ECG-Based Classification of Resuscitation Cardiac Rhythms for Retrospective Data Analysis

Ali Bahrami Rad*, *Student Member, IEEE*, Trygve Eftestøl, *Member, IEEE*, Kjersti Engan, *Member, IEEE*, Unai Irusta, Jan Terje Kvaløy, Jo Kramer-Johansen, Lars Wik, and Aggelos K. Katsaggelos, *Fellow, IEEE*

***Abstract*— *Objective:* There is a need to monitor the heart rhythm in resuscitation to improve treatment quality. Resuscitation rhythms are categorized into: ventricular tachycardia (VT), ventricular fibrillation (VF), pulseless electrical activity (PEA), asystole (AS), and pulse-generating rhythm (PR). Manual annotation of rhythms is time-consuming and infeasible for large datasets. Our objective was to develop ECG-based algorithms for the retrospective and automatic classification of resuscitation cardiac rhythms. *Methods:* The dataset consisted of 1631 3-s ECG segments with clinical rhythm annotations, obtained from 298 out-of-hospital cardiac arrest patients. In total, 47 wavelet- and time-domain-based features were computed from the ECG. Features were selected using a wrapper-based feature selection architecture. Classifiers based on Bayesian decision theory, k-nearest neighbor, k-local hyperplane distance nearest neighbor, artificial neural network (ANN), and ensemble of decision trees were studied. *Results:* The best results were obtained for ANN classifier with Bayesian regularization backpropagation training algorithm with 14 features, which forms the proposed algorithm. The overall accuracy for the proposed algorithm was 78.5%. The sensitivities (and positive-predictive-values) for AS, PEA, PR, VF, and VT were 88.7% (91.0%), 68.9% (70.4%), 65.9% (69.0%), 86.2% (83.8%), and 78.8% (72.9%), respectively. *Conclusions:* The results demonstrate that it is possible to classify resuscitation cardiac rhythms automatically, but the accuracy for the organized rhythms (PEA and PR) is low. *Significance:* We have made an important step toward making classification of resuscitation rhythms more efficient in the sense of minimal feedback from human experts.**

***Index Terms*—Cardiac arrest, cardiopulmonary resuscitation, cardiac rhythm classification, feature extraction/selection, nested cross-validation.**

I. Introduction

DIFFERENT resuscitation actions such as chest compressions, ventilations, drugs, and electrical shocks are key elements in the treatment of cardiac arrest victims [1], [2]. The therapeutic efforts during resuscitation frequently have an effect on the patients' cardiac rhythm, for instance a defibrillation shock may convert a lethal ventricular fibrillation into a normal perfusing heart rhythm; or provision of chest compressions can revitalize the heart causing within-rhythm changes and also transitions between rhythms. Analysis of the patterns in rhythm changes may provide important information about the quality of the therapy [3], [4] contributing to identify factors that influence survival. Thus, rhythm classification is crucial for retrospective evaluation of treatment decisions, and ultimately to improve the quality of treatment [5].

Currently, cardiac rhythm classification for retrospective review of resuscitation episodes relies on a cumbersome, time-consuming, and error-prone manual annotation process. Resuscitation datasets frequently contain hundreds of cases and several hundred hours of recordings, making manual annotation inefficient and time-consuming. An automatic or semi-automatic rhythm classification system would substantially ease the burden of manually inspecting and annotating such large databases. Rhythm classification is based on the analysis of biomedical signals, particularly of the electrocardiogram (ECG). Other signals recorded by external defibrillators, such as transthoracic impedance (TTI), chest compression depth related acceleration, and/or the capnogram are used to extract additional information about therapeutic interventions like chest compressions and ventilations during cardiopulmonary resuscitation (CPR) [6]–[9]. Although the TTI or the capnogram may help in rhythm classification, these signals are unavailable in many external defibrillators. And even if available they are not recorded during long intervals (capnogram), or are recorded without sufficient amplitude and/or time resolution required for rhythm classification (TTI).

Manuscript received February 3, 2017; accepted March 8, 2017. Date of publication March 30, 2017; date of current version September 18, 2017. This work was supported by the Spanish Ministerio de Economia y Competitividad under Project TEC201564678. *Asterisk indicates corresponding author.*

*A. B. Rad was with the Department of Electrical Engineering and Computer Science, University of Stavanger, Stavanger 4036, Norway. He is now with the NeuroGroup, BioMediTech and Faculty of Medicine and Life Sciences, University of Tampere, Tampere 33520, Finland (e-mail: ali.bahrami.rad@uta.fi).

T. Eftestøl and K. Engan are with the Department of Electrical Engineering and Computer Science, University of Stavanger.

U. Irusta is with the Communications Engineering Department, University of the Basque Country.

J. T. Kvaløy is with the Department of Mathematics and Natural Sciences, University of Stavanger.

J. Kramer-Johansen and L. Wik are with the Norwegian National Advisory Unit on Prehospital Emergency Medicine (NAKOS) and Department Of Anaesthesiology, Oslo University Hospital and University of Oslo.

