

A window-based time series feature extraction method



Deniz Katircioglu-Öztürk^{a,*}, H. Altay Güvenir^b, Ursula Ravens^{c,1}, Nazife Baykal^a

^a Middle East Technical University, Institute of Informatics, Medical Informatics Department, 06800 Ankara, Turkey

^b Bilkent University, Computer Engineering Department, 06800 Ankara, Turkey

^c Technische Universität Dresden, Institut für Pharmakologie und Toxikologie, 01187 Dresden, Germany

ARTICLE INFO

Keywords:

Time series analysis
Feature extraction
Cardiac action potential
Atrial fibrillation
Electrocardiography
Myocardial infarction

ABSTRACT

This study proposes a robust similarity score-based time series feature extraction method that is termed as **Window-based Time series Feature ExtraCtion (WTC)**. Specifically, WTC generates domain-interpretable results and involves significantly low computational complexity thereby rendering itself useful for densely sampled and populated time series datasets. In this study, WTC is applied to a proprietary action potential (AP) time series dataset on human cardiomyocytes and three precordial leads from a publicly available electrocardiogram (ECG) dataset. This is followed by comparing WTC in terms of predictive accuracy and computational complexity with shapelet transform and fast shapelet transform (which constitutes an accelerated variant of the shapelet transform). The results indicate that WTC achieves a slightly higher classification performance with significantly lower execution time when compared to its shapelet-based alternatives. With respect to its interpretable features, WTC has a potential to enable medical experts to explore definitive common trends in novel datasets.

1. Introduction

In the medical domain, biophysical signals in the form of time series are frequently used for diagnostic and prognostic purposes and are specifically relevant to documenting the history [1] and clinical course of a disease [2,3]. Many decisive traits emerge from biophysical signals as rules of thumb suggested by health professionals based on a visual inspection [4]. Conversely, computer-aided methods extract common patterns among time series and establish a more objective data assessment framework. Thus, they are of immense practical value in interpreting data with respect to diagnostic, prognostic, and therapeutic perspectives [5].

Computerized analysis of medical data is an active research area. In the computational geometry domain, it is assumed that “B-spline” constitutes one of the most efficient surface representation methods [6,7]. Significant curve reconstruction efforts focused on using various spline paradigms [8], [9] especially in the cardiac domain [10]. However, using spline-based methods on biophysical signals is not very helpful to professionals in diagnosis or assessment of indicators for diseased conditions; because splines do not produce human-readable summarizing models. Previous studies examined the construction of local and piece-wise polynomial models to represent time series [11,12]. Recently, the use of Gaussian Process modeling is considered as an attractive option

for analyzing time series data, especially in the medical domain. Within a Bayesian framework, [13] conducted an analysis by solving a regression problem by assuming that a dataset consisted of observations and outcome variables. This approach is suitable for predicting missing outcome values of biological time series that typically suffer from irregularities in sampling and missing observations [13,14]. Another feature extraction method was recently been developed to discover significant subsequences termed as “shapelets” within a dataset [15,16]. A shapelet is a time series subsequence identified as representative of a certain dataset. The discovery of shapelets requires brute-force traversal of all overlapping subsequences wherein the task of determining the individual length and number of these subsequences is left in an unguided manner to the algorithm users. A shapelet transformation produces human readable outputs, and therefore they are used for benchmarking purposes in the present study.

This study addresses the problem of extracting definitive time-domain features from a given time series dataset to be used in data mining applications with the ultimate goal of devising a computationally efficient algorithm. A time series feature extraction method is designed to summarize class-dependent behavior within consecutive time windows and to derive an overall similarity score for this behavior. Specifically, the proposed method involves identifying temporal features that define the behavior of instances in a certain target class that are later used to

* Corresponding author.

E-mail address: denizkatircioglu@gmail.com (D. Katircioglu-Öztürk).

¹ Present address: Freiburg University, Institute for Experimental Cardiovascular Medicine, University Heart Centre, 79110 Freiburg, Germany.

Table 1
Symbols in alphabetical order.

Symbol	Explanation
α_j	The trajectory-based weight of SCB_j based on $I_{j,k}$
β_j	The distance-based weight of SCB_j
c	The PMF resolution
$d_{j,k}$	The Euclidean distance of T_k to T^A within SCB_j
δ	DCT energy threshold
\mathbb{D}	PMF of the deterministic distribution with atom at zero
$\mathcal{D}(\cdot, \cdot)$	Distance between two PMFs
DT^A	One-dimensional DCT of T^A
f_N	Nyquist rate of T^A
f_s	Sampling rate of S
γ_j	The weight of the distance-based component of the similarity score for SCB_j
H	The maximum amplitude of the time series instances in T
$I_{j,k}$	The indicator of inclusion for T_k within SCB_j
J	Time window index
J	Number of shape confidence bands
k	Time series instance index
K	Number of time series instances
l^*	Cut-off index in DCT domain
L	Shapelet length for the brute-force and fast shapelet transform
M	Number of perceptually important time points (PIPs) in T^A
n	Sample point index of a registered time series instance
N	Number of sample points in a registered time series instance
p	Confidence level for the confidence interval
\mathbb{P}_j	Distance PMF for SCB_j
r	PIP oversampling ratio
S	Time series dataset
S_k	k^{th} time series instance in the dataset S
SCB_j	j^{th} shape confidence band
Z_j	The ensemble average similarity score for SCB_j
$Z_{j,k}$	The individual similarity score of T_k for SCB_j
$Z_{j,k}^{(d)}$	Distance-based component of the similarity score of T_k for SCB_j
$Z_{j,k}^{(t)}$	Trajectory-based component of the similarity score of T_k for SCB_j
T	Registered time series dataset
T^A	Class representative average time series
T_k	k^{th} time series instance in the registered dataset T
T^L	Lower bound of the confidence interval with the level p around T^A
T^U	Upper bound of the confidence interval with the level p around T^A
\mathbb{U}	PMF of the uniform distribution
$\phi_{j,k}$	The reward associated with the proximity of T_k to T^A for J
W	Number of sample points in an W_j where $j \neq J$
W_j	j^{th} time window

discriminate between instances among other classes. In order to demonstrate the effectiveness of the proposed method, a human cardiac action potential (AP) dataset and an ECG dataset consisting of recordings from three different leads are employed.

Cardiac APs are bioelectrical signals that are recorded in cardiac tissues obtained during heart surgery from patients with defined heart rhythms. Based on a patient's heart rhythm at the time of surgery, APs may either exhibit a “spike-and-dome” shape for physiological sinus rhythm (SR) or a triangular shape for atrial fibrillation (AF) [17]. Specifically, AF is considered as the most common arrhythmia in clinical practice with an approximate prevalence of 0.4 – 1% in the general population and is usually associated with stroke, heart failure, and a significant increase in all-cause mortality [18]. Patients with AF are diagnosed based on thorough clinical examinations and ECG recordings. However, the resulting AP signals may exhibit SR characteristics based on the stage of the disease. Cardiac AP is a highly informative signal that attracts significant attention from researchers with a key focus of revealing dynamics in cardiac ion channels that govern electrical activity within the heart. An understanding of ion channels is essential in identifying the effects of certain agents in related drug studies [17]. Additionally, AP signal acquisition is a labor-intensive task, and this is considered as a limiting factor for widespread studies especially in the area of statistical time series analysis. Extant studies produced desired AP properties by fitting empirical models, [8], to simulate the real-life

behavior of AP. The AP dataset used in the present study in addition to validating the proposed model also serves the purpose of stress testing the proposed method for a dataset consisting of crowded and densely sampled time series. To the best of the authors' knowledge, this is the first study in which a data mining analysis is conducted on a cardiac AP time series dataset.

Conversely, the ECG dataset, consisting of patients with acute myocardial infarction and control subjects, is specifically selected to present a clinical application of the proposed method. Myocardial infarction (MI) constitutes a fatal cardiac disease that corresponds to irreversible loss of heart muscle due to ischemia caused by blood flow interruption [19]. Prominent guidelines recommend the use of ECGs especially for emergency cases in addition to certain cardiac biomarkers with their superior sensitivity to MI [20]. Additionally, ECG is an effective and non-invasive technique to confirm MI diagnosis during an initial evaluation. With respect to ST elevation myocardial infarction (STEMI), 12-lead ECGs exhibit certain morphological changes in the expected waveform such as ST-segment elevations or depressions and wave inversions or losses in different leads depending on the localization of MI [21]. Public repositories publish anonymized digital datasets in the form of time series. Examples include “PhysioNet” that is a repository consisting of a large and balanced amount of MI and control subjects [22]. The present study examines a specific selection of a PTB Diagnostic ECG Database available in the PhysioNet repository [23].

The rest of the paper is organized as follows: Section 2 describes the proposed method, WTC. Section 3 presents the performance evaluation results for the WTC method in conjunction with two other shapelet-based benchmark methods that use the cardiac AP and the three leads of the ECG time series datasets. The medical and technical aspects of these results are discussed in Section 4, and this is followed by the final section that presents the conclusions of the study.

2. WTC method

In this section, the proposed Window-based Time Series Feature Extraction (WTC) method is explained in a stepwise manner. Initially, Section 2.1 describes the necessity of the pre-processing performed on the time series instances and how important this preliminary step is to the proposed method. Following that, Section 2.2 entails the proposed method by elucidating its components that constitute a basis for it. The mathematical expressions for the major components are also predicated with a pseudocode for a further grasp of the computational complexity calculated at the end of this section. Table 1 summarizes the notation presumed throughout this section.

2.1. Dataset pre-processing

The proposed method requires processing local features in the time series of interest. Thus, time series instances are registered both in time and signal amplitudes. Registration in time refers to aligning time series to a fixed and known position such as a time instant at which a certain event is expected to occur for a single-cycle phenomenon (e.g., ex-vivo action potential recordings) or the epoch of a cycle for a cyclic phenomenon (e.g., ECG and phonocardiogram recordings). In contrast, registration in signal amplitude facilitates performing time series analysis irrespective of the offset and scale of the recorded signal that may well depend on an experimental setup. In this study, each instance in a time series dataset S is denoted by S_k where $k \in [1, K]$ and K denotes the total number of instances available. For the rest of this section, it is assumed that the whole dataset S is registered to obtain the dataset T in which instances are denoted by T_k . Registration constitutes a data-dependent step, and thus, the method to obtain T from S is described in Section 3 that introduces AP and ECG time series datasets used to validate the proposed method.

2.2. Proposed analysis

For the purpose of notational simplicity, the proposed WTC method is described with respect to a single class label throughout this section. It is noted that it is possible to independently run WTC for each class for a multiple class scenario as shown in Section 3 in which datasets with two class labels are examined.

2.2.1. Determining the local time window

Time series instances continue to exhibit a certain degree of inter-subject variation even after registration. The averaging of registered time series instances makes the proposed method robust against noise and distortions that stem from trajectory variability and possible artifacts [24]. In this aspect, the registered time series instances, T_k , are averaged out to obtain a class representative average time series that is denoted by T^A as follows:

$$T^A(n) = \frac{1}{K} \sum_{k=1}^K T_k(n) \quad (1)$$

where n denotes the sample point index of the time series. Restrictions on the averaging step to time series instances bearing the same class label emphasize local dynamics and their significance in defining class behavior.

The resulting representative trajectory constitutes a base time series from which the local features of the time series are extracted. The determination of suitable time windows, preferably extracted from the time series itself, is necessary to obtain local features. The domain of locality is defined as non-overlapping and contiguous time windows denoted by W_j with a fixed length w where $j \in \{1, 2, \dots, J\}$ and J corresponds to the total number windows along T^A . J is equal to $\lfloor N/w \rfloor$ where N denotes the total number of sample points in T^A . It is noted that these time windows span the entire time series, and the length of the last time window, W_J , may be lesser than w .

In order to choose w , this study considers “Perceptually Important Points” (PIPs) [25,26] and well-known Discrete Cosine Transform (DCT) as heuristic tools. This is followed by briefly reviewing the concept of PIP and subsequently describing the manner in which it is complemented by DCT to select w . Specifically, PIPs are defined as observational points in a time series that exert a relatively high influence on human vision [27]. The PIP approach is useful in reducing dimensionality while indicating the hierarchy among data points based on their visual importance. The identification of PIPs is as follows [26]: The first and the last data points in the time series are identified as the first two important points in the importance hierarchy, and thus the next PIP is determined as the data point with the largest distance to the points. Recursively, each candidate PIP at a specific iteration corresponds to the data point with the largest distance to its two adjacent PIPs that is determined so far. The PIP selection process continues until all data points in the time series are exhausted. The original study, [26], proposes that the distance between PIPs is measured by using any of the following metrics: Euclidean, vertical, or perpendicular distances. The present study employs Euclidean distance.

In the context of the present study, determination of PIPs in a time series fits closely with the aim of extracting class-representative patterns from a time series dataset. Essentially, the PIP algorithm sorts the data points of a given time series of total length N based on their perceptual importance. Next, the “most important” M data points are selected from the N points. For this purpose, it is assumed that $M = \lfloor N/r \rfloor$, where r denotes the oversampling ratio of the signal that is defined as follows:

$$r = f_s/f_N \quad (2)$$

The terms f_s and f_N in Eq. (2) denote sampling and Nyquist rates, respectively, of the signal of interest. The widely known one-dimensional DCT is employed to calculate $r = N/l^*$ where $r \in [1, N]$ and

$l^* \in \{1, 2, \dots, N\}$. The term l^* denotes the lowest frequency index in the DCT domain representation of the signal below in which δ percentage of its total energy is contained. It is assumed that DT^A denotes the one-dimensional DCT of the time series T^A , and then l^* is determined based on the following expression:

$$\arg \min_{l^*} \frac{\sum_{l=1}^{l^*} |DT_l^A|^2}{\sum_{l=1}^N |DT_l^A|^2} \geq \delta \quad (3)$$

The best choice for the threshold parameter δ is explored in Section 3 in conjunction with the remaining parameters of the proposed method.

This is followed by extracting M number of PIPs from the time series to finally set w to the maximum horizontal length between the consecutive PIPs such that it is guaranteed that at least one PIP appears in W_j for all j .

2.2.2. Distance-based similarity

Local time windows W_j are determined to calculate a distance-based similarity score for each time series instance, T_k . The Euclidean distance of a specific time series instance T_k to T^A within the support of W_j is denoted by $d_{j,k} = \|T^A(n) - T_k(n)\|$ where $n \in W_j$ and $d_{j,k} \in [0, \sqrt{w}H]$. H indicates the maximum amplitude value of the time series instances in T . The distance-based weight β_j corresponding to W_j is expressed as follows:

$$\beta_j = \frac{\sum_{k=1}^K \phi_{j,k}}{\sum_{j=1}^J \sum_{k=1}^K \phi_{j,k}} \quad (4)$$

where,

$$\phi_{j,k} = \sqrt{w}H - d_{j,k} \quad (5)$$

The term $\phi_{j,k}$ in Eq. (5) that falls in the interval $[0, \sqrt{w}H]$ corresponds to the reward associated with the proximity of the time series instance T_k to T^A for the implied time window W_j . The term $\phi_{j,k}$ provides a standardized scale for all local time windows that span the time series. Thus, they are weighted against each other as shown in Eq. (4). It is noted that an increase in the proximity of the time series instances to T^A for a particular W_j increases the associated weight. The distance-based similarity score denoted by $Z_{j,k}^{(d)}$ for time series instance T_k and time window W_j is calculated based on $\phi_{j,k}$ as follows:

$$Z_{j,k}^{(d)} = \beta_j \frac{\phi_{j,k}}{\sqrt{w}H} \quad (6)$$

2.2.3. Trajectory-based similarity

Distance-based similarity is complemented and reinforced with trajectory-based similarity. The average class representative time series, T^A , is used to obtain a similarity band as a class representative trajectory. It is assumed that each sample point in the registered time series follows a Gaussian distribution as shown in Ref. [28], and an upper (lower) time series denoted by T^U (T^L) is formed in which sample points are higher (lower) than those in T^A dictated by a confidence level, p . Consequently, the aforementioned similarity band is defined as a band bounded by T^U and T^L that is subsequently partitioned into time localized bandlets termed as Shape Confidence Bands (SCBs). Each SCB denoted by SCB_j is related to the corresponding W_j as follows:

$$SCB_j := \{(n, y) | n \in W_j \wedge y \in \mathbb{R} \wedge T^L(n) < y < T^U(n)\} \quad (7)$$

Specifically, the SCB_j is essential in determining as to whether the temporally coinciding part of a time series of interest aligns with the class representative average time series within W_j . The rationale of SCBs involves quantifying the similarity of a trajectory by restricting the orientation of an instance such that it is aligned with that of T^A allowing for a certain margin of error. For this purpose, each subsequence of a time

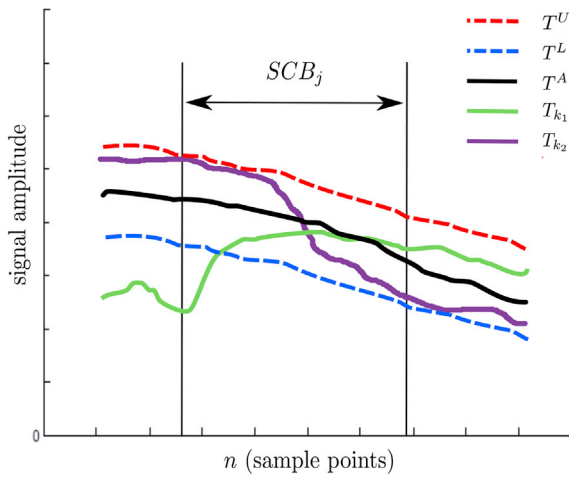


Fig. 1. An example of a case with two time series instances T_{k1} and T_{k2} with similar Euclidean distances from the representative average time series T^A albeit following different trajectories.

series is tested as to whether it completely resides within the boundaries of the corresponding SCB_j . It is assumed that $I_{j,k}$ denotes the indicator of inclusion for a time series instance T_k within SCB_j , which is defined as follows:

$$I_{j,k} = \begin{cases} 1, & T_k(n) \in SCB_j \forall n \in W_j, \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

Finally, the trajectory-based similarity score denoted by $Z_{j,k}^{(t)}$ for each time series instance T_k and SCB_j is expressed as follows:

$$Z_{j,k}^{(t)} = I_{j,k} \alpha_j \quad (9)$$

where,

$$\alpha_j = \frac{\sum_{k=1}^K I_{j,k}}{\sum_{j=1}^J \sum_{k=1}^K I_{j,k}} \quad (10)$$

The term α_j in Eq. (10) is defined as the trajectory-based weight of SCB_j calculated over all the time series instances that reside within SCB_j . It is noted that an increase in the number of time series instances that follow the trajectory of SCB_j increases the associated weight. Fig. 1 shows an example of a scenario that emphasizes the difference between the previously discussed distance and trajectory metrics. Time series instances T_{k1} and T_{k2} involve similar distances to T^A . However, the trajectory of T_{k1} falls out of the band SCB_j , and this is in contrast to T_{k2} .

