A Machine Learning approach for Atrial Fibrillation detection in telemonitored patients

Pedro L. Barrera[1*], L. G. Vecino Schandy[1*], M. P. Bonomini[2,3], C. Mateos[4], M. Hirsch[2,4], L. R. Grana[5] and S. Liberczuk[6,7]

[1] Instituto Tecnológico de Buenos Aires (ITBA), Argentina
[2] Instituto Argentino de Matemáticas Alberto P Calderon (IAM), CONICET, Argentina
[3] Departamento de Electrónica, Tecnología de Computadoras y Proyectos, Universidad Politécnica de Cartagena, Spain
[4] Instituto Superior de Ingeniería de Software Tandil (ISISTAN), Argentina
[5] Virtual Sense S.A., Argentina
[6] Centro de Altos Estudios en Tecnología Informática - Facultad de Tecnología Informática - Universidad Abierta Interamericana, Argentina
[7] Instituto de Ingeniería y Agronomía - Universidad Nacional Arturo Jauretche, Argentina

[*] These authors contributed equally to this work

pbarrera@itba.edu.ar
lvecino@itba.edu.ar

Abstract. Atrial fibrillation (AF) is the most common type of cardiac arrhythmia. As it is typically asymptomatic, it often goes undiagnosed until major complications arise, such as stroke. Therefore, the development of rapid, economical, and widely accessible diagnostic tools for detecting AF at an early stage is crucial. Telemonitoring with machine learning-assisted devices shows promise in achieving this goal. This paper presents an algorithm that automatically detects AF in signals obtained by portable electrocardiographs connected to a telemonitoring platform via smartphones. The algorithm consists of three stages: a noise detection, ectopic beat removal and an AF detection. The noise detection involves analyzing the ECG signals using 5-second windows with a 1-second shift. A K-nearest neighbors (KNN) classifier predicts the presence or absence of noise in each window, allowing for the detection of noisy and non-noisy segments of the signal. The non-noisy segments are processed using a Pan-Tompkins algorithm to find the R peaks of the signal, and the corresponding RR interval series. Then ectopic beats are removed using an XGBoost classifier, generating the NN series. In the AF detection stage, X features are obtained from this series, which serve as input features of an XGBoost classifier that predicts the presence or absence of AF in the ECG signal. The algorithm was trained and tested using the Physionet Short Single-Lead AF Database (SSLAFDB) and achieved an accuracy of 90.87% and an F1-score of 90.91%. Further validation was performed by an external partner using two other databases, reporting an accuracy of 90.41% and 89.61% respectively.

Keywords: Machine Learning, Biomedical Signal Processing, ECG, Atrial Fibrillation, Telemonitoring.

1 Introduction

Atrial fibrillation (AF) is the most common type of cardiac arrhythmia and is associated with a significantly increased risk of stroke and other complications [1]. Due to its often asymptomatic nature, AF usually goes undiagnosed until major complications arise [2]. Timely and accurate detection of AF is, therefore, critical to prevent serious health consequences. Telemonitoring using machine learning-assisted devices has emerged as a promising solution to achieve these objectives. In this paper, we present a machine learning-based algorithm that automatically detects AF in signals obtained by portable electrocardiographs (ECG) connected to a telemonitoring platform via smartphones.

The algorithm was commissioned by Virtual Sense, a company dedicated to developing digital platforms for remote health monitoring, with the aim of facilitating access to telemedicine and improving patients' quality of life.

ECG signals with AF episodes exhibit three distinctive properties: highly irregular rhythm, absence of P waves, and presence of f waves [3]. Since ECG signals from telemonitored patients are expected to be of low quality with a high level of noise, morphology analysis (i.e., assessing the presence of P and f waves) is not reliable. As a result, the AF detection method presented in this paper utilizes a rhythm- based approach. The algorithm consists of three stages: a noise detection stage, an ectopic beat elimination phase and an AF detection stage. The noise detection stage uses a KNN classifier to identify noisy segments of the signal and the ectopic beat elimination phase utilizes an XGBoost classifier to identify the beats. Finally, the AF detection stage uses a logistic regression classifier trained on indices calculated from the RR interval series of the ECG signal. The complete algorithm pipeline is shown in **Fig. 1** and will be explained in section 2.

