Combining Rhythmic and Morphological ECG Features for Automatic Detection of Atrial Fibrillation

Gennaro Laudato[1], Franco Boldi[2], Angela Rita Colavita[3], Giovanni Rosa[1], Simone Scalabrino[1], Paolo Torchitti[2], Aldo Lazich[4] and Rocco Oliveto[1]

[1]*STAKE Lab, University of Molise, Pesche (IS), Italy*

[2]*XEOS, Roncadelle (BS), Italy*

[3]*ASREM, Campobasso (CB), Italy*

[4]*Ministero della Difesa, Roma (RM), Italy*

Keywords: ECG Analysis, Atrial Fibrillation, Arrhythmia, Decision Support System, Machine Learning.

Abstract: Atrial Fibrillation (AF) is a common cardiac disease which can be diagnosed by analyzing a full electrocardiogram (ECG) layout. The main features that cardiologists observe in the process of AF diagnosis are (i) the morphology of heart beats and (ii) a simultaneous arrhythmia. In the last decades, a lot of effort has been devoted for the definition of approaches aiming to automatic detect such a pathology. The majority of AF detection approaches focus on R-R Intervals (RRI) analysis, neglecting the other side of the coin, *i.e.,* the morphology of heart beats. In this paper, we aim at bridging this gap. First, we present some novel features that can be extracted from an ECG. Then, we combine such features with other classical rhythmic and morphological features in a machine learning based approach to improve the detection accuracy of AF events. The proposed approach, namely MORPHYTHM, has been validated on the Physionet MIT-BIH AF Database. The results of our experiment show that MORPHYTHM improves the classification accuracy of AF events by correctly classifying about 4,400 additional instances compared to the best state of the art approach.

1 INTRODUCTION

Atrial Fibrillation (AF) is a quite common yet dangerous cardiac pathological condition. The numbers say that in the UK, almost 534k people have contracted this disease, in 1995 (Stewart et al., 2004). In 2010, the estimated numbers of men and women who were affected by AF world-wide were respectively 20.9 and 12.6 million. Moreover, the incidence was higher in developed countries, such as Europe and US. Indeed, it is expected that - by 2030 - the number of AF patients will be between 14 and 17 million only in Europe (Kirchhof et al., 2016). Besides, such a condition is very expensive: the direct cost of healthcare for patients affected by AF was about $\sim$655M in 2000, equivalent to 0.97% of the total UK National Health System (NHS) expenditure (Stewart et al., 2004). While in US, it has been estimated that the medical cost caused by AF is $26 billion annually (January et al., 2014). Also, the prevalence of the disease is expected to more than double in the next 50 years as the population grows older (Miyasaka, 2006).

Most of the cost of healthcare for patients affected by AF is due to hospitalizations and home nursing. In this context, telemedicine would be very helpful. Indeed, telemedicine would allow to remotely and continuously monitoring thousands of patients. However, telemedicine alone is not enough: remote monitoring could help reducing the global cost, but physicians and nurses would be still required to perform such a task.

The best way for reducing the cost of AF for NHS through telemedicine would be by employing automated approaches for AF detection: a software system constantly acquires data from the patient and, when an anomalous condition is detected, physicians are warned (Balestrieri et al., 2019). This would allow to reduce the number of specialized personnel that is required to monitor the patients.

Many automated approaches for AF detection were proposed in the literature (Asgari et al., 2015; Lee et al., 2013; Petrėnas et al., 2015; Zhou et al., 2014, 2015). Such approaches acquire and transform the electrocardiogram (ECG) signal to detect, for each

Laudato, G., Boldi, F., Colavita, A., Rosa, G., Scalabrino, S., Torchitti, P., Lazich, A. and Oliveto, R.
Combining Rhythmic and Morphological ECG Features for Automatic Detection of Atrial Fibrillation.
DOI: 10.5220/0008982301560165
In *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2020) - Volume 5: HEALTHINF*, pages 156-165
ISBN: 978-989-758-398-8; ISSN: 2184-4305

heart beat, if it is *fibrillating* or *non-fibrillating*. One of the best approaches available, *i.e.,* the one introduced by Zhou et al. (2015), still classifies about 20k *fibrillant* heart beat signals as *non-fibrillant*. This means that if we assign half a second to each misclassified beat—which implies 120 BPM, *i.e.,* a quite high heart rate—three hours of *fibrillating* recordings were completely ignored by the approach. This shows that, even if the accuracy of AF detectors is very high, there is still room for improvement. Indeed, in this context, even a small advance would be important and it would possibly help saving human lives.

Since arrhythmia is one of the most prominent symptoms of AF, most of the state of the art approaches are based on rhythmic features, *i.e.,* measures that try to capture the regularity of the heart beat. On the other hand, another main indicator of AF is the absence of the P-wave, which is visible through the ECG. However, only a minority of studies considered morphological features, which aim at capturing the shape of a single heart beat.

In this paper, we present MORPHYTHM, an approach based on machine learning techniques that combines rhythmic and morphological features to detect AF events. MORPHYTHM uses the most promising state of the art rhythmic and morphological features and some novel features.

We compare MORPHYTHM with the approach introduced by Zhou et al. (2015), the best state of the art approach. The results show that MORPHYTHM can improve the classification accuracy, reducing the number of false negatives (*i.e.,* instances classified as *non-fibrillating* that are, actually, *fibrillating*) by about 4.4k instances, which corresponds to about 35 minutes of ECG (at 120 BPM).