2.2.4. Overall similarity

The trajectory-based and the distance-based similarity scores $Z_{j,k}^{(t)}$ and $Z_{j,k}^{(d)}$, respectively, are linearly combined to devise a single similarity score of time series T_k for each SCB_j that is denoted by $Z_{j,k}$ as follows:

$$Z_{j,k} = (1 - \gamma_j) Z_{j,k}^{(t)} + \gamma_j Z_{j,k}^{(d)}, \quad (11)$$

where $(1 - \gamma_j)$ and γ_j correspond to the associated weights of the trajectory-based and distance-based components, respectively. Thus, the average similarity score for each SCB_j is defined as follows:

$$Z_j = \sum_{k=1}^K Z_{j,k}. \quad (12)$$

It is proposed to calculate γ_j based on the probability mass function (PMF) of time series distances within SCB_j denoted by \mathbb{P}_j as follows:

$$\gamma_j = \frac{\mathcal{D}(\mathbb{D}, \mathbb{P}_j)}{\mathcal{D}(\mathbb{D}, \mathbb{P}_j) + \mathcal{D}(\mathbb{U}, \mathbb{P}_j)}, \quad (13)$$

The terms \mathbb{U} and \mathbb{D} in Eq. (13) denote PMFs corresponding to the uniform and deterministic distance distributions, respectively. Additionally, \mathcal{D} denotes an operator that computes the distance between two PMFs based on Algorithm 1 presented in Ref. [29]. The aforementioned PMFs are constructed with a signal level resolution denoted by c . Throughout this study, c is set to 0.001. Furthermore, PMF \mathbb{U} is defined over the support $[0, \sqrt{WH}]$, and \mathbb{D} corresponds to an impulse at zero. A random variable associated with the PMF \mathbb{U} (\mathbb{D}) possesses maximum (minimum) entropy for its support. The rationale of Eq. (13) involves assigning a weight to the distance-based (trajectory-based) similarity score of SCB_j inversely proportional to the distance of \mathbb{P}_j to \mathbb{U} (\mathbb{D}). Trajectory-based similarity scores are calculated over time series instances residing within each SCB_j . Conversely, distance-based similarity scores are calculated over the entire set of time series instances that are more dispersed with respect to the confidence band of interest. With respect to extreme cases where \mathbb{P}_j tends to \mathbb{U} (\mathbb{D}), distance-based (trajectory-based) similarity scores dominate the overall similarity score.

The overall algorithm is outlined in Algorithm 1, and the related source codes are accessible online via [30].

Algorithm 1 WTC (\mathcal{T}, p)

1: $[T^A, T^U, T^L] \leftarrow \text{RepresentativeTimeSeries}(\mathcal{T}, p)$	Eq. (1)
2: $r \leftarrow \text{PIPCompressionRatio}(T^A)$	Eq. (2)
3: $PIPs \leftarrow \text{PIPIndices}(T^A)$	[26]
4: $W_j \leftarrow \text{LocalTimeWindows}(PIPs, r)$	Section 2.2.1
5: $SCB_j \leftarrow \text{extractSCBs}(T^U, T^L, W_j)$	Eq. (7)
6: $\alpha_j \leftarrow \text{SCBWeightsAlpha}(\mathcal{T}, SCB_j, W_j)$	Eq. (10)
7: $\beta_j \leftarrow \text{SCBWeightsBeta}(\mathcal{T}, T^A, W_j)$	Eq. (4)
8: $\gamma_j \leftarrow \text{SCBWeightsGamma}(\mathcal{T}, SCB_j, W_j)$	Eq. (13)
9: $Z_j \leftarrow \text{SimilarityScores}(\mathcal{T}, \alpha_j, \beta_j, \gamma_j, W_j)$	Eq. (12)
10: return $[SCB_j, Z_j]$	

A prominent advantage of Algorithm 1 corresponds to its relatively lower complexity, which is analyzed next. Step 1 of Algorithm 1 involves point-wise mean and confidence level computations that involve $O(KN)$ complexity. Steps 2 and 3 require the computation of DCT and the execution of the PIP algorithm, respectively, over a single times series T^A , which corresponds to $O(N \log N)$ when N corresponds to a power of 2 ($O(N^2)$ otherwise) [31] and $O(N^2)$ [26]. Evidently, steps 4 and 5 involve $O(N)$ complexity. As suggested by the corresponding equations, Steps 6, 7, 8, and 9 correspond to $O(KN)$ complexity. It is concluded that Algorithm 1 possesses $O(KN + N^2)$ quadratic complexity, thereby summing up the individual complexities of its constituent steps. It is noted that it is possible to improve the efficiency of Step 3 that constitutes the most computationally intensive part of the overall algorithm by abandoning the related iterations immediately after determining the 100/ r percent-age of PIPs.

This section is concluded by stating that WTC generates Z_j which corresponds to the intra-class average similarity score for each SCB_j for a given time series dataset. Specifically, Z_j denotes an efficient representation of the dataset that it pertains to and is therefore expected to increase the interpretability of the dataset by domain experts.

3. Results

In this section, results of the proposed WTC method are presented for AP and ECG time series datasets. First, in Section 3.1, time series features are extracted from these datasets with WTC and two other benchmark methods, namely, the shapelet transform [16] (hereafter referred to as “brute-force shapelet transform”) and one of its variants that is termed as “fast shapelet transform” [32]. Fast shapelet transform is proposed to

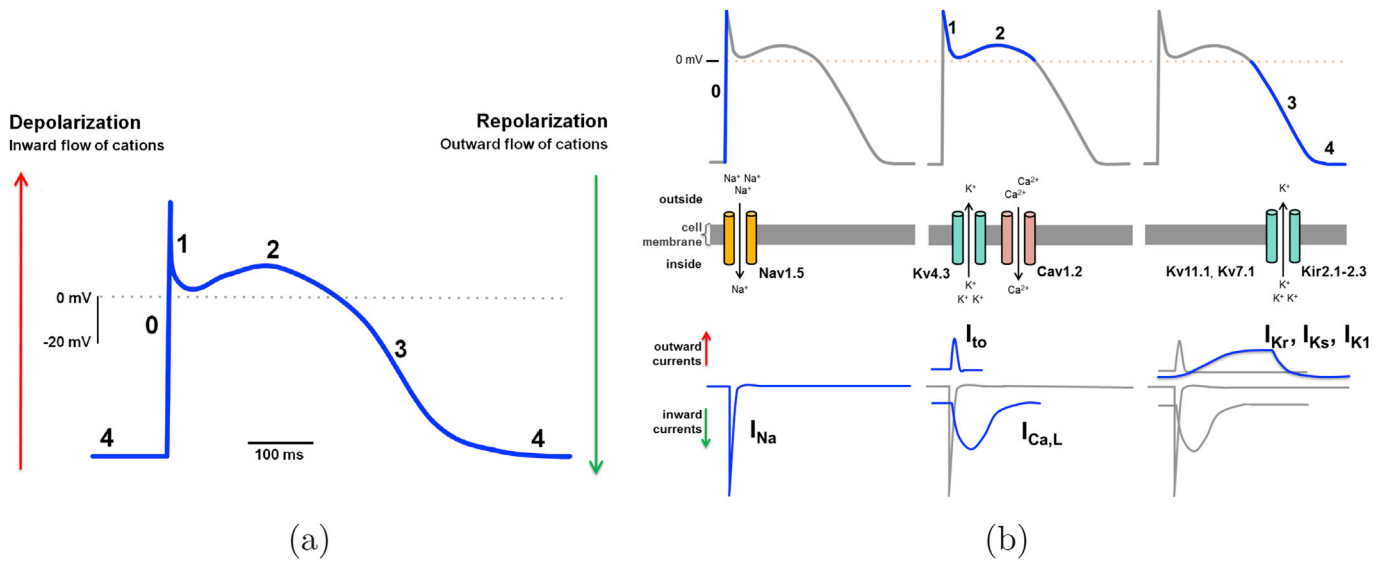


Fig. 2. (a) Schematic of a human cardiac AP in five phases, namely phase 0: rapid depolarization; phase 1: early rapid repolarization; phase 2: “plateau” phase; phase 3: final repolarization; and phase 4: resting membrane potential. (b) Schematic of human cardiac APs (upper row), ion channels (middle row), and major ion current flows (lower row). Nav1.5, cardiac Na⁺ channel conducting Na⁺ current (I_{Na}); Kv4.3, K⁺ channel conducting transient outward current (I_{to}); Cav1.2, L-type Ca²⁺ channel conducting L-type Ca²⁺ current (I_{Ca,L}); Kv11.1, K⁺ channel conducting rapidly activating, delayed outward rectifier K⁺ current (I_{Kr}); Kv7.1, slowly activating outward rectifier K⁺ current (I_{Ks}); Kir2.1, Kir2.3, inward rectifier K⁺ current (I_{K1}).

reduce the computational complexity of the legacy brute-force shapelet transform without sacrificing much from accuracy. The details for the shapelet-based methods are given in Section 3.1.1 before moving on to the feature extraction results provided in Sections 3.1.2 and 3.1.3. Next, the extracted features are evaluated with a series of classifiers and the corresponding results of WTC are compared with those of the brute-force and fast shapelet transforms in Section 3.2. Finally, statistical results regarding the ranking of WTC among the benchmark methods and the robustness aspect of WTC are presented in Section 3.3.

3.1. Feature extraction

3.1.1. Benchmark methods for feature extraction

Brute-force shapelet transform is a feature extraction method to discover the so-called shapelets [16] from a given time series dataset. Shapelets are essentially subsequences that are extracted from the time series instances themselves through an exhaustive sliding window-based search operation. It is assumed that L denotes a desired shapelet length specified by a user of the brute-force shapelet transform. For a given K number of time series instances of length N , there exist $(N - L + 1)$ subsequences, i.e., shapelet candidates, from each time series that result in a total of $(N - L + 1)K$ candidates. The distance between each subsequence - time series pair is computed. It is noted that a single distance computation requires sliding the subsequence across the time series of interest in all possible ways to determine the closest possible match. Next, a distance vector to all time series instances (with mixed class labels) is obtained for each subsequence. A distance threshold is then determined for each subsequence to yield the most homogenous separation of instances, i.e., the highest information gain, in terms of class labels. Finally, a user specified number of subsequences (with the highest information gain) is designated as shapelets. The corresponding distance vectors of the shapelets then correspond to transformed feature vectors that are generated by the brute-force shapelet transform. It is stated that the computational complexity of the shapelet extraction routine corresponds to $O(K^2N^4)$ [16,15]. The current state-of-the-art shapelet discovery method [33] reflects an improvement with respect to the brute-force method as it reduces the worst-case running time to $O(K^2N^3)$, which is still deemed as inconvenient for large datasets. More importantly, the temporal location of a shapelet is disregarded by the

very nature of the shapelet extraction process, and this could be of significant importance for biological interpretations.

A part of the latest efforts to improve shapelet extraction includes the fast shapelet transform [32]. In this approach, a complete dataset is projected to a lower-dimensional representation that is referred to as SAX [28] such that the entire shapelet discovery is completed in $O(KN^2)$ time at the expense of a degradation in performance in terms of accuracy. Although the fast shapelet transform avoids exhaustive distance computations, it does not guarantee that the same set of shapelets is reached as that reached by its brute-force counterpart. The proposed WTC method with $O(KN + N^2)$ complexity clearly outperforms the family of shapelet methods in the aforementioned aspect. In the present study, an open-source Java solution [15] is employed to obtain results for the brute-force shapelet transform, and a proprietary C++ based implementation [32] is used for fast shapelet transform.

3.1.2. Cardiac action potential (AP) dataset

As an application of the proposed method to a real-life dataset, AP recordings obtained from human right atrial biopsies are used [34]. Within the scope of a project entitled “The European Network for Translational Research in Atrial Fibrillation (EUTRAF)” [35] that is funded by the European Community's Seventh Framework Programme, the AP dataset was collected in the period from January 2006 to February 2014. Each patient's written informed consent was obtained, and thus the study conforms to the Declaration of Helsinki and was approved by the ethics committee of Dresden University of Technology (No. EK790799).

3.1.2.1. AP dataset description. The time series instances of the AP dataset are clinically attributed to the class label AF if it is considered that the corresponding patient is in chronic atrial fibrillation (ICD-10 code I48.2) and to the class label SR with respect to control patients with a normal sinus rhythm. Patients with paroxysmal or intermittent AF are excluded. The presence of AF was confirmed with pre-operative ECGs throughout the data collection. Eligible AP recordings were obtained from a total of 341 unique patients that comprise 142 AF (aged from 47 years to 85 years with a mean of 72.29 years and 37.68% female, 62.32% male) and 199 SR (aged 22 years–86 years with a mean of 67.12 years and 23.98% female, 76.02% male) patients with a temporal sampling rate of 10 kHz. In some cases, where an additional trabecula could be

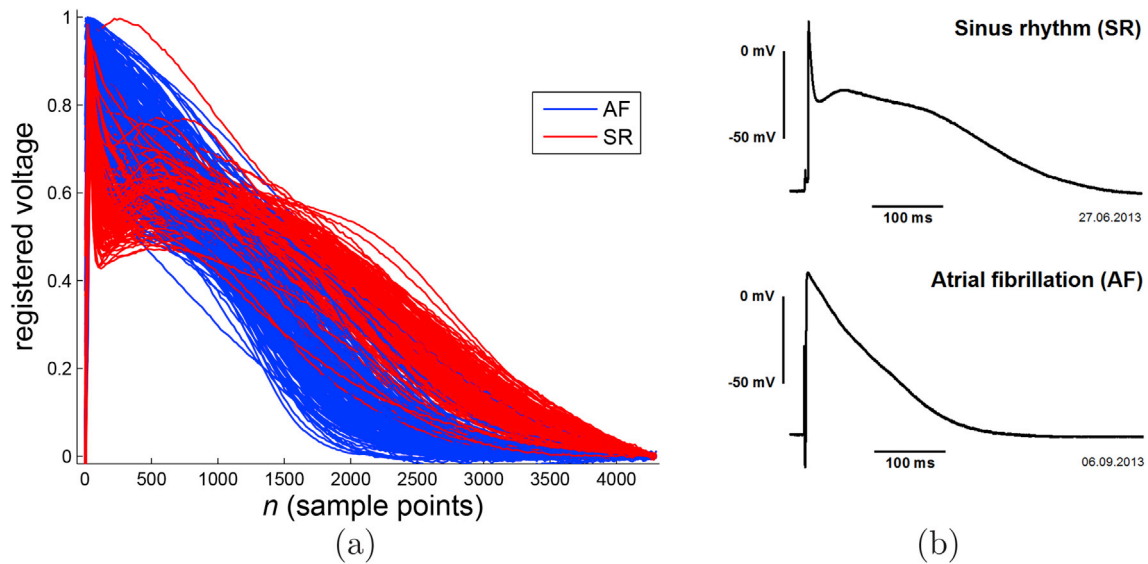


Fig. 3. (a) Registered AP time series instances of AF (202 instances) and SR (219 instances) with 4303 sample points each. (b) Typical “spike-and-dome” shaped AP from right atrial tissue of an SR patient (top) and triangular shaped AP from an AF patient (bottom).

dissected for a separate experiment, more than one AP sequence per patient were recorded. For multiple recordings from the same patient, experiments were conducted either in parallel (but on different set-ups) or one after the other, where the remaining preparations were stored in oxygenated Tyrode solution until use. With respect to the class AF, 86 patients have a single recording, 52 patients have 2 recordings and 4 patients have 3 recordings each, making a total of 202 time series instances. For the class SR, 179 patients have a single recording and 20 patients have 2 recordings each, making a total of 219 time series instances. Therefore, $K = 202$ AF time series instances were obtained from 142 AF patients and $K = 219$ SR time series instances were obtained from 199 SR patients.

In order to pump blood, it is necessary for all parts of the heart to contract in concert. Additionally, electrical excitation of the cardiomyocytes is a prerequisite for coordinated contraction. From the pacemaker cells in which electrical activity is generated, APs propagate throughout the heart and trigger contraction [36].

Each AP exhibits the following distinct phases (as shown in Fig. 2a): the resting membrane potential (phase 4), the rapid depolarization

(“upstroke”, phase 0), early rapid repolarization (phase 1), the “plateau” (phase 2); final repolarization (phase 3), and return to resting membrane potential (phase 4). Excitable cells possess an inside-negative resting membrane potential that is maintained by electrochemical gradients for the cations Na^+ and K^+ across the cell membrane due to ion pumps as well as voltage-dependent ion channels that open and close in a voltage-dependent and time-dependent manner (see Fig. 2b).

When the threshold for activation is reached by a small depolarization of the cell membrane, Na^+ channels open rapidly and allow Na^+ to enter the cell. The influx of positive charge further depolarizes the cell membrane. Repolarization is caused by an efflux of K^+ through various K^+ channels of different kinetics. An influx of Ca^{2+} via L-type Ca^{2+} channels triggers the release of further Ca^{2+} from intracellular stores, and this is necessary for the activation of contractile machinery. Thus, the shape of an AP is governed by the sum of all membrane currents listed along the lower row of Fig. 2b that flow through various ion channels.

Atrial APs from SR patients have a characteristic “spike-and-dome” shape. During AF, electrical activity becomes very fast and uncoordinated, leading to remodeling processes that change - amongst others - the expression of ion channels in the atrial cardiomyocytes and giving rise to the typical triangular-shaped AP (see Fig. 3b; [17,37]).