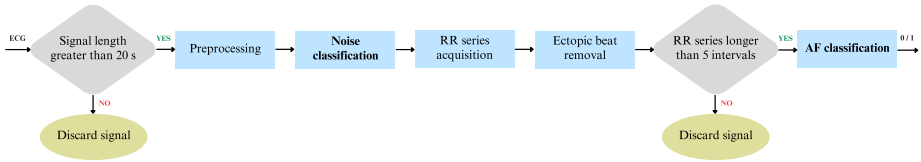


Fig. 1. Algorithm pipeline

2 Materials and Methods

2.1 Data

The Short Single-Lead Atrial Fibrillation Database (SSLAFDB), from Physionet [4] was used to develop and train the algorithm. This database consists of 8528 lead I ECG signals, with a sampling frequency of 300 Hz and duration ranging between 9 and 60 seconds. These signals were obtained with KardiaMobile AliveCor devices, which are portable ECG devices for smartphones [5]. The cardiac rhythms annotated are divided

into 4 classes: N (normal/sinus rhythm), A (AF), O (other arrhythmias) and ~ (noisy signals). This database was selected due to similarity with the acquisition method to be used by Virtual Sense.

Testing was performed using a subset of the PTB-XL database [6] selected and modified by Kaggle [7] and a synthetic database. This database consists of 6528 ECG 12-lead signals, with sampling frequency of 500 Hz and 10-second duration. The cardiac rhythms annotated by Kaggle are divided into 3 classes: SR (normal/sinus rhythm), AF (AF and atrial flutter), and VA (other arrhythmias). In this study, lead I was chosen since single-lead ECG devices usually acquire this lead. The synthetic database was generated using a PhysioNet Model for Simulating ECG and PPG Signals with Arrhythmia Episodes [8]. The parameters used for the simulation include ECG lead I, sampling frequency of 250 Hz, mix of ECG noises with RMS equal to 0.06 mV. The resulting database consists of 10000 signals, with balanced rhythms N and A.

For the ectopic beat automatic detection stage, the MIT-BIH database was used [9]. The database consists of 48 half-hour ECG recordings, with each record containing two channels of digitized ECG data sampled at 360 Hz. The benefit for ectopic beat detection is that it contains beat annotations. The database contains various annotations within the record, including different types of beats, as well as P-waves and U-waves, among others. According to Khalaf et al. [10], there are often challenges in achieving accurate validation when using this database, mostly due to heterogeneities in the beat annotations. Based on the recommendation of annotations considered as beats and supported by our own judgment, the following annotations were chosen: ['N', 'L', 'R', 'A', 'J', 'S', 'V', 'E', 'F', 'Q', '/', 'a', 'e', 'j', 'f']. This was done to achieve a more robust and consistent algorithm in line with the existing literature. In this case, 'V' represents a ventricular ectopic beat, while 'A' represents an atrial ectopic beat (or supraventricular). All other beats were labeled as 'O' (others).

2.2 ECG Preprocessing

Firstly, signals shorter than 20 seconds were discarded. An exception was made for the PTB-XL database, where due to superior signal quality, reliable AF detection is possible even in 10-second signals. Subsequently, a two-step moving median filter was applied to the signals, with window sizes of 200 ms and 600 ms respectively. The 200 ms filter was applied to the signal and the 600 ms filter applied to its output. Then, the filtered signal was obtained by subtracting the output of the second filter to the original signal. This proved to be very effective in the elimination of baseline wander and some residual noise [11,12]. The power line interference was also removed applying a low-pass filter with cutoff frequency of 40 Hz.