A. K. Katsaggelos is with the Department of Electrical Engineering and Computer Science, Northwestern University.

Digital Object Identifier 10.1109/TBME.2017.2688380

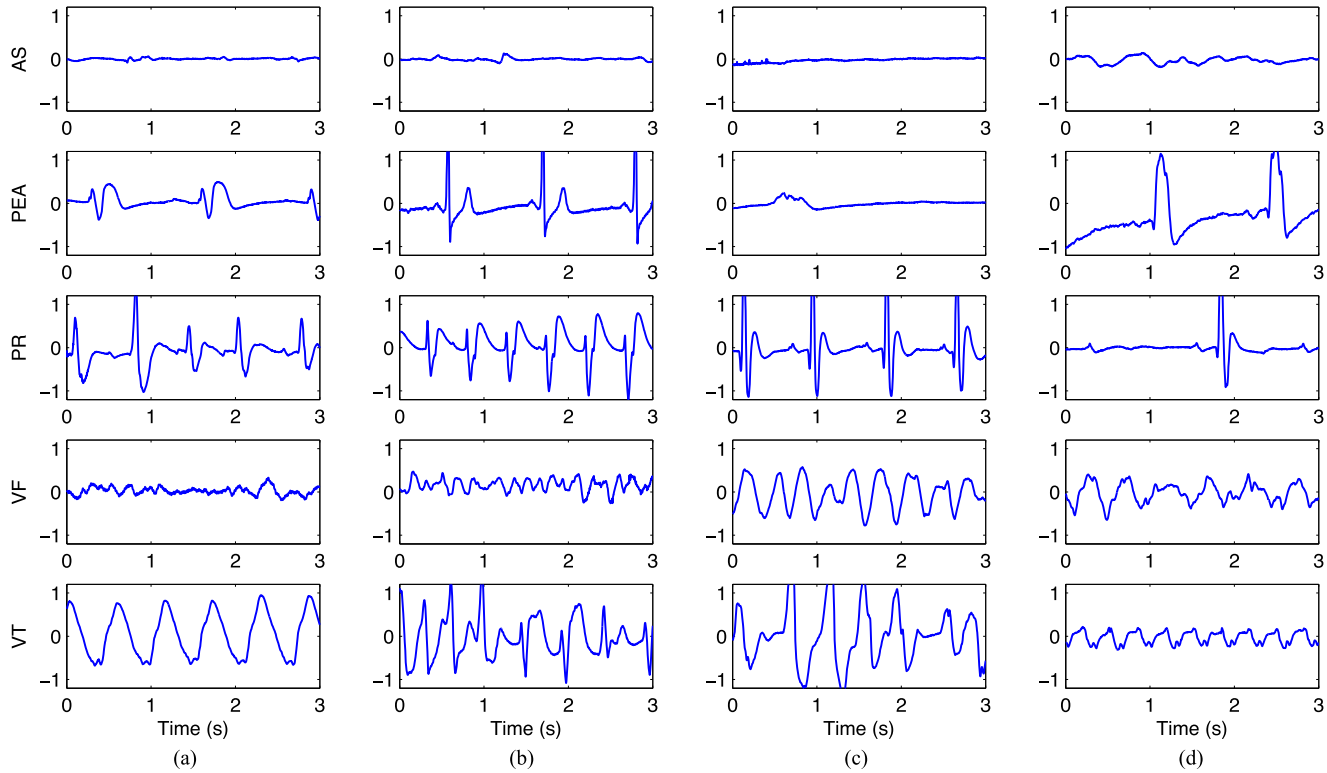

Fig. 1. Examples of 3 s segments of the five different rhythm types (AS, PEA, PR, VF, VT) observed during resuscitation of cardiac arrest patients. The y axes are the amplitudes of the ECG in mV. This figure illustrates the high variability within the same class (intraclass variability) and the similarities between distinct classes (interclass similarity).

Current resuscitation guidelines and treatment recommendations [2] group cardiac rhythms during resuscitation into five categories: ventricular fibrillation (VF), pulseless ventricular tachycardia (VT), pulseless electrical activity (PEA), asystole (AS), and pulse-generating rhythms (PR). VF and VT are lethal ventricular arrhythmias that can be effectively treated through an electrical defibrillation shock. In VT the ventricular activity is more regular and during VF more chaotic. During PEA the heart has an organized electrical activity but no myocardial muscle activity and no palpable pulse. During AS the heart has neither electrical nor mechanical activity. PR rhythms are those observed when the cardiac arrest patient recovers spontaneous circulation.

Shock advice algorithms for defibrillators [10]–[14] require only a gross rhythm categorization into shockable (VF, VT) and non-shockable (PR, PEA, AS). However, other therapeutic interventions such as the evaluation of CPR quality [5], [15], the analysis of spontaneous or therapy induced rhythm transitions [3], [4], [16], and the detection of pulse [17]–[19] involve the five rhythm types mentioned above.

An accurate five class resuscitation rhythm classification based on the ECG only is particularly challenging. As shown in Fig. 1, the ECG of rhythms within a class can be very different (intraclass variability), and there are many borderline cases in which the ECGs of rhythms from different classes are very similar (interclass similarity). So far, clinical applications such as shock advice algorithms [10]–[14], or pulse detection (PEA/PR discrimination) [17]–[19] have been the focus of rhythm classification during resuscitation. However, these applications pose simpler problems based on dichotomous decisions. To the best of our knowledge only our previous attempts [20]–[22] have addressed the detailed classification of resuscitation rhythms based on the ECG.

This work is a comprehensive technical description of ECG-based five class resuscitation rhythm classifiers. It thoroughly describes features derived from the wavelet analysis of the ECG, and combines them with classical shock advice algorithmic features. It introduces an improved feature selection scheme based on a nested cross-validation (CV) technique, and multiple experiments with different classification approaches. Finally, an algorithm is proposed using an optimal set of features and the best classifier.

II. ECG Dataset

ECG data were extracted from 298 cases of out-of-hospital cardiac arrest (OHCA). The original study was conducted by Wik *et al.* [15] to measure the quality of CPR in three European locations: Akershus (Norway), Stockholm (Sweden), and London (UK), between March 2002 and September 2004. Modified defibrillators based on the Heartstart 4000 (Philips Medical Systems, Andover, MA, USA) were placed in six ambulances in each location. Data from each resuscitation case were collected in data cards. The raw data consisted of the ECG, TTI, compression depth related acceleration, and pad pressure signals, all sampled at 500 Hz ($f_s = 500$) with 16 bit resolution. The ECG had a resolution of 1.031 μV per least significant bit.

All recordings were originally annotated by expert reviewers into the five resuscitation rhythm categories [15]. AS was

defined as peak-to-peak amplitude below 100 μV, and/or rates under 12 bpm. Rhythms with supraventricular activity (QRS complexes) and rates above 12 bpm were labeled as either PR or PEA. Pulse annotations (PR) were based on clinical annotations of return of spontaneous circulation (ROSC) made in patient charts during CPR, and on the observation of fluctuations in the TTI signal aligned with QRS complexes. Irregular ventricular rhythms were annotated as VF (coarse VF was defined for peak-to-peak amplitudes above 200 μV). Fast and regular ventricular rhythms without pulse, and rates above 120 bpm were annotated as VT.