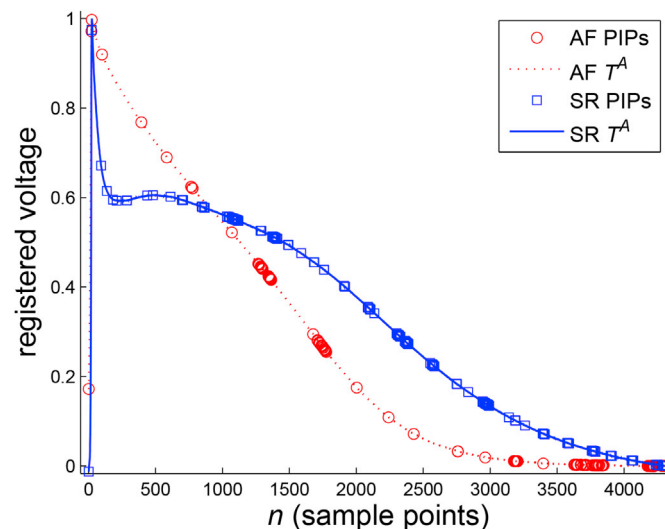


Fig. 4. Identified PIPs from the class representative average time series of AF and SR classes.

3.1.2.2. AP dataset registration. AP time series instances are single event-triggered and acyclic, and therefore do not require segmentation for registration. All AF and SR instances are registered in time based on the instant at which membrane potential of each time series peak occurs, i.e., end of phase 0. The time window length of a time series from its peak membrane potential to the resting membrane potential actually constitutes a differentiating factor between AF and SR classes due to ion channel activities under investigation. Thus, in the study, AP registration is done with respect to the time course of the acquired signal by avoiding techniques such as re-sampling and dynamic time warping. Time domain alignment provides human-readability and brings localized time windows into prominence, thereby easing decision making in real-life practice. First, the peak membrane potentials of all time series instances are aligned in time. Then, the largest leftward and rightward supports are found for which each time series instance is guaranteed to have all corresponding sample points. For this particular AP dataset, the leftward and rightward supports turn out to be 24 and 4278 sample points, respectively. Adding the single sample point of the peak membrane potential, this amounts to an AP time series length of $N = 4303$

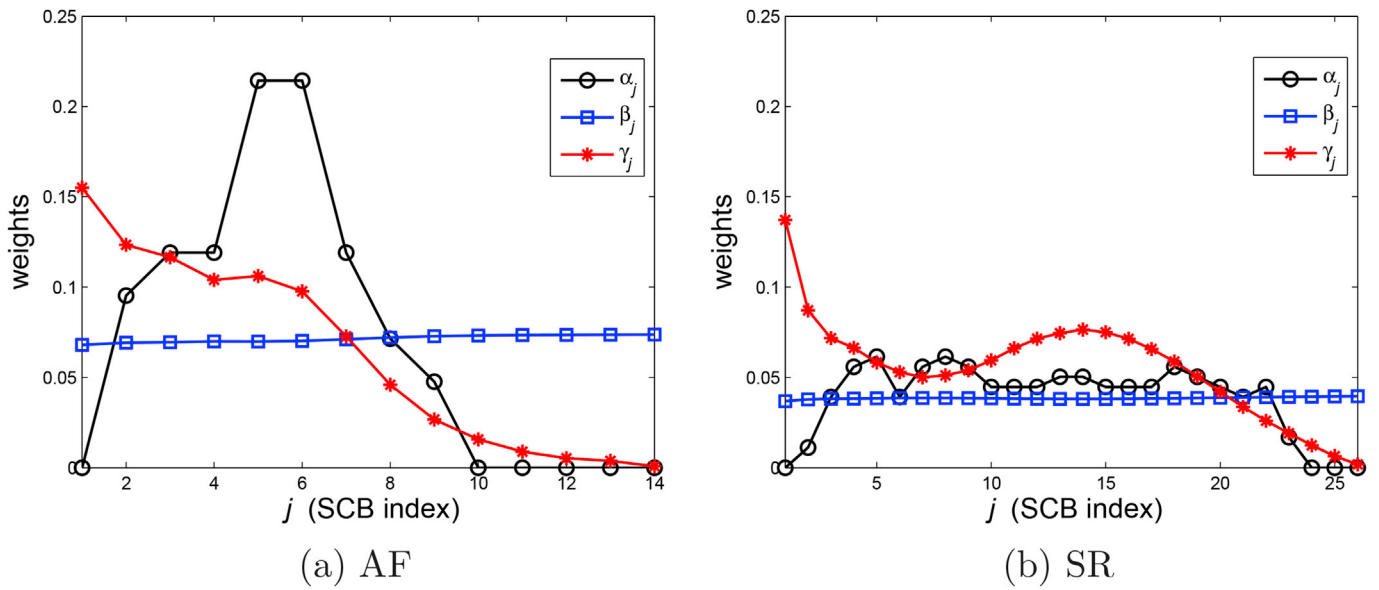


Fig. 5. The weights α_j , β_j and γ_j for SCBs in (a) AF and (b) SR.

sample points for both classes.

The next step comprises of registering signal amplitudes. With respect to the AP dataset, a single cycle begins with the voltage level that indicates a resting membrane potential and ends with almost the same ground value after an AP is triggered. Given this fact, the peak membrane potential occurring at the end of the rapid depolarization phase, i.e., phase 0, and the resting membrane potential phase, i.e., phase 4, of the experiment are considered as reference points for signal amplitude registration. Specifically, the registered time series T_k is obtained as follows:

$$T_k(n) = \frac{S_k(n) - S_k(N)}{\max\{S_k(n)\} - S_k(N)} \quad (14)$$

In Fig. 3a, all registered instances of AF (blue) and SR (red) classes are shown.

Time series instances in Fig. 3a manifest substantial intra-class variability for both classes that corresponds to a coherent observation with

the study in Ref. [38].

3.1.2.3. WTC on AP dataset. An arbitrarily chosen set of 50 AF and 50 SR time series instances are used to extract features with WTC method. The same set is also used for the brute-force and fast shapelet transform methods in the following section. A complementary set with 152 AF and 169 SR instances is spared to evaluate the classification performances of all three methods in Section 3.2.1.

The proposed WTC method is executed for the energy threshold parameter $\delta \in \{0.9900, 0.9990, 0.9999\}$ and confidence level parameter $p \in \{0.95, 0.99\}$. However, only the results corresponding to $\delta = 0.9990$ and $p = 0.95$ are presented in this section. For these values of δ and p , WTC is shown in Section 3.2.1.1 to have peak performance in terms of classification accuracy. This selection of δ results in an oversampling ratio r corresponding to 27.0629 and 21.4078 for AF and SR, respectively. Fig. 4 shows selected PIPs from the class-representative average time series based on these values of r for both AF and SR classes.

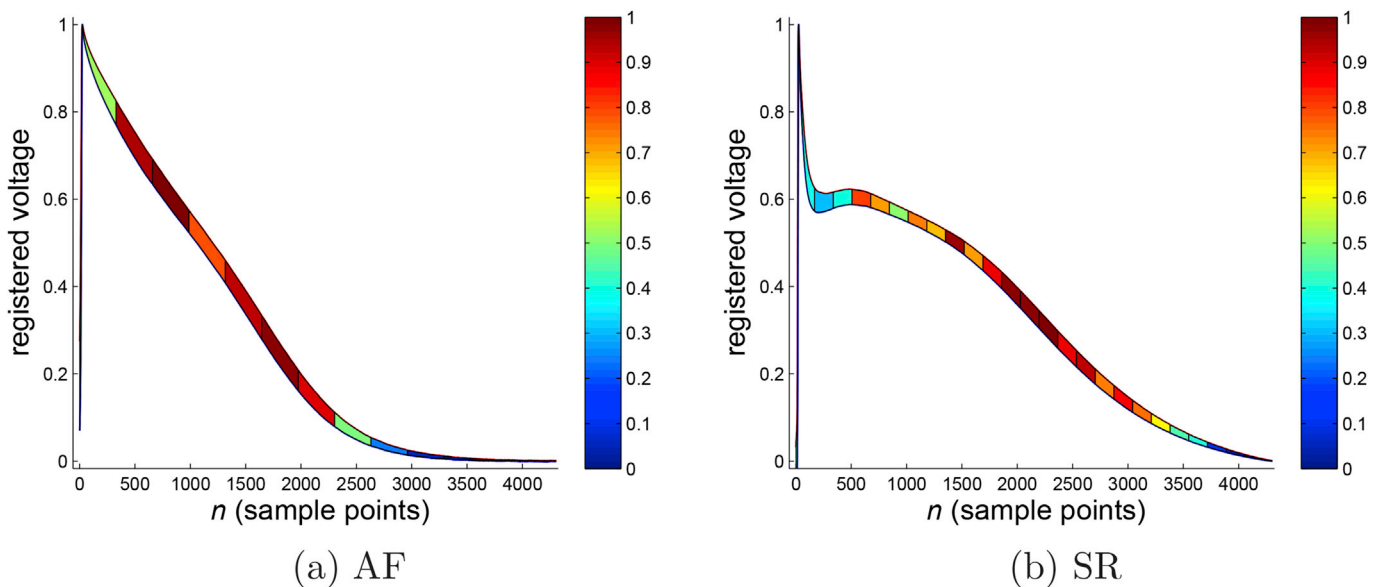


Fig. 6. Color map plots for the average values of Z_j for the classification set. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

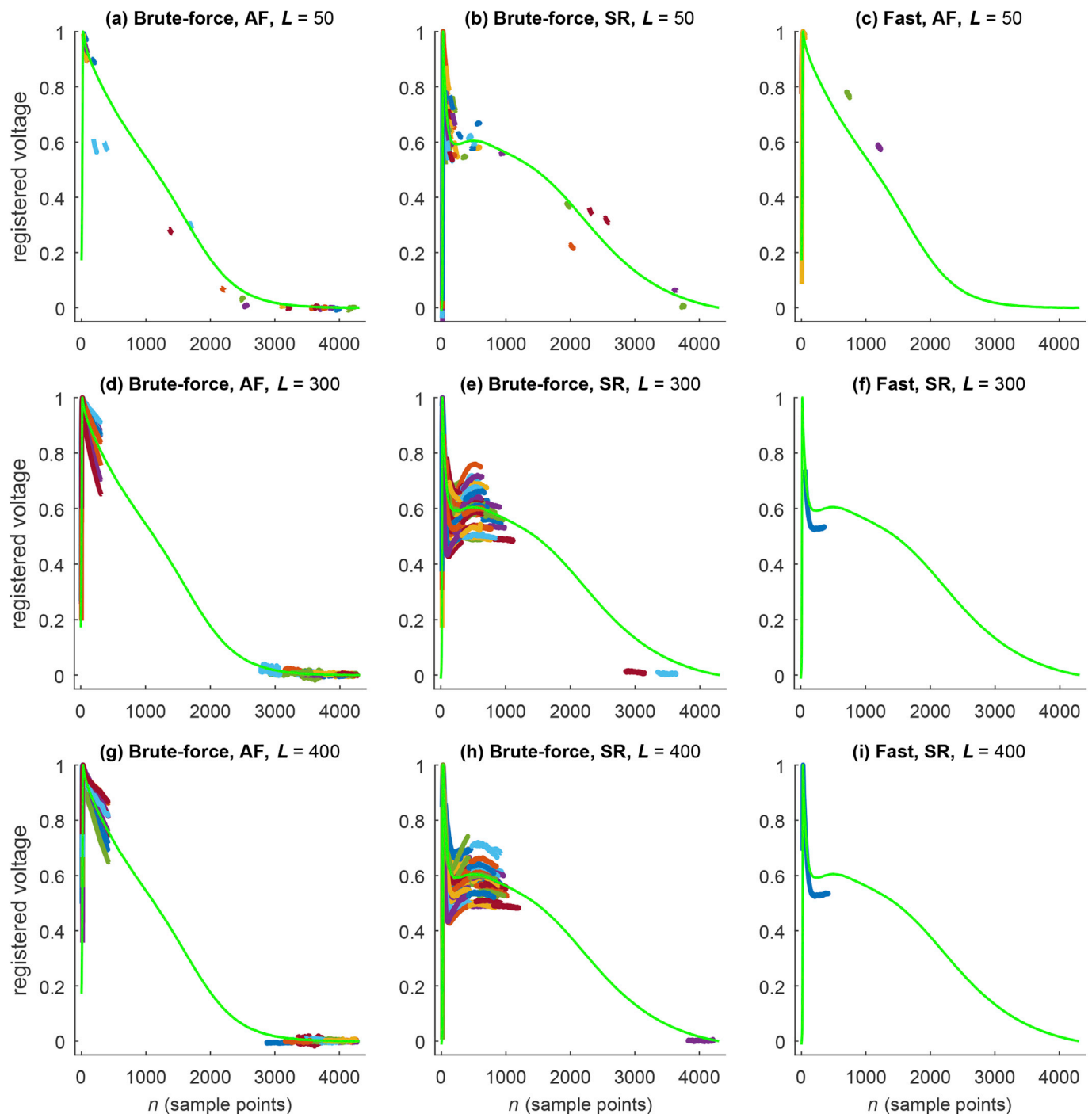


Fig. 7. Top 100 brute-force shapelets extracted from AF and SR instances with (a, b) $L = 50$, (d, e) $L = 300$, (g, h) $L = 400$ and fast shapelets with (c) $L = 50$ from AF instances, (f) $L = 300$ and (i) $L = 400$ from SR instances that are overlaid with the associated class representative average time series T^A .

The values of the localized time window length, w , which corresponds to the maximum horizontal separation between the adjacent PIPs, involve 329 and 169 sample points corresponding to 32.9 ms and 16.9 ms for AF and SR, respectively. The window length determination step is immediately followed by constructing confidence bands with a chosen confidence level corresponding to $p = 0.95$.

This is followed by segmenting the confidence bands as suggested by the localized time windows W_j into SCBs for each class. The calculated weights, namely α_j , β_j , and γ_j , for the corresponding W_j are depicted in Fig. 5a and b for AF and SR, respectively.

With respect to both classes, the results indicate that the distance-based weights β_j are close to each other, and this implies that the

rewards given based on the proximity to the class average for the corresponding SCB_j exhibit a similar trend. Conversely, trajectory-based weights α_j exhibit a more significant variation with respect to the variations in the SCB_j and even decrease to zero for higher indexed SCBs for both classes. This behavior indicates the failure of all time series instances to be completely encapsulated by the corresponding SCBs. The calculated weights γ_j are much closer to 0 than to 1, thereby indicating that they favor the trajectory-based component of the similarity score for both classes.

In order to visualize the predictive power of WTC in terms of Z_j , color map plots are presented in Fig. 6a and b for AF and SR, respectively.

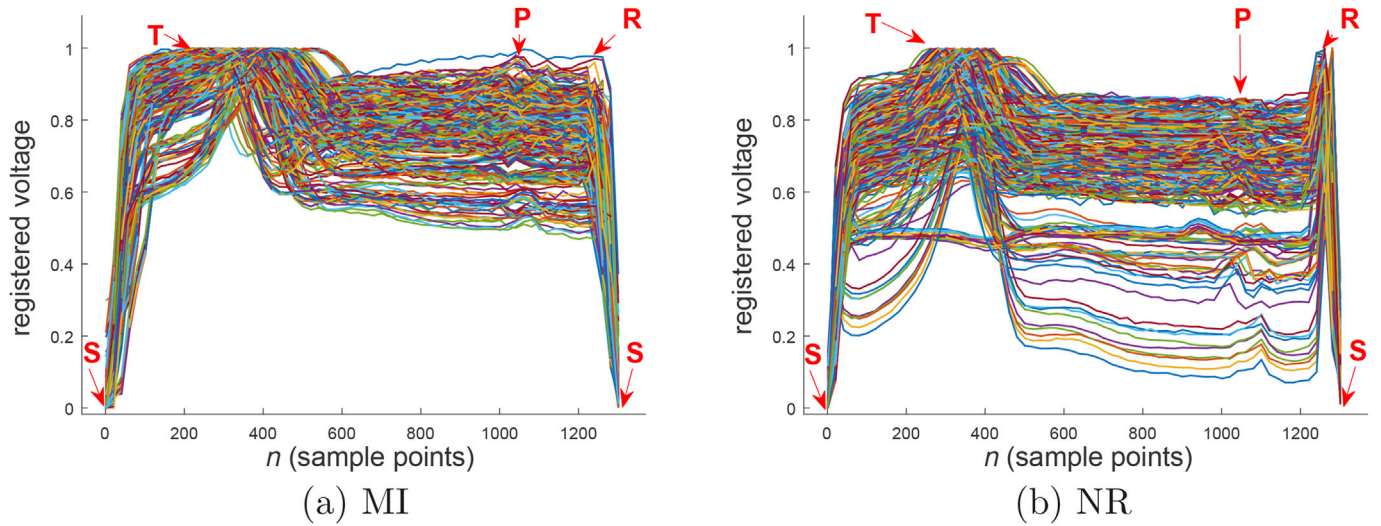


Fig. 8. Registered and annotated ECG lead v2 time series instances of (a) MI (40 instances) and (b) NR (40 instances) that each include 1308 sample points.

These figures constitute the most interpretable output of WTC and show the similarity scores Z_j calculated by averaging $Z_{j,k}$ over all AF and SR time series instances in the classification set. The more reddish tones imply regions with higher similarity, and thereby SCBs with higher predictive power. This observation coincides with the expected change in morphology for phases 2 and 3 of the AP signal (Fig. 2a) in the case of AF [39]. These figures also depict the double-sided confidence bands of the class representative average time series and the extracted SCBs.

3.1.2.4. Benchmark methods on AP dataset. The following feature extraction results are obtained by shapelet-based methods for the aforementioned subsets of AP dataset (refer to Section 3.1.2.3). In order to achieve a meaningful level of parameter exploration for the brute-force shapelet transform in a reasonable time, a range for the shapelet length L (in units of sample points) is quantized as $L \in \{50, 100, 150, 200, 250, 300, 350, 400\}$. It is noted that the chosen values for L span a wider range when compared to the values reached by the proposed WTC method for its time localized window length, w , with the aid of DCT domain analysis. The default value of the Java implementation [15] that corresponds to 100 is used for the number of shapelets to be extracted. A set of extracted brute-force shapelets are shown in Fig. 7a, b, d, e, g and h along with the class average time series, T^A , for both AF and SR.