The ECG signals from lead I have a waveform in which the R peaks of normal beats are positive. Therefore, the polarity was analyzed by detecting the R peaks of the signal, obtaining the QRS templates, and calculating their median template. If the magnitude of the maximum value of the median QRS template was greater than the magnitude of the minimum value, the polarity was considered positive. Otherwise, the polarity was considered negative, and the signal was corrected by multiplying it by -1.

Finally, the amplitude of the signals was normalized windows so that the amplitude of the QRS peaks ranges between -1 and 1. This was done by calculating the new median QRS template and finding its maximum value. The final signal was obtained by dividing the signal by this maximum value.

2.3 Noise Detection

The noise detection phase was treated as a binary classification problem, with 'noisy' and 'non-noisy' as the two classes. The features used for this problem were Signal Quality Indices (SQI) calculated from 5-second windows. Training data was gathered through manual selection of these windows from the SSLAFDB. Specifically, 364 noisy windows were chosen from the signals annotated as 'noisy', and 468 non-noisy windows were selected from signals annotated with rhythms 'N', 'A', and 'O', ensuring the inclusion of all available rhythms. Some examples of the selected windows are shown in **Fig. 2**.

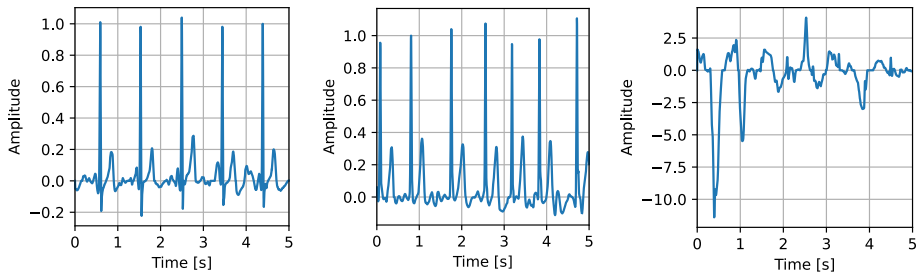


Fig. 2. Example of manually selected, preprocessed SSLAFDB windows. On the left, a noisy window. In the middle, a non-noisy signal with rhythm A. On the right, a non-noisy window with rhythm N.

For every window, 4 SQIs were obtained [13,14]:

- bSQI: Compares the detection of two well-documented open-source QRS complex detectors with different noise sensitivities, named *ep_limited* and *wqrs*.
- pSQI: Compares the energy present in the frequency bands containing most of the QRS complex energy (5-15 Hz) and the whole ECG excluding low frequency (5-40 Hz). Used to detect EMG interference.
- sSQI: Based on the skewness of the signal.
- kSQI: Based on the kurtosis of the signal.
- basSQI: Compares the energy present in the frequency bands containing the baseline (0-1 Hz) and the whole ECG (0-40 Hz).
- fSQI: The proportion of the signal that appears to be a flat line. It is calculated as the percentage of the signal whose gradient is less than 1% of the maximum.

In this rhythm-based AF detection approach, the primary goal of noise filtering is to accurately identify of R peaks. A 5th order Butterworth bandpass filter with cutoff frequencies between 8 and 25 Hz was used. Any apparent noise with frequency components outside the range of the band-pass filter will be removed and will not interfere with R peak detection. As such, it is logical to calculate these indices post band-pass filtering. However, the presence of energy in a broader frequency range could accentuate the contrast of noise with frequency components between 8 and 25 Hz, thereby enhancing noise detection. Consequently, sSQI, kSQI, fSQI and bSQI are also calculated prior to the application of the band-pass filter. For pSQI and basSQI, the band-pass filter isn't applied, as these indices estimate the energy in frequency bands that would be removed by the band-pass filter.

To choose the noise classifier, four widely used machine learning algorithms were evaluated: Support Vector Machines (SVM), K Nearest Neighbors (KNN), Logistic Regression (LR) and XGBoost.

The optimal combination of indices and hyperparameters for the classifier was determined through iterative evaluation of all possible combinations for each of the four classifiers. The final selection of classifier, features, and hyperparameters was based on overall performance, analyzing accuracy, sensitivity, specificity and F1-score. This process was performed using 80% of the data, reserving 20% for final validation.

