Abnormal Rhythm Detection from a Single-lead ECG via a Recurrent Neural Network

Jérôme Van Zaen[1], Guillaume Bonnier[1], Jakub Parak[2], Mikko Salonen[2], Yara-Maria Proust[1], Luisa Marques[3], Alia Lemkaddem[1], Cyril Pellaton[3], Mathieu Lemay[1]

[1] CSEM, Neuchâtel, Switzerland
[2] Firstbeat Technologies, Jyväskylä, Finland
[3] Cardiology Unit, Réseau Hospitalier Neuchâtelois, Neuchâtel, Switzerland

Abstract

Cardiac arrhythmias affect millions of individuals worldwide and can lead to severe complications such as stroke or heart failure. They can be difficult to diagnose with ambulatory electrocardiogram monitors due to their transient nature. We propose a system for long-term arrhythmia monitoring that takes single-lead electrocardiogram and tri-axis acceleration signals as inputs. It is composed of a beat detector to extract interbeat intervals and a classifier to detect arrhythmias. This system is evaluated on two datasets including 42 patients and achieves an accuracy of 0.988 for the abnormal class, 0.967 for the normal class, and 0.979 for the tachycardia class.

1. Introduction

Millions of individuals suffer from cardiac arrhythmias (CAs). While most CAs are not directly life-threatening, they can lead to serious complications. Atrial fibrillation (AF), the most common CA, affects between 2 % and 4 % of the adult population and is associated with substantial morbidity and mortality [1]. In addition, ventricular arrhythmias are the cause of approximately 80 % of sudden cardiac deaths [2]. An abnormal heart rate (HR) may cause symptoms such as palpitations, shortness of breath, or fainting. However, symptoms are not necessarily present and many patients are asymptomatic [3] which prevents early detection.

A 12-lead electrocardiogram (ECG) reviewed by a trained cardiologist is the gold standard to diagnose CAs. However, it is not applicable for long-term monitoring. In addition, it might miss transient CAs. Therefore, the common solution is to use a Holter for long-term monitoring with the drawback that reviewing recordings of 24 hours or more is expensive and time-consuming. Usually, automatic systems are used to tag parts that most likely include CAs to be reviewed by a specialist [4]. We propose such a system to detect CAs from a single-lead ECG and tri-axis acceleration signals recorded by a sensor that can be worn over long periods of time. This system combines a beat detector to extract interbeat intervals (IBIs) and a recurrent neural network to classify CAs.

2. Materials and Methods

2.1. Datasets

The proposed system is evaluated on two datasets including single-lead ECG and acceleration data acquired with the Bodyguard 3 device (Firstbeat Technologies, Jyväskylä, Finland) shown in Figure 1. The first dataset was collected at Hôpital Pourtalès, Neuchâtel, Switzerland from 20 patients with suspected CA during 24-hours Holter monitoring (BASEC ID: 2022-00986). The second dataset was collected at KNF-Laboratoriot Oy, Helsinki, Finland from 26 patients with apnea during overnight respiratory polygraphy (NCT05235984). Both studies were approved by local ethics committees. The reference rhythm annotations were obtained automatically with the Cardiac Navigator software (Bittium, Oulu, Finland) for the two datasets and include the following rhythms: atrial bigeminy, AF, normal sinus rhythm, second degree heart block, sinoatrial block, sinus bradycardia, supraventricular tachyarrhythmia, ventricular bigeminy, and ventricular tachycardia. The only preprocessing applied before detecting CAs was to resample the acceleration and ECG signals to 128 Hz.

2.2. System

We propose a system composed of the following elements to detect abnormal rhythms from a single-lead ECG and acceleration signals:

1. Beat detection from ECG
2. Interbeat interval windows extraction
3. Window filtering using acceleration

Figure 1: Bodyguard 3 device.

4. Interbeat interval outlier detection
5. Abnormal rhythm classification

Cardiac beats are detected by finding the ECG extremum in segments where the curvature of the ECG signal is above an adaptive threshold. The IBIs are computed from the detected beats. Then, we extract IBI windows without overlap by grouping successive IBIs until their cumulative sum exceeds 30 s.