This is followed by the execution of the fast shapelet transform. The fast shapelet procedure acts as an extension to the brute-force shapelet transform method and builds an internal decision tree termed as a “fast shapelet tree” with nodes that are ordered based on the information gain that they individually offer. It is assumed that L possesses the same range of values as those for the brute-force shapelet transform for the purpose of fairness. A set of extracted fast shapelets are depicted in Fig. 7c, f and i along with the class average time series, T^A .

3.1.3. ECG dataset

In this section, three chest leads of an ECG dataset are examined to demonstrate the potential of the proposed WTC method from a clinical perspective. The ECG is highly clinically relevant, and therefore constitutes one of the most studied biophysical signal types.

3.1.3.1. ECG dataset description. The instances of the ECG dataset used in this study are compiled from Physikalisch-Technische Bundesanstalt DataBase (PTBDB), which is a publicly available repository of physiological signals in PhysioNet [23]. The PTBDB contains 549 12-lead ECG time series recordings from 290 unique patients. Each time series is

digitized with a 1 kHz sampling rate and 16-bit resolution over a signal amplitude range of ± 16.384 mV. The PTBDB contains instances with a wide variety of labels including myocardial infarctions, cardiomyopathy/heart failures, dysrhythmia, myocardial hypertrophy, and myocarditis. The largest subset of PTBDB is comprised of time series instances belonging to patients diagnosed with acute myocardial infarction. There are identified differences in the expression of P-QRS-T sequences of the ECG signal recorded from precordial leads during myocardial infarction [40]. Furthermore, it is known that ECG contains noise originating from different sources [41–43]. Thus, this dataset is considered as suitable to test the robustness of the proposed WTC method with respect to noise. Hereafter, the time series instances of this dataset is referred to as “MI” and “NR” corresponding to patients diagnosed with acute myocardial infarction and control patients with normal heart rhythms, respectively. In the present study, out of 148 instances of MI patients, a sub-group of 40 unique patients (aged from 37 years to 85 years with a mean age of 61.03 years and 22.50% female, 77.50% male) are randomly selected to obtain a balanced set relative to 40 available NR patients (aged from 17 years to 81 years with a mean age corresponding to 45.59 years and 25.00% female, 75.00% male).

3.1.3.2. ECG dataset registration. All MI and NR time series instances are registered for the proposed method. Each cyclic ECG time series instance is first segmented into multiple single-cycle P-QRS-T sequences beginning from an S point to a subsequent point. In contrast to the acyclic recordings in the AP dataset, ECG recordings are self-repeating and their periodicities are determined by the heart rate of the patients. However, it is the signal variation among the P-QRS-T sequences as opposed to the heart rate variability that differentiates MI patients from NR patients. Hence, heart rate variability (i.e., length of the time series from an S point to a subsequent point) is compensated by incorporating a signal processing technique termed as re-sampling. Re-sampling essentially warps the specified time series, which constitutes a single-cycle ECG signal in this specific case. Each single-cycle P-QRS-T time series is re-sampled in order to extend its length to that of the longest one. It is assumed that $S_k(n)$ denotes the re-sampled single-cycle P-QRS-T time series, and the registered time series $T_k(n)$ is obtained by normalizing their amplitude as follows:

$$T_k(n) = \frac{S_k(n) - \min\{S_k(n)\}}{\max\{S_k(n)\} - \min\{S_k(n)\}}. \quad (15)$$

The three precordial leads of v2, v3, and v4 are separately examined. Fig. 8a and b shows $T_k(n)$ of MI and NR classes, respectively, for lead v2.

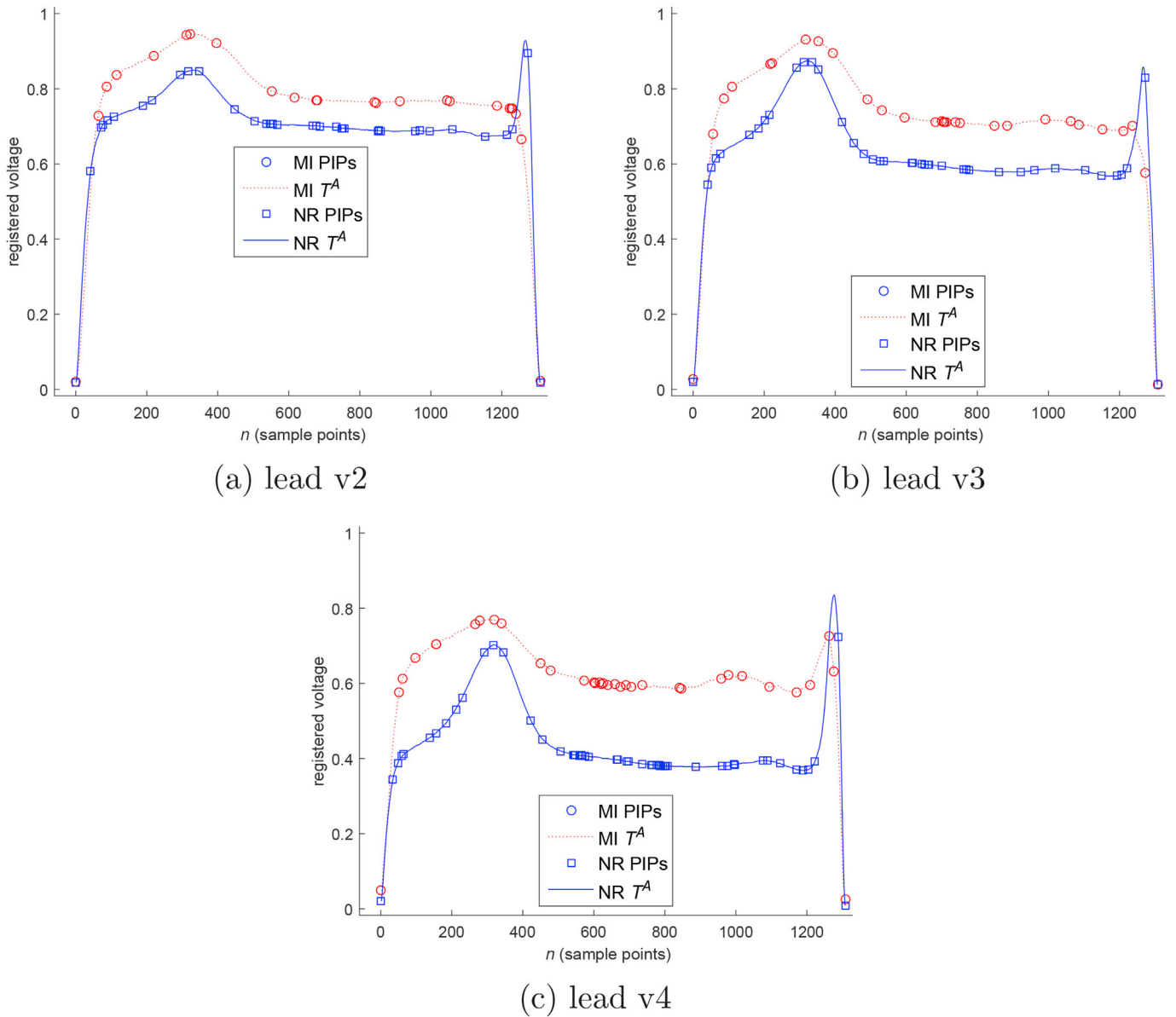


Fig. 9. Identified PIPs from the MI and NR class representative average time series from (a) v2 (b) v3, and (c) v4 leads of the ECG dataset.

3.1.3.3. WTC on ECG dataset. Out of 40 MI (NR) unique patients, 25 are reserved for the feature extraction phase, and this leaves 15 patients to evaluate the classification performance in Section 3.2.2. With respect to the feature extraction set, a single P-QRS-T sequence is allowed from each patient resulting in 25 time series instances. However, for the classification set, multiple (either 3 or 4) sequences are collected from each patient to reach a total of 50 time series instances.

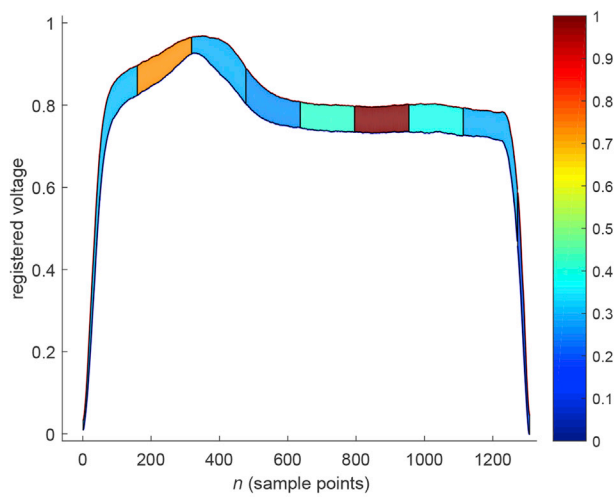
For all leads of the ECG dataset, the same set of parameter values $\delta \in \{0.9900, 0.9990, 0.9999\}$ and confidence level parameter $p \in \{0.95, 0.99\}$ as those of the AP dataset in Section 3.1.2.3 are used. In a similar manner, the presented results in this section are limited to those of $\delta = 0.9990$ and $p = 0.95$, which are shown to yield the best classification accuracy results in Section 3.2.2.1. Accordingly, Fig. 9a, b, and c show selected PIPs in conjunction with the class-representative average time series for both MI (red) and NR (blue) classes for leads v2, v3, and v4, respectively.

The resulting similarity scores, Z_j , are shown in the form of color map plots for both MI and NR and for the three studied leads in Fig. 10a through f. These figures depict the degree of descriptiveness of the WTC method within each time window as discussed in Section 3.1.2.3. The

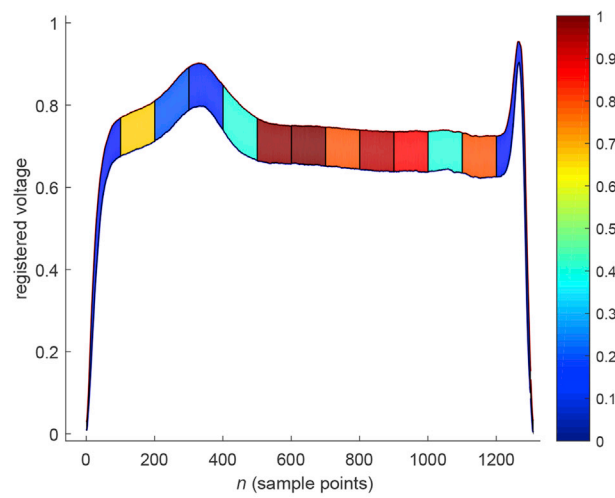
obvious difference in the thickness of the confidence bands when compared to those of the AP dataset indicates a higher variability of amplitude among the ECG instances.

Thus, WTC forms intra-class similarity along the temporal dimension for each SCB and points out morphological differences between the time series classes of interest. Prevalent cardiological literature indicates that the earliest signs of acute MI include increased T-wave amplitude (defined as “hyper-acute”) over the affected area. These are termed as hyper-acute T-waves and are most evident in the anterior precordial chest leads [44] (as shown in Fig. 11a with respect to the fore-mentioned MI behavior in leads v2, v3, and v4). Additionally, MI SCBs around the T-wave (Fig. 10a, c and e) possess a relatively deeper red color. This observation aligns with a characteristic hyper-acute T-wave pattern that is expected to appear in the early phases of MI and is accompanied by a loss of the R-wave amplitude in the anterior chest leading to v2, v3, and v4 [21,45].

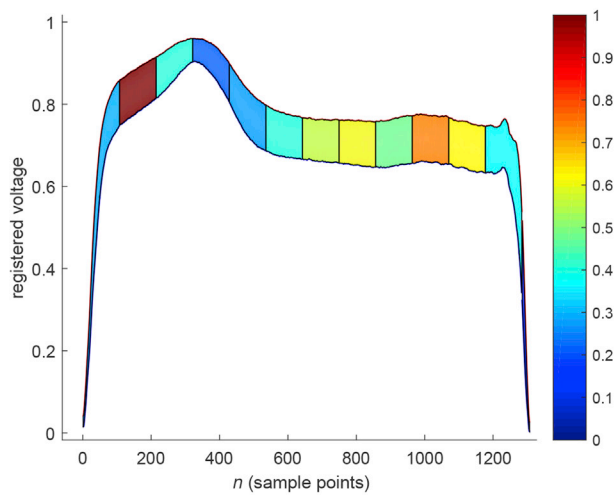
3.1.3.4. Benchmark methods on ECG dataset. The following shapelet-based executions are performed by using the aforementioned feature extraction set comprising of 25 MI and 25 NR time series instances of



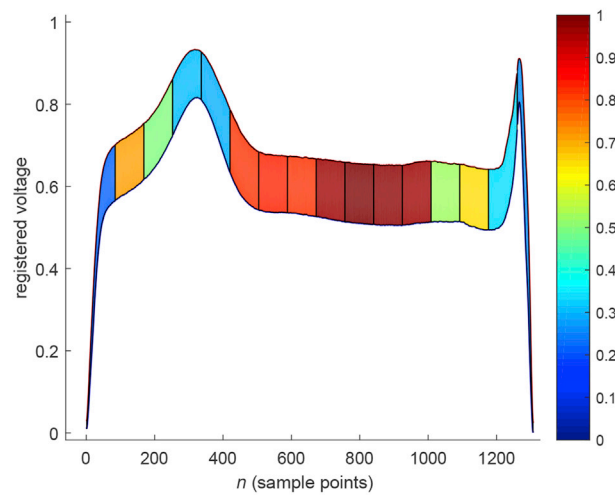
(a) MI(lead v2)



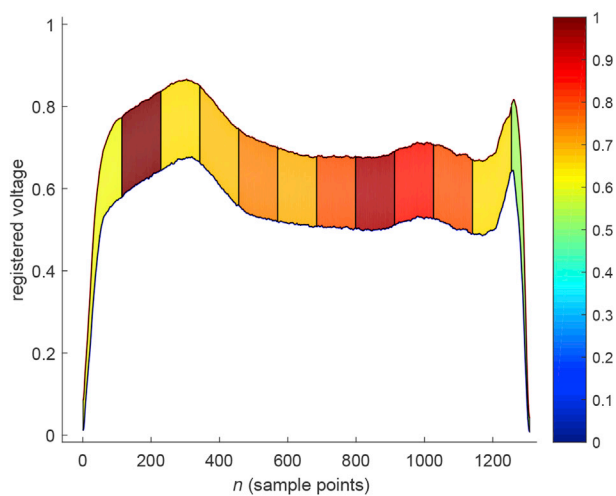
(b) NR(lead v2)



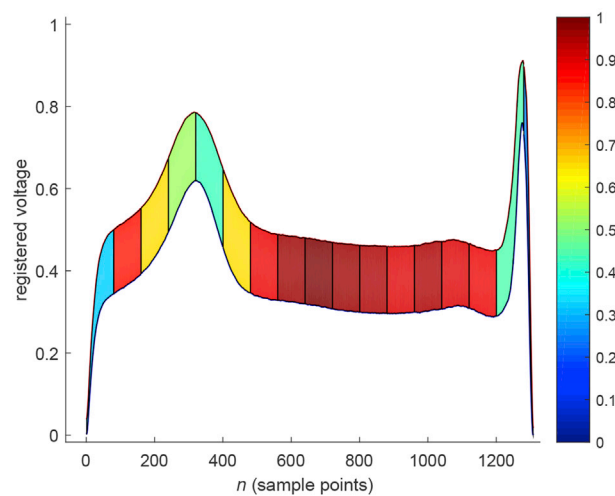
(c) MI(lead v3)



(d) NR(lead v3)



(e) MI(lead v4)



(f) NR(lead v4)

Fig. 10. Color map plots for the average values of Z_j of leads v2, v3 and v4 for (a,c,e) MI and (b,d,f) NR. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

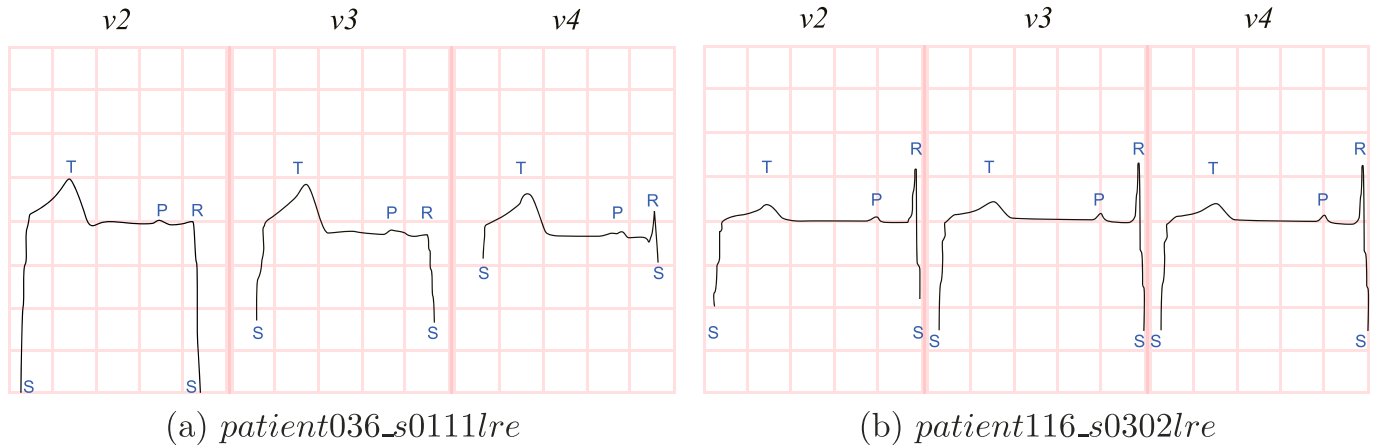


Fig. 11. (a) Typical “hyper acute” T-waves in ECG recordings collected from precordial leads v2, v3, and v4 indicating the condition of MI during early stages (record id: patient036_s0111lre) and (b) waveforms from corresponding leads of a control patient (record id: patient116_s0302lre). Each unit cell represents 0.2 s in time and 0.5 mV in amplitude. (Source: PhysioBank ATM).

leads v2, v3, and v4 separately. The desired shapelet length is sampled in the same manner with Section 3.1.2.4, while obtaining results for the brute-force and fast shapelet transform and limiting the number of shapelets to be extracted to 100. For the purpose of compactness, a limited set of extracted brute-force and fast shapelets from lead v2 is shown in Fig. 12.