The achieved results confirm our initial conjecture: *morphological features combined with rhythmic ones should be considered for AF detection*. Thus, this paper does the first step towards the definition of novel approaches for AF detection that simultaneously combine morphological and rhythmic features.

The rest of the paper is structured as follows. Section 2 provides details on AF and its ECG diagnostic features. Section 3 presents MORPHYTHM, our novel approach for AF detection. Section 4 reports the design and the results of the empirical study we conducted to evaluate MORPHYTHM. Finally, Section 5 concludes the paper and provides suggestions for possible future research directions.

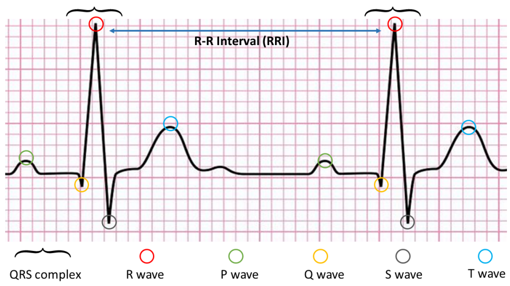

Figure 1: ECG theoretical waveform.

2 BACKGROUND

This section provides details on AF and how it can be identified through the manual analysis of an electrocardiogram. We also discuss the approaches proposed in the literature for AF automatic detection.

2.1 Atrial Fibrillation

AF is a pathological heart rhythm which results in a rapid and irregular beating of the atria. The consequences of this cardiac disease are very adverse. Indeed, contracting AF may lead to stroke, dementia, and death. Thus, a precise diagnosis of this pathology needs to become a priority (Schnabel et al., 2015).

During AF, the hearts atria are quicker than normal beating. This leads to the condition that the blood is not ejected completely out of atria and there might be chances of formation of blood clots in the atria. The result is an increased risk of stroke. Electrocardiograms (ECGs) are useful tools for AF detection. ECGs are recordings of heart's electrical activity and are widely used by physicians to diagnose pathologies related to the heart. Patients with or at risk of cardiovascular diseases often present ECGs that are irregular in rate and in morphology of the signal (Chou et al., 2016).

According to the official international guidelines (Kirchhof et al., 2016), AF can be detected by observing three main features in the ECG, as shown in Figure 1, *i.e.,* (i) absence of the P-wave, (ii) presence of fluctuating waveforms (f-waves) instead of the P-wave, and (iii) heart rate irregularity. The first two features can be defined as *morphological*, while the third one is *rhythmic*.

Since AF is often asymptomatic (Camm et al., 2012; Kearley et al., 2014), a reliable device combined with an accurate, real-time, and automatic AF detection algorithm is desirable for improving detection of AF (Camm et al., 2012; Capucci et al., 2012; Censi et al., 2013; Kearley et al., 2014).

Table 1: Literature Detector Performances on MIT-BIH AFDB. In $AFDB_1$ records “00735” and “03665” excluded, while in $AFDB_2$ records “04936” and “05091” excluded.

Method	Year	DB	Se[%]	Sp[%]
Zhou et al. (2015)	2015	AFDB	97.4	98.4
Petrėnas et al. (2015)	2015	AFDB	97.1	98.3
Asgari et al. (2015)	2015	$AFDB_2$	97.0	97.1
Zhou et al. (2014)	2014	AFDB	96.9	98.3
Lee et al. (2013)	2013	$AFDB_1$	98.2	97.7
Huang et al. (2010)	2011	AFDB	96.1	98.1

2.2 Automatic Detection of AF

In the last decade, several methods have been proposed for the automatic detection of AF. Most of them have shown good results by exploiting only the analysis of heart beat rhythm (Colloca et al., 2013; Mohebbi and Ghassemian, 2008; Sepulveda-Suescun et al., 2017; Xiong et al., 2017; Yuan et al., 2016). Morphological features were used in the patent by Kurzweil et al. (2016) and, even not specifically focused only on AF detection, in the work by Xu et al. (2018).

For sake of space limitation, in the following we focus the attention on the most accurate methods reported in the literature, *i.e.,* the ones summarized in Table 1. These methods represent our baseline, due to the common evaluation on the Physionet MIT-BIH AF Database (Goldberger et al., 2000). This database includes 25 long-term ECG recordings of patients with atrial fibrillation (mostly paroxysmal[1]). Of these 25 long-term ECG recordings, 23 include the ECG signals while for records (*i.e.,* patients) 00735 and 03665 only information on the rhythm are available. The individual recordings are 10 hours each in duration and contain two ECG signals each sampled at 250 samples per second with 12-bit resolution over a range of ± 10 millivolts.

Huang et al. (2010) propose a method to detect the transition between AF and sinus rhythm, based on RRI. In the proposed method the authors first obtain the delta RR interval distribution difference curve from the density histogram of delta RRI, and then detect its peaks, which represent the AF events. Once an AF event was detected, four successive steps have been used to classify its type.

[1]AF can be classified into specific types depending on the duration and ability to self-terminate or to be terminated by some therapeutic technique (Kirchhof et al., 2016). AF is named as paroxysmal when it is self-terminating (in most cases within 48 hours). Some AF paroxysmal episodes may continue up to 7 days. Thus, also AF episodes that are cardioverted within 7 days are considered paroxysmal.

Lee et al. (2013) introduce a method for automatic detection of AF using time-varying coherence functions (TVCF). The TVCF is estimated by the multiplication of two time-varying transfer functions (TVTFs). The first TVTF is obtained by considering two adjacent data segments (as input and output signals); the second TVTF is computed by reversing these signals. They found that the resultant TVCF between two adjacent normal sinus rhythm segments shows high coherence values (near 1) while lower than 1 if either or both segments partially or fully contain AF, throughout the entire frequency range. They have also combined TVCF with Shannon entropy. In this case, the approach shows even more accurate AF detection rate: 97.9% for the MIT-BIH AF database (considering 23 records) with 128 beat segments.