The IBI windows are filtered to exclude the ones with motion. Motion is detected by computing a motion signal $m(n)$ from the tri-axis acceleration signals $a(n)$ as follows,

$$m(n) = ||a(n) - a(n-1)||$$

where $||\cdot||$ denotes the Euclidean norm. A threshold of 0.01 g (g = 9.81 m s^{-2}) is then applied to this motion signal and IBI windows where the proportion of values above the threshold is greater than 0.05 are labeled motion.

Valid IBIs are then detected in each window not labeled as motion. An IBI is considered valid if it falls in the range [0.25 s, 3 s] and the amplitude of the corresponding R wave is less than 6 mV. All windows where the number of valid IBIs is less than 12 or the proportion of valid IBIs is less than 0.9 are labeled as undecidable.

A classifier is then applied to the valid IBIs in each remaining window. The classifier is a neural network with two layers: a gated recurrent unit layer [5] that takes sequences of IBIs as input and a sigmoid layer to output the probability of abnormal rhythm. This classifier was trained on five separate dataset: three datasets from PhysioNet [6] (European ST-T Database [7], Long Term AF Database [8], and MIT-BIH Arrhythmia Database [9]) and two internal datasets. The valid IBIs are also used to compute the mean IBI in each window. Finally, each window is assigned a label with the following procedure:

- If the mean IBI is greater than 1 s (corresponding to a HR under 60 bpm), the label is bradycardia.
- If the abnormal rhythm probability is greater than 0.7, the label is abnormal.
- If the mean IBI is less than 0.6 s (corresponding to a HR above 100 bpm), the label is tachycardia.
- If the abnormal rhythm probability is less than 0.3, the label is normal.
- Otherwise, the label is undecidable.

The undecidable label indicates IBI windows for which a reliable decision is not possible.

Table 1: Performance metrics for each class.

Class	Accuracy	TPR	TNR	PPV	NPV	F_1 score
Abnormal	0.944	0.691	0.976	0.787	0.961	0.736
Bradycardia	0.892	0.985	0.883	0.429	0.999	0.598
Normal	0.861	0.836	0.964	0.990	0.587	0.907
Tachycardia	0.967	N/A	0.967	0	1	0

2.3. Performance Evaluation

To evaluate performance, we mapped the reference annotations for all windows to one of the labels outputted by the proposed system using the following rules. Normal sinus rhythm is mapped to normal, sinus bradycardia to bradycardia, and ventricular tachycardia to tachycardia. Atrial bigeminy, AF, second degree heart block, sinoatrial block, supraventricular tachyarrhythmia, and ventricular bigeminy are all mapped to abnormal.

Windows labeled as motion or undecidable by the system or with multiple reference annotations were excluded before computing the following classification metrics for each class: accuracy, true positive rate (TPR), true negative rate (TNR), positive predictive value (PPV), negative predictive value (NPV), and F_1 score.

We also reviewed records where the system achieved poor performance to better understand its limitations.

3. Results

We obtained 101 246 windows after applying the proposed system to the two datasets. We excluded 45 355 windows due to motion (35 757), multiple reference annotations (8206), and undecidable labels (1392). This resulted in a set of 55 891 windows with the following reference annotations: normal (45 010), abnormal (6319), and bradycardia (4562). No window was annotated as tachycardia. The global accuracy computed on these windows was 0.832 and the metrics for each class are reported in Table 1.

A few observations can be derived from these metrics. First, the number of false negatives is non-negligible for abnormal rhythms meaning that several cases are not detected. Second, most bradycardia episodes are correctly detected at the cost of many false positives. Third, there are many false negative for normal rhythms. Finally, some

tachycardia episodes were detected although there were no tachycardia in the annotated references. To better understand these discrepancies, we reviewed the records with the poorest performance for the different cardiac rhythms.