3.2. Performance evaluation

A series of classifiers are used to facilitate a comparison of the proposed WTC and the benchmark shapelet-based methods for AP and ECG datasets, in Sections 3.2.1 and 3.2.2, respectively. The following methods within WEKA framework [46] are selected to maintain a broad coverage among the available classifiers: the well-known Naive Bayes classifier, J48 pruned tree (an implementation of the well-known C4.5 algorithm), random decision forest [47], adaptive boosting (AdaBoost.M1) [48], classification via regression (employing a type of decision tree with linear regression functions at the leaves [49]), bagging [50], multi-boosting (MultiBoostAB) [51], locally weighted learning (LWL) [52], partial decision tree classifier (PART) [53], ensemble of nested dichotomies (END) [54], decision stump [55], simple classification and regression tree (CART) [56], a proprietary algorithm termed as “ranking instances by maximizing the area under the ROC curve (RIMARC)” [57], Bayesian network learning, dagging [58], random subspace method [59], decision table majority classifier, ripple-down rule learner (RIDOR) [60], alternating decision tree (ADTree) [61], and random tree construction and multi-class alternating decision tree with a logit-boost strategy (LAD-Tree). The similarity scores $Z_{j,k}$ and the shapelet distances to time series instances constitute transformed feature vectors for WTC and the benchmark methods, respectively. The classification accuracy results are obtained with a 10-fold cross validation.

3.2.1. AP dataset

3.2.1.1. Classification performance of WTC on AP dataset. In order to assess its performance, the proposed WTC method is executed for the Cartesian product of the parameter sets $\delta \in \{0.9900, 0.9990, 0.9999\}$, and $p \in \{0.95, 0.99\}$. The resulting classification accuracies based on the feature extraction of WTC are tabulated in Table 2 for the spared classification set of 152 AF and 169 SR instances and for various classifiers offered by WEKA. With the exception of a few entries in Table 2, WTC yields accuracies within 5% vicinity of each other, and this underlines the relative insensitivity of WTC to its parameters. The success rate of WTC exceeds 94% for a majority of the cases. The highest average classification accuracy is achieved by the parameter pair $\delta = 0.9990$ and $p = 0.95$.

The proposed WTC method is self-sufficient in predicting a time localized window length, w . However, classification results are also obtained by artificially enforcing a window length. With respect the specific case, SCB extraction is performed by “by-passing” the prediction part for w . Table 3 lists the accuracies for variations in the enforced SCB lengths and for the selected classifiers.

A comparison of Tables 2 and 3 for the columns corresponding to ($\delta = 0.9990$, $p = 0.95$) and $w = 200$ (also for $w = 150$), respectively, reveals the success of the PIP and DCT-based heuristics pursued by the proposed WTC method to determine the value of time localized window length w .

3.2.1.2. Classification performances of benchmark methods on AP dataset. The individual and average classification accuracies of the brute-force shapelet transform for varying values of shapelet length L and the selected classifiers are given in Table 4. The brute-force shapelet transform is shown to yield varying classification accuracies for variations in L irrespective of the deployed classification method. Shapelet length $L = 400$ yields the most favorable ensemble average accuracy level (93.057%) and the corresponding first 100 brute-force shapelets are already shown in Fig. 7g and h along with the class average time series, T^A , for both AF and SR. The figures reveal that the selected shapelets from different instances substantially overlap with each other for both classes. This observation is also interpreted as a redundancy in the computations since majority of subsequences correspond to a narrow time interval.

The resulting classification accuracies of the fast shapelet transform are presented in Table 5. The distance values obtained from the “fast shapelet tree” are considered as inputs for the listed classifiers of WEKA.

As shown in Table 5, the ensemble average classification accuracy (90.269%) peaks at $L = 300$ for which the single discovered shapelet among the instances of SR is shown in Fig. 7f.

Based on the results in Tables 2, 3 and 5 the classification accuracies calculated for WTC approximately exceed those of the fast shapelet transform by 4%. Conversely, despite the exhaustive search with an impractical computational complexity, the peak ensemble average classification accuracy of the brute-force shapelet transform (93.057%) for $L = 400$ (shown in Table 4) is slightly lower than that of WTC for the enforced case with $w = 200$ (also for $w = 150$) (94.125%) as well as for the non-enforced, i.e., the original case (94.333%).

3.2.2. ECG dataset

3.2.2.1. Classification performance of WTC on ECG dataset. The energy threshold δ and the confidence level p parameters are explored within the

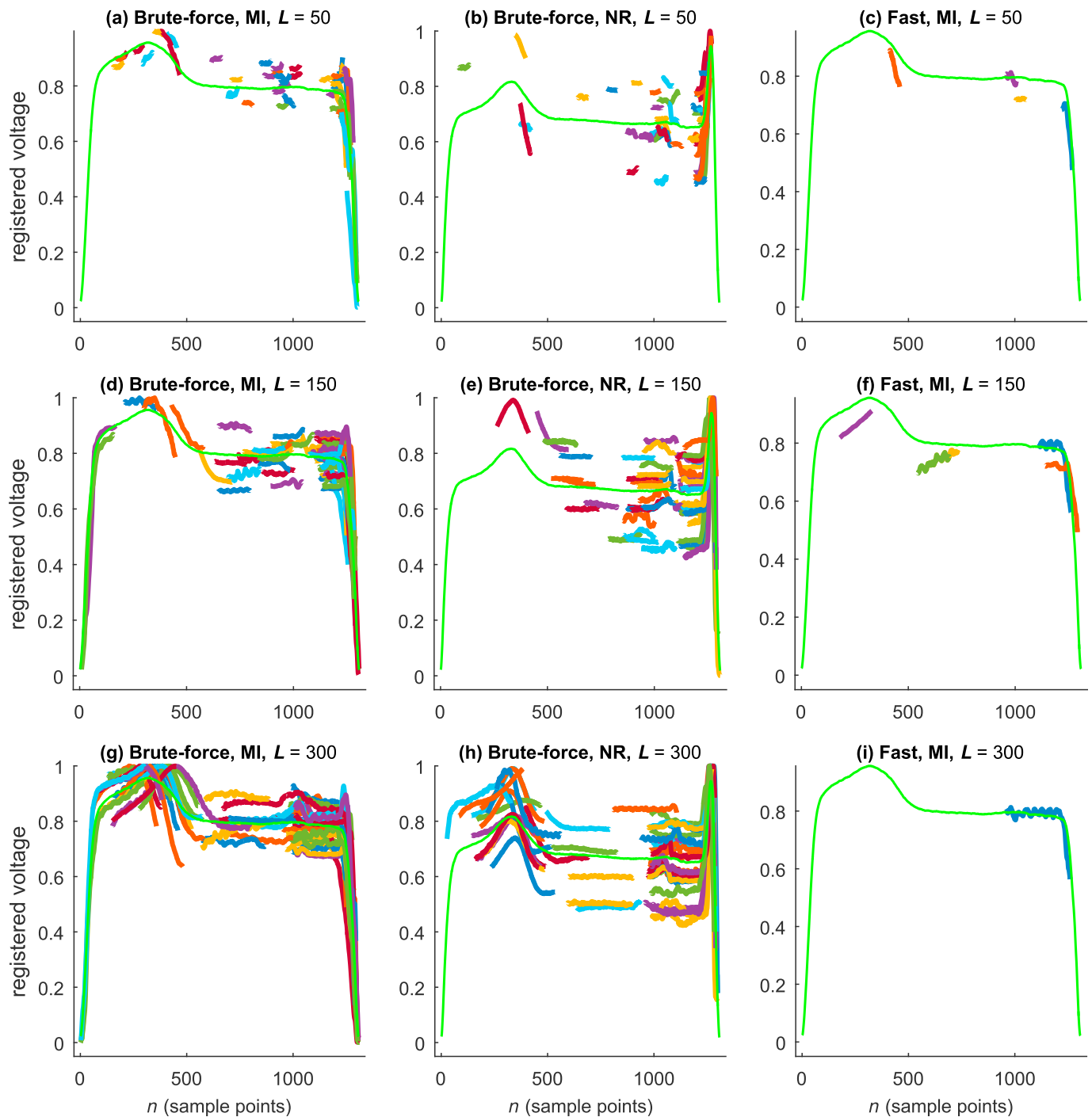


Fig. 12. Top 100 brute-force shapelets extracted from MI and NR instances with (a, b) $L = 50$, (d, e) $L = 150$, (g, h) $L = 300$ and fast shapelets selected only among MI instances with (c) $L = 50$, (f) $L = 150$ and (i) $L = 300$ that are overlaid with the associated class representative average time series T^A of lead v2.

same range of values as those in Section 3.2.1.1, to determine the pair that yields the highest average classification with respect to the aforementioned 21 classifiers within WEKA. Due to space limitations, only classification accuracies for the lead v2 are presented in Table 6 for variations in δ and p .

Table 6 reveals that the average accuracy for lead v2 of ECG dataset peaks (98.095%) at the same set of parameter values as those of the AP dataset in Section 3.2.1. Therefore, $\delta = 0.9990$ and $p = 0.95$ are also fixed for the leads v3 and v4. For the purpose of completeness, the corresponding classification results are listed in Table 7 for both of the leads.

3.2.2.2. Classification performances of benchmark methods on ECG dataset. The individual accuracies achieved by the chosen set of WEKA classifiers on lead v2 by using features extracted by the brute-force and fast shapelet transform methods are listed in Tables 8 and 9, respectively.

As shown in Tables 8 and 9, the classification success of both brute-force and fast shapelet transforms are highly sensitive to L . Shapelet length $L = 300$ yields the highest ensemble average accuracy (94.381%) for the brute-force shapelet transform. The first 100 shapelets with $L = 300$ on lead v2 are already depicted in Fig. 12g and h overlaid with the class average time series, T^A , for MI and NR, respectively. Fig. 12f also

Table 2

Individual and ensemble average classification accuracies of the proposed WTC method that is applied to the AP dataset for different values of the energy threshold δ , the confidence level p , and for the selected classifiers. The highest average classification accuracy is shown in boldface.

	$\delta = 0.9900$		$\delta = 0.9990$		$\delta = 0.9999$	
	$p = 0.95$	$p = 0.99$	$p = 0.95$	$p = 0.99$	$p = 0.95$	$p = 0.99$
Naive Bayes	74.766%	81.620%	92.212%	92.212%	89.720%	90.031%
J48	91.589%	92.212%	93.770%	93.458%	93.458%	94.081%
Random forest	93.770%	93.770%	94.704%	94.704%	93.146%	94.081%
AdaBoost.M1	93.146%	93.146%	95.639%	95.639%	95.639%	95.950%
Classif. via regr.	95.327%	95.327%	95.327%	95.016%	93.458%	93.146%
Bagging	94.393%	94.393%	94.704%	94.704%	94.704%	94.704%
MultiBoostAB	93.458%	93.458%	94.704%	94.704%	94.704%	94.704%
LWL	90.654%	90.654%	94.081%	94.704%	94.393%	94.704%
PART	94.081%	94.081%	95.639%	95.639%	94.081%	93.770%
END	91.900%	92.212%	93.458%	93.146%	92.835%	93.458%
Decision stump	90.654%	90.654%	95.016%	95.016%	95.016%	95.016%
Simple CART	92.835%	92.835%	94.704%	94.704%	94.393%	94.704%
RIMARC	94.081%	94.081%	95.016%	95.016%	94.704%	94.704%
Bayes NET	92.835%	92.835%	93.458%	93.458%	94.081%	94.081%
Dagging	94.393%	93.770%	94.704%	94.081%	92.835%	94.704%
Random SubSpace	93.770%	93.458%	93.146%	94.081%	94.081%	93.458%
Decision Table	92.523%	93.770%	92.523%	92.212%	93.146%	93.458%
Ridor	92.523%	92.523%	94.704%	94.393%	93.770%	94.081%
ADTree	94.393%	94.081%	94.393%	94.393%	95.016%	95.016%
LAD Tree	93.146%	93.458%	94.393%	94.393%	95.016%	95.327%
Random Tree	90.343%	91.277%	94.704%	92.523%	93.146%	90.966%
AVERAGE	92.123%	92.553%	94.333%	94.200%	93.873%	94.007%

shows the fast shapelets with $L = 150$ resulting in the highest ensemble average accuracy level (95.143%). For the purpose of completeness, the resulting classification accuracies obtained for lead v3 and v4 are presented in Table 10 through 13 for the benchmark methods by noting that similar observations with lead v2 are made.

As the classification accuracies of the proposed WTC method presented in Section 3.2.2.1 are considered, it is apparent that the brute-force and fast shapelet transforms yield relatively lower accuracies for all three leads of the ECG dataset. The brute-force and fast shapelet transform methods attain accuracies of (94.381%, 95.190%, 95.190%) and (95.143%, 94.667%, 88.286%) for leads (v2, v3, v4), respectively, and these are exceeded by those of the proposed WTC method that correspond to (98.095%, 98.429%, 96.095%).

3.3. Statistical evaluation

Evaluation of the classification performance of the proposed WTC

Table 3

Individual and ensemble average classification accuracies of the proposed WTC method applied to the AP dataset given variations in the enforced SCB length, w , and the selected classifiers. The highest average classification accuracy is shown in boldface.

	$w = 50$	$w = 100$	$w = 150$	$w = 200$	$w = 250$	$w = 300$	$w = 350$	$w = 400$
Naive Bayes	91.277%	90.966%	92.212%	91.589%	91.277%	90.654%	90.654%	90.031%
J48	93.770%	90.654%	93.458%	93.770%	94.081%	94.393%	93.458%	93.146%
Random forest	94.081%	94.704%	94.393%	94.393%	94.081%	93.770%	94.081%	93.458%
AdaBoost.M1	94.704%	95.016%	95.950%	94.393%	95.639%	94.393%	96.262%	95.327%
Classif. via regr.	94.393%	95.016%	95.950%	96.885%	95.327%	95.327%	95.950%	95.639%
Bagging	94.704%	95.016%	95.016%	94.704%	94.393%	94.704%	94.393%	94.393%
MultiBoostAB	95.016%	95.016%	94.704%	95.016%	94.704%	94.704%	93.770%	93.458%
LWL	93.770%	94.704%	93.770%	94.393%	92.835%	94.081%	91.589%	91.900%
PART	93.146%	92.212%	91.589%	91.900%	94.081%	92.835%	94.704%	94.704%
END	93.458%	90.031%	93.458%	93.458%	94.081%	94.704%	93.146%	93.146%
Decision stump	94.393%	95.016%	93.770%	95.016%	94.081%	95.016%	91.900%	91.900%
Simple CART	95.327%	95.016%	94.081%	93.770%	93.458%	96.262%	93.146%	94.704%
RIMARC	95.327%	95.327%	95.327%	95.327%	95.327%	95.327%	95.327%	95.327%
Bayes NET	94.081%	93.770%	94.393%	94.081%	94.081%	94.081%	94.393%	94.081%
Dagging	93.458%	92.212%	94.393%	94.081%	91.900%	93.146%	92.835%	92.835%
Random SubSpace	93.770%	93.770%	94.393%	94.081%	93.770%	93.458%	94.393%	94.081%
Decision Table	94.081%	91.277%	90.343%	92.212%	93.146%	93.770%	92.523%	94.393%
Ridor	94.081%	91.900%	95.016%	94.704%	94.393%	94.081%	94.081%	94.393%
ADTree	93.146%	95.016%	96.573%	95.327%	94.704%	94.393%	94.393%	94.393%
LAD Tree	94.704%	94.704%	95.639%	94.704%	93.770%	94.704%	95.016%	94.393%
Random Tree	92.835%	90.654%	92.212%	92.835%	94.704%	92.212%	92.835%	92.835%
AVERAGE	93.977%	93.428%	94.125%	94.125%	93.992%	94.096%	93.755%	93.740%

method is concluded by presenting a critical difference (CD) diagram [62]. The CD diagram is a representation that enables statistical evaluation of multiple methods over multiple datasets. Briefly, methods of interest are ranked for each dataset to obtain corresponding rank vectors that are subsequently averaged to determine average ranks. Methods whose rank differences exceed the value of CD are termed as “critically different”. The CD is calculated as follows:

$$CD = q_\alpha \sqrt{\frac{F(F+1)}{6R}}, \quad (16)$$

where F , R , and q_α denote the number of methods, number of datasets, and “critical value”, respectively. In the present study, $F = 3$ methods, $R = 4$ datasets, and $q_{0.1}$, which is equal to 2.052 is used for a two-tailed Nemenyi test. Further details on the CD diagram are available in Ref. [62]. The analysis in the present study involves ranking methods based on their average classification accuracies considered over 21

Table 4

Individual and ensemble average classification accuracies of the brute-force shapelet transform applied to the AP dataset for variations in the shapelet length, L , and the selected classifiers. The highest average classification accuracy is shown in boldface.