Zhou et al. (2014) devise a method for real-time, automated detection of AF episodes in ECGs. This method utilizes RR intervals, and it involves several basic operations of nonlinear/linear integer filters, symbolic dynamics and the calculation of Shannon entropy.

Asgari et al. (2015) employ a stationary wavelet transform and a support vector machine to detect AF episodes. The proposed method eliminates the need for P-peak or R-Peak detection (a pre-processing step required by many existing algorithms), and hence its performance (sensitivity, specificity) does not depend on the performance of beat detection.

Petrėnas et al. (2015) propose a RR-based AF detector with a low complexity structure. The detector involves blocks for pre-processing, bigeminal suppression, characterization of RR irregularity, signal fusion and threshold detection.

Zhou et al. (2015) adopt heart rate sequence and apply symbolic dynamics and Shannon entropy. Using novel recursive algorithms, a low-computational complexity can be obtained. With this approach, the authors were able to slightly improves their previous work. The approach proposed by Zhou et al. (2015) is the most accurate method of AF detection on the MIT-BIH AF Database proposed so far.

The method proposed by Zhou et al. (2015) will be deeply explained in the next subsection for two main reasons: (i) the method represents our baseline in the evaluation of MORPHYTHM; (ii) the entropy measure used in Zhou et al. (2015) has been exploited as feature in MORPHYTHM.

2.3 Baseline Method for AF Detection

This section provides details on the method proposed by Zhou et al. (2015), *i.e.,* our baseline in the evaluation of MORPHYTHM. Such an approach consists in the following steps (see Figure 2):

- The HR sequence is converted to a symbolic sequence in a fixed interval;
- A probability distribution is constructed from the word sequence which is transformed from the symbolic sequence;
- A coarser version of Shannon entropy is employed to quantify the information size of HR sequence using the probability distribution of word sequence;
- Discrimination of the heart beat type (AF or no-AF) using a threshold.

Step 1: Converting the HR Sequence. Considering a preliminary stage of RRI analysis and thus known the HR sequence, the first step expected in the method is to evaluate a symbolic dynamic. This quantity encodes the information of hr_n to a series with fewer symbols, with each symbol aims at representing an instantaneous state of heart beating. The mapping function is the following:

$$sy_n = \begin{cases} 63, & \text{if } n \text{ hr} \geq 315 \\ \lfloor hr_n \rfloor, & \text{other cases} \end{cases}$$

where $[\cdot]$ represents a floor operator.

Step 2: Building the Symbolic Sequence. The authors apply a 3-symbols template in order to explore the entropic properties of the symbolic series sy_n. Thus, to examine the chaotic behavior, the word value can then be calculated by the operator as defined below:

$$wv_n = (sy_{n-2} \times 2^{1}2) + (sy_{n-1} \times 2^{6}) + sy_n$$

Step 3: Computing the Entropy. The authors define a coarser version of Shannon entropy $H''(A)$ to quantitatively calculate the information size of wv_n. In this study, the dynamic A comprises of 127 consecutive word elements from wv_{n-126} to wv_n, as proposed in the function below:

$$H''(A) = -\frac{k}{N log_2 N} \sum_{i=1}^{k} p_i log_2 p_i$$

where N and k are total number of the elements and characteristic elements in space A, respectively.

Step 4: Classification. Based on the obtained entropy value, a final beat-to-beat classification (*fibrillant* or *non-fibrillant*) is presented by applying a threshold discrimination. The optimal threshold was empirically identified at 0.639.

2.4 Usage Scenarios of AF Detectors

AF detection methods might be useful in two different scenarios: *offline* and *online*. In an *offline* scenario, the ECG of a patient is recorded and, later, an AF detection method is used to find possible AF events occurred in the recorded period. This can help physicians discovering AF events in possibly long EGCs.

AF detection methods could be also particularly valuable in an *online* scenario: while the ECG is acquired, it is immediately passed to the AF detector, which promptly detects AF events. Online AF detection can be useful in tele-medical applications, where patients are constantly monitored.

The application context we take into account in this paper is the *online* monitoring. In other words, we assume that we have chunks of ECG incrementally available. Therefore, we specifically focus on real-time (or near real-time) approaches. Online monitoring is useful for tele-medical applications.

Several tele-medical projects were proposed in the literature. Zhu et al. (2015) introduced the SPHERE system, which combines several sensors which acquire data through wearable, environment, and video devices. Villar et al. (2015) introduced Hexoskin, a line of cutting-edge smart clothing that include body sensors into garments for health monitoring. Balestrieri *et al.* Balestrieri et al. (2019) recently introduced ATTICUS, an innovative Internet of Medical Things (IoMT) system for implementing personalized health services.

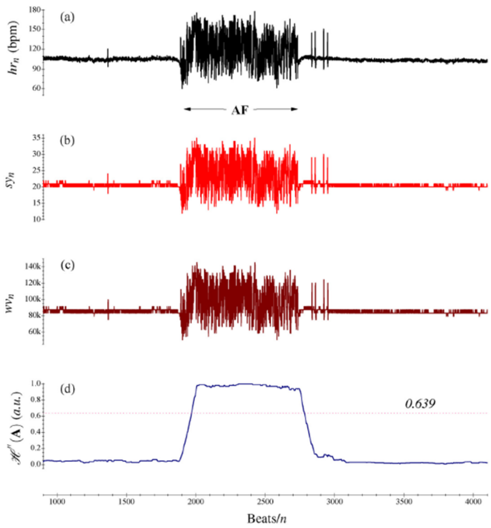

Figure 2: Graphical representation of the main steps in the method by Zhou et al. (2015).

