Detection of atrial fibrillation using RR interval parameters and k-nearest neighbour classification

Nilén, Emil (BME14, har13en2@student.lu.se); Oswald, Olle (BME14, mte14oos@student.lu.se)

***Abstract*—The time intervals between the heart beats, known as RR intervals, are useful for detection of atrial fibrillation (AF), which is an irregular heart rhythm. Several algorithms already use RR intervals to find parameters which can distinguish AF from normal rhythm. This paper shows how two previously published methods can be combined together with filters and a k-nearest neighbour classifier to classify sequences of the RR signal. The coefficient of variation measures the distribution of the time intervals in the RR signal, and a histogram based parameter examines the variation of the ΔRR signal. Only using the coefficient of variation results in a high specificity and sensitivity, averaging 94% and 92% respectively. The classifier performs less accurately when this method is combined with the histogram parameter, indicating that it is a poor parameter.**

I. Introduction

The medical disorder atrial fibrillation is an occasional irregularity of the heart beat. The irregularity, or quivering, leads to a less efficient pumping of the blood, which can result in heart-related complications such as blood clots and stroke [1]. Detecting the AF is of high importance for future treatment of the patient, but there are several issues to deal with.

Manual diagnosis of atrial fibrillation is typically facilitated by examining the P-waves in the electrocardiogram (ECG), since the absence of P-waves is an indication of AF. Determining the absence of P-waves automatically is however difficult, hence for the development of automatic AF detectors attention is also aimed at variations in the ventricular rate, which is another indicator of AF. The ventricular rate is the number of QRS-complexes (or RR intervals) in one minute, and an irregular rate is an indicator of atrial fibrillation. There have been several algorithms proposed in literature which make use of RR intervals in different ways to discover the irregular heart beats and classify them as AF [2].

One example is the Poincaré plot, which plots successive pairs of RR intervals (x(n),x(n+1)) on the x- and y-axis [3]. A very scattered plot is indicative of large variability and is more likely to be linked to AF than a plot where the points are gathered.

Different variants of entropy estimation is another common measurement. Shannon Entropy is widely used in information theory, and is a measurement of how predictable the information content of a message or signal is [4]. A higher value of the entropy indicates a higher heart rate variability which is a typical effect of AF. Rather than using the probability of a certain RR interval to occur, one might instead opt for quantifying how repetitive the signal is, using the Sample Entropy [5]. A high value corresponds to a low degree of repetition, which in turn is indicative of AF.

The classification can be distracted by noise corrupting the measured signal, and solely depending on RR intervals will make it hard to distinguish between irregularities caused by AF and those caused by other arrhythmias [6]. In this report, focus is aimed at a measurement of statistical dispersion in the RR signal known as the coefficient of variation and a histogram based parameter quantifying the variation in the ΔRR signal. Classification is then performed by using these signal parameters with a machine learning classifier known as the k-nearest neighbours technique. The aim is to have a detector with a sensitivity and specificity that can compare to previously published detectors.

II. Data

The data consisted of seven datasets of RR-intervals collected from the MIT-BIH AF database (AFDB) [7]. Figure 1 shows the first dataset. Four of these datasets, referred here to as AFDB_1 to AFDB_4, where used in development of the classifier, whereas the remaining three (AFDB_5 to AFDB_7) where saved for final performance evaluation.
RR-intervals are the time intervals between heart beats, defined as the distance between subsequent R peaks that can be found in the QRS complex. The QRS complex appears in an ECG and corresponds to the depolarization of the heart during one beat, where the R peak is the maximum potential as is shown in figure 2.

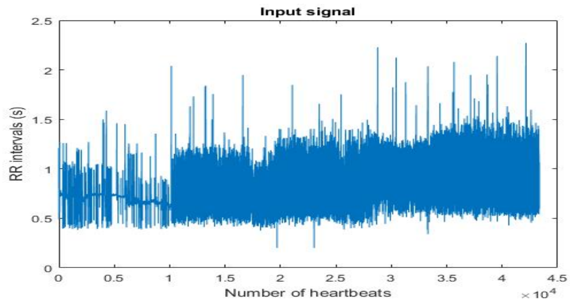

Fig. 1. RR intervals from the first dataset.