	$L = 50$	$L = 100$	$L = 150$	$L = 200$	$L = 250$	$L = 300$	$L = 350$	$L = 400$
Naive Bayes	74.766%	86.293%	89.408%	90.654%	91.900%	92.835%	93.146%	93.458%
J48	85.358%	87.851%	90.966%	89.408%	94.393%	93.770%	92.835%	92.523%
Random forest	86.293%	90.654%	93.458%	92.835%	94.393%	95.639%	94.393%	94.704%
AdaBoost.M1	81.308%	88.785%	90.966%	91.277%	94.081%	94.081%	94.704%	95.016%
Classif. via regr.	83.489%	90.654%	90.966%	92.835%	95.327%	94.704%	94.081%	95.327%
Bagging	87.851%	90.654%	92.212%	92.212%	94.393%	92.835%	93.146%	93.770%
MultiBoostAB	84.424%	88.474%	91.589%	92.212%	94.393%	94.081%	95.016%	94.704%
LWL	81.620%	81.932%	88.474%	87.851%	91.277%	87.227%	87.227%	90.031%
PART	81.308%	90.031%	90.654%	89.408%	94.081%	92.212%	93.458%	93.770%
END	85.358%	87.851%	90.966%	89.408%	94.393%	93.770%	92.835%	92.523%
Decision stump	81.620%	81.620%	88.785%	88.162%	88.785%	89.097%	90.654%	89.097%
Simple CART	84.112%	86.916%	91.277%	89.097%	94.081%	91.900%	92.835%	92.835%
RIMARC	83.801%	90.654%	92.523%	91.900%	91.900%	93.458%	94.393%	94.393%
Bayes NET	82.866%	89.097%	90.966%	90.654%	90.031%	91.900%	91.589%	93.146%
Dagging	83.489%	92.212%	93.146%	92.523%	95.016%	95.950%	94.393%	95.950%
Random SubSpace	80.685%	88.785%	91.589%	90.966%	90.343%	91.589%	91.589%	93.458%
Decision Table	67.601%	72.586%	79.128%	78.505%	83.801%	77.882%	80.685%	83.178%
Ridor	84.424%	92.523%	90.966%	89.097%	94.393%	92.212%	94.704%	92.835%
ADTree	83.489%	89.720%	91.589%	91.277%	94.081%	95.016%	95.327%	96.573%
LAD Tree	87.851%	87.851%	90.966%	91.277%	94.393%	93.770%	94.704%	93.146%
Random Tree	80.685%	88.474%	88.785%	89.408%	90.031%	90.654%	92.835%	93.770%
AVERAGE	82.495%	87.791%	90.447%	90.046%	92.642%	92.123%	92.598%	93.057%

Table 5

Individual and ensemble average classification accuracies of the fast shapelet transform that is applied to the AP dataset for variations in shapelet length, L , and the selected classifiers. The highest average classification accuracy is shown in boldface.

	$L = 50$	$L = 100$	$L = 150$	$L = 200$	$L = 250$	$L = 300$	$L = 350$	$L = 400$
Naive Bayes	80.062%	81.308%	85.358%	86.604%	88.474%	90.654%	92.523%	88.474%
J48	78.193%	82.555%	86.293%	87.851%	88.474%	90.031%	90.966%	87.227%
Random forest	72.274%	82.866%	79.751%	81.620%	82.866%	87.227%	90.031%	78.816%
AdaBoost.M1	77.882%	80.997%	86.916%	87.851%	88.785%	90.654%	90.031%	86.604%
Classif. via regr.	79.439%	79.751%	85.358%	85.981%	89.720%	90.031%	90.966%	88.474%
Bagging	78.193%	84.112%	86.293%	87.227%	89.097%	91.277%	89.720%	87.851%
MultiBoostAB	77.570%	82.555%	87.851%	88.474%	88.785%	90.654%	89.720%	86.916%
LWL	78.816%	80.997%	85.047%	87.851%	88.785%	90.031%	89.408%	87.227%
PART	77.259%	82.555%	85.981%	87.851%	88.474%	90.343%	91.589%	87.227%
END	78.193%	82.555%	86.293%	87.851%	88.474%	90.654%	90.966%	87.227%
Decision Stump	78.816%	78.505%	87.227%	87.851%	88.785%	90.654%	90.343%	86.293%
Simple CART	78.505%	81.620%	86.604%	84.424%	89.097%	90.031%	90.343%	87.227%
RIMARC	80.062%	85.981%	87.227%	88.162%	90.654%	91.589%	93.458%	89.720%
Bayes NET	76.324%	82.866%	86.916%	87.851%	88.785%	90.654%	89.408%	86.604%
Dagging	78.505%	78.816%	78.193%	86.604%	89.408%	90.031%	90.654%	86.293%
Random SubSpace	77.570%	71.651%	86.916%	87.851%	88.785%	90.654%	84.735%	86.604%
Decision Table	78.816%	81.932%	86.916%	87.851%	88.785%	90.654%	88.474%	86.604%
Ridor	77.570%	82.243%	83.489%	85.358%	87.227%	90.031%	88.162%	85.670%
ADTree	77.259%	82.866%	84.735%	87.227%	88.474%	91.277%	90.654%	87.539%
LAD Tree	75.389%	82.243%	85.047%	85.981%	87.227%	91.277%	90.654%	87.851%
Random Tree	71.963%	80.685%	78.505%	81.620%	82.555%	87.227%	89.097%	79.128%
AVERAGE	77.555%	81.412%	85.091%	86.664%	88.177%	90.269%	90.090%	86.456%

classification algorithms. The shapelet-based benchmark methods are favored by selecting their “best” L that yields the highest average classification accuracy for each dataset as tabulated in Table 14.

The resulting CD diagram is shown in Fig. 13. The horizontal axis in the diagram represents the average ranks of each feature extraction method. The average rank based on classification accuracy improves from left to right. The CD is marked above the axis and corresponds to an indicator in rank magnitude that is required for the compared methods to differ such that they are termed as “critically different”. The connecting blue lines in the diagram depict the groups of methods that are not critically different. With respect to the datasets of the present study, the performance of the proposed WTC method significantly exceeds that of the fast shapelet transform. The average rank of WTC exceeds that of the brute-force shapelet transform. However, the difference is not critically different.

Finally, the robustness aspect of WTC is demonstrated in a statistical sense. Robustness of WTC is attributed to the averaging step taken in

obtaining T^A . T^A is the class representative average time series from which the local time window length w is extracted. Making the computation of w robust against intra-class subject variability and noise is vital to the success of WTC.

In order to evaluate robustness of WTC, random subsets are generated to obtain the corresponding w for each class of each dataset presented in this study, following the very procedure described in Section 2.2.1. Size of the random subsets are chosen to be the same with those used in Section 3.1, which are 50 and 25 for AP and ECG datasets, respectively. A total of 100 random subsets are generated for each class of each dataset. As a measure of robustness, Coefficient of Variation (CoV) of the random variable w denoted by V_w is chosen. V_w is defined as follows:

$$V_w = \frac{\sigma_w}{\mu_w}, \quad (17)$$

where σ_w and μ_w are the standard deviation and mean of w , respectively. The resulting values of V_w are compared with those of a similar

Table 6

Individual and ensemble average classification accuracies of the proposed WTC method applied to lead v2 of the ECG dataset for different values of the energy threshold δ , confidence level p , and selected classifiers. The highest average classification accuracy is shown in boldface.

	$\delta = 0.9900$		$\delta = 0.9990$		$\delta = 0.9999$	
	$p = 0.95$	$p = 0.99$	$p = 0.95$	$p = 0.99$	$p = 0.95$	$p = 0.99$
Naive Bayes	95.000%	99.000%	96.000%	95.000%	96.000%	94.000%
J48	97.000%	97.000%	98.000%	98.000%	98.000%	98.000%
Random forest	100.000%	100.000%	100.000%	99.000%	100.000%	99.000%
AdaBoost.M1	98.000%	98.000%	98.000%	98.000%	100.000%	100.000%
Classif. via regr.	99.000%	99.000%	100.000%	100.000%	94.000%	93.000%
Bagging	100.000%	99.000%	97.000%	97.000%	97.000%	97.000%
MultiBoostAB	100.000%	99.000%	98.000%	98.000%	98.000%	98.000%
LWL	93.000%	95.000%	97.000%	96.000%	99.000%	99.000%
PART	97.000%	97.000%	98.000%	98.000%	98.000%	98.000%
END	97.000%	97.000%	98.000%	98.000%	98.000%	98.000%
Decision stump	93.000%	93.000%	93.000%	93.000%	93.000%	93.000%
Simple CART	99.000%	99.000%	100.000%	100.000%	100.000%	100.000%
RIMARC	100.000%	100.000%	100.000%	100.000%	100.000%	100.000%
Bayes NET	100.000%	100.000%	98.000%	98.000%	98.000%	98.000%
Dagging	91.000%	94.000%	94.000%	94.000%	96.000%	90.000%
Random SubSpace	96.000%	96.000%	99.000%	98.000%	98.000%	98.000%
Decision Table	98.000%	99.000%	98.000%	97.000%	99.000%	99.000%
Ridor	99.000%	99.000%	100.000%	100.000%	100.000%	100.000%
ADTree	99.000%	99.000%	100.000%	100.000%	100.000%	100.000%
LAD Tree	99.000%	99.000%	100.000%	100.000%	100.000%	100.000%
Random Tree	98.000%	96.000%	98.000%	86.000%	92.000%	95.000%
AVERAGE	97.524%	97.810%	98.095%	97.286%	97.810%	97.476%

Table 7

Individual and ensemble average classification accuracies of the WTC method applied to leads v3 and v4 of the ECG dataset for the selected classifiers. The highest average classification accuracy is shown in boldface.

	LeadV3	LeadV4
Naive Bayes	92.000%	96.000%
J48	98.000%	97.000%
Random forest	99.000%	99.000%
AdaBoost.M1	99.000%	97.000%
Classif. via regr.	99.000%	94.000%
Bagging	99.000%	95.000%
MultiBoostAB	99.000%	94.000%
LWL	99.000%	97.000%
PART	98.000%	97.000%
END	98.000%	97.000%
Decision stump	99.000%	94.000%
Simple CART	99.000%	95.000%
RIMARC	99.000%	99.000%
Bayes NET	99.000%	94.000%
Dagging	93.000%	94.000%
Random SubSpace	99.000%	93.000%
Decision Table	99.000%	97.000%
Ridor	100.000%	98.000%
ADTree	100.000%	99.000%
LAD Tree	100.000%	99.000%
Random Tree	100.000%	93.000%
AVERAGE	98.429%	96.095%

experiment in which w is extracted directly from the time series instances, omitting the averaging step. The number of values for w is equal to the number of instances. It is noted that, there exist a total 202 AF and 219 SR instances for the AP dataset and 75 instances each for MI and NR classes of the studied ECG dataset. As evident in Eq. (17), CoV normalizes standard deviation with mean, and thus is a suitable candidate for comparing standard deviation of two different random variables with different mean values. As seen from Table 15, the averaging step considerably decreases CoV of w which implies an improved level of robustness for WTC.

4. Discussion

In the study, a method termed as WTC is proposed to extract features from the signals in the form of time series. WTC is applied to a cardiac AP

dataset with labeled SR and AF patients and three precordial leads of an ECG dataset that consist of control subjects and patients diagnosed with acute MI. Extracted feature vectors from these datasets are then examined for their classification accuracies that yield favorable results. Subsequently, the brute-force and fast shapelet transforms are used to compare the performance of WTC in terms of predictive accuracy and computational complexity. Critical Difference (CD) analysis is performed on the datasets to reveal that WTC is “critically better” than fast shapelet transform. Although WTC and brute-force shapelet transforms are not “critically different”, the performance of the former is slightly better.

Analog signal acquisition and digitization involve thermal and quantization noise and biophysical signals that are not exceptions in this aspect. Noise caused by motion artifacts resulting from electrode, probe, sensor, equipment, and even from patient movements constitutes disturbances in the desired signal in addition to thermal noise. Furthermore, almost all biophysical signals exhibit individualized polymorphism among instances to a certain extent as mentioned by Ref. [63] for ECGs, and this may hamper interpretation. WTC constructs the class representative time series as an ensemble average of all instances sharing a common class label and devises a mean trajectory that the population statistically follows within a confidence bound to mitigate the noise originated effects and variability among individual instances. Hence, as opposed to directly extracting features from the instances themselves, WTC uses a class representative trajectory for this purpose. The study findings indicate that this approach yields favorable results for AP and ECG datasets, and the latter is recognized as a signal that is usually contaminated with high levels of noise [41–43].

The DCT representation of the time series in conjunction with the extraction of perceptually important points (PIPs) constitutes a central part of the proposed WTC method since it allows the determination of a proper window length for local features. Briefly, DCT is utilized as a robust heuristic to feed a cut-off percentage to select the PIPs from the class representative average time series. Additionally, DCT is a widely used digital processing technique [64] that serves the purpose of eliminating redundancy in the processed data. It is noted that WTC determines the time windows and also assigns scores to indicating an order based on their descriptive power.

The presented results from completely different datasets reveal that the values of the energy threshold (δ) and the confidence level (p) parameters that yield favorable results for the WTC method are almost the

Table 8

Individual and ensemble average classification accuracies of the brute-force shapelet transform applied to lead v2 of ECG dataset for the variations in the shapelet length, L , and the selected classifiers. The highest average classification accuracy is shown in boldface.

	$L = 50$	$L = 100$	$L = 150$	$L = 200$	$L = 250$	$L = 300$	$L = 350$	$L = 400$
Naive Bayes	92.000%	95.000%	94.000%	90.000%	92.000%	90.000%	90.000%	90.000%
J48	91.000%	93.000%	94.000%	93.000%	93.000%	96.000%	93.000%	94.000%
Random forest	92.000%	96.000%	97.000%	97.000%	98.000%	99.000%	98.000%	97.000%
AdaBoost.M1	91.000%	96.000%	97.000%	96.000%	98.000%	97.000%	95.000%	97.000%
Classif. via regr.	92.000%	95.000%	95.000%	91.000%	94.000%	95.000%	93.000%	94.000%
Bagging	92.000%	93.000%	91.000%	93.000%	94.000%	93.000%	93.000%	92.000%
MultiBoostAB	92.000%	93.000%	91.000%	93.000%	93.000%	93.000%	93.000%	93.000%
LWL	87.000%	94.000%	98.000%	95.000%	96.000%	90.000%	95.000%	94.000%
PART	92.000%	93.000%	94.000%	93.000%	94.000%	96.000%	93.000%	94.000%
END	91.000%	93.000%	94.000%	93.000%	93.000%	96.000%	93.000%	94.000%
Decision stump	85.000%	93.000%	95.000%	90.000%	94.000%	93.000%	92.000%	93.000%
Simple CART	90.000%	93.000%	95.000%	91.000%	96.000%	96.000%	92.000%	95.000%
RIMARC	95.000%	94.000%	94.000%	93.000%	94.000%	93.000%	93.000%	93.000%
Bayes NET	92.000%	93.000%	92.000%	93.000%	93.000%	93.000%	93.000%	93.000%
Dagging	91.000%	93.000%	93.000%	93.000%	93.000%	92.000%	89.000%	92.000%
Random SubSpace	92.000%	93.000%	93.000%	94.000%	93.000%	93.000%	92.000%	92.000%
Decision Table	83.000%	88.000%	84.000%	86.000%	86.000%	87.000%	87.000%	85.000%
Ridor	86.000%	93.000%	95.000%	92.000%	91.000%	96.000%	96.000%	95.000%
ADTree	91.000%	96.000%	97.000%	98.000%	98.000%	97.000%	98.000%	97.000%
LAD Tree	90.000%	97.000%	97.000%	96.000%	97.000%	98.000%	97.000%	98.000%
Random Tree	87.000%	90.000%	91.000%	94.000%	93.000%	99.000%	98.000%	96.000%
AVERAGE	90.190%	93.524%	93.857%	93.048%	93.952%	94.381%	93.476%	93.714%

Table 9

Individual and ensemble average classification accuracies of the fast shapelet transform applied to lead v2 in the ECG dataset for variations in the fast shapelet lengths, L , and for the selected classifiers. The highest average classification accuracy is shown in boldface.

	$L = 50$	$L = 100$	$L = 150$	$L = 200$	$L = 250$	$L = 300$	$L = 350$	$L = 400$
Naive Bayes	81.000%	90.000%	96.000%	91.000%	90.000%	86.000%	86.000%	93.000%
J48	82.000%	88.000%	95.000%	92.000%	92.000%	92.000%	92.000%	93.000%
Random forest	80.000%	85.000%	98.000%	92.000%	92.000%	93.000%	88.000%	87.000%
AdaBoost.M1	85.000%	89.000%	98.000%	91.000%	92.000%	92.000%	92.000%	93.000%
Classif. via regr.	81.000%	90.000%	92.000%	93.000%	93.000%	93.000%	93.000%	93.000%
Bagging	85.000%	90.000%	90.000%	93.000%	93.000%	93.000%	93.000%	94.000%
MultiBoostAB	85.000%	91.000%	93.000%	93.000%	93.000%	93.000%	93.000%	94.000%
LWL	80.000%	90.000%	98.000%	93.000%	93.000%	93.000%	92.000%	94.000%
PART	80.000%	88.000%	95.000%	92.000%	92.000%	92.000%	92.000%	93.000%
END	82.000%	88.000%	95.000%	92.000%	92.000%	92.000%	92.000%	93.000%
Decision stump	85.000%	90.000%	98.000%	93.000%	93.000%	93.000%	92.000%	94.000%
Simple CART	82.000%	90.000%	98.000%	91.000%	93.000%	94.000%	88.000%	94.000%
RIMARC	85.000%	92.000%	99.000%	93.000%	93.000%	93.000%	93.000%	94.000%
Bayes NET	85.000%	91.000%	91.000%	93.000%	93.000%	93.000%	93.000%	94.000%
Dagging	80.000%	88.000%	93.000%	93.000%	86.000%	83.000%	89.000%	91.000%
Random SubSpace	76.000%	87.000%	84.000%	84.000%	81.000%	93.000%	75.000%	84.000%
Decision Table	85.000%	90.000%	98.000%	93.000%	93.000%	93.000%	93.000%	94.000%
Ridor	83.000%	90.000%	98.000%	91.000%	90.000%	92.000%	91.000%	89.000%
ADTree	84.000%	85.000%	96.000%	92.000%	93.000%	95.000%	89.000%	88.000%
LAD Tree	81.000%	85.000%	97.000%	92.000%	93.000%	94.000%	90.000%	89.000%
Random Tree	83.000%	83.000%	96.000%	92.000%	92.000%	94.000%	88.000%	87.000%
AVERAGE	82.381%	88.571%	95.143%	91.857%	91.571%	92.190%	90.190%	91.667%

same, thereby demonstrating its consistency. In contrast to the length parameter (L) of the shapelet-based methods, the parameters of the WTC method are unitless and relatively less sensitive to the dataset of interest. Additionally, the accuracy cost of deviation from the values is relatively less when compared to those involved in the benchmark methods. The results of the study indicate that it is possible for prospective users of the WTC method to use the recommended values of the parameters ($\delta = 0.9990$ and $p = 0.95$) for different datasets without engaging in the parameter exploration step.