For this study, segments from the original OHCA episodes were automatically extracted using the manually annotated rhythm labels and information about the time intervals of defibrillation and chest compression sequences. The following criteria were used: 3-second duration, a single rhythm type, and no chest compression artifacts. These criteria were established to mimic segment durations used in shock advice algorithms [10], [23], and to secure a unique rhythm label on an artifact-free segment, as recommended by the American Heart Association (AHA) for shock advice algorithms [24]. Rhythm analysis during chest compressions is currently unreliable even for shock advice algorithms [25], and is therefore outside the scope of this work.

All segments were reviewed and audited to be compliant with the rhythm definitions, and to the AHA recommendations. Two biomedical engineers, experienced in resuscitation data processing, and not involved in the development of the classification algorithms, double blindedly reviewed the segments. Segments with severe noise, ongoing chest compressions, incorrect rhythm annotations (transitional rhythms), or low ECG signal quality were either relabeled, removed, or re-extracted. The final annotated dataset of segments consisted of 388 AS (n = 217 patients), 366 PEA (n = 200), 264 PR (n = 111), 377 VF (n = 157), and 236 VT (n = 32).

III. Evaluation Criteria

Sensitivity (Sen) and positive predictive value (PPV) were used to evaluate the performance of the classifiers. Since these metrics are defined for binary classification the approach presented in [26] was used to generalize them to our multiclass problem. The new definitions are based on the contingency table shown in Table I, where $N_{i,j}$ represents the number of samples of class i (expert label) classified as class j (algorithmic result). In our case $i, j \in \{\text{AS, PEA, PR, VF, VT}\} \equiv \{1, 2, 3, 4, 5\}$. Consequently, the total number of samples R_i in class i and the total number of algorithmic decisions C_j for class j can be expressed as:

$$R_i = \sum_{j=1}^{5} N_{i,j} \quad \text{and} \quad C_j = \sum_{i=1}^{5} N_{i,j}. \tag{1}$$

The total number of samples is then:

$$N_{\text{tot}} = \sum_{i=1}^{5}\sum_{j=1}^{5} N_{i,j} = \sum_{i=1}^{5} R_i = \sum_{j=1}^{5} C_j. \tag{2}$$

TABLE I
CONTINGENCY TABLE FOR CLASSIFICATION OF RESUSCITATION CARDIAC RHYTHMS

		Algorithm Label					$\sum$
		AS	PEA	PR	VF	VT	
Expert Reviewers' Label	AS	$N_{1,1}$	$N_{1,2}$	$N_{1,3}$	$N_{1,4}$	$N_{1,5}$	R_1
	PEA	$N_{2,1}$	$N_{2,2}$	$N_{2,3}$	$N_{2,4}$	$N_{2,5}$	R_2
	PR	$N_{3,1}$	$N_{3,2}$	$N_{3,3}$	$N_{3,4}$	$N_{3,5}$	R_3
	VF	$N_{4,1}$	$N_{4,2}$	$N_{4,3}$	$N_{4,4}$	$N_{4,5}$	R_4
	VT	$N_{5,1}$	$N_{5,2}$	$N_{5,3}$	$N_{5,4}$	$N_{5,5}$	R_5
$\sum$		C_1	C_2	C_3	C_4	C_5	N_{tot}

Numerical class equivalences are: 1 ≡ AS, 2 ≡ PEA, 3 ≡ PR, 4 ≡ VF, 5 ≡ VT.

The definitions for true positive (TP), false positive (FP), false negative (FN), and true negative (TN) are generalized for class i as follows:

$$\begin{aligned} \text{TP}_i &= N_{i,i} \\ \text{FP}_i &= C_i - N_{i,i} \\ \text{FN}_i &= R_i - N_{i,i} \\ \text{TN}_i &= N_{\text{tot}} - \text{TP}_i - \text{FP}_i - \text{FN}_i \\ &= N_{\text{tot}} - R_i - C_i + N_{i,i}. \end{aligned} \tag{3}$$

Therefore, the performance metrics for class i are defined as:

$$\text{Sen}_i = \frac{\text{TP}_i}{R_i} = \frac{N_{i,i}}{R_i} \quad \text{and} \quad \text{PPV}_i = \frac{\text{TP}_i}{C_i} = \frac{N_{i,i}}{C_i}. \tag{4}$$

Finally two global measures of performance were used to summarize the full confusion matrix into a single parameter [26]. These parameters are the multiway accuracy (*MulAcc*) and the unweighted mean of sensitivities (*UMS*) defined as:

$$\text{MulAcc} = \frac{1}{N_{\text{tot}}}\sum_{i=1}^{5} N_{i,i} \quad \text{and} \quad \text{UMS} = \frac{1}{5}\sum_{i=1}^{5} \text{Sen}_i. \tag{5}$$

Multiway accuracy measures the total accuracy of the classifier, but is sensitive to class imbalance, because it weights every sample equally. Classes were only slightly imbalanced in our dataset and *MulAcc* was therefore adopted as the main performance criterion. We also report *UMS* because by weighting all classes equally, it is unaffected by class imbalance.

IV. Feature Engineering and Model Assessment

A. Feature Extraction

In our previous studies [20]–[22] the wavelet approach was identified as the most promising feature extraction method. In the current study a set of 32 features based on the discrete wavelet transform (DWT) was extracted and combined with 12 classical features for shock/no-shock classification, and 3 features derived from R wave detection.

The DWT was employed for multiresolution analysis of the ECG [27]. The 3-second ECG segments were decomposed into its subbands with the Daubechies 4 wavelet [28] and 8 levels of

decomposition. At each level of decomposition the DWT can be implemented by a pair of lowpass/highpass filters which decompose the signal into the lower and upper halves of the subband under analysis (dyadic wavelet transforms). The decomposition process of the original ECG segment, $s_{\text{ecg}}(n)$, in J levels can be represented by Mallat's fast algorithm [29]:

$$\begin{aligned} a_0(n) &= s_{\text{ecg}}(n) \\ a_{j+1}(p) &= \sum_n h_0(n-2p)a_j(n) \\ d_{j+1}(p) &= \sum_n h_1(n-2p)a_j(n) \end{aligned} \tag{6}$$

in which h_0 and h_1 are conjugate mirror filters corresponding to the scaling and wavelet functions, and a_{j+1} and d_{j+1} are approximation and detail coefficients of level $j+1$ ($j = 0, 1, 2, \cdots, J-1$). In addition, the reconstruction process can be formulated by the following equation:

