



HAL
open science

Antidictionary-Based Cardiac Arrhythmia Classification

Julien Duforest, Benoit Larras, Deepu John, Olev Märtens, Antoine Frappé

► **To cite this version:**

Julien Duforest, Benoit Larras, Deepu John, Olev Märtens, Antoine Frappé. Antidictionary-Based Cardiac Arrhythmia Classification. 16ème Colloque National du GDR SOC2, Jun 2022, Strasbourg, France. hal-03760895

HAL Id: hal-03760895

<https://hal.science/hal-03760895>

Submitted on 25 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Antidictionary-Based Cardiac Arrhythmia Classification

Julien Duforest¹, Benoît Larras¹, Deepu John², Olev Märtens³, and Antoine Frappé¹

¹ Univ. Lille, CNRS, Centrale Lille, Junia, Univ. Polytechnique Hauts-de-France, UMR 8520 - IEMN, Lille, France

² University College Dublin, Dublin, Ireland

³ Tallinn University of Technology, Institute of electronics, Tallinn, Estonia

Abstract— Cardiovascular diseases can be detected early by analyzing the electrocardiogram of a patient using wearable systems. In the context of smart sensors, detecting arrhythmias with good accuracy and ultra-low power consumption is required for long-term monitoring. This paper presents a novel cardiac arrhythmia classification method based on antidictionaries. The features are sequences of consecutive slopes generated from the input signal's event-driven processing. The proposed system shows an average detection accuracy of 98% while offering an ultra-low complexity. This antidictionary-based method is also particularly suited to imbalanced datasets since the antidictionaries are created exclusively from heartbeats classified as normal beats.

Keywords—Cardiac Arrhythmia Classification, Electrocardiogram, Event-driven, Antidictionary

I. INTRODUCTION

Cardiovascular diseases (CVD) are the leading cause of death globally [1]. Electrocardiogram (ECG) is used for early CVD diagnosis and wearable at-home monitoring systems can avoid patient examinations at the hospital. To enable wearable health monitoring systems with long battery life, their power consumption must be extremely low. Smart ECG sensors rely on the near-sensor computing paradigm, in which the ECG signal is directly processed in the sensor, by extracting features and determining if arrhythmias have occurred. This reduces the volume of data to send to a data center for further analysis, and thus the power consumption of the embedded device. Furthermore, ECG datasets are imbalanced containing a majority of non-arrhythmic heartbeats. To be used with a neural network, the dataset must be biased by duplicating the arrhythmic heartbeats as previously demonstrated in the literature [2], [3].

In this work, we investigate a novel classification method based on Antidictionaries (AD). An AD represents a set of non-appearing patterns from an original dataset. If the AD is created from normal beats, it can be used to detect arrhythmias without training on arrhythmic beats, thus avoiding the insertion of a bias [4].

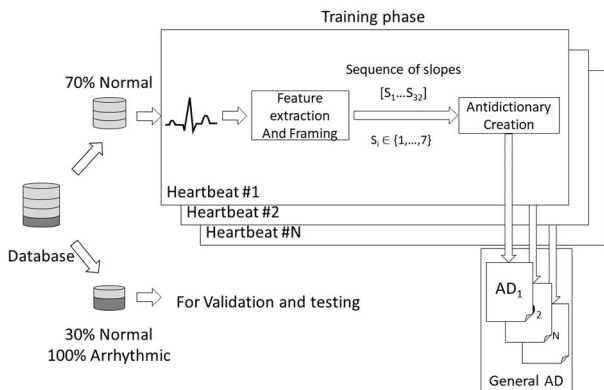


Fig. 1. Process to create a general antidictionary from a single record

Additionally, performing the inference with an offline-learned antidictionary has the potential to be extremely low power, thus compatible with a wearable solution. It has been demonstrated in [3] that the event-driven extraction of a sequence of slopes from an ECG signal is a relevant feature for cardiac arrhythmia classification (CAC). A finite sequence of slopes is therefore suitable for the creation of ADs. The present work exploits the combination of event-driven quantized slopes and antidictionary-based classification to provide an ultra-low complexity solution for CAC with decent accuracies.

This paper is organized as follows. In section II, the process of creation of ADs is described and inference for the validation and testing phase. This section also details the required reduction of the size of the AD. The classification results are presented in section III and the efficiency and interest of the proposed classification method are discussed.

II. CLASSIFICATION USING ANTIDictionaries

A. Definition and creation of antidictionaries

An AD is a list of forbidden words (FBW) produced from negative information of a sequence of symbols, each symbol being part of a predefined alphabet [5]. The FBWs are formed following three rules: 1) the FBW is not a subpart of the sequence of symbols; 2) the 1-letter shorter prefix of the FBW is a subpart of the sequence of symbols; 3) the 1-letter shorter suffix is a subpart of the sequence of symbols.

To apply ADs to ECG signals, ECG signals can be represented by a sequence of coarse-quantized slopes between two consecutive event-driven samples. Those sequences have been proven to be effective features for CAC with a state-of-the-art classifier and constitute a limited set of distinct values [3].

The ADs were tested on the MIT-BIH arrhythmia database. The process to make an AD for a specific patient is shown in Fig. 1. 70% of the normal beats of the dataset are used for AD creation. The rest of the database is later used for validation and testing. Each heartbeat is converted to a sequence of slopes, and 70% of the sequences from normal beats are used to create a single AD. A general AD is then made by concatenating all the ADs.