Execution time is an important criterion for online applications and especially for those involving big datasets. The $O(KN + N^2)$ complexity of the proposed WTC method outperforms the benchmark methods of brute-force, [16] and fast shapelet, [32] transforms in terms of computational complexity. In addition to the time-consuming feature extraction phase of shapelet-based transforms, [16], the discovered shapelets may not belong to the time windows associated with the expected morphologies of the class label of interest. Unconstrained selection of

subsequences based solely on their minimum distance to other subsequences from arbitrary time intervals may hinder the accurate assessment of the underlying dynamics. Moreover, the temporal location of a subsequence is typically associated with a specific importance for clinical interpretation that may be disregarded by the very nature of time series shapelets. Additionally, the necessity of providing the shapelet length and total number of shapelets requires preliminary knowledge as well as forming assumptions about the dataset of interest.

The study indicates that time windows of WTC that coincide with phases 2 and 3 of the AP signal (Fig. 2a) are most predictive for both classes. The plateau of AP (phase 2) is governed by the balance between calcium influx through L-type Ca^{2+} channels and potassium efflux via a plethora of K^+ channels. Conversely, late repolarization (phase 3) is dominated by K^+ current (I_{K1}) through inward rectifier channels (when compared to Fig. 2b). The more positive potentials during early repolarization and the rapid final repolarization in AF correspond to reduced transient outward current and enhanced I_{K1} , respectively as induced by

Table 10

Individual and ensemble average classification accuracies of the brute-force shapelet transform applied to lead v3 in ECG dataset for variations in the shapelet length, L , and the selected classifiers. The highest average classification accuracy is shown in boldface.

	$L = 50$	$L = 100$	$L = 150$	$L = 200$	$L = 250$	$L = 300$	$L = 350$	$L = 400$
Naive Bayes	95.000%	95.000%	96.000%	95.000%	96.000%	100.000%	100.000%	100.000%
J48	95.000%	90.000%	94.000%	99.000%	98.000%	97.000%	99.000%	99.000%
Random forest	96.000%	94.000%	97.000%	99.000%	100.000%	100.000%	100.000%	100.000%
AdaBoost.M1	97.000%	95.000%	96.000%	100.000%	98.000%	98.000%	98.000%	98.000%
Classif. via regr.	97.000%	93.000%	93.000%	100.000%	98.000%	98.000%	98.000%	98.000%
Bagging	95.000%	95.000%	94.000%	96.000%	96.000%	100.000%	100.000%	100.000%
MultiBoostAB	95.000%	95.000%	94.000%	96.000%	97.000%	100.000%	100.000%	100.000%
LWL	93.000%	90.000%	93.000%	100.000%	97.000%	100.000%	99.000%	99.000%
PART	95.000%	90.000%	94.000%	99.000%	98.000%	97.000%	99.000%	99.000%
END	95.000%	90.000%	94.000%	99.000%	98.000%	97.000%	99.000%	99.000%
Decision stump	96.000%	92.000%	93.000%	100.000%	98.000%	98.000%	98.000%	98.000%
Simple CART	96.000%	92.000%	93.000%	100.000%	98.000%	98.000%	98.000%	98.000%
RIMARC	98.000%	96.000%	96.000%	96.000%	97.000%	100.000%	100.000%	100.000%
Bayes NET	95.000%	95.000%	94.000%	96.000%	97.000%	100.000%	100.000%	100.000%
Dagging	90.000%	91.000%	84.000%	90.000%	90.000%	92.000%	89.000%	90.000%
Random SubSpace	95.000%	95.000%	96.000%	96.000%	97.000%	100.000%	100.000%	100.000%
Decision Table	85.000%	77.000%	85.000%	79.000%	71.000%	79.000%	83.000%	85.000%
Ridor	97.000%	91.000%	94.000%	99.000%	97.000%	98.000%	98.000%	98.000%
ADTree	97.000%	95.000%	94.000%	100.000%	98.000%	98.000%	98.000%	98.000%
LAD Tree	96.000%	92.000%	94.000%	100.000%	98.000%	98.000%	98.000%	98.000%
Random Tree	91.000%	91.000%	90.000%	94.000%	95.000%	99.000%	99.000%	98.000%
AVERAGE	94.714%	92.095%	93.238%	96.809%	95.809%	97.476%	97.762%	97.857%

Table 11

Individual and ensemble average classification accuracies of the fast shapelet transform applied to lead v3 in ECG dataset for variations in the fast shapelet length, L , and the selected classifiers. The highest average classification accuracy is shown in boldface.

	$L = 50$	$L = 100$	$L = 150$	$L = 200$	$L = 250$	$L = 300$	$L = 350$	$L = 400$
Naive Bayes	95.000%	92.000%	84.000%	81.000%	69.000%	81.000%	74.000%	76.000%
J48	94.000%	89.000%	82.000%	87.000%	77.000%	74.000%	74.000%	77.000%
Random forest	94.000%	87.000%	82.000%	91.000%	73.000%	68.000%	67.000%	72.000%
AdaBoost.M1	96.000%	91.000%	86.000%	87.000%	75.000%	80.000%	71.000%	78.000%
Classif. via regr.	95.000%	88.000%	81.000%	88.000%	76.000%	65.000%	73.000%	76.000%
Bagging	96.000%	88.000%	83.000%	87.000%	69.000%	78.000%	69.000%	71.000%
MultiBoostAB	97.000%	89.000%	82.000%	88.000%	73.000%	71.000%	70.000%	74.000%
LWL	97.000%	89.000%	82.000%	88.000%	72.000%	70.000%	74.000%	73.000%
PART	94.000%	89.000%	83.000%	86.000%	72.000%	74.000%	73.000%	77.000%
END	94.000%	89.000%	82.000%	87.000%	77.000%	74.000%	74.000%	77.000%
Decision stump	97.000%	89.000%	84.000%	88.000%	66.000%	69.000%	73.000%	76.000%
Simple CART	96.000%	91.000%	83.000%	88.000%	69.000%	76.000%	75.000%	76.000%
RIMARC	98.000%	94.000%	85.000%	93.000%	87.000%	86.000%	81.000%	81.000%
Bayes NET	97.000%	89.000%	83.000%	88.000%	71.000%	71.000%	73.000%	77.000%
Dagging	74.000%	73.000%	67.000%	73.000%	62.000%	60.000%	59.000%	67.000%
Random SubSpace	97.000%	71.000%	73.000%	86.000%	65.000%	71.000%	72.000%	73.000%
Decision Table	97.000%	89.000%	83.000%	88.000%	75.000%	80.000%	73.000%	78.000%
Ridor	94.000%	88.000%	83.000%	83.000%	71.000%	76.000%	72.000%	72.000%
ADTree	96.000%	89.000%	84.000%	88.000%	80.000%	75.000%	74.000%	71.000%
LAD Tree	96.000%	90.000%	83.000%	88.000%	71.000%	79.000%	76.000%	72.000%
Random Tree	94.000%	89.000%	85.000%	87.000%	76.000%	75.000%	70.000%	70.000%
AVERAGE	94.667%	87.762%	81.905%	86.667%	72.667%	73.952%	72.238%	74.476%

remodeling in chronic AF [17]. The high predictive power of phases 2 and 3 as shown in the present study (with reference to color maps in Fig. 6a and b) is in good agreement with a previous study on human atrial APs that introduced the analysis parameter “plateau potential” (PLT₂₀) that is defined as the average membrane potential of a time window between 20% and 30% of the AP duration at 90% of repolarization (APD₉₀) [39]. Specifically, PLT₂₀ (mV) and APD₉₀ (ms) were ranked 5th and 1st, respectively, out of 62 features in classifying an instance as positive for AF [34]. The ability to classify an AP as AF or SR allows the formation of conclusions with respect to the underlying changes in ion channel properties and is of significance to predict antiarrhythmic drug action [65]. The findings on AP dataset do not point to a clinical outcome, but they hold an experimental relevance. Thus, it provides good real-life biophysical data to test the discriminative power of the WTC method aided by the preserved locality of the extracted features.

In addition to the AP dataset, the analysis of the three chest leads of an ECG dataset is suitable for demonstrating the usability of the proposed

method for clinical applications. With respect to the MI that corresponds to the condition of interest, certain leads in ECG recordings are expected to undergo predictable morphological and temporal changes including an elevation of the ST-segment (during STEMI) [44] and is usually preceded by subtle changes in the T-wave shape and size [66] and succeeded by a loss of the R-wave amplitude [21]. The distribution of similarity scores across the time windows as shown in Fig. 10a, c, and e is consistent with the aforementioned morphologies that are expected to be observed in anterior chest leads of v2, v3, and v4, respectively. In contrast, the resulting shapelets of the brute-force shapelet transform for lead v2 of MI time series instances (Fig. 12) do not particularly highlight any clinically relevant time windows. In a manner similar to the AP dataset, the resulting classification accuracies of the brute-force shapelet transform for the ECG dataset (Tables 8, 10 and 12) show considerable variation with respect to the input parameter of shapelet length.

It is conjectured that WTC has potential clinical benefits including its integration with emergency, ambulatory, inpatient, and home care

Table 12

Individual and ensemble average classification accuracies of the brute-force shapelet transform applied to lead v4 in ECG dataset for variations in the shapelet length, L , and the selected classifiers. The highest average classification accuracy is shown in boldface.

	$L = 50$	$L = 100$	$L = 150$	$L = 200$	$L = 250$	$L = 300$	$L = 350$	$L = 400$
Naive Bayes	84.000%	74.000%	70.000%	80.000%	77.000%	89.000%	95.000%	95.000%
J48	97.000%	92.000%	88.000%	87.000%	94.000%	96.000%	99.000%	96.000%
Random forest	89.000%	92.000%	88.000%	90.000%	96.000%	99.000%	99.000%	99.000%
AdaBoost.M1	96.000%	94.000%	82.000%	84.000%	87.000%	99.000%	99.000%	97.000%
Classif. via regr.	94.000%	93.000%	83.000%	85.000%	88.000%	93.000%	99.000%	96.000%
Bagging	73.000%	72.000%	81.000%	79.000%	82.000%	84.000%	90.000%	90.000%
MultiBoostAB	89.000%	74.000%	78.000%	78.000%	80.000%	86.000%	91.000%	91.000%
LWL	98.000%	91.000%	85.000%	84.000%	90.000%	94.000%	100.000%	95.000%
PART	97.000%	92.000%	87.000%	89.000%	92.000%	96.000%	99.000%	96.000%
END	97.000%	92.000%	88.000%	87.000%	94.000%	96.000%	99.000%	96.000%
Decision stump	98.000%	91.000%	83.000%	84.000%	89.000%	93.000%	99.000%	96.000%
Simple CART	98.000%	94.000%	83.000%	84.000%	89.000%	97.000%	99.000%	96.000%
RIMARC	91.000%	94.000%	86.000%	84.000%	85.000%	87.000%	95.000%	95.000%
Bayes NET	78.000%	74.000%	80.000%	78.000%	81.000%	86.000%	92.000%	91.000%
Dagging	72.000%	74.000%	76.000%	81.000%	80.000%	83.000%	85.000%	86.000%
Random SubSpace	77.000%	74.000%	79.000%	77.000%	81.000%	85.000%	89.000%	91.000%
Decision Table	91.000%	67.000%	72.000%	72.000%	74.000%	76.000%	83.000%	81.000%
Ridor	96.000%	93.000%	84.000%	88.000%	89.000%	92.000%	99.000%	94.000%
ADTree	96.000%	92.000%	88.000%	90.000%	93.000%	97.000%	99.000%	97.000%
LAD Tree	95.000%	91.000%	86.000%	87.000%	95.000%	97.000%	99.000%	98.000%
Random Tree	79.000%	71.000%	73.000%	77.000%	70.000%	91.000%	90.000%	95.000%
AVERAGE	89.762%	84.810%	81.905%	83.095%	86.000%	91.238%	95.190%	93.857%

Table 13

Individual and ensemble average classification accuracies of the fast shapelet transform applied to lead v4 in ECG dataset for variations in the fast shapelet length, L , and the selected classifiers. The highest average classification accuracy is shown in boldface.

	$L = 50$	$L = 100$	$L = 150$	$L = 200$	$L = 250$	$L = 300$	$L = 350$	$L = 400$
Naive Bayes	58.000%	71.000%	88.000%	81.000%	85.000%	77.000%	77.000%	77.000%
J48	58.000%	75.000%	90.000%	79.000%	84.000%	74.000%	81.000%	82.000%
Random forest	59.000%	70.000%	91.000%	71.000%	81.000%	65.000%	64.000%	70.000%
AdaBoost.M1	64.000%	70.000%	87.000%	81.000%	86.000%	74.000%	77.000%	83.000%
Classif. via regr.	57.000%	72.000%	90.000%	81.000%	81.000%	75.000%	80.000%	83.000%
Bagging	53.000%	70.000%	88.000%	76.000%	83.000%	70.000%	79.000%	83.000%
MultiBoostAB	49.000%	70.000%	88.000%	77.000%	82.000%	74.000%	80.000%	83.000%
LWL	60.000%	74.000%	90.000%	76.000%	87.000%	73.000%	80.000%	83.000%
PART	58.000%	75.000%	90.000%	78.000%	83.000%	74.000%	81.000%	82.000%
END	58.000%	75.000%	90.000%	79.000%	84.000%	74.000%	81.000%	82.000%
Decision stump	55.000%	74.000%	90.000%	78.000%	87.000%	75.000%	80.000%	83.000%
Simple CART	61.000%	74.000%	88.000%	79.000%	82.000%	72.000%	80.000%	83.000%
RIMARC	77.000%	78.000%	97.000%	86.000%	89.000%	80.000%	83.000%	85.000%
Bayes NET	49.000%	68.000%	87.000%	78.000%	84.000%	75.000%	80.000%	83.000%
Dagging	57.000%	68.000%	82.000%	80.000%	80.000%	74.000%	77.000%	75.000%
Random SubSpace	49.000%	67.000%	78.000%	74.000%	76.000%	75.000%	80.000%	83.000%
Decision Table	58.000%	74.000%	87.000%	79.000%	87.000%	75.000%	80.000%	83.000%
Ridor	58.000%	71.000%	81.000%	77.000%	87.000%	72.000%	71.000%	81.000%
ADTree	68.000%	73.000%	93.000%	85.000%	77.000%	71.000%	77.000%	82.000%
LAD Tree	64.000%	73.000%	93.000%	82.000%	82.000%	72.000%	75.000%	81.000%
Random Tree	50.000%	70.000%	86.000%	70.000%	69.000%	72.000%	70.000%	67.000%
AVERAGE	58.095%	72.000%	88.286%	78.429%	82.667%	73.476%	77.762%	80.667%

Table 14

Values of the shapelet length, L , which are used for the critical difference diagram for the brute-force and fast shapelet transforms.

	Brute-force shapelet	Fast shapelet
AP	400	300
ECG lead v2	300	150
ECG lead v3	400	50
ECG lead v4	350	150

expert systems that commonly suffer from the misinterpretation of different symptoms [67]. Cases demanding urgent decision making, such as MI, require a digital assessment of the available measurements and decision support for triage. Early diagnosis and determining the severity of MI is vital for patients, [68]. Prior to in-hospital interventions, it is advantageous to perform a preliminary analysis of the prehospital ECGs through mobile expert systems to alert patients, caregivers, or primary

healthcare professionals [45]. Recently, ECG devices are equipped with various diagnostic detection mechanisms such as ST-segment monitoring to detect silent ischemia. The importance of automatic detection of acute MI and its place in clinical decision-making are discussed in Refs. [69] and [70] along with different types of implementations. Specifically for the case of ischemia, it is possible for the course of the disease to change rapidly and therefore a single randomly performed ECG may be insufficient to describe its extent. In order to form a reasonable plan for treatment, extant studies such as [71] indicate that a continuum of ECGs should be assessed, so that unnecessary cath-lab activation is avoided [72]. All these studies emphasize the need for fast and lightweight algorithms that produce accurate results.

In addition to pointing out the incidence of a disorder, WTC is also suitable for distinguishing the various types of morphologies. For example, it is necessary to differentiate tall and symmetrical T waves that are observed during hyperkalemia from broad and rather skewed hyperacute T waves that typically occur in the early stages (within the first

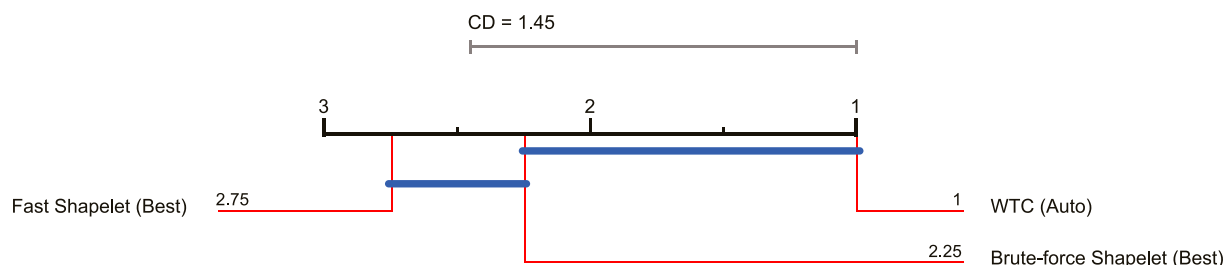


Fig. 13. Critical difference diagram based on the average classification accuracies achieved by WTC, brute-force and fast shapelet transforms.

Table 15

CoV values of w , V_w , with and without averaging for each dataset.

Dataset	Class Label	with Averaging	without Averaging
AP	AF	0.187	1.019
	SR	0.120	0.486
ECG lead v2	MI	0.186	0.378
	NR	0.153	0.381
ECG lead v3	MI	0.121	0.380
	NR	0.168	0.457
ECG lead v4	MI	0.164	0.433
	NR	0.143	0.460

30 min) of STEMI [73]. Furthermore, WTC could contribute to the education of medical professionals in the course of their training. The “human readability” aspect of the algorithm allows practitioners to recognize patterns more easily and study underlying dynamics and possible “cause-and-effect” relations. Moreover, its modest computational complexity is such that it can be integrated with any kind of mobile or centralized online framework as well as stand-alone offline units that require the rapid examination of densely sampled time series data.