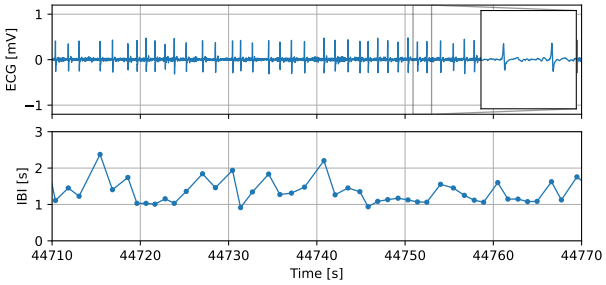
In the record with the worst classification performance for AF, more than 73 % of windows annotated as AF are classified as bradycardia by our system. However, the HR is quite low (below 50 bpm in most cases). An excerpt of this record is shown in Figure 2a where the mean IBI per window is above 1 s resulting in a bradycardia classification. This patient was under beta blocker medication at the time of the recording which explains the very low HR for AF. One can argue that patients under medication affecting cardiac rhythms are already monitored by their physician for cardiac anomalies. Therefore, these patients are outside the scope of our approach for long-term monitoring.

There are two records with many windows annotated as normal sinus rhythm that are classified as abnormal by our system. Reviewing the ECG signals and the corresponding IBIs suggests that the reference annotations are incorrect for these records. The IBIs are very irregular during these misclassified episodes and the proposed system seems to be correct when predicting an abnormal rhythm as illustrated in the example shown in Figure 2b.

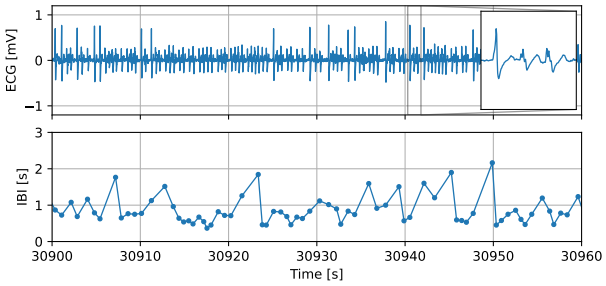
In the only record with a non-negligible number of windows annotated as ventricular bigeminy, the proposed system outputted a normal label for most of these windows. Further investigation of the ECG signal and corresponding IBIs revealed that the rhythm is indeed abnormal. However, it appears to be premature ventricular contractions (PVCs) instead of bigeminy. Nonetheless, the system fails to recognize this rhythm as abnormal. A short example of such incorrect classification is shown in Figure 2c. This is a clear limitation of our system which is most likely caused by the very limited number of examples with many PVCs included in the datasets used to train the neural network.

We also observed several records with many windows annotated with normal sinus rhythm references but classified as bradycardia or tachycardia by our system. However, in these cases, the mean IBI per window was always either above the threshold for bradycardia (1 s) or under the threshold for tachycardia (0.6 s). We even found several windows with a HR above 120 bpm which were annotated as normal sinus rhythm. Therefore, such classification errors most likely arise because the software used to obtain reference annotations uses a different approach to estimate HR (and maybe slightly different thresholds). In particular, our system does not discriminate between normal and abnormal beats (beyond the validity check for IBIs described in Section 2) before estimating the HR.

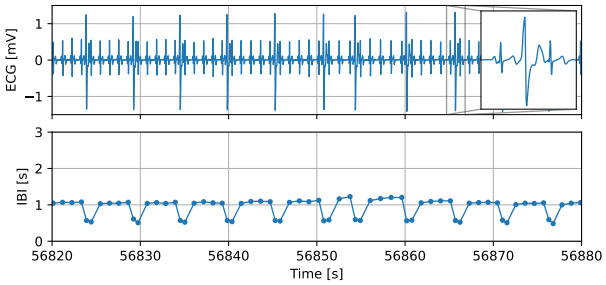
Lastly, the only record with sinoatrial block automatic reference annotations included more than 10 hours of such annotations. The windows in this part of the record were



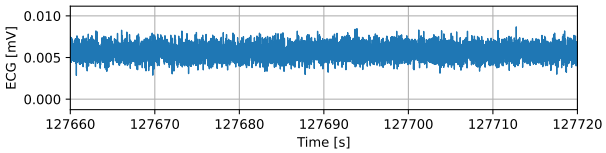
(a) ECG annotated as AF and classified as bradycardia.



(b) ECG annotated as normal sinus rhythm and classied as abnormal.



(c) ECG annotated as ventricular bigeminy and classified as normal.



(d) ECG annotated as sinoatrial block and classified as undecidable.