When determining the noise level of a new signal, each second of the signal is assessed for quality. This involves analyzing the preprocessed signal using 5-second windows with a 1-second shift. The probability of each window being noisy is determined using the trained classifier. Then, the noise probability for each second of the signal is calculated using a weighted average of the probabilities from the five encompassing windows. The weighting for this average is performed using a weight vector of (1, 2, 3, 2, 1).

Once the probability is calculated for every second of the signal, this value is rounded to the nearest integer, resulting in either 0 (non-noisy) or 1 (noisy).

2.4 RR series acquisition

The QRS time occurrences of a signal (also known as R peaks) were obtained by applying a Pan-Tompkins algorithm to the non-noisy segments of the signal. The RR interval series for the signal was created by listing the intervals defined by these R peaks, which depicts the structure of the signal based on the time elapsed between two consecutive heartbeats. If the obtained RR series was shorter than 5 intervals, its analysis was considered not reliable and was therefore discarded.

While this R peak detection method is mostly accurate, there are cases where peaks are missed or detected twice. To address this issue, the RR series is processed to remove excessively long intervals or merge intervals that are too short with the adjacent smallest one.

2.5 Ectopic beat handling

AF is characterized by irregular RR intervals due to the absence of a consistent atrial contraction pattern. Ectopic beats can further contribute to irregularities in the RR interval series, making it challenging to identify and classify AF episodes accurately. By excluding ectopic beats, the RR interval series better represents the underlying rhythm, enhancing AF detection.

To accomplish this, the RR interval series were initially constructed from the annotated beats. Given that rhythm changes are also annotated, separate RR series were assembled for each rhythm present. Twelve parameters were derived from each RR interval of each series: the prior, current, subsequent, and second-next RR interval, all normalized by the median RR interval in the series; the mean of the current and next normalized interval; two ratios obtained from the RR ratio series, defined as the quotient of the n-th interval and the previous interval; the 10th, 35th, and 90th percentiles of the ratio series, and an irregularity parameter derived from the Poincaré plot. This last parameter was added to mitigate confusion in the classifier due to AF irregularity.

After identifying the ectopic beats, the premature and compensatory RR intervals surrounding them were removed, resulting in the NN series. Only patient files with annotated ectopic beats were utilized. For training and hyperparameter tuning, 80% of these patient files were employed, with K-fold cross-validation used to evaluate the performance of each hyperparameter combination. The remaining 20% of patient files were reserved for final validation.

The classifier selection will prioritize maximizing precision, while maintaining a reasonable level of sensitivity. This approach is chosen due to the dataset's high imbalance, where normal beats significantly outnumber the ectopic ones.

2.6 AF detection

As mentioned earlier, this AF detection method utilizes a rhythm-based approach. Therefore, the input features for the AF classifier are irregularity and rythm indices obtained from the NN interval series. Several indices were explored, including those based on statistical dispersion, Shannon and sample entropy, the number of inflection points, Poincaré plots, and coherence function. Histogram-based indices were not assessed since they are recommended for signals of long duration with a minimum of 100 beats [3]. Finally, 16 indices were assessed: three based on statistical dispersion, commonly known as NMASD, RMSSD and the coefficient of variation, PNN50 and PNN100, Shannon entropy, three based on sample entropy, the relative number of turning points, and five based on the Poincaré plots (x(n), x(n+1)), (Δx(n), Δx(n+1)) and (x(n), Δx(n)).

To select the optimal classifier, hyperparameters, and features, the same iterative method used in the noise detection stage was applied. In addition to the four previously mentioned classifiers, AdaBoost was included in the evaluation. The process involved 5-fold cross-validation on 80% of the data for tuning and training, while the remaining 20% was used for validation.