3 MORPHYTHM OVERVIEW

In this section we present MORPHYTHM, a novel approach for the detection of AF events. The proposed approach combines rhythmic and morphological features through machine learning techniques.

3.1 Preprocessing of ECG Data

Before extracting features, the ECG data (from the AFDB) have to be pre-processed according to Pan and Tompkins (1985) and Clifford et al. (2006). The main steps involved in this phase are the followings:

- **Detrend of ECG Signal:** the offset has been removed from the raw ECG signal by removing the mean from the signal.
- **Filtering Stage:** first, a low and high pass filters have been applied to get rid of baseline wander and discard high frequency noise, respectively. Subsequently, a derivative filtering has been operated on the signal aiming at emphasizing the high frequency components of the ECG.
- **Sample Amplitude Normalization:** each recording has been normalized in terms of sample amplitude around the maximum.

3.2 Rhythmic Features

Rhythmic features are based on one or more heart beats and they aim at capturing aspects that mostly regard the regularity of the heart beat signal. Zhou et al. (2015) state that the detection methods based on RRI are more useful to produce a precise and accurate identification of AF because the R-wave peak of the QRS complex is the most prominent characteristic feature of an ECG recording. Such a characteristic is less subject to noise (Huang et al., 2010; Lake and Moorman, 2010; Lee et al., 2012; Lian et al., 2011).

In MORPHYTHM we use two features based on the observation of a single heart beat signal, *i.e.,* HBL and HBDL, and two additional rhythmic features that consider the information of a sequence of consecutive heart beats, *i.e.,* HBR and Entropy:

- **Heart Beat Length (HBL).** This feature represents how long a single heart beat signal lasts. We measure *HBL* as the number of samples from a peak R to the next peak R;
- **Heart Beat Discrete Length (HBDL).** Such a feature is a classification of the heart beat signal in three classes, based on its length. A beat is (i) *short* if it takes less than 0.5 seconds, (ii) *long* if it takes more than 1.2 seconds, and (iii) *regular* otherwise;
- **Heart Beat Regularity (HBR).** This feature is based on HBDL. It considers a rhythmic pattern of 10 consecutive discrete heart beats lengths. Once obtained the pattern, we compute HBR simply counting the number of regular heart beats. It is worth noting that there are approaches in the literature which consider a very short windowed sequence of heart beats (Petrėnas et al., 2015);
- **Entropy**, as defined by Zhou et al. (2015) and described in Section 2.3.

While HBL and Entropy have been previously used in the literature (Zhou et al., 2015), HBDL and HBR are two new rhythmic features defined in this paper.

3.3 Morphological Features

Even if the acquisition of rhythmic features can be very reliable, such features can only help detecting arrhythmia, which is just one of the possible signs of AF. On the other hand, morphological features are necessary to detect anomalies in the shape of a single heart beat signal.

In MORPHYTHM we propose three different measures that—given a sequence of samples provided for a heart beat signal[2]—return a single numeric value:

- **Mean Signal Intensity (MSI).** Such a feature is measured as the mean of all the samples acquired in a heart beat signal. The mean signal intensity, alone, provides a very rough indication of regularity of the heart beat signal. If there is any anomaly in any part of the heart beat signal, such a feature may help identifying it. For example, if the P-wave is missing, the MSI may be slightly affected;
- **Signal Intensity Variance (SIV).** This feature is measured as the variance of all the samples acquired in a heart beat signal. The SIV helps characterizing the heart beat signal: again, a low SIV might indicate the absence of the P-wave.
- **Signal Intensity Entropy (SIE).** This feature is computed as the entropy (Moddemeijer, 1989) of the distribution of the sample values in a heart beat signal. This feature is similar to SIV, *i.e.,* it is aimed at representing the variations in the signal of a heart beat.

It is worth noting that extracting features by considering a whole heart beat might compress too much the information in the ECG data. To extract richer information, we also propose a novel descriptor of a heart beat signal by (i) dividing the whole heart beat in n segments; and (ii) computing the above defined features on each segment:

[2] We consider as a *heart beat* a digital signal which goes from a R-peak to the successive. Such an interpretation is very suitable for AF detection, because it highlights the atrial activity.

- **Segmented Mean Signal Intensity (S-MSI$_i$):** given the i-th segment of the heart beat signal, S-MSI$_i$ is computed as the mean of the sample values of such a segment.
- **Segmented Signal Intensity Variance (S-SIV$_i$):** given the i-th segment of the heart beat signal, S-SIV$_i$ is computed as the variance of the sample values of such a segment.
- **Segmented Signal Intensity Variance (S-SIE$_i$):** given the i-th segment of the heart beat signal, S-SIE$_i$ is computed as the entropy (Moddemeijer, 1989) of the sample values of such a segment.

All such features allow to roughly represent the shape of the signal of the heart beat. We reduce the resolution of the heart beat signal to just 30 values (n=10 for each feature) to reduce the noise of the samples.

Besides the aforementioned features, we also integrate in MORPHYTHM other state of the art morphological features:

- **Fast Fourier Transform (FFT$_i$):** we include the features introduced by Haque et al. (2009) by calculating the Fast Fourier Transform of the heart beat signal on 32 points.
- **Auto-Regressive Model (ARM$_i$):** we include the features introduced by Zhao and Zhang (2005) by estimating the coefficients of the Auto-Regressive model of order 16.