$$a_j(p) = \sum_n h_0(p-2n)a_{j+1}(n) + \sum_n h_1(p-2n)d_{j+1}(n). \tag{7}$$

In the current study, each 3-second ECG segment was decomposed into eight-octaves ($J = 8$) to generate nine sets of coefficients (a_8, and d_8 to d_1) roughly corresponding to the following frequency bands: 0-0.98 Hz, 0.98-1.95 Hz, 1.95-3.91 Hz, 3.91-7.81 Hz, 7.81-15.62 Hz, 15.62-31.25 Hz, 31.25-62.5 Hz, 62.5-125 Hz, and 125-250 Hz. For feature extraction, only detail coefficients of level 4-8 (d_4-d_8) were used. This process is equivalent to retaining the spectral components in the 0.98-31.25 Hz band, which is similar to the typical automated external defibrillator (AED) analysis band. The other coefficients (d_1-d_3, and a_8) either do not contain useful information or they correspond to high-frequency/low-frequency noise. Thus, the clean ECG, $s(n)$, was obtained by using (7) and setting d_1, d_2, d_3, and a_8, equal to zero.

The 32 DWT features were computed using the clean ECG, $s(n)$, and its 5 detail coefficients $d_m(n)$, $m = 4, 5, \cdots, 8$. The first 15 features statistically characterize the amplitudes of $d_m(n)$ in terms of interquartile ranges (IQR) (1-5), variances (6-10) and first quartiles (11-15). The next 11 features are statistical and morphological descriptors of the amplitudes of $s(n)$ and its first and second forward differences:

$$\dot{s}(n) = s(n+1) - s(n), \ddot{s}(n) = s(n+2) - 2s(n+1) + s(n). \tag{8}$$

Features 16-18 are the IQRs of $s(n)$, $\dot{s}(n)$ and $\ddot{s}(n)$. Features 19-21 are based on the central moments of $s(n)$:

$$\mu_{q,s} = \left| E\left[(s(n) - \mu_s)^q \right] \right|^{\frac{1}{q}} \quad \text{for} \quad q = 2, 3, 4 \tag{9}$$

where μ_s is the mean value of the signal, and the absolute value of the q-th central moment is raised to the $1/q$ power so that features are in the amplitude units of $s(n)$. Features 22 and 23 were obtained by applying (9) to $\dot{s}(n)$ for $q = 3, 4$. Features 24-26 are the amplitude range, and the positive and negative area counts of $s(n)$ over the 3-second segment:

$$\frac{1}{2}\sum_{n=0}^{3f_s} \left(s(n) + |s(n)| \right) \quad \text{and} \quad \frac{1}{2}\sum_{n=0}^{3f_s} \left(s(n) - |s(n)| \right). \tag{10}$$

Then, $s(n)$ was modeled as an order 4 autoregressive (AR) process:

$$s(n) = -\sum_{k=1}^{4} a_k s(n-k) + v(n). \tag{11}$$

The coefficients of the model, a_k, and the variance of the white noise term $v(n)$, σ_v^2, were used as features 27-31. These values were estimated using Burg's method [30], and completely characterize the power spectral density of the ECG when modeled as an AR(4) process. Finally feature 32 is the number of peaks with values larger than 0.2 of $\tilde{R}(n)$, the normalized autocorrelation function of $s(n)$:

$$\begin{aligned} R(\ell) &= \sum_{n=0}^{3f_s - \ell} s(n+\ell)s(n), \ \ell = 0, 1, \cdots, 3f_s \\ \tilde{R}(\ell) &= \frac{R(\ell)}{R(0)}. \end{aligned} \tag{12}$$

The final set of 47 features was built adding 12 classical shock/no-shock features, and 3 R wave detection features, all computed using $s(n)$. Shock/no-shock decision features have been comprehensively studied by Figuera *et al.* in the context of OHCA [31]. For this study the best performing 12 features were added:

1) Time domain features: bWT, bCP and ECG power [10], count3 [32], MAV [33], TCSC [34] and x1 [23].
2) Spectral features: vFleak [35] and x3 [23].
3) Time-frequency feature: Li [36].
4) Complexity features: sample entropy (SampEn) and phase space reconstruction (PSR) [37].

In addition QRS complexes were detected using a QRS detector based on the Hamilton-Tompkins approach [38], and three features were added to measure the number of beats, and the mean duration and dispersion of the cardiac cycle, $\overline{\text{RR}}$ and σ_{RR}, respectively.

B. Feature Selection and Model Assessment

Fig. 2 shows the nested cross-validation (CV) architecture [39] used for feature selection (FS) and model assessment. In the CV loops data were always partitioned to secure that different patients were used in the development and testing sets (patient-wise CV). In the model assessment loop (outer loop) 10-fold CV was applied to evaluate the performance of the main classifier. In this loop the three-sigma rule [40] was used to remove outliers from the training data (inner loop), but not from the testing data. The features used in the main classifier were selected in the FS loop (inner loop), using 5-fold CV and the wrapper-based approach [41]. In the wrapper-based approach an internal FS classifier was used for feature selection, and the feature subset that minimized its error rate ($\text{Err} = 1 - \text{MulAcc}$) was selected. In this work we used a linear discriminant

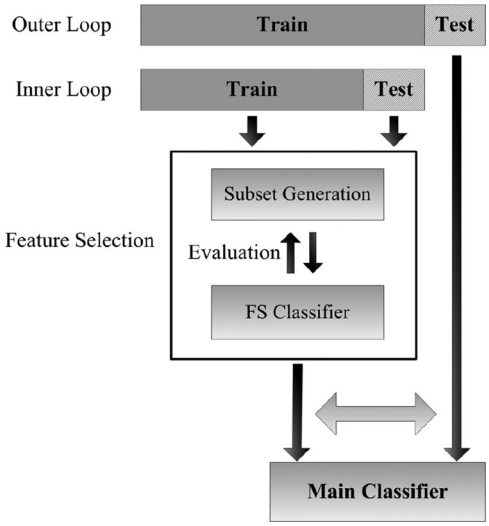

Fig. 2. Nested cross-validation architecture used for feature selection and model assessment.

analysis (LDA) classifier in the inner loop (FS classifier) and "plus ℓ-take away r," PTA(ℓ, r), search strategy for feature subset selection [42].