The performance of the AF classifier trained on SSLAFDB was internally assessed, testing on N (class 0) and A rhythms (class 1), as well as on N, A, and O rhythms where N and O were combined into the AF absence class (class 0). Furthermore, the trained AF classification model was also validated by external researchers from Virtual Sense, using both the Kaggle-modified PTB-XL database and the synthetic ECG database.

3 Results

For the noise classification, a KNN classifier with 8 neighbors, inverse distance weighting, and Manhattan distances achieved an accuracy, sensitivity, specificity, and F1-score of 94.28%, 94.10%, 94.67% and 94.33%, respectively on the training set. On the validation set, the classifier obtained an accuracy, sensitivity, specificity, and F1-score of 91.62%, 90.41%, 92.55% and 90.41%, respectively.

The results of the noise classifier indicate that approximately 7 in 100 noisy windows will be misclassified as non- noisy, and vice-versa. When analyzing the noise of a new signal, the noise algorithm performs a weighted average of the noise probability of the neighboring windows. Therefore, some of these misclassifications are likely to be corrected when analyzing the complete signal, thus increasing the accuracy of the noise classifier.

For the ectopic detection, the XGBoost classifier with 300 decision trees, each with a maximum depth of 7, achieved a balanced accuracy, precision, specificity, sensitivity, and F1-score of 91.85%, 89.35%, 84.41%, 99.29%, and 86.81%, respectively on the training set. On the validation set, the classifier attained a balanced accuracy, precision, sensitivity, specificity, and F1-score of 93.87%, 92.55%, 88.23%, 99.51%, and 90.34%, respectively. Normal accuracy was not reported due to its diminished relevance in the highly imbalanced dataset.

For AF detection, the AdaBoost classifier, consisting of 150 decision trees with a maximum depth of 5, trained on 8 AF features, exhibited superior performance. This classifier not only provides the predicted class (AF or no AF) but also estimates the probability of being AF. Figure 3 illustrates examples of fully processed signals, including noise and beat detection, along with the predicted AF probability. The evaluation metrics for the AF classifier, obtained using the SSLAFDB, PTB-XL, and synthetic databases, are summarized in Table 1.

Table 1. Internal and external validation results of AF classification (in %).

Database	**SSLAFDB**		**PTB-XL**	**Synthetic**
Set	Training	Validation	All	All
Accuracy	90.10	90.87	90.41	89.61
Sensitivity	90.55	91.27	85.35	82.02
Specificity	91.50	90.48	92.04	85.35
F1-Score	90.98	90.91	81.30	90.29

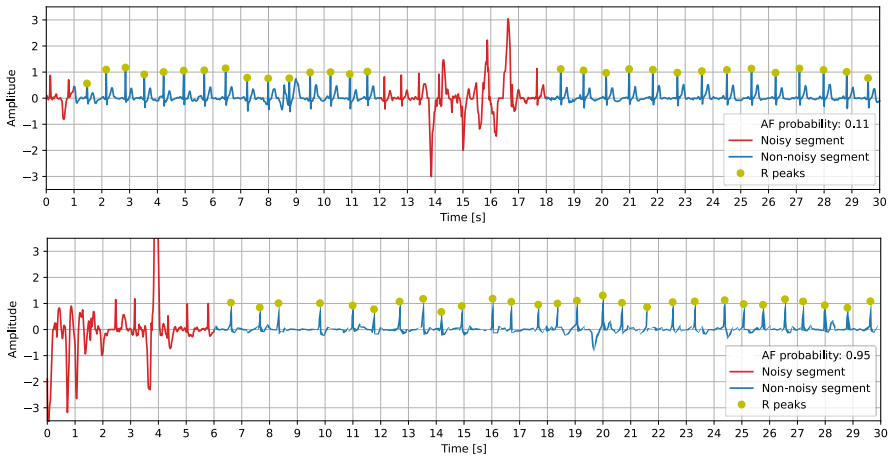


Fig. 3. Example of preprocessed SSLAFDB signals. Noisy segments are in red, non-noisy in blue, R peaks in green. The top signal (rhythm N) has an AF probability of 0.11, while the bottom signal (rhythm A) has an AF probability of 0.95.