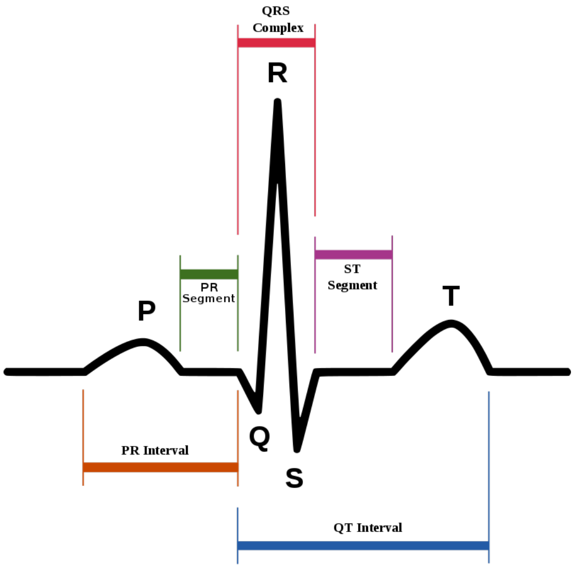

Fig. 2. The electrocardiogram of a heart beat, with the maximum potential at the R peak in the QRS complex.

III. Methods

The proposed AF detection system consists of extracting a set of features from the AFDB signals and training a k-nearest neighbour classifier upon these features. This machine learning approach is highly dependent upon the quality of the features selected for training. [https://machinelearningmastery.com/an-introduction-to-feature-selection/] While it might be tempting to select several features to improve classification performance, adding features of poor quality may instead have an adverse effect. For this reason, two systems with different feature sets were tested. The first uses the coefficient of variation as the only feature, whereas the second system uses both the coefficient of variation as well as a histogram based feature. Figure 3 shows a block diagram of the two systems.

A. Median filter

Single peaks that appear in areas without any nearby peaks where considered as outliers, e.g. due to the presence of ectopic beats. A median filter was used to erase these. The median filter loops through the signal and replaces the value x(n) with the median of x(n+1), x(n) and x(n-1).

B. Feature extraction

Two features where extracted from the AFDB signals: the coefficient of variation and a histogram based feature.

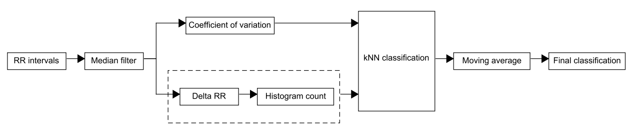

Fig. 3. Block diagram of the proposed classification system. The blocks surrounded by a dashed line box were part of the second classification system only.

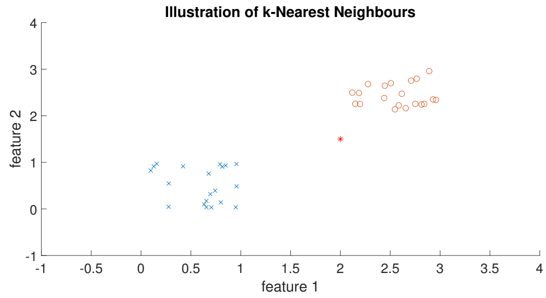

Fig. 4. Example of the k-nearest neighbours technique. There are two possible classes, "x" and "o". Two features have been extracted from the training data, and the data points are placed in the feature space accordingly. The distance from the new observation "*" to the training points is calculated, and the class of the k-nearest neighbours determines the classification of the new observation. Any N number of features can be used, resulting in an N-dimensional feature space.

The coefficient of variation is defined as

$$P_{CV} = \frac{\sigma_x}{m_x}$$

where σ_x is the standard deviation and m_x is the mean of the sequence x. The value of this feature was estimated for each RR sample by surrounding it with a sliding window of width 17 and performing the calculation. The window was then shifted to the next sample in the RR sequence until the feature had been extracted for the entire signal. The beginning and end of the RR signal was zero padded to enable the sliding window to operate in these signal regions.