3.4 Putting All Together

MORPHYTHM combines all the features we previously described using supervised machine learning techniques. After the training phase, MORPHYTHM is able—given a heart beat signal—to classify it as *fibrillating* or *not fibrillating*. In the MORPHYTHM evaluation, we experimented several classifiers.

4 EMPIRICAL EVALUATION

The *goal* of this study is to evaluate the accuracy of MORPHYTHM is classifying AF events in a patient. The *perspective* is both (i) of a researcher who wants to understand if combining rhythmic and morphological features is useful for detecting AF events, and (ii) of a practitioner who wants to use the most accurate and precise approach in a telemedicine application. Thus, the study is steered by the following research question:

Can the combination of rhythmic and morphological features improve the classification accuracy of Atrial Fibrillation events?

4.1 Context Selection

The context of this study is represented by MIT-BIH AF Database (Goldberger et al., 2000), a commonly used benchmark which contains recordings of 25 patients. Due to the embedding of morphology descriptors, our overall study has been performed on the AFDB$_1$, *i.e.,* the AFDB without records 00735 and 03665 because, for such records, only information on the rhythm is available (Goldberger et al., 2000). Each recording in the dataset lasts 10 hours and contains two ECG signals sampled at 250 samples per second (12-bit resolution).

In the context of our study, we experimented a large set of machine learning technique to train MORPHYTHM. Especially, we experimented tree-based classifiers, *i.e.,* J48 (Quinlan, 2014), Replication Tree (Devasena, 2014), and Random Forest (Barandiaran, 1998). Such approaches, indeed, can build models that are also easy to understand by a human. We also experimented Logistic regression (Cramer, 2002) and AdaBoost M1 (Freund and Schapire, 1997).

4.2 Experimental Procedure

To evaluate the accuracy of MORPHYTHM, we used a classical Leave-1-Person Out (L1PO) cross-validation: we divided all the data in n folds, one for each patient, and we use one at a time each of such folds as test set and the union of the remaining folds as training set. This means that the data related to a single patient were embedded once in the test dataset and n-1 times in the training dataset. This technique allows to build a classifier which is not trained and tested on the data belonging to the same patient. We did this to evaluate the technique in the most challenging scenario: the ECG of different patients can be very different.

We compared MORPHYTHM to the approach proposed by Zhou et al. (2015), previously presented in Section 2.3. The instances to be classified were all the single heart beat signals provided in the dataset, labeled as *fibrillating* or *non-fibrillating*. The work by Zhou et al. (2015) just reported the performance of the approach globally, *i.e.,* for all the patients. Instead, we provide the performance of the approaches with a finer grain, *i.e.,* on patient-by-patient base. Since we do not have the patient-by-patient results for the baseline, it was necessary to re-implement the approach and to re-compute the results.

To answer our research question we compared two critical aspects: True Positives (TP), *i.e.,* the number of instances classified as *fibrillating* by the approach and that were actually *fibrillating*, and the False Neg-

Table 2: Comparison of MORPHYTHM and the approach proposed by Zhou et al. (2015). In boldface the results achieved by MORPHYTHM that are better than the baseline.

Approach	TP	TN	FP	FN	ΔTP	ΔFN
Zhou et al. (2015)	489,834	603,216	17,188	19,911		
MORPHYTHM — Random Forest	490,810	584,692	35,612	18,935	**+976**	**-976**
MORPHYTHM — J48	479,411	560,567	57,049	33,122	-10,423	+13,211
MORPHYTHM — Logistic	494,255	595,664	24,789	15,445	**+4,421**	**-4,466**
MORPHYTHM — AdaBoost M1	494,384	601,974	18,430	22,362	**+4,550**	+2,451
MORPHYTHM — RepTree	481,397	571,262	49,142	32,348	-8,437	+12,437

atives (FN), *i.e.,* the number of instances classified as *non-fibrillating* which were, actually, *fibrillating*. A high number of TP is desirable, because it indicates the number of AF episodes correctly detected. Also, ideally, a perfect approach does not lose any AF episode: thus, keeping the number of FN low is very important.

We use a Wilcoxon signed-rank test to verify if MORPHYTHM achieves statistically significant better results than the approach proposed by Zhou et al. (2015). To do this, we use the results achieved patient by patient in terms of TP and FN. Formally, our null hypotheses are:

- $H0_1$: MORPHYTHM *does not identify a higher number of TP as compared to the approach proposed by Zhou et al. (2015)*;
- $H0_2$: MORPHYTHM *does not identify a lower number of FN as compared to the approach proposed by Zhou et al. (2015)*;

We reject a null hypothesis if the p-value is lower than $\alpha = 0.05$.

Even if we evaluate the possible improvement only on TP and FN, we also report the global results in terms of True Negatives (TN — *i.e.,* instances correctly classified as *non-fibrillating*) and False Positives (FP — *i.e.,* instances classified as *fibrillating* that are, actually, *non-fibrillating*).

4.3 Analysis of the Results

We show the global performance of the compared approaches in Table 2. For MORPHYTHM, we also specifically report the difference in terms of TP and FN with the baseline, *i.e.,* ΔTP (the higher, the better) and ΔFN (the lower, the better), and we put in boldface the cases in which MORPHYTHM achieves better results.