B. Inference with antidictionaries

The sequence of slopes from the tested beat is analyzed regarding the general AD. Using a simple identification process, with *e.g.* a content-addressable memory (CAM), a flag is generated for each subset of the sequence that is present in the AD. Each heartbeat will then generate several flags that determine the number of times FBWs have appeared in the sequence. This process is applied to the validation dataset, composed of 15% of the remaining normal beats and 85% of the arrhythmic beats, randomly selected. On this validation set, a threshold can be set to differentiate

normal and arrhythmic beats depending on the number of flags generated.

However, when using the general AD created as the union of all the ADs, no clear separation can be envisioned for all records. This is because few FBWs in the AD are too specific to a particular normal beat and flags are detected even for other normal beats. The number of FBWs in the general AD must be reduced to enable a clear delimitation between arrhythmic and normal beats and reduce the necessary memory to store the AD.

C. Antidictionary Reduction

The size of the general antidictionary is reduced in two steps. The first step is to ignore all FBWs with a length greater than 5 symbols. The longer the FBW, the more heartbeat-specific it tends to be. If the FBW is too specific, it is likely to generate flags even for normal beats.

The second step is to reduce the antidictionaries depending on how FBWs react to the validation beats. Only FBWs flagging more in arrhythmic beats than in normal beats are aggregated in the general antidictionary.

To reduce even further the number of FBWs in the AD during the validation phase, FBWs can be ordered using the ratio between the number of times they appear in the arrhythmic beats and the number of times they appear in normal beats. The most relevant FBWs (*i.e.* with the highest ratio) are then used to create the reduced AD. The resulting ADs and the optimal classification thresholds can then be used during the testing phase.

III. CLASSIFICATION RESULTS

A. Antidictionary classification with the test dataset

The proposed AD-based classification is tested on records from the MIT-BIH database containing mainly Premature Ventricular Complex (PVC) arrhythmias. The testing subset is composed of the remaining 15% normal beats and 15% arrhythmic beats. The accuracies and F1 scores are calculated using reduced ADs and associated classification thresholds. The classification was simulated using Matlab.

The maximum accuracy for each record varies between 94.40% and 100% with an average of 97.97% and the F1 scores vary from 69.23% and 100% with an average of 86.80%.

B. Comparison with the state of the art in CAC

Table I gives a comparison of the proposed approach with recent works. The system presented in this paper can

TABLE I. COMPARISON WITH EVENT-DRIVEN PVCs CLASSIFICATION SYSTEM

	[4] Bodynets 2019	[2] TBioCAS 2020	[8] BioCAS 2021	<i>This work</i>
Classification Accuracy	97%	99%	98%	98%
Classification F1-Score	-	95%*	-	87%
Extracted features	Slopes between points in Nyquist sampling	74-bit event-driven direction + 2×11-bit RR interval	32×3-bit event-driven slopes	32×3-bit event-driven slopes
Inference method	Compression rate	32x16x5 ANN	8x4x2 ANN	4 to 32 FBW AD

reach an accuracy similar to the state of the art (1% lower than the best-reported accuracy), with no timing information on the RR-intervals and no arrhythmic data augmentation for the biased training. The inference method is also less complex than the other works, with a maximum of 32 FBWs, composed of a maximum of 5 3-bit letters (480 bits total), that can be easily addressed and identified through a simple CAM. The classification threshold used in the counter can be coded on 6 bits since the highest classification threshold needed is 61. With all the information needed for the classification stored on 486 bits and a completely clockless event-driven system, ultra-low energy consumption can be envisioned for a future wearable health monitoring system.

IV. CONCLUSION

In this paper, a novel classification method based on ADs is presented. The ADs are created using extracted features from normal heartbeats, in the form of quantized event-driven slopes. The proposed classification method allows a system with low complexity for inference, while the training (creation of the ADs) benefits from the imbalanced nature of the datasets. This solution shows promising results with an average 97.97% accuracy for PVC CAC. This system offers an accuracy similar to the state of the art while being extremely simple and easily adaptable from patient to patient.

ACKNOWLEDGMENT

This work was supported by the Nano2022 program, by the CHIST-ERA grant JEDAI CHIST-ERA-18-ACAI-003, by the French Research Agency project ANR-19-CHR3-0005-01, EU Regional Development Fund (Estonian Centre of Excellence in ICT Research EXCITE TAR16013), Mobilitas+ project Mobera20, and Irish Research Council.

REFERENCES

- [1] “Cardiovascular diseases (CVDs).” [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)) (accessed Apr. 22, 2021).
- [2] Y. Zhao, Z. Shang, and Y. Lian, “A 13.34μW Event-driven Patient-specific ANN Cardiac Arrhythmia Classifier for Wearable ECG Sensors,” *IEEE Trans. Biomed. Circuits Syst.*, pp. 1–1, 2019, DOI: 10.1109/TBCAS.2019.2954479.
- [3] J. Duforest, B. Larras, D. John, O. Martens, and A. Frappé, “Slope-Based Event-Driven Feature Extraction For Cardiac Arrhythmia Classification,” in *2021 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, Oct. 2021, pp. 01–04. DOI: 10.1109/BioCAS49922.2021.9644957.
- [4] G. Frias, H. Morita, and T. Ota, “Arrhythmia Detection with Antidictionary Coding and Its Application on Mobile Platforms,” in *Body Area Networks: Smart IoT and Big Data for Intelligent Health Management*, Cham, 2019, pp. 50–67. DOI: 10.1007/978-3-030-34833-5_5.
- [5] “Crochemore et al. - 1999 - Text Compression Using Antidictionaries.pdf.” Accessed: Mar. 30, 2021. [Online]. Available: <https://hal-upec-upem.archives-ouvertes.fr/hal-00619991/document>