4 Discussion

The noise detection algorithm under consideration demonstrated solid overall performance. Numerous noise detection and signal quality assessment algorithms have been documented before. Among the initial efforts to employ the SQIs utilized in this work was a study by Li et al. [14]. However, the performance of their work was gauged indirectly via the heart rate estimation error, making a direct comparison of results impossible. Zhao et al. [13] also developed a noise detection system based SQI thresholds and fuzzy logic This system was trained and validated using the SSLAFDB, and it achieved an accuracy, sensitivity, and specificity of 92.67%, 97.33%, and 88.67% respectively, results that align closely with those achieved in this work. Nonetheless, the metrics reported in this work and in Zhao's are hard to compare. In their work, they classified between noisy and non-noisy signals, whereas in this case the objective is the quality assessment of noisy segments, using the classifier to estimate the probability of 5 second windows being noisy and performing a weighted average to assess the quality of each individual second of the signal.

Ectopic removal is a stage often incorporated into AF detection algorithms. Among the varied approaches, one simple rule-based method was developed by Dash et al. [15], which solely depends on three RR series ratios and associated thresholds. However, like many ectopic beat removal techniques used in AF detection algorithms, the actual performance of their approach remains unreported, leaving its effectiveness uncertain. In contrast, the ectopic beat detector developed within this study has shown satisfactory results. García et al. [16] developed a more sophisticated technique, creating an ectopic classifier for Holter records with superior performance. This method fuses morphology

features with RR series features, resulting in a low incidence of false negatives. Yet, given the unreliability of morphology analysis in low-quality signals, such as those captured by smartphone devices, their approach may not be optimal or directly comparable when applied to ECG data from smartphone devices.

The AF detection algorithm demonstrated solid performance across various databases, suggesting that the model achieved a substantial level of generalization. A key advantage of the presented method is its simplicity; it proposes a straightforward algorithm that only requires eight features for classification, employing a relatively uncomplicated model. In contrast, many other methods designed for accurate AF detection in low-quality signals propose the use of complex models like convolutional neural networks, as in Hong et al.'s [17] work. Alternatively, they use feature-based classifiers similar to this study but with a larger feature set, as in Datta et al. [18]'s work, which used over 20 features. These two studies were co-winners of the first place in the 2017 Physionet Computing in Cardiology challenge [19].

The Apple Watch 4 (AW4) rhythm detection system provides a comparative standpoint, with two rhythm assessment methods: real-time notification and offline PDF waveform analysis. According to the AHA journal [20], the real-time notification method demonstrated a sensitivity of 41% for AF detection and a 61% agreement with telemetry. However, it resulted in inconclusive findings in 31% of cases. In contrast, the PDF waveform analysis achieved a sensitivity of 96% and a telemetry agreement of 98.9%. It is important to note that the PDF waveform analysis is not directly comparable to the real-time algorithm developed in this study. This study's algorithm prioritizes robust performance, low feature requirements, and operational simplicity on simpler devices, potentially offering advantages over commercial systems like AW4, particularly for real-time applications.

5 Conclusion

In conclusion, this paper details a robust machine learning-based algorithm for atrial fibrillation detection, utilizing a three-stage approach: noise detection, ectopic beat removal, and AF detection. The algorithm shows strong performance metrics and is particularly robust, having been trained on low-quality signals acquired by smartphone devices. The results suggest high accuracy in AF detection even amidst noise and other arrhythmias. With the rise of telemonitoring in healthcare, the algorithm could potentially play a crucial role in improving telemedicine access and patient care. However, further validation is required on signals from telemonitored patients via the Virtual Sense platform. Future studies will assess its performance on this platform, providing valuable insight into real-world applicability. The results could also guide further refinements to the algorithm, including the addition of features and training data, advancing its utility in AF detection and telemonitoring.