PTA(ℓ, r) combines sequential forward selection (SFS) [43] and sequential backward selection (SBS) [44] to avoid the "nesting effect" of SFS and SBS. Each step of PTA(ℓ, r) is divided into two substeps. First SFS is run to sequentially add the ℓ best new features, and then SBS is run to sequentially remove the r least useful ones. If $\ell > r$, PTA(ℓ, r) starts with SFS and the initial empty feature subset. For $\ell < r$, PTA(ℓ, r) starts with SBS and the initial full feature set [45]. In both cases, these two substeps are iteratively applied until a specific number of features is selected. After trying several combinations of ℓ and r we adopted a PTA(4,5) search strategy with 14 selected features on each fold of the outer CV loop.

V. Classifier Models

A. Classifiers Based on Bayesian Decision Theory

Bayes classifiers are based on selecting the most probable class for a given observation, thus minimizing the probability of erroneous classifications [46]. The probability of having a class ω_i given a new observation (feature vector) $\boldsymbol{x}$, $P(\omega_i|\boldsymbol{x})$, is derived from the class conditional probability densities (CCPD), $p(\boldsymbol{x}|\omega_i)$, and the class prior probabilities $P(\omega_i)$ using Bayes' theorem. CCPDs can be estimated from the observations using different models:

1) LDA: CCPDs are multivariate Gaussian distributions with a fixed covariance matrix (homoscedasticity).
2) Quadratic discriminant analysis (QDA): CCPDs with different covariance matrices (heteroscedasticity).
3) Gaussian Mixture Model (GMM): CCPDs are linear combinations of Gaussians [47]. We minimized the Akaike information criterion (AIC) [48] to determine the optimal number of Gaussian distributions per class.

TABLE II
Features Ranked by N, the Number of Times They Were Selected in 10 Random Repetitions of the 10-fold CV

Feature	N	Feature	N	Feature	N	Feature	N
SampEn	100	TCSC	66	$\mu_{2,s}$	52	a_3	37
Var(d_7)	79	Li	58	$\mu_{4,s}$	47	MAV	36
a_1	78	a_2	55	Var(d_5)	46	IQR(d_6)	35
σ_{RR}	71	IQR($\dot{s}(n)$)	54	IQR(d_7)	43	$\overline{\mathrm{RR}}$	31
Var(d_6)	67	Var(d_8)	52	IQR($\ddot{s}(n)$)	40	IQR(d_8)	30

Var is variance, a_i and d_i are AR and detail coefficients, respectively.

B. K-Nearest-Neighbor (KNN) Rule

In KNN each query sample is assigned to the most frequent class of its K closest neighbors. Proximity between two observations was measured using the Euclidean distance.

C. K-Local Hyperplane Distance Nearest-Neighbor (HKNN)

HKNN is an extension of KNN that improves accuracy to the level of kernel-based support vector machines [49]. In HKNN each class is modeled as a low dimensional manifold embedded in a high dimensional feature space, and the class of a query sample is that of the nearest class-specific manifold [50]. Class-specific manifolds are represented by a set of local hyperplanes determined by the K-nearest points to a query sample. We used a regularized Euclidean distance to determine proximity. A decay penalty term, controlled by the regularization parameter λ, was added to penalize moving away from the centroid of the K-nearest points.

D. Artificial Neural Network (ANN)

ANNs consist of a set of interconnected nodes (neurons), in which the output of the node is a non-linear function of a linear combination of its inputs. The weights of the linear combinations are determined during the network's training phase. In this study we used feedforward ANNs with two hidden layers, a hyperbolic tangent activation function for the neurons, and the same number of hidden neurons per layer: $N_h = 5, 10, 15, 20, 25, 30, 35$. The number of neurons in the output layer is 5 for the five-class classification task. We followed two strategies to train the ANN, resilient backpropagation [51] and Bayesian regularization backpropagation [52], ANNRprop and ANNBreg, respectively in what follows.

E. Ensemble of Decision Trees

Ensemble algorithms combine the outputs of multiple classifiers to improve performance. We used ensembled decision trees (EDT), and the ensemble's diversity was achieved using bagging [53], i.e., bootstrapped replicas of the training data. Each replica, the size of the training set, was obtained by random selection with replacement. Thus, training subsets had significant overlap, with many samples present in most subsets, and some samples present multiple times in a subset [54]. Each training data-subset was fed into a decision tree, which generates sufficiently different decision boundaries. The test

TABLE III
THE RESULTS OF ALL EIGHT CLASSIFIERS WITH DIFFERENT PARAMETERS

			AS		PEA		PR		VF		VT	
Algorithm	MulAcc	UMS	Sen	*PPV*	Sen	*PPV*	Sen	*PPV*	Sen	*PPV*	Sen	*PPV*
LDA	76.27	75.72	88.14	83.41	61.75	70.40	64.39	68.55	83.82	82.72	80.51	70.37
QDA	74.00	72.89	89.95	86.39	67.76	67.76	57.95	65.38	75.07	79.49	73.73	64.21
GMM	74.74	73.56	91.75	88.56	60.66	67.68	67.80	59.27	80.64	81.28	66.95	70.22
KNN, $K = 25$	76.09	74.73	88.66	84.94	66.39	67.69	57.95	70.18	86.47	80.30	74.15	72.02
HKNN, $K = 13, \lambda = 100$	77.01	76.44	89.18	87.37	64.48	67.82	65.53	66.54	82.49	85.21	80.51	75.52
ANNRprop, $N_{h_X} = 25$	77.81	76.98	86.86	90.35	69.67	70.44	65.53	67.32	84.88	82.69	77.97	73.03
ANNBreg, $N_{h_X} = 25$	78.48	77.57	88.66	91.25	69.40	71.95	66.29	68.09	85.94	82.86	77.54	72.33
EDT+SBS, $N_{DT} = 100$	77.99	77.14	89.18	92.27	67.76	68.32	67.05	67.30	84.62	82.22	77.12	75.21
Proposed Algorithm[a]	**78.54**	**77.69**	88.66	91.01	68.85	70.39	65.91	69.05	86.21	83.76	78.81	72.94

[a]In the proposed algorithm we used ANNBreg classifier with the top 14 ranked features, as described in Table II.

data were classified by taking the majority vote of all classifiers. In this work we used EDTs with different number of trees ($N_{DT} = 20, 50, 100, 200, 300, 500$), and a deep architecture (i.e., decision trees with many leaves and branch nodes).