Figure 2: Excerpts of different ECG recordings.

systematically labeled undecidable by the proposed system. Visualizing the ECG signal reveals that it is composed exclusively of noise and that the recording device was most likely not worn during this part of the record. A short excerpt is shown in Figure 2d. Therefore, the reference annotations for sinoatrial block apprear to be incorrect and this example shows that our system is able to reject noisy ECG signals.

After considering all the discrepancies between the reference annotations and the predictions of the proposed system, we re-computed the performance metrics with two

Table 2: Performance metrics for each class when considering bradycardia as normal and excluding four records.

Class	Accuracy	TPR	TNR	PPV	NPV	F_1 score
Abnormal	0.988	0.820	0.994	0.854	0.993	0.837
Normal	0.967	0.973	0.820	0.993	0.545	0.983
Tachycardia	0.979	N/A	0.979	0	1	0

major changes. First, we considered bradycardia as normal in both the annotations and predictions to avoid classification errors for windows annotated as normal sinus rhythm despite a HR under 60 bpm. Second, we excluded four records: one record from a patient under beta blocker medication with AF and a very low HR (Figure 2a), one record with many normal sinus rhythm annotations despite a HR above 100 bpm, and two records with abnormal rhythms clearly visible in the ECG and IBIs but with normal sinus rhythm annotations (Figure 2b). It is worth mentioning that we did not exclude the patient with PVC windows misclassified as normal reported above (Figure 2c) to avoid over-inflating performance metrics. With these modifications, the number of windows was reduced to 49 610 and the global accuracy increased to 0.967. The metrics per class are reported in Table 2. As expected, the performance increased. However, the NPV for normal rhythm is still low indicating too many false negatives for this class. In addition, the TPR and PPV for abnormal rhythm are still under 0.9 showing that there is still room for improvement.

4. Conclusion

We proposed a system for long-term monitoring of abnormal cardiac rhythm from a single-lead ECG and tri-axis acceleration signals. This study has two main limitations. First, the reference rhythm annotations were obtained with a commercial software without verification by a specialist. Consequently, some annotations were incorrect which resulted in underestimated performance metrics. In particular, bradycardia and tachycardia seem to use different thresholds on HR than the one used in our system. This was partially corrected by considering bradycardia as normal and excluding four records. Future assessments of our system should be performed on datasets with reference annotations reviewed by a specialist. Second, the neural network classifier in the proposed system was trained on datasets including few examples with many PVCs. The classifier was thus unable to recognize PVC windows as abnormal. We plan to retrain the classifier with datasets including more examples with many PVCs to correct this issue. Furthermore, false positives for abnormal rhythms could be reduced by considering the class of successive windows (e.g. four abnormal windows out of four) to identify them as abnormal. Notwithstanding these limitations, the proposed system provides a robust solution to process daily life data, automatically rejecting low quality ECG (due to motion or noise) and, detecting abnormal rhythm with a global accuracy above 95 %.

Acknowledgements

This work was supported by the Eurostars E114929 Screenbeat project.

References

[1] Hindricks G, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation. European Heart Journal 2021;42(5):373–498.

[2] Mehra R. Global public health problem of sudden cardiac death. Journal of Electrocardiology 2007;40(6):S118–S122.

[3] Timmermans C, et al. Spontaneous episodes of atrial fibrillation after implantation of the Metrix Atrioverter: observations on treated and nontreated episodes. Journal of the American College of Cardiology 2000;35(6):1428–1433.

[4] De Chazal P, et al. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. IEEE Transactions on Biomedical Engineering 2004;51(7):1196–1206.

[5] Cho K, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation, 2014. arXiv:1406.1078.

[6] Goldberger AL, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 2000;101(23):e215–e220.

[7] Taddei A, et al. The European ST-T database: standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography. European Heart Journal 1992; 13(9):1164–1172.

[8] Petrutiu S, et al. Abrupt changes in fibrillatory wave characteristics at the termination of paroxysmal atrial fibrillation in humans. Europace 2007;9(7):466–470.

[9] Moody GB, Mark RG. The impact of the MIT-BIH arrhythmia database. IEEE Engineering in Medicine and Biology Magazine 2001;20(3):45–50.

Address for correspondence:

Jérôme Van Zaen
CSEM, Rue Jaquet-Droz 1, 2002 Neuchâtel, Switzerland
jerome.vanzaen@csem.ch