The histogram based method consisted of differentiating the RR signal ($RR \longrightarrow \Delta RR$) and then studying the distribution of ΔRR values within a sliding window of width 17. The values within the sliding window where sorted into bins varying from -3 to 3 with a width of 0.1. To quantify the distribution of each window, the ratio of non-empty bins divided by the total amount was calculated

$$h = \frac{\text{\#non-empty bins}}{\text{\#bins}}$$

During AF episodes the RR intervals are expected to exhibit a larger variability, and the distribution within a time window is therefore expected to be broader and more flat, resulting in a larger value of h. In contrast, when AF is not prevalent the distribution is expected to be more narrow, resulting in a lower value of h.

C. AF Classification

A machine learning approach was adopted for the classification of atrial fibrillation episodes, using an implementation of the k-nearest neighbours (kNN) technique. The implementation which was used is from the MATLAB Statistics and Machine Learning Toolbox, with a value of $k = 4$ [8].
The training of a kNN classifier consists of storing the training data in a feature space, based on the features extracted from each datapoint. When a new observation is classified, a set of features is extracted from it. The distance between the observation and the training data points in the feature

space is calculated and the k nearest neighbours (the training datapoints) decide upon the classification of the new observation [9]. An illustration of the kNN technique is shown in figure 4. The output of the kNN classifier is a sequence of 0's and 1's, denoting non-AF and AF respectively.

D. Detection averaging

To reduce outliers in the detection sequence, the output from the classifier is smoothed with an averaging filter. It calculates the average within a sliding window of width 11 centered on each detection, and uses a threshold of 0.6 to set it to 0 or 1. Thus, if a 0 is surrounded by many 1's, it is replaced by a 1 (and vice versa). This is motivated by the guideline definitions of AF stating that an AF-episode should last at least 30 seconds in order to be regarded as clinically significant [6].

IV. Results

The performance of each classifier was estimated by calculating the sensitivity and specificity of the classifier on each of the four development datasets (AFDB_1 to AFDB_4). Sensitivity is defined as

$$\frac{\#\text{True positives}}{\#\text{True positives} + \#\text{False negatives}}$$

and is a measurement of how good the classifier is at detecting AF when there is AF.
Specificity is defined as

$$\frac{\#\text{True negatives}}{\#\text{True negatives} + \#\text{False positives}}$$

and is a measurement of how good the classifier is at rejecting episodes where AF is not prevalent. To make the performance measurement as fair as possible, a leave-one-out approach was adopted where the classifier was trained upon three of the AFDB datasets, and then validated with the fourth one. The procedure was repeated for all four AFDB development datasets. The performance of the system using only P_{CV} is shown in table I and the performance of the system using P_{CV} and the histogram based feature is shown in table II. In figure 5 the classification performance of the P_{CV}-only classifier is shown as color coded segments of the RR signals.

	Sensitivity	Specificity
Subject 1	0.95	0.99
Subject 2	0.84	0.92
Subject 3	0.98	0.92
Subject 4	0.99	0.86
Average	**0.94**	**0.92**

TABLE I
THE SENSITIVITY AND SPECIFICITY OF FOUR OF THE FOUR SUBJECTS USING THE P_{CV}-ONLY CLASSIFIER.

	Sensitivity	Specificity
Subject 1	0.74	0.99
Subject 2	0.67	0.91
Subject 3	0.98	0.92
Subject 4	0.96	0.86
Average	**0.84**	**0.92**

TABLE II
THE SENSITIVITY AND SPECIFICITY OF FOUR OF THE FOUR SUBJECTS USING THE P_{CV} AND HISTOGRAM CLASSIFIER.