VI. RESULTS

The 8 classifiers described above were evaluated. All features were normalized to the $[-1, 1]$ range, and in each internal CV loop 14 features were obtained using the PTA(4,5) FS strategy. Then classifiers were trained with those 14 features on the training data of the external CV loop, and tested on the testing data, as shown in Fig. 2. All results refer to the test set.

Table II shows the features ranked by the number of times they were selected in 10 random repetitions of the 10-fold CV procedure. The most frequently selections include classical shock/no-shock decision features such as SampEn, TCSC, Li, or MAV, features from the analysis of the cardiac cycle, σ_{RR} or $\overline{RR}$, and several of the features introduced in this paper. These new features are the statistical descriptors of detail coefficients $d_5 - d_8$, the AR model coefficients and the statistical moments of the denoised signal, $s(n)$, and statistical descriptors of its first and second forward differences, $\dot{s}(n)$ and $\ddot{s}(n)$.

The optimal configuration, multiway accuracy, and unweighted mean of sensitivities of the eight classifiers are shown in Table III. The table also shows the sensitivity and positive predictive value of the algorithms for each rhythm type. The best results were obtained for an ANN with Bayesian regularization (ANNBreg) and 25 hidden neurons per layer.

A final algorithm was designed, the proposed algorithm, based on the results shown in Tables II and III. Only those features that were selected more than 40 times in the feature ranking procedure (see Table II) were used in an ANNBreg classifier with $N_h = 25$. The performance of the proposed algorithm was tested using 10-fold CV scheme. Table IV shows the detailed results in the form of a confusion matrix and the accuracies are reported in Table III. *MulAcc* was increased to 78.54%. As shown in the confusion matrix most classification errors occured between PEA/PR, PEA/AS, AS/VF, and VT/VF.

TABLE IV
CONFUSION MATRIX OF THE PROPOSED ALGORITHM

		Proposed Algorithm Label				
		AS	PEA	PR	VF	VT
Expert Reviewers' Label	AS	344	30	1	13	0
	PEA	21	252	65	15	13
	PR	0	67	174	1	22
	VF	12	5	1	325	34
	VT	1	4	11	34	186

VII. DISCUSSION

This paper provides the comprehensive technical description of ECG-based resuscitation rhythm classifiers using different machine learning techniques. These algorithms are intended for offline annotation, in a system for the comprehensive retrospective analysis of resuscitation episodes. Resuscitation datasets can easily grow to the thousands of cases, and each case is normally between 30 minutes and one hour long. In this period several rhythm transitions occur, either spontaneous or as a consequence of therapeutic interventions such as defibrillation or CPR. The proposed algorithm will ease the annotation process, i.e. the review process of such datasets. The rhythm classifier would process the ECG continuously and could be used along with chest compression detection [6] and CPR artifact removal filters [55] for rhythm annotation during CPR. This will allow fully automatic review of resuscitation episodes as described in [56].

The multiyway accuracy of the proposed algorithm is 78.5%, a considerable advance because the baseline random guess accuracy for a balanced five class problem is only 20%. Compared to our previous work [20], evaluation on a quality controlled dataset, the use of advanced classifiers, extended feature sets, and a robust feature selection increased the accuracy by 10 percentage points. The current version of our system is semi-automatic since it would involve a final quality assurance review of the annotations by expert clinicians. However, the workload would be substantially reduced opening the possibility of annotating large cardiac arrest databases.

A. Performance of the Classifiers and Feature Selection

The best classification results were obtained for ANNBreg, although comparable results were obtained for EDT, ANNRprop, or HKNN classifiers. Classifiers based on Bayesian decision theory had lower accuracy, probably because of simplistic

CCPD model assumptions and/or insufficient data to estimate the CCPDs. The features selected for the proposed algorithm show the importance of a multidomain analysis of the ECG [31]. Classical quantitative measures of complexity (SampEn), heart rate (TCSC), morphological consistence (Li), and dispersion of the cardiac cycle (σ_{RR}) were selected along five statistical measures of detail coefficients $d_5 - d_8$, two AR coefficients (a_1 and a_2), two modified central moments of the denoised signal $s(n)$ ($\mu_{2,s}$ and $\mu_{4,s}$), and one statistical descriptor (IQR) of the signal's first difference (slope). The statistical measures of $d_5 - d_8$ quantify variability of the ECG components in four subbands of the 1-16 Hz frequency range. Remarkably features related to d_4 (16-31 Hz) were not selected probably because resuscitation rhythms are either ventricular or supraventricular with low heart rates and aberrant QRS complexes. Two of the four AR model coefficients were included, suggesting that the shape of the spectral distribution of the rhythms is a distinctive characteristic.

The classifier is not conceived for real-time analysis of the ECG, so processing time is not an issue. Feature selection was therefore addressed using a wrapper-based method instead of filter based methods, to favor accuracy over speed [57].

B. Sources of Misclassification

The confusion matrix for the proposed algorithm, Table IV, shows the main sources of misclassification. Classification errors can be grouped into four categories [20]:

1) Errors in PEA/PR discrimination, providing evidence of the limitations of detecting pulse using the ECG alone [18], [19].
2) Errors due to borderline rhythms such as: VF with low amplitudes and/or dominant frequency (AS/VF), bradycardic rhythms (AS/PEA), tachycardias of either ventricular or supraventricular origin (VT/PEA, VT/PR), and irregular or polymorphic VT (VF/VT).
3) Errors due to transitional states between rhythms.
4) Errors in rhythm interpretation immediately after stopping chest compressions.

PEA/PR discrimination deserves special attention. Pulse detection during resuscitation is of paramount importance. In fact, resuscitation efforts focus on restoring pulse. Currently PEA/PR discrimination uses other signals combined with the ECG, such as impedance [17]–[19], or the restoration of normal values in the capnogram. This paper explores the limits of using the ECG alone, and therefore Sen and *PPV* values for PEA and PR are lower than for other rhythms (see Table III). However, by combining PEA and PR into an organized rhythm class (ORG) [23] the overall accuracy increases to 86.6% and the Sen/*PPV* values for ORG would be 88.6%/91.5%.

