from two prospective mass screening studies for AF, Strokestop I (SSI) [1] and Strokestop II (SSII) [11], were employed. Both studies focused on 75-year-old adults, whose clinical condition was assessed at the time of screening and monitored in the following years. During screening, single-lead, 30-s ECG measurements were performed autonomously by the study participants 2–4 times daily during 2–4 weeks using a Zenicor handheld device (www.zenicor.com). ECG recordings from 6 145 participants in Strokestop I and from 6 054 participants in Strokestop II with no known history of AF were employed for training and testing, respectively. Endpoint data in the form of AF diagnosis outcome within 5.5 years after screening were acquired from the Swedish nation-wide registries.

The proposed method consists of three analysis steps - SVA detection, feature extraction, and AF prediction - followed by performance evaluation. Details about each step are provided in the respective subsections below.

SVA detection: Seven groups of SVAs are taken into account in the present work: single and multiple supraventricular ectopic beats with compensatory pauses ($sSVEB_C$, $mSVEB_C$), episodes of consecutive short-duration beats not followed by a compensatory long-duration beat ($SVEB_{NC}$), bigeminy episodes (Bige), trigeminy episodes (Trige), non-sustained SVT (nsSVT) and sustained SVT (sSVT). For detailed definitions, see Table 1. Detection of these arrhythmias is performed by first applying the commercially available analysis software *ECG Parser* (www.cardiolund.com). For each recording, this software provides a first set of detected SVAs together with the series of RR intervals, the median RR duration and their ratio for each respective RR interval.

Table 1: SVA groups defined for the present study.

Group	Characteristic
$sSVEB_C$	arrhythmic episodes of 1 short-duration beat followed by a compensatory long-duration beat
$mSVEB_C$	arrhythmic episodes of 2-4 consecutive short-duration beats followed by a compensatory long-duration beat
$SVEB_{NC}$	arrhythmic episodes of 2-4 consecutive short-duration beats not followed by a compensatory long-duration beat
Bige	arrhythmic episodes characterised by a pattern where every second beat is a SVEB
Trige	arrhythmic episodes characterised by a pattern where every third beat is a SVEB
nsSVT	series of at least 4 short-duration beats with an average heart rate $\geq$ 100bpm
sSVT	30-s ECG recording with a heart rate $\geq$ 100bpm

In order to further improve SVA detection, quality control is employed to ensure that all detected SVA episodes only include reliable beat detections. Here, quality control is achieved through a CNN-based detector of transient noise [12], and an SVM-based detector of sequences of consecutive beats of similar beat morphology [13]. Detected SVAs are approved only if all included beats are deemed to not be transient noise and to be morphologically similar to surrounding detections. It is noted that the similar beat sequence criterion also ensures that no ventricular beat is detected as a SVEB due to its deviating morphology compared to that of the SVEB.

Following quality control, the ECG recordings are subject to RR-based pattern correlation with the goal of increasing the sensitivity of the SVA detection. In this step, an RR-based pattern template per arrhythmia is generated for each subject using already detected episodes. If no previously detected episode exists, no pattern template is generated, and no correlation is performed. Pattern correlation is employed to find SVAs that have gone undetected, but whose RR pattern is similar to those of the detected ones. If a match is found and all beats involved pass quality control, the corresponding SVA detection is added to the set of identified SVAs.

Feature extraction: In order to characterise the detected SVAs, a novel set of features is proposed. Features are extracted for each arrhythmia group in Table 1 and across each subject's recordings and include: the minimum degree of prematurity of the short-duration beats involved, the minimum interval between the arrhythmic event and the next short-duration beat in the recording, and the arrhythmia group burden. Hence, given the seven arrhythmia groups, a total of twenty-one features are used in this work, and their individual importance is assessed via feature permutation.

AF prediction: In the last step of the analysis, a 1D-CNN is trained for long-term prediction of AF. The CNN is composed of a 1D convolutional layer with kernel size and stride equal to 3 (equal to the number of features for each arrhythmia) and 4 channels, a flattening layer, a dense layer with 28 nodes, a dropout layer with dropout rate of 0.2 and two dense layers with seven and one nodes/node, respectively. Binary prediction with a 5.5-years horizon is achieved by setting a threshold through Youden's J statistics.