V. Discussion

The results from the detector only using P_{CV} (table I) are better than combining P_{CV} with the histogram based feature (table II). This indicates that using the histogram and bin counting in the way attempted was not a good approach and resulted in a feature which is more harmful than beneficial. Experiments were we used only the histogram based method resulted in very poor performance, which further supports this theory. Previous literature has suggested that a window length of at least 100 is necessary in order for a histogram to effectively distinguish different cardiac rhythms [6]. However, modifying the method with increased window length resulted in only a marginal increase in performance for our classifier.

It is interesting to note that the performance of both classifiers is noticeably worse for AFDB_2 than for the other datasets. This is a bit surprising, considering that visually the RR series for AFDB_1 and AFDB_2 are rather similar.

Something to note regarding computational complexity is that a kNN classification approach requires that distance is calculated to each training point. If the classifier is heavily trained, there are quite many distances to calculate. Using a lighter classification method, e.g. thresholding, is less demanding in terms of computation. If the detector is implemented in a wearable device or smartphone, this translates to battery life and computational speed which is worth to consider. Computational complexity also is another reason for rejecting the histogram based feature from the detector system, as it took a considerable amount of time to compute.

Comparing the performance of our detector with the performance of previous publications, presented in table III, shows that it does not fall far behind in terms of sensitivity and specificity. Machine learning classifiers such as k-nearest neighbour typically perform better when there is a large amount of data available for training, and the limitation of four datasets in this project may provide too little data for the method to reach its full potential. The addition of more features to distinguish AF from non-AF is another modification which could increase performance.

Method by	Year	Database	Se(%)	Sp(%)
Zhou et al.	2014	AFDB	96.9	98.3
Asgari et al.	2015	AFDB	97.0	97.1
Petrénas et al.	2015	AFDB	97.1	98.3
Zhou et al.	2015	AFDB	97.4	98.4

TABLE III
PREVIOUSLY PUBLISHED RESULTS ON THE SPECIFICITY AND THE SENSITIVITY. [6]

VI. CONCLUSIONS

The aim of this paper was to create a detector which could come close to previously published performance results (some of which can be seen in table III), and we think that averaging 94% in specificity and 92% in sensitivity fulfills that. Perhaps we were lucky with the datasets, but the results show that the usage of the P_{CV} parameter is not completely wrong. The histogram based parameter was very off and consequently rejected. The problem using only P_{CV} is the difficulty in separating atrial fibrillation from other arrhythmias, and we think it is a good idea to research more on how to automate the process of determining P-waves. Combining P_{CV} with morphology as a priori information will perhaps increase the accuracy of the classifier.

REFERENCES

[1] American Heart Association, *What is Atrial Fibrillation (AFib or AF)?*, http://www.heart.org/HEARTORG/Conditions/Arrhythmia/AboutArrhythmia/What-is-Atrial-Fibrillation-AFib-or-AF_UCM_423748_Article.jsp#.Wt7wEMiFPIU [[Accessed 2018-04-24]].

[2] K. Tateno and L. Glass, *Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and deltaRR intervals,*. Med. Biol. Eng. Comput., vol. 39, pp. 664–671, 2001.

[3] T. Anan, K. Sunagawa, H. Araki and N. Nakamura, *Arrythmia analysis by successive RR plotting*, J. Electrocardiol, vol 23, pp. 243-248, 1990.

[4] C. E. Shannon, *A mathematical theory of communication*, Bell. Sys. Tech. J., vol. 27, pp. 379-423, 1948.

[5] D. E. Lake, J. S. Richman, M. P. Griffin and J. R. Moorman, *Sample entropy analysis of neonatal heart rate variability*, Am. J. Physiol. Regul. Integr. Comp. Physiol. vol 283, pp. R789-R797, 2002.

[6] L. Sörnmo, *Atrial fibrillation from an engineering perspective*, Springer, 2018.

[7] The MIT-BIH Atrial Fibrillation Database, https://www.physionet.org/physiobank/database/afdb/

[8] MathWorks, *Nearest Neighbour*, https://se.mathworks.com/help/stats/classification-nearest-neighbors.html

[9] P. Langley, W. Iba, *Average-Case Analysis of Nearest Neighbour algorithm*, Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence, 1993.