5. Conclusion

The findings of this study indicate that WTC emerges a robust, fast, interpretable, and accurate time series feature extraction method based on the results presented for the atrial AP and ECG time series datasets. In addition to achieving relatively higher classification accuracies and lower execution times, the proposed method highlights the underlying dynamics that are typical of the inspected class label. WTC yields results in accordance with extant studies concerning atrial AP and ECG. In conclusion, WTC emerges as a promising analysis tool with an unbiased parameter determination for biomedical researchers to conduct objective data mining analysis. Future studies will focus on the application of WTC to other biophysical time series such as heart auscultation, electromyogram and electroneurogram recordings, event-related brain potentials, and regular measurements of blood pressure and blood sugar (for detecting hypo/hyper-glycemic periods).

Conflict of interest statement

None declared.

Acknowledgements

The original research leading to the results for the cardiac action potential dataset received funding from the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreement no HEALTH-F2-2010-261057.

References

- [1] B.Y. Reis, K.D. Mandl, Time series modeling for syndromic surveillance, *BMC Med. Inf. Decis. Mak.* 3 (1) (2003) 2.

- [2] C.E. Kennedy, J.P. Turley, Time series analysis as input for clinical predictive modeling: modeling cardiac arrest in a pediatric ICU, *Theor. Biol. Med. Model.* 8 (1) (2011) 40.
- [3] N. Channouf, P. Lécuyer, A. Ingolfsson, A.N. Avramidis, The application of forecasting techniques to modeling emergency medical system calls in Calgary, Alberta, *Health Care Manag. Sci.* 10 (1) (2007) 25–45.
- [4] S.K. Sahoo, W. Lu, S.D. Teddy, D. Kim, M. Feng, Detection of atrial fibrillation from non-episodic ECG data: a review of methods, in: 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2011, pp. 4992–4995.
- [5] I.N. Soyiri, D.D. Reidpath, An overview of health forecasting, *Environ. Health Prev. Med.* 18 (1) (2013) 1–9.
- [6] F.S. Cohen, Z. Huang, Z. Yang, Curve recognition using B-spline representation, in: *Applications of Computer Vision*, IEEE, 1992, pp. 213–220. *Proceedings, 1992., IEEE Workshop on.*
- [7] Y. Wang, E.K. Teoh, A novel 2D shape matching algorithm based on B-spline modeling, in: *Image Processing, IEEE, 2004*, pp. 409–412. *ICIP'04. 2004 International Conference on*, vol. 1.
- [8] J.M. Rogers, Modeling the cardiac action potential using B-spline surfaces, *Biomed. Eng., IEEE Trans.* 47 (6) (2000) 784–791.
- [9] T. Toe, T. Van To, Curve matching by using B-spline curves, *AU J. Technol.* 5(3).
- [10] F. Badilini, A.J. Moss, E.L. Titlebaum, Cubic spline baseline estimation in ambulatory ECG recordings for the measurement of ST segment displacements, in: *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 13, IEEE, 1991, pp. 584–585.
- [11] E. Masry, Multivariate local polynomial regression for time series: uniform strong consistency and rates, *J. Time Ser. Anal.* 17 (6) (1996) 571–599.
- [12] S. Manganaris, Supervised Classification with Temporal Data, Vanderbilt University, 1997.
- [13] Z. Liu, L. Wu, M. Hauskrecht, Modeling clinical time series using Gaussian process sequences, in: *SIAM International Conference on Data Mining*, SIAM, 2013, pp. 623–631.
- [14] S. Roberts, M. Osborne, M. Ebdon, S. Reece, N. Gibson, S. Aigrain, Gaussian processes for time-series modelling, *Philos. Trans. R. Soc. Lond. A Math. Phys. Eng. Sci.* 371 (1984) (2013) 20110550.
- [15] J. Hills, J. Lines, E. Baranauskas, J. Mapp, A. Bagnall, Classification of time series by shapelet transformation, *Data Min. Knowl. Discov.* 28 (4) (2014) 851–881.
- [16] L. Ye, E. Keogh, Time series shapelets: a novel technique that allows accurate, interpretable and fast classification, *Data Min. Knowl. Discov.* 22 (1–2) (2011) 149–182.
- [17] D. Dobrev, U. Ravens, Remodeling of cardiomyocyte ion channels in human atrial fibrillation, *Basic Res. Cardiol.* 98 (3) (2003) 137–148.
- [18] H. Dai, S. Jiang, Y. Li, Atrial activity extraction from single lead ECG recordings: evaluation of two novel methods, *Comput. Biol. Med.* 43 (3) (2013) 176–183.
- [19] Gabriel Steg, Stefan K. James, Dan Atar, Luigi P. Badano, Carina Blomstrom-Lundqvist, Michael A. Borger, Carlo Di Mario, Kenneth Dickstein, Gregory Ducrocq, Francisco Fernandez-Aviles, et al., ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation, *Eur. Heart J.* 33 (20) (2012) 2569–2619.
- [20] P.T. O'Gara, F.G. Kushner, D.D. Ascheim, D.E. Casey, M.K. Chung, J.A. De Lemos, S.M. Ettinger, J.C. Fang, F.M. Fesmire, B.A. Franklin, et al., 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: a report of the American College of Cardiology Foundation/American heart association task force on practice guidelines, *J. Am. Coll. Cardiol.* 61 (4) (2013) e78–e140.
- [21] J.V. Nable, W. Brady, The evolution of electrocardiographic changes in ST-segment elevation myocardial infarction, *Am. J. Emerg. Med.* 27 (6) (2009) 734–746.
- [22] A.L. Goldberger, L.A. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet components of a new research resource for complex physiologic signals, *Circulation* 101 (23) (2000) e215–e220. <https://physionet.org/>.
- [23] R. Boussefjot, D. Kreiseler, A. Schnabel, Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet, *Biomed. Technik/Biomedical Eng.* 40 (s1) (1995) 317–318. <https://physionet.org/physiobank/database/ptbdb/>.
- [24] S. Soheily-Khah, A. Douzal-Chouakria, E. Gaussier, Progressive and iterative approaches for time series averaging, *Proc. AALTD 2015* (2015) 123.
- [25] T.-c. Fu, F.-l. Chung, K.-y. Kwok, C.-m. Ng, Stock time series visualization based on data point importance, *Eng. Appl. Artif. Intell.* 21 (8) (2008) 1217–1232.
- [26] F.-l. Chung, T.-c. Fu, R. Luk, V. Ng, Flexible time series pattern matching based on perceptually important points, in: *International Joint Conference on Artificial*

- Intelligence IJCAI-01 Workshop on Learning from Temporal and Spatial Data, 2001, pp. 1–7.
- [27] G.-Z. Yu, H. Peng, Q.-L. Zheng, Pattern distance of time series based on segmentation by important points, in: *Machine Learning and Cybernetics*, IEEE, 2005, pp. 1563–1567. *Proceedings of 2005 International Conference on*, vol. 3.
- [28] J. Lin, E. Keogh, L. Wei, S. Lonardi, Experiencing SAX: a novel symbolic representation of time series, *Data Min. Knowl. Discov.* 15 (2) (2007) 107–144.
- [29] S.-H. Cha, S.N. Srihari, On measuring the distance between histograms, *Pattern Recognit.* 35 (6) (2002) 1355–1370.
- [30] GitHub Repository [online], WTC Source Codes, 2017.
- [31] M. Narasimha, A. Peterson, On the computation of the discrete cosine transform, *IEEE Trans. Commun.* 26 (6) (1978) 934–936.
- [32] T. Rakthanmanon, E. Keogh, Fast shapelets: a scalable algorithm for discovering time series shapelets, in: *Proceedings of the 13th SIAM Conference on Data Mining (SDM)*, SIAM, 2013, pp. 668–676.
- [33] A. Mueen, E. Keogh, N. Young, Logical-shapelets: an expressive primitive for time series classification, in: *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2011, pp. 1154–1162.
- [34] U. Ravens, D. Katircioglu-Öztürk, E. Wettwer, et al., Application of the RIMARC algorithm to a large data set of action potentials and clinical parameters for risk prediction of atrial fibrillation, *Med. Biol. Eng. Comput.* 53 (3) (2015) 263–273. [online], European Network for Translational Research in Atrial Fibrillation (EUTRAF), 2015.
- [35] D.M. Bers, Cardiac excitation–contraction coupling, *Nature* 415 (6868) (2002) 198–205.
- [36] U. Schotten, S. Verheule, P. Kirchhof, A. Goette, Pathophysiological mechanisms of atrial fibrillation: a translational appraisal, *Physiol. Rev.* 91 (1) (2011) 265–325.
- [37] C. Sánchez, A. Bueno-Orovio, E. Wettwer, S. Loose, J. Simon, U. Ravens, E. Pueyo, B. Rodriguez, Inter-subject variability in human atrial action potential in sinus rhythm versus chronic atrial fibrillation, *PLoS ONE* 9 (8) (2013) e105897.
- [38] E. Wettwer, O. Hála, T. Christ, J.F. Heubach, D. Dobrev, M. Knaut, A. Varró, U. Ravens, Role of IKur in controlling action potential shape and contractility in the human atrium influence of chronic atrial fibrillation, *Circulation* 110 (16) (2004) 2299–2306.
- [39] F. Morris, W.J. Brady, ABC of clinical electrocardiography: acute myocardial infarction, Part I, *BMJ* 324 (7341) (2002) 831–834.
- [40] J.R. Mou, S.M.R. Islam, X. Huang, K.L. Ou, Noise removal and QRS detection of ECG signal, *J. Biomed. Eng. Med. Imaging* 3 (4) (2016) 01.
- [41] S. Pongponsoi, X.-H. Yu, An adaptive filtering approach for electrocardiogram (ECG) signal noise reduction using neural networks, *Neurocomputing* 117 (2013) 206–213.
- [42] U. Satija, B. Ramkumar, M.S. Manikandan, Low-complexity detection and classification of ECG noises for automated ECG analysis system, in: *International Conference on Signal Processing and Communication*, IEEE, 2016, pp. 1–5.
- [43] K. Thygesen, J.S. Alpert, A.S. Jaffe, M.L. Simoons, B.R. Chaitman, H.D. White, Third universal definition of myocardial infarction, *Circulation* 126 (16) (2012) 2020–2035.
- [44] M.P. Somers, W.J. Brady, A.D. Perron, A. Mattu, The prominent T wave: electrocardiographic differential diagnosis, *Am. J. Emerg. Med.* 20 (3) (2002) 243–251.
- [45] [online], Weka 3: Data Mining Software in Java, 2015.
- [46] L. Breiman, Random forests, *Mach. Learn.* 45 (1) (2001) 5–32.
- [47] Y. Freund, R.E. Schapire, et al., Experiments with a new boosting algorithm, in: *ICML*, vol. 96, 1996, pp. 148–156.
- [48] E. Frank, Y. Wang, S. Inglis, G. Holmes, I.H. Witten, Using model trees for classification, *Mach. Learn.* 32 (1) (1998) 63–76.
- [49] L. Breiman, Bagging predictors, *Mach. Learn.* 24 (2) (1996) 123–140.
- [50] G.I. Webb, Multiboosting: a technique for combining boosting and wagging, *Mach. Learn.* 40 (2) (2000).
- [51] C.G. Atkeson, A.W. Moore, S. Schaal, Locally weighted learning for control, in: *Lazy Learning*, Springer, 1997, pp. 75–113.
- [52] E. Frank, I.H. Witten, Generating Accurate Rule Sets without Global Optimization, 1998.
- [53] E. Frank, S. Kramer, Ensembles of nested dichotomies for multi-class problems, in: *Proceedings of the 21st International Conference on Machine Learning*, ACM, 2004, p. 39.
- [54] W. Iba, P. Langley, Induction of one-level decision trees, in: *Proceedings of the 9th International Conference on Machine Learning*, 1992, pp. 233–240.
- [55] L. Breiman, J. Friedman, C.J. Stone, R.A. Olshen, *Classification and Regression Trees*, CRC press, 1984.
- [56] H.A. Güvenir, M. Kurtcephe, Ranking instances by maximizing the area under ROC curve, *IEEE Trans. Knowl. Data Eng.* 25 (10) (2013) 2356–2366.
- [57] K.M. Ting, I.H. Witten, Stacking bagged and dagged models, in: *14th International Conference on Machine Learning*, 1997.
- [58] T.K. Ho, The random subspace method for constructing decision forests, *IEEE Trans. Pattern Anal. Mach. Intell.* 20 (8) (1998) 832–844.
- [59] P. Compton, R. Jansen, Knowledge in context: a strategy for expert system maintenance, in: *Australian Joint Conference on Artificial Intelligence*, Springer, 1988, pp. 292–306.
- [60] Y. Freund, L. Mason, The alternating decision tree learning algorithm, in: *16th International Conference on Machine Learning (ICML)*, vol. 99, 1999, pp. 124–133.
- [61] J. Demšar, Statistical comparisons of classifiers over multiple data sets, *J. Mach. Learn. Res.* 7 (Jan 2006) 1–30.
- [62] B. Liu, J. Liu, G. Wang, K. Huang, F. Li, Y. Zheng, Y. Luo, F. Zhou, A novel electrocardiogram parameterization algorithm and its application in myocardial infarction detection, *Comput. Biol. Med.* 61 (2015) 178–184.
- [63] K.R. Rao, P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications*, Academic Press, 2014.
- [64] U. Ravens, E. Wettwer, O. Hála, Pharmacological modulation of ion channels and transporters, *Cell Calcium* 35 (6) (2004) 575–582.
- [65] W. Dressler, R. Hugo, High T waves in the earliest stage of myocardial infarction, *Am. Heart J.* 34 (5) (1947) 627–645.
- [66] K. Kroenke, A.D. Mangelsdorff, Common symptoms in ambulatory care: incidence, evaluation, therapy, and outcome, *Am. J. Med.* 86 (3) (1989) 262–266.
- [67] S.D. Turnipseed, A.E. Bair, J.D. Kirk, D.B. Diercks, P. Tabar, E.A. Amsterdam, Electrocardiogram differentiation of benign early repolarization versus acute myocardial infarction by emergency physicians and cardiologists, *Acad. Emerg. Med.* 13 (9) (2006) 961–966.
- [68] M. Karim, M. El Kouache, M.-Y. Amarouch, et al., An fpga-based system for real-time electrocardiographic detection of STEMI, in: *Advanced Technologies for Signal and Image Processing (ATSIP)*, 2016 2nd International Conference on, IEEE, 2016, pp. 830–835.
- [69] R.E. Gregg, S. Babaeizadeh, Improving automatic detection of acute myocardial infarction in the presence of confounders, in: *Computing in Cardiology Conference (CinC)*, 2014, IEEE, 2014, pp. 637–640.
- [70] C.P. Cannon, *Management of Acute Coronary Syndromes*, Springer Science & Business Media, 1998.
- [71] B. Potter, A. Matteau, S. Mansour, C. Naim, M. Riahi, R. Essiambre, M. Montigny, I. Sareault, F. Gobeil, Sustained performance of a physicianless system of automated prehospital STEMI diagnosis and catheterization laboratory activation, *Can. J. Cardiol.* 33 (1) (2017) 148.
- [72] C. Van Mieghem, M. Sabbe, D. Knockaert, The clinical value of the ecg in noncardiac conditions, *Chest J.* 125 (4) (2004) 1561–1576.

Deniz Katircioglu-Öztürk (PhD) received her BSc degree in Computer Engineering, MSc and PhD degrees in Medical Informatics from Middle East Technical University (METU) in 2005, 2009 and 2017, respectively. Currently, she works as a software team director in MITS (Cyberpark, Ankara) where she leads major data mining projects of modelling clinical and experimental databases in cardiology with the aim of developing new score systems as well as contributing to existing guidelines. Her research and professional interests both lie in the area of health informatics with particular focus on bio-statistical data mining, time series analysis, medical decision support and expert systems. <https://scholar.google.com.tr/citations?user=PKNfQRIAAA&hl=tr>

Prof. H. Altay Güvenir (PhD) received his BSc and MSc degrees in Electronics & Communications Engineering from Istanbul Technical University, in 1979 and 1981, respectively, and PhD degree in Computer Engineering & Science from Case Western Reserve University in 1987. He joined the Department of Computer Engineering at Bilkent University in 1987. He has been a professor and serving as the chairman of this department since 2001. His research interests include machine learning, data mining, intelligent data analysis and artificial intelligence particularly focused on medical data. https://www.researchgate.net/profile/Halil_Altay_Guevenir

Prof. Ursula Ravens (MD, PhD, FESC, FAHA, HonFBPHS, FISHR) received her degree in Medicine from the University of Freiburg, Germany in 1969 and obtained her post-doctoral training in Pharmacology at Kiel University, Germany in 1970. Following her “Habilitation” in 1979, she obtained her first tenure position as a Professor of Cardiovascular Pharmacology at Essen University in 1985. After a sabbatical at Imperial College London in 1995, she was appointed full Professor of Pharmacology and Toxicology at the Medical Faculty of TU Dresden from 1997 to 2014. Since 2016 she joined the Institute of Experimental Cardiovascular Medicine at the University Heart Center Freiburg-Bad Krozingen. Her research, published in over 280 peer-reviewed articles, has focused on cardiovascular pharmacology with a special interest in cardiac cellular electrophysiology and electrical remodelling in atrial fibrillation. https://www.researchgate.net/profile/Ursula_Ravens2

Prof. Nazife Baykal (PhD) received her BSc degree in Mathematics, MSc and PhD (1996) degrees in Computer Engineering from the Middle East Technical University (METU) in Ankara, Turkey. Between 1993 and 2003, she held visiting research positions at University of Maryland and University of Texas, School of Health Information Sciences in USA. In 2003, she was appointed as the head of the Health Informatics Department in Institute of Informatics of METU and since 2010 she has been serving as the director of this institute. Her research interests include medical informatics, data mining, fuzzy systems and computer networks. <https://www.ncc.metu.edu.tr/tr/node/2692>