Borderline rhythms are intrinsically difficult to detect. These cases could be automatically identified and labeled as dubious, for instance by analyzing the amplitude of VF/AS rhythms. The final decision would be left to the clinician in the quality assurance process. Other borderline cases may require ad hoc algorithms such as the discrimination of supraventricular and ventricular rhythms during resuscitation [58].

Rhythm dynamics during resuscitation are complex and rhythm transitions may be frequent [4]. These transitional states are challenging for a short 3-second analysis window. Misclassifications caused by rhythm transitions can be addressed using sequential classification algorithms based on the analysis of consecutive windows [10]. Approaches based on hidden Markov models (HMM) [59] and conditional random fields (CRFs) [60] will be further investigated.

C. Future Work

Quality assurance in resuscitation episode review requires higher accuracy. Future directions of work will involve: (1) the use of additional signals, such as TTI or the capnogram, to detect circulation; (2) sequential algorithms to combine consecutive analysis (transitional states); and (3) ad hoc algorithms for borderline rhythms. Adding signals like the TTI or the capnogram may limit the applicability of the algorithm, but would make it more accurate for those datasets that have those signals available.

VIII. Conclusion

This paper lays the groundwork for automatic resuscitation rhythm classification using only ECG, for use in the resuscitation data review process. The proposed algorithm has an accuracy of 78.5%. Further work is needed for quality assurance of resuscitation data review. Research directions towards this goal have been identified.

Acknowledgment

The authors would like to thank the anonymous reviewers for their insightful critique which helped and contributed to improving this body of work.

References

[1] G. D. Perkins *et al.*, "European resuscitation council guidelines for resuscitation 2015: Section 2. Adult basic life support and automated external defibrillation," *Resuscitation*, vol. 95, pp. 81–99, 2015.
[2] J. Soar *et al.*, "European resuscitation council guidelines for resuscitation 2015: Section 3. Adult advanced life support," *Resuscitation*, vol. 95, pp. 100–147, 2015.
[3] T. Nordseth *et al.*, "Clinical state transitions during advanced life support (ALS) in in-hospital cardiac arrest," *Resuscitation*, vol. 84, no. 9, pp. 1238–1244, 2013.
[4] E. Skogvoll *et al.*, "Dynamics and state transitions during resuscitation in out-of-hospital cardiac arrest," *Resuscitation*, vol. 78, no. 1, pp. 30–37, 2008.
[5] T. Eftestøl *et al.*, "Representing resuscitation data-considerations on efficient analysis of quality of cardiopulmonary resuscitation," *Resuscitation*, vol. 80, no. 3, pp. 311–317, 2009.
[6] U. Ayala *et al.*, "Automatic detection of chest compressions for the assessment of CPR-quality parameters," *Resuscitation*, vol. 85, no. 7, pp. 957–963, 2014.
[7] E. Alonso *et al.*, "Reliability and accuracy of the thoracic impedance signal for measuring cardiopulmonary resuscitation quality metrics," *Resuscitation*, vol. 88, pp. 28–34, 2015.
[8] M. Risdal *et al.*, "Impedance-based ventilation detection during cardiopulmonary resuscitation," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 12, pp. 2237–2245, Dec. 2007.
[9] E. Aramendi *et al.*, "Feasibility of the capnogram to monitor ventilation rate during cardiopulmonary resuscitation," *Resuscitation*, vol. 110, pp. 162–168, Jan. 2017.
[10] U. Irusta *et al.*, "A high-temporal resolution algorithm to discriminate shockable from nonshockable rhythms in adults and children," *Resuscitation*, vol. 83, no. 9, pp. 1090–1097, 2012.