References

1. Argentina Health Ministery. “Glosario de salud”, https://www.argentina.gob.ar/salud/glosario/fibrilacionauricular, last accessed 2023/06/23.
2. Turakhia Mintu P., Shafrin, J., Bognar, K., et al. Estimated prevalence of undiagnosed atrial fibrillation in the United States. PLOS ONE 13,1-11 (2018).
3. Sörnmo, L.: Atrial Fibrillation from an Engineering Perspective. 1st end. Springer (2018).
4. Goldberger, A., Amaral, L., Glass, L., et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation 101 (23), e215–e220 (2000).
5. Clifford, G. D., Liu, C., Moody, B., et al. AF Classification from a Short Single Lead ECG Recording: The PhysioNet/Computing in Cardiology Challenge 2017. Computing in cardiology, 44, 2 (2017).
6. Wagner, P., Strodthoff, N., Bousseljot, RD. et al. PTB-XL, a large publicly available electrocardiography dataset. Sci Data 7, 154 (2020).
7. Inc. Kaggle. “PTB-XL - Atrial Fibrillation Detection”, https://www.kaggle.com/datasets/arjunascagnetto/ptbxl-atrial-fibrillation-detection?resource=download, last accessed 2023/04/23.
8. Sološenko, A., Petrėnas, A., Paliakaitė, et al. Model for Simulating ECG and PPG Signals with Arrhythmia Episodes. (version 1.3.0). PhysioNet (2021).
9. Moody, G. B., & Mark, R. G. The impact of the MIT-BIH arrhythmia database. IEEE engineering in medicine and biology magazine: the quarterly magazine of the Engineering in Medicine & Biology Society, 20(3), 45–50 (2001).
10. Khalaf, A. & Mohammed, S. Verification and comparison of MIT-BIH arrhythmia database based on number of beats. International Journal of Electrical and Computer Engineering (IJECE). 11. 4950. pp 4950-4961 (2021).
11. Hossain, M. B., Bashar, S. K., Walkey, A., J., et al. An Accurate QRS Complex and P Wave Detection in ECG Signals Using Complete Ensemble Empirical Mode Decomposition with Adaptive Noise Approach. IEEE Access, 128869-128880 (2019).
12. Panigrahy, D., Sahu, P. P and T wave detection and delineation of ECG signal using differential evolution (DE) optimization strategy. Australasian Physical Engineering Sciences in Medicine 41(1), 225–241 (2018).
13. Zhao, Z., Zhang, Y. SQI Quality Evaluation Mechanism of Single-Lead ECG Signal Based on Simple Heuristic Fusion and Fuzzy Comprehensive Evaluation. Frontiers in Physiology 9, 727. (2018).
14. Li, Q., Mark, R. G., Clifford, G. D. Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter. Physiological Measurement 29(1), 15–32. (2007).
15. Hong, S., Wu, M., Zhou, Y., et al. (2017). ENCASE: an ENsemble ClASsifiEr for ECG Classification Using Expert Features and Deep Neural Networks. Computing in Cardiology (CinC), 1-4 (2017).
16. Dash, S., Chon K. H., Lu S., Raeder E. A. Automatic Real Time Detection of Atrial Fibrillation. Annals of Biomedical Engineering 37(9), 1701-1708 (2009).
17. García, A., Romano, H., Laciar, E., Correa, R. Development of an algorithm for heartbeats detection and classification in Holter records based on temporal and morphological features. Journal of Physics: Conference Series, 332(1), 12023 (2011).
18. Datta, P., Puri, C., Mukherjee, A., et al. Identifying normal, AF and other abnormal ECG rhythms using a cascaded binary classifier. Computing in Cardiology (CinC), pp. 1-4 (2017).

19. Physionet “2017 computing in cardiology”, https://physionet.org/content/challenge-2017/1.0.0/, last accessed 2023/06/20.
20. Seshadri, D. R., Bittel, B., Browsky, D., et al. Accuracy of Apple Watch for Detection of Atrial Fibrillation. Circulation 141(8), 702-703 (2020).