Model optimization and evaluation: Fifteen models with the same structure were trained to derive average measures of performance. The proposed model is evaluated by first deriving confusion matrices providing the number of true and false positives and negatives in terms

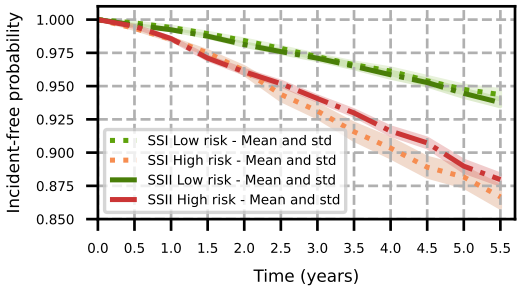

Figure 2: Risk stratification curves for the low- and high-risk patients based on the Kaplan-Meier estimator for the training (dotted line, SSI) and test (dash-dotted line, SSII).

of the mean and standard deviation over the 15 models. Secondly, by dividing the subjects into high- and low-risk groups through the model's binary prediction, risk stratification curves showing the proportion of incident-free subjects for each group over time are obtained via the Kaplan-Meier estimator. Lastly, through the comparison of the 5.5-year model predictions to the cumulative AF diagnoses over time, a measure of the resulting AUC for each time stamp over the 5.5 years is obtained, indicating how the model performs when applied over the entire observation period.

3. Results

The average performances across the 15 models predicting long-term AF with a 5.5 year prediction horizon yielded an AUC of 0.60 ± 0.01, corresponding to a weighted F1 score of and 0.72 ± 0.03. The summarised confusion matrices across the models for the training (SSI) and test (SSII) sets, respectively, are shown in Fig. 1, where the trade-off between false positives and false negatives in relation to true ones can be observed.

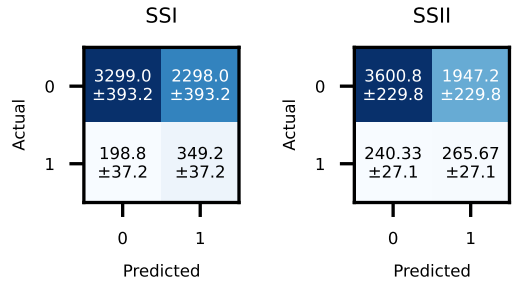

Figure 1: Confusion matrices for the training (SSI) and test (SSII) sets showing the mean and standard deviation over the 15 trained models.

The risk stratification curves in Fig. 2 show that, in both datasets, the probability of not being diagnosed with AF of the low-risk group is consistently higher than that of the high-risk group. This confirms that the novel feature set holds valuable information. In the test set, high-risk subjects had an 88 ± 0.5 % probability of not developing AF at the end of the observation period, while this percentage increased to 94 ± 0.2 % for the low-risk group.

The average feature importance, when summarised over the 15 models, is presented in Fig. 4, showing that the three most important features are the burden, prematurity and interval of $sSVEB_C$. In addition, among the most important features, several are related to the degree of prematurity of other SVA groups.

Figure 3 shows the average AUC of the 15 models computed for each year together with their overall average, corresponding to 0.66 and 0.60 for the training and test sets, respectively. It can be observed that, in the training set, the general trend is that the AUC decreases with time. This decreasing trend is also observed in the test set starting from 1.5 years after screening.

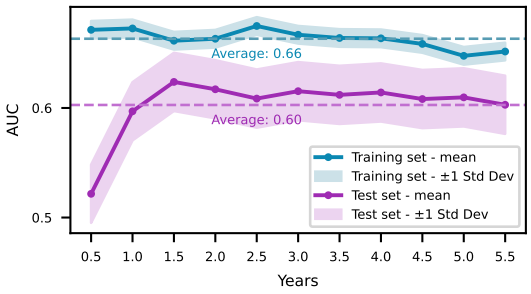

Figure 3: AUC when applying the proposed 5.5-year prediction model to endpoint data up until each half-year during the time span of the study.

4. Discussion

A new model for long-term AF prediction was developed based on a novel feature set describing various characteristics of SVA in the ECG. The results shown by the risk stratification curves and the confusion matrices in Figs. 1 and 2 confirm the predictive ability of the 1D-CNN model and, consequently, of the novel arrhythmia-based feature set for AF prediction. For both databases, a high number of subjects that did not get an AF diagnosis were misassigned in the high-risk group. These misassignments can be explained by that rhythm deviations are common in the elderly, even in those who do not develop AF in the future.