The first consideration that can be derived from the analysis of Table 2 is that three (*Random Forest, Logistic and AdaBoost M1*) of the five chosen machine learning are able to achieve better results than the baseline. Furthermore, the Logistic method performs definitely better than its competitors by showing an improvement of around 4,400 heart beats compared to the method by Zhou et al. (2015). If we would assign an inter-beat interval of 0.5 seconds, an improvement of 4,400 indicates more than *35 minutes* of AF rhythm improved in the classification with respect to the baseline. Even if this could appear as a negligible result, it should be noticed that the accuracy level achieved by such approaches is very high and, therefore, even achieving a small improvement is very difficult.

It can be noticed that the global accuracy of the approach by Zhou et al. (2015) slightly differs from the global accuracy reported in the original paper. Especially, in the original paper the authors reported the following values for *sensitivity*, *specificity*, and *accuracy*—they just report aggregated measures: 97.31%, 98.28%, and 97.89%. With our replication of the approach by Zhou et al. (2015) we achieve the following results: 96.09% of sensitivity, 97.22% of specificity, and 96.71% of accuracy. We are confident that the different results are not due to implementation errors, but to different choices in the evaluation design. The different results could be due to the following reasons:

- the transient: to avoid any error due to interpretation, the first 128 (126 coming from the entropy compression + 2 from the word sequence evaluation) beats have not been considered in the replication of the work by Zhou et al. (2015). Unfortunately, in the paper by Zhou et al. (2015) there is no clear indication on how the authors deal with the initial 128 beats;
- the timestamps: Physionet offers two different timestamps: one for each beat classification another one for each AF events (rhythm annotation). There are cases where there is a mismatch between the two timestamps, *i.e.,*, the AF event does not start (or does not end) with the beginning of a (or the end) of beat. In other words beats and AF events are not always synchronized. This causes an ambiguity regarding the interpretation of "hy-

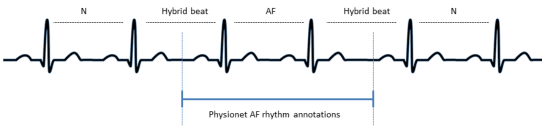

Figure 3: A graphical example of hybrid heart beats.

Table 3: Patient-level comparison between MORPHYTHM and Zhou et al. (2015). In boldface the best results for each patient.

Record	Zhou et al. (2015) TP	FN	MORPHYTHM TP	FN
04015	478	40	**491**	**27**
04043	8,690	5,862	**9,608**	**4,944**
04048	419	387	**443**	**363**
04126	3,082	204	**3,154**	**132**
04746	30,731	**137**	**30,624**	244
04908	5,443	359	**5,557**	**245**
04936	32,833	6,812	**33,725**	**5,920**
05091	0	133	0	133
05121	32,575	1,164	**33,563**	**176**
05261	655	268	**766**	**157**
06426	52,104	1,006	**52,633**	**477**
06453	126	313	**130**	**309**
06995	27,072	448	**27,240**	**280**
07162	39,297	0	39,297	0
07859	61,891	0	61,891	0
07879	**39,944**	89	39,939	**49**
07910	**6,499**	**266**	6,440	325
08215	**32,958**	**170**	32,912	216
08219	12,627	1,528	**13,420**	**735**
08378	**10,995**	**478**	10,969	504
08405	45,005	88	**45,041**	**52**
08434	**2,307**	**0**	2,301	6
08455	44,103	159	**44,111**	**151**

brid beats", *i.e.,* beats that are not aligned with an AF event (see Figure 3).

It is worth noting that the different results achieved does not represent a threat for the final message of this paper. Indeed, improving the accuracy of the approach by Zhou et al. (2015) by performing a different evaluation design likely results in an improvement of MORPHYTHM as well, since the approach by Zhou et al. (2015) is one of the features exploited by MORPHYTHM.

Table 3 shows a patient-by-patient comparison between the approach by Zhou et al. (2015) and MORPHYTHM (with the classifier that achieves the best results globally, *i.e.,* Logistic Regression). Even if the method proposed by Zhou et al. (2015) is incredibly accurate, as it can be observed from Table 3, MORPHYTHM achieves much better results for some patients and comparable results on some other patients. Specifically, MORPHYTHM identifies a higher number of TPs for 15 out of 23 patients and it identifies a lower number of FNs for 15 out of 23 patients. The results of the Wilcoxon signed-rank test show that we

Table 4: Comparison between the proposed classifier on the record 05091.

Approach	TP	TN	FP	FN
Zhou et al. (2015)	0	36,644	0	133
MORPHYTHM — Random Forest	25	36,633	11	108
MORPHYTHM — J48	19	36,592	52	114
MORPHYTHM — Logistic	0	36,640	4	133
MORPHYTHM — AdaBoost M1	0	36,644	0	133
MORPHYTHM — RepTree	14	36,620	24	119

Table 5: Features ranking using Information Gain.

InfoGain	Attribute	Type
0.86	Entropy from Zhou et al. (2015)	Rhythmic
0.20	Entropy from the rhythmic pattern	Rhythmic
0.18	Heart beat absolute length	Rhythmic
0.14	coeff. no. 10 from AR model	Morphological
0.13	coeff. no. 11 from AR model	Morphological
0.12	coeff. no. 7 from AR model	Morphological
0.12	coeff. no. 9 from AR model	Morphological
0.11	coeff. no. 8 from AR model	Morphological

can reject both our null hypotheses: MORPHYTHM identifies a *significantly* higher number of TPs ($p = 0.021$) and a *significantly* lower number of FNs ($p = 0.014$).