[11] J.-P. Didon *et al.*, "Shock advisory system with minimal delay triggering after end of chest compressions: Accuracy and gained hands-off time," *Resuscitation*, vol. 82, pp. S8–S15, 2011.
[12] F. Alonso-Atienza *et al.*, "Detection of life-threatening arrhythmias using feature selection and support vector machines," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 3, pp. 832–840, Mar. 2014.
[13] Q. Li *et al.*, "Ventricular fibrillation and tachycardia classification using a machine learning approach," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 6, pp. 1607–1613, Jun. 2014.
[14] Y. Li *et al.*, "An algorithm used for ventricular fibrillation detection without interrupting chest compression," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 1, pp. 78–86, Jan. 2012.
[15] L. Wik *et al.*, "Quality of cardiopulmonary resuscitation during out-of-hospital cardiac arrest," *JAMA*, vol. 293, no. 3, pp. 299–304, 2005.
[16] J. T. Kvaløy *et al.*, "Which factors influence spontaneous state transitions during resuscitation?" *Resuscitation*, vol. 80, no. 8, pp. 863–869, 2009.
[17] M. Risdal *et al.*, "Automatic identification of return of spontaneous circulation during cardiopulmonary resuscitation," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 1, pp. 60–68, 2008.
[18] J. Ruiz *et al.*, "Reliable extraction of the circulation component in the thoracic impedance measured by defibrillation pads," *Resuscitation*, vol. 84, no. 10, pp. 1345–1352, 2013.
[19] E. Alonso *et al.*, "Circulation detection using the electrocardiogram and the thoracic impedance acquired by defibrillation pads," *Resuscitation*, vol. 99, pp. 56–62, 2016.
[20] A. B. Rad *et al.*, "Automatic cardiac rhythm interpretation during resuscitation," *Resuscitation*, vol. 102, pp. 44–50, 2016.
[21] A. B. Rad *et al.*, "Probabilistic classification approaches for cardiac arrest rhythm interpretation during resuscitation," in *Proc. Comput. Cardiol.*, 2013, pp. 125–128.
[22] A. B. Rad *et al.*, "Nearest-manifold classification approach for cardiac arrest rhythm interpretation during resuscitation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2014, pp. 3621–3625.
[23] U. Ayala *et al.*, "A reliable method for rhythm analysis during cardiopulmonary resuscitation," *BioMed Res. Int.*, vol. 2014, 2014, Art. no. 872470.
[24] R. E. Kerber *et al.*, "Automatic external defibrillators for public access defibrillation: Recommendations for specifying and reporting arrhythmia analysis algorithm performance, incorporating new waveforms, and enhancing safety," *Circulation*, vol. 95, no. 6, pp. 1677–1682, 1997.
[25] S. Ruiz de Gauna *et al.*, "Rhythm analysis during cardiopulmonary resuscitation: past, present, and future," *BioMed Res. Int.*, vol. 2014, 2014, Art. no. 872470.
[26] T. Mar *et al.*, "Optimization of ECG classification by means of feature selection," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 8, pp. 2168–2177, 2011.
[27] S. Mallat, "Multiresolution approximations and wavelet orthonormal bases of $L^2(R)$," *Trans. Amer. Math. Soc.*, vol. 315, pp. 69–87, 1989.
[28] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA, USA: SIAM, 1992.
[29] S. Mallat, *A Wavelet Tour of Signal Processing*. New York, NY, USA: Academic, 1999.
[30] J. P. Burg, "A new analysis technique for time series data," in *NATO Advanced Study Institute of Signal Processing With Emphasis on Underwater Acoustics*. New York, NY, USA: IEEE Press, 1968.
[31] C. Figuera *et al.*, "Machine learning techniques for the detection of shockable rhythms in automated external defibrillators." *PLoS One*, vol. 11, 2016, Art. no. e0159654.
[32] I. Jekova and V. Krasteva, "Real time detection of ventricular fibrillation and tachycardia." *Physiol. Meas.*, vol. 25, pp. 1167–1178, Oct. 2004.
[33] E. M. A. Anas *et al.*, "Sequential algorithm for life threatening cardiac pathologies detection based on mean signal strength and EMD functions." *Biomed. Eng. Online*, vol. 9, no. 1, p. 43, Sep. 2010.
[34] M. A. Arafat *et al.*, "A simple time domain algorithm for the detection of ventricular fibrillation in electrocardiogram," *Signal, Image Video Process.*, vol. 5, no. 1, pp. 1–10, 2011.
[35] S. Kuo and R. Dillman, "Computer detection of ventricular fibrillation," in *Proc. Comput. Cardiol.*, 1978, pp. 347–349.
[36] Y. Li *et al.*, "An algorithm used for ventricular fibrillation detection without interrupting chest compression," *IEEE Trans. Bio-Med. Eng.*, vol. 59, pp. 78–86, Jan. 2012.
[37] A. Amann *et al.*, "Detecting ventricular fibrillation by time-delay methods," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 1, pp. 174–177, Jan. 2007.
[38] P. S. Hamilton and W. J. Tompkins, "Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database," *IEEE Trans. Biomed. Eng.*, vol. BME-33, no. 12, pp. 1157–1165, Dec. 1986.
[39] S. Varma and R. Simon, "Bias in error estimation when using cross-validation for model selection," *BMC Bioinformat.*, vol. 7, no. 1, p. 91, Feb. 2006.
[40] R. A. Maronna *et al.*, *Robust Statistics: Theory and Methods*. New York, NY, USA: Wiley, 2006.
[41] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, nos. 1/2, pp. 273–324, 1997.
[42] S. D. Stearns, "On selecting features for pattern classifiers," in *Proc. 3rd Int. Joint Conf. Pattern Recognit.*, 1976, pp. 71–75.
[43] A. W. Whitney, "A direct method of nonparametric measurement selection," *IEEE Trans. Comput.*, vol. C-20, no. 9, pp. 1100–1103, Sep. 1971.
[44] T. Marill and D. Green, "On the effectiveness of receptors in recognition systems," *IEEE Trans. Inf. Theory*, vol. IT-9, no. 1, pp. 11–17, Jan. 1963.
[45] J. Reunanen, "Search strategies," in *Feature Extraction*. Berlin, Germany: Springer, 2006, pp. 119–136.
[46] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed., New York, NY, USA: Academic, 2008.
[47] G. McLachlan and D. Peel, *Finite Mixture Models*. New York, NY, USA: Wiley, 2000.
[48] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.
[49] P. Vincent and Y. Bengio, "K-local hyperplane and convex distance nearest neighbor algorithms," in *Advances in Neural Information Processing Systems*, vol. 14, Cambridge, MA, USA: MIT Press, 2001, pp. 985–992.
[50] O. Okun, "Protein fold recognition with k-local hyperplane distance nearest neighbor algorithm," in *Proc. Second Eur. Workshop Data Mining Text Mining Bioinformat.*, 2004, vol. 1, pp. 51–57.
[51] M. Riedmiller and H. Braun, "A direct adaptive method for faster backpropagation learning: The RPROP algorithm," in *Proc. IEEE Int. Conf. Neural Netw.*, 1993, pp. 586–591.
[52] D. J. C. MacKay, "Bayesian interpolation," *Neural Comput.*, vol. 4, no. 3, pp. 415–447, May 1992.
[53] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996.
[54] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits Syst. Mag.*, vol. 6, no. 3, pp. 21–45, Jul.–Sep. 2006.
[55] U. Irusta *et al.*, "A least mean-square filter for the estimation of the cardiopulmonary resuscitation artifact based on the frequency of the compressions," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 4, pp. 1052–1062, Apr. 2009.
[56] T. Eftestøl and L. D. Sherman, "Towards the automated analysis and database development of defibrillator data from cardiac arrest," *BioMed Res. Int.*, vol. 2014, 2014, Art. no. 276965.
[57] A. Wang *et al.*, "Accelerating wrapper-based feature selection with k-nearest-neighbor," *Knowl.-Based Syst.*, vol. 83, pp. 81–91, 2015.
[58] U. Irusta and J. Ruiz, "An algorithm to discriminate supraventricular from ventricular tachycardia in automated external defibrillators valid for adult and paediatric patients," *Resuscitation*, vol. 80, no. 11, pp. 1229–1233, 2009.
[59] H. Kwok *et al.*, "Adaptive rhythm sequencing: A method for dynamic rhythm classification during CPR," *Resuscitation*, vol. 91, pp. 26–31, 2015.
[60] J. D. Lafferty *et al.*, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. Eighteenth Int. Conf. Mach. Learn.*, 2001, pp. 282–289.

Authors, photographs and biographies not available at the time of publication.