The feature importance ranking in Fig. 4 provides valuable information about SVA characteristics used by the

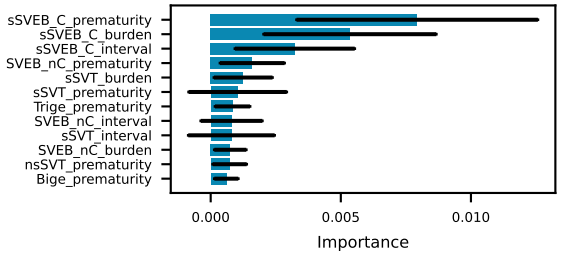

Figure 4: Feature importance ranking for the 12 most important features.

model. First of all, the 3 most important features are all related to the $sSVEB_C$ arrhythmia group. These features are related to deviations common also in combination with other SVAs. Prematurity is an important feature for most SVAs, which is a reasonable result as the higher degree of prematurity, the more the short-duration beats are deviating from the median RR-interval duration. The proposed model achieved an AUC of 0.60 ± 0.01, corresponding to a weighted F1 score of 0.72 ± 0.03 in the test set. When compared to previous studies, in particular Khurshid et al. who obtained an AUC of 0.82 for 5-year prediction [4], the difference in performance may be explained by differences in data, both in terms of the number of measurements and their type, i.e., >40 000 individuals and in-hospital 12-lead recordings for training in [4] and 6 145 patients and handheld single-lead ECGs in the present work.

5. Conclusion

In this study, a CNN model for long-term prediction of AF was trained with a novel set of features that characterises SVAs detected in 30-s ECG recordings. The developed model can distinguish between high- and low-risk patients with an average weighted F1 score of 0.72 ± 0.03. Hence these features, i.e., the degree of prematurity of SVA events, their respective burdens, and the intervals between such events, as well as the risk prediction models based on these features, contribute to a personalised risk assessment after AF screening.

Acknowledgements

This work was supported by the European Union Horizon 2022 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 101119941 (INSIDE-HEART).

References

[1] Svennberg E, Engdahl J, Al-Khalili F, et al. Mass screening for untreated atrial fibrillation: the Strokestop study. Circulation 2015;131(25):2176–2184.

[2] 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):1092–1101.

[3] Raghunath S, Pfeifer JM, Ulloa-Cerna AE, et al. 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. Circulation 2021;143(13):1287–1298.

[4] Khurshid S, Friedman S, Reeder C, et al. ECG-based deep learning and clinical risk factors to predict atrial fibrillation. Circulation 2022;145(2):122–133.

[5] Hirota N, Suzuki S, Arita T, et al. Prediction of current and new development of atrial fibrillation on electrocardiogram with sinus rhythm in patients without structural heart disease. Int J Cardiol 2021;327:93–99.

[6] Farinha JM, Gupta D, Lip GY. Frequent premature atrial contractions as a signalling marker of atrial cardiomyopathy, incident atrial fibrillation, and stroke. Cardiovasc Res 2023;119(2):429–439.

[7] Binici Z, Intzilakis T, Nielsen OW, et al. Excessive supraventricular ectopic activity and increased risk of atrial fibrillation and stroke. Circulation 2010;121(17):1904–1911.

[8] Hygrell T, Stridh M, Friberg L, Svennberg E. Prognostic implications of supraventricular arrhythmias. Am J Cardiol 2021;151:57–63.

[9] Johnson LS, Persson AP, Wollmer P, et al. Irregularity and lack of p waves in short tachycardia episodes predict atrial fibrillation and ischemic stroke. Heart Rhythm 2018;15(6):805–811.

[10] Fredriksson T, Gudmundsdottir KK, Frykman V, et al. Brief episodes of rapid irregular atrial activity (micro-AF) are a risk marker for atrial fibrillation: a prospective cohort study. BMC Cardiovasc Disor 2020;20:1–8.

[11] Kemp Gudmundsdottir K, Fredriksson T, Svennberg E, et al. Stepwise mass screening for atrial fibrillation using n-terminal b-type natriuretic peptide: the Strokestop II study. EP Europace 2020;22(1):24–32.

[12] Halvaei H, Svennberg E, Sörnmo L, Stridh M. Identification of transient noise to reduce false detections in screening for atrial fibrillation. Front Physiol 2021;12:672875.

[13] Halvaei H, Hygrell T, Svennberg E, et al. Detection of non-sustained supraventricular tachycardia in atrial fibrillation screening. IEEE J Transl Eng Health Med 2024;12:480–487.

Address for correspondence:
Livia Colucci
Department of Biomedical Engineering, Lund University, Lund, Sweden
livia.colucci@bme.lth.se