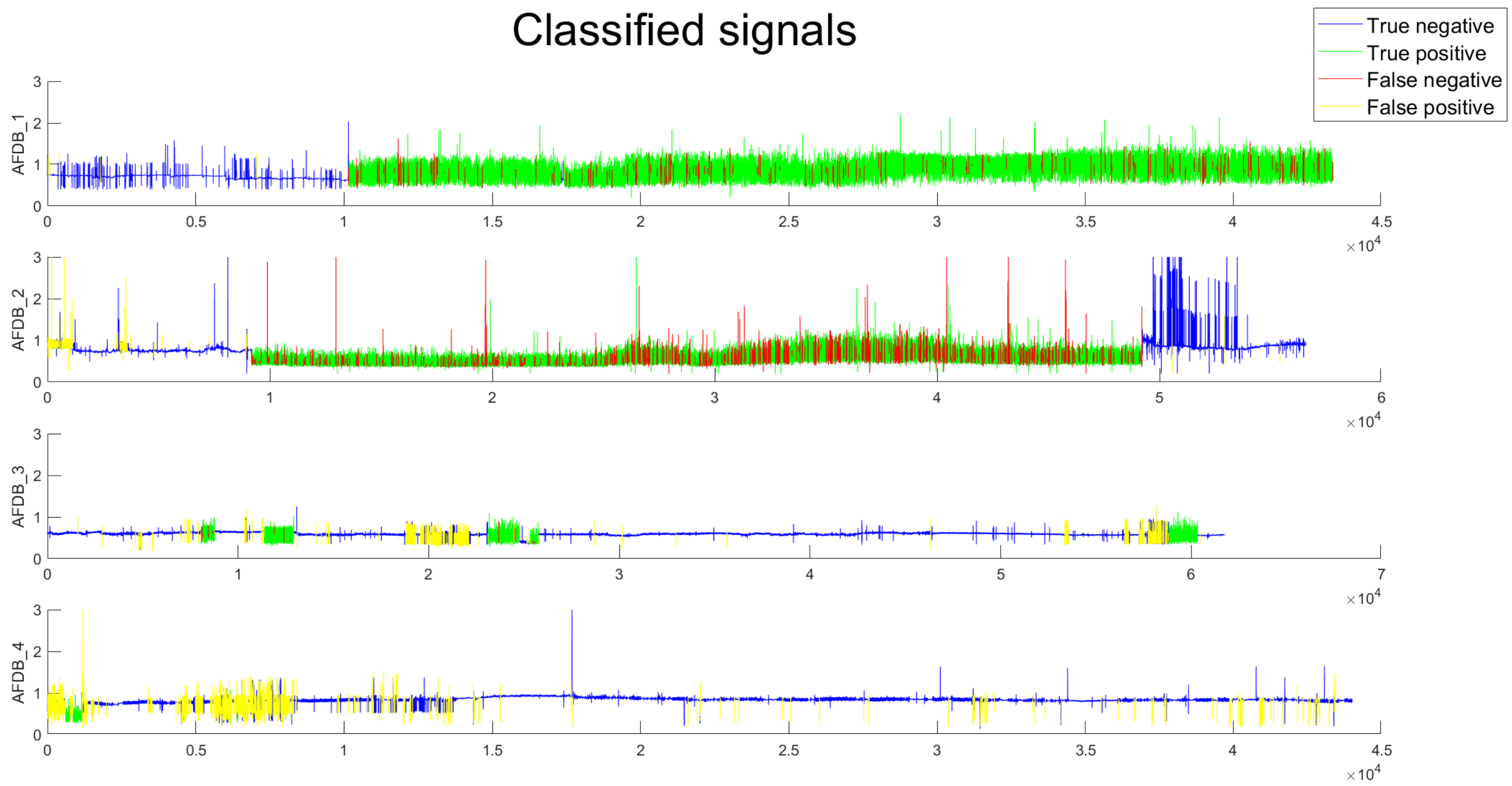

Fig. 5. The classification results coded in different colors. Blue = true negative, green = true positive, red = false negative, yellow = false positive.