Table 4 shows the comparison between MORPHYTHM and the approach by Zhou et al. (2015) for one of the patients, *i.e.,* 05091. For such a patient, the baseline has never detected any *fibrillating* event. This means that this record has been classified as not affected by AF, overall, even if it was. On the other hand, most of the classifiers we consider are able to detect some *fibrillating* events. However, unfortunately, this is not true for the best classifier (*i.e.,* Logistic regression), which, similarly to the baseline, does not identify any heart beat of the specific patient as *fibrillating*.

Finally, Table 5 shows the best attributes and their worth computed with Information Gain. From the achieved ranking, we can observe that, as expected, the classifiers use the entropy values as their main source of information. After that, the length of the signal and the central coefficient obtained from the AR model of order 4 seem to be the features which help the classifiers to slightly improve its accuracy as compared to the baseline. Very interesting is the result on the AR model because it seems that the central coefficient, which should describe the atrial activity, are mostly taken in consideration.

5 CONCLUSIONS

We have presented MORPHYTHM, an approach based on machine learning that combines rhythmic and morphological features extracted from ECG data to detect AF events. We compared MORPHYTHM with the method introduced by Zhou et al. (2015), the most accurate on the AFDB in the literature. The results show that (i) MORPHYTHM globally achieves better results compared to the baseline, since it is able to correctly classify about 4,400 more heart beats, and (ii) that some of the patients for which all the *fibrillating* heart beat were mis-classified by the baseline were correctly classified by MORPHYTHM.

The improvement achieved is promising; however, there is still much room for improving the accuracy. In order to increase the generalizability of our results, we aim at apply in the future at least one classifier for each family. In this work, for example, we did not consider Bayesian networks, Rules-based classifiers, and Neural Networks. Also, to maximize the accuracy of MORPHYTHM, it would be desirable to use feature selection, to remove useless features that could decrease the classification accuracy. Specifically, since morphological features can be very patient-dependent, it could be useful performing feature selection for each single patient rather than globally. Finally, we plan to perform a cost-benefit analysis. Indeed, in some online applications, it could be necessary to have some constraints, such as the total reduction of FN, even if the FP rate increases. Thus, we would like to study this specific scenario and observe if the application of a cost-benefit analysis can suite some specific constraints.

REFERENCES

Asgari, S., Mehrnia, A., and Moussavi, M. (2015). Automatic detection of atrial fibrillation using stationary wavelet transform and support vector machine. *Computers in biology and medicine*, 60:132–142.

Balestrieri, E., Boldi, F., Colavita, A. R., De Vito, L., Laudato, G., Oliveto, R., Picariello, F., Rivaldi, S., Scalabrino, S., Torchitti, P., et al. (2019). The architecture of an innovative smart t-shirt based on the internet of medical things paradigm. In *2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pages 1–6. IEEE.

Barandiaran, I. (1998). The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell*, 20(8):1–22.

Camm, A. J., Corbucci, G., and Padeletti, L. (2012). Usefulness of continuous electrocardiographic monitoring for atrial fibrillation. *The American journal of cardiology*, 110(2):270–276.

Capucci, A., Calcagnini, G., Mattei, E., Triventi, M., Bartolini, P., Biancalana, G., Gargaro, A., Puglisi, A., and Censi, F. (2012). Daily distribution of atrial arrhythmic episodes in sick sinus syndrome patients: implications for atrial arrhythmia monitoring. *Europace*, 14(8):1117–1124.

Censi, F., Calcagnini, G., Mattei, E., Gargaro, A., Biancalana, G., and Capucci, A. (2013). Simulation of monitoring strategies for atrial arrhythmia detection. *Annali dell'Istituto superiore di sanita*, 49:176–182.

Chou, T., Tamura, Y., and Wong, I. (2016). Detection of atrial fibrillation in ecgs. *Computer Methods and Programs in Biomedicine*, 136:143–50.

Clifford, G. D., Azuaje, F., McSharry, P., et al. (2006). *Advanced methods and tools for ECG data analysis*. Artech house Boston.

Colloca, R., Johnson, A. E., Mainardi, L., and Clifford, G. D. (2013). A support vector machine approach for reliable detection of atrial fibrillation events. In *Computing in Cardiology 2013*, pages 1047–1050. IEEE.

Cramer, J. S. (2002). The origins of logistic regression (technical report). In *Tinbergen Institute*.

Devasena, C. L. (2014). Comparative analysis of random forest rep tree and j48 classifiers for credit risk prediction. In *International Conference on Communication, Computing and Information Technology (ICCCMIT-2014)*.

Freund, Y. and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139.

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K., and Stanley, H. E. (2000). Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220.

Haque, A. F., Ali, M. H., Kiber, M. A., and Hasan+, M. T. (2009). Automatic feature extraction of ecg signal using fast fourier transform.

Huang, C., Ye, S., Chen, H., Li, D., He, F., and Tu, Y. (2010). A novel method for detection of the transition between atrial fibrillation and sinus rhythm. *IEEE Transactions on Biomedical Engineering*, 58(4):1113–1119.

January, C. T., Wann, L. S., Alpert, J. S., Calkins, H., Cigarroa, J. E., Cleveland, J. C., Conti, J. B., Ellinor, P. T., Ezekowitz, M. D., Field, M. E., et al. (2014). 2014 aha/acc/hrs guideline for the management of patients with atrial fibrillation: a report of the american college of cardiology/american heart association task force on practice guidelines and the heart rhythm society. *Journal of the American College of Cardiology*, 64(21):e1–e76.

Kearley, K., Selwood, M., Van den Bruel, A., Thompson, M., Mant, D., Hobbs, F. R., Fitzmaurice, D., and Heneghan, C. (2014). Triage tests for identifying atrial fibrillation in primary care: a diagnostic accuracy study comparing single-lead ecg and modified bp monitors. *BMJ open*, 4(5):e004565.

Kirchhof, P., Benussi, S., Kotecha, D., Ahlsson, A., Atar, D., Casadei, B., Castella, M., Diener, H.-C., Heidbuchel, H., Hendriks, J., et al. (2016). 2016 esc guidelines for the management of atrial fibrillation developed in collaboration with eacts. *European journal of cardio-thoracic surgery*, 50(5):e1–e88.

Kurzweil, R. C., Gibson, L., Albrecht, P., and Grimshaw, P. (2016). Atrial fibrillation detection and associated methods. US Patent 9,456,762.

Lake, D. E. and Moorman, J. R. (2010). Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices. *American Journal of Physiology-Heart and Circulatory Physiology*, 300(1):H319–H325.

Lee, J., Nam, Y., McManus, D. D., and Chon, K. H. (2013). Time-varying coherence function for atrial fibrillation detection. *IEEE Transactions on Biomedical Engineering*, 60(10):2783–2793.

Lee, J., Reyes, B. A., McManus, D. D., Maitas, O., and Chon, K. H. (2012). Atrial fibrillation detection using an iphone 4s. *IEEE Transactions on Biomedical Engineering*, 60(1):203–206.

Lian, J., Wang, L., and Muessig, D. (2011). A simple method to detect atrial fibrillation using rr intervals. *The American journal of cardiology*, 107(10):1494–1497.

Miyasaka (2006). Secular trends in incidence of atrial fibrillation in olmsted county, minnesota, 1980 to 2000, and implications on the projections for future prevalence (vol 114, pg 119, 2006). *Circulation*, 114(11):E498–E498.

Moddemeijer, R. (1989). On estimation of entropy and mutual information of continuous distributions. *Signal processing*, 16(3):233–248.

Mohebbi, M. and Ghassemian, H. (2008). Detection of atrial fibrillation episodes using svm. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 177–180. IEEE.

Pan, J. and Tompkins, W. J. (1985). A real-time qrs detection algorithm. *IEEE Trans. Biomed. Eng*, 32(3):230–236.

Petrėnas, A., Marozas, V., and Sörnmo, L. (2015). Low-complexity detection of atrial fibrillation in continuous long-term monitoring. *Computers in biology and medicine*, 65:184–191.

Quinlan, J. R. (2014). *C4. 5: programs for machine learning*. Elsevier.

Schnabel, R. B., Yin, X., Gona, P., Larson, M. G., Beiser, A. S., McManus, D. D., Newton-Cheh, C., Lubitz, S. A., Magnani, J. W., Ellinor, P. T., et al. (2015). 50 year trends in atrial fibrillation prevalence, incidence, risk factors, and mortality in the framingham heart study: a cohort study. *The Lancet*, 386(9989):154–162.

Sepulveda-Suescun, J., Murillo-Escobar, J., Urda-Benitez, R., Orrego-Metaute, D., and Orozco-Duque, A. (2017). Atrial fibrillation detection through heart rate variability using a machine learning approach and poincare plot features. In *VII Latin American Congress on Biomedical Engineering CLAIB 2016, Bucaramanga, Santander, Colombia, October 26th-28th, 2016*, pages 565–568. Springer.

Stewart, S., Murphy, N., Walker, A., McGuire, A., and McMurray, J. (2004). Cost of an emerging epidemic: an economic analysis of atrial fibrillation in the uk. *Heart*, 90(3):286–292.

Villar, R., Beltrame, T., and Hughson, R. L. (2015). Validation of the hexoskin wearable vest during lying, sitting, standing, and walking activities. *Applied Physiology, Nutrition, and Metabolism*, 40(10):1019–1024.

Xiong, Z., Stiles, M. K., and Zhao, J. (2017). Robust ecg signal classification for detection of atrial fibrillation using a novel neural network. In *2017 Computing in Cardiology (CinC)*, pages 1–4. IEEE.

Xu, S. S., Mak, M.-W., and Cheung, C.-C. (2018). Towards end-to-end ecg classification with raw signal extraction and deep neural networks. *IEEE journal of biomedical and health informatics*.

Yuan, C., Yan, Y., Zhou, L., Bai, J., and Wang, L. (2016). Automated atrial fibrillation detection based on deep learning network. In *2016 IEEE International Conference on Information and Automation (ICIA)*, pages 1159–1164. IEEE.

Zhao, Q. and Zhang, L. (2005). Ecg feature extraction and classification using wavelet transform and support vector machines. In *2005 International Conference on Neural Networks and Brain*, volume 2, pages 1089–1092. IEEE.

Zhou, X., Ding, H., Ung, B., Pickwell-MacPherson, E., and Zhang, Y. (2014). Automatic online detection of atrial fibrillation based on symbolic dynamics and shannon entropy. *Biomedical engineering online*, 13(1):18.

Zhou, X., Ding, H., Wu, W., and Zhang, Y. (2015). A real-time atrial fibrillation detection algorithm based on the instantaneous state of heart rate. *PloS one*, 10(9):e0136544.

Zhu, N., Diethe, T., Camplani, M., Tao, L., Burrows, A., Twomey, N., Kaleshi, D., Mirmehdi, M., Flach, P., and Craddock, I. (2015). Bridging e-health and the internet of things: The sphere project. *IEEE Intelligent Systems*, 30(4):39–46.