Ventricular Fibrillation and Tachycardia Classification Using a Machine Learning Approach

Qiao Li, Cadathur Rajagopalan, *Senior Member, IEEE*, and Gari D. Clifford*, *Senior Member, IEEE*

***Abstract*—Correct detection and classification of ventricular fibrillation (VF) and rapid ventricular tachycardia (VT) is of pivotal importance for an automatic external defibrillator and patient monitoring. In this paper, a VF/VT classification algorithm using a machine learning method, a support vector machine, is proposed. A total of 14 metrics were extracted from a specific window length of the electrocardiogram (ECG). A genetic algorithm was then used to select the optimal variable combinations. Three annotated public domain ECG databases (the American Heart Association Database, the Creighton University Ventricular Tachyarrhythmia Database, and the MIT-BIH Malignant Ventricular Arrhythmia Database) were used as training, test, and validation datasets. Different window sizes, varying from 1 to 10 s were tested. An accuracy (Ac) of 98.1%, sensitivity (Se) of 98.4%, and specificity (Sp) of 98.0% were obtained on the in-sample training data with 5 s-window size and two selected metrics. On the out-of-sample validation data, an Ac of 96.3% $\pm$ 3.4%, Se of 96.2% $\pm$ 2.7%, and Sp of 96.2% $\pm$ 4.6% were obtained by fivefold cross validation. The results surpass those of current reported methods.**

***Index Terms*—Machine learning, public domain electrocardiogram (ECG) database, support vector machine (SVM), ventricular fibrillation (VF) detection.**

I. Introduction

VENTRICULAR fibrillation (VF) and rapid ventricular tachycardia (VT) are dangerous arrhythmic events leading to inevitable death if no defibrillation shock is applied to the subject within a few minutes [1].

In the last decades, a number of algorithms for VF and VT detection [1]–[10] have been proposed. Methods based on processing of the electrocardiogram (ECG) signal in the time domain have included the use of threshold crossing criteria [2], autocorrelation function [3], and conversion of the ECG signal into a binary signal and assessing its complexity [5]. In the frequency domain, studies have proposed the use of bandstop filtering of the signal and estimation of the leakage [6], and spectrum analysis methods [7]. Some novel methods to detect VF include using wavelet transform [8], neural networks [9], and support vector machines (SVMs) [10]. A combination of parameters, which reflect the frequency and morphological ECG characteristics, was found to be an efficient approach for VF detection [11], [12].

Some of the algorithms have been evaluated in [13] and [14]. Amann *et al*. [15] reviewed ten VF detection algorithms and compared their performances under equal conditions using open, published annotated databases. Among the ten algorithms, the signal comparison algorithm (SCA) achieved the best performance with an accuracy (Ac) of 96.2%, a sensitivity (Se) of 71.2%, and specificity (Sp) of 98.5%. Amann's results showed that no algorithm achieved its values for Se or Sp claimed in the original papers because the original researchers made a preselection of the signals by hand. Therefore, using public domain databases is essential to evaluate true algorithm performance.

Some studies have shown that the combination of ECG features extracted from different algorithms may enhance the performance of VF detection [1], [9], [16], [17]. Jekova extracted a set of ten parameters of the ECG signals and employed discriminant analysis to select variables [1]. Four parameters were selected by this approach and the best performance achieved had a Se of 94.1% and a Sp of 93.8%. However, almost all the reported studies used the entire dataset for the development and performance reporting of their algorithms without splitting the dataset into a training and test dataset for use during the development and validation of the algorithms.

In this paper, a VF classification algorithm using a machine learning method is proposed. The algorithm design included a development phase and a validation phase. Three annotated public domain ECG databases were used; the Creighton University Ventricular Tachyarrhythmia Database (CUDB), the MIT-BIH Malignant Ventricular Arrhythmia Database (VFDB), and the American Heart Association Database (AHADB).

In the development phase, 14 VF metrics were extracted from a specific window length (5 s) of ECG. We then used a feature selection technique, a genetic algorithm (GA), to select the optimal combination of variables. The GA mimics the principles of natural selection to "breed" possible successful combinations of parameters, and "kills off" poorly performing combinations of parameters. Every possible combination of selected features was trained and tested using an SVM classifier on the study dataset, which was split further into training and test set. The combinations of different number of features which result in the best performance were selected.

In the validation phase, fivefold cross validation was performed on the study dataset with different window lengths (from 1 to 10 s) to evaluate the algorithm.

Manuscript received May 30, 2013; revised July 16, 2013; accepted July 22, 2013. Date of publication July 26, 2013; date of current version May 15, 2014. This work was supported by Mindray DS USA. *Asterisk indicates corresponding author*.

Q. Li is with the Institute of Biomedical Engineering, School of Medicine, Shandong University, Jinan 250012, China, and also with the Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford, OX1 3PJ, U.K. (e-mail: liqiao@sdu.edu.cn).

C. Rajagopalan is with Mindray DS USA, Mahwah, NJ 07430 USA (e-mail: C.Rajagopalan@mindray.com).

*G. D. Clifford is with the Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford, OX1 3PJ, U.K. (e-mail: gari@robots.ox.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2013.2275000

TABLE I
DETAILS OF TRAINING AND TEST SET

DB	Training			Test			Total	
	include	Nb records	Nb channels	include	Nb records	Nb channels	Nb records	Nb channels
CUDB	odd order	18	18	even order	17	17	35	35
VFDB	odd order	11	22	even order	11	22	22	44
AHADB	odd order	5	10	even order	5	10	10	20
Total		34	50		33	49	67	99

Nb = Number.

TABLE II
NUMBER OF SEGMENTS OF VF AND NON-VF WITH DIFFERENT WINDOW LENGTHS

Window length (s)	Training		Test		Total	
	VF	Non-VF	VF	Non-VF	VF	Non-VF
1	17376	50158	13869	53707	31245	103865
2	8610	24993	6855	26739	15465	51732
3	5694	16596	4530	17769	10224	34365
4	4232	12415	3352	13287	7584	25702
5	3362	9893	2656	10585	6018	20478
8	2114	6134	1656	6568	3770	12702
10	1689	4876	1328	5221	3017	10097

II. METHODS

A. Database

All the records including VF and VT signals from CUDB, VFDB, and AHADB were used in this study. The CUDB includes 35 records of 8 min single-channel ECG data. The VFDB includes 22 records of 35 min two-channel ECG data. The AHADB VT and VF records (No. 8201–8210) include ten records of 30 min two-channel ECG data. There are a total of 67 records and 99 channels of full length annotated ECG in the entire dataset (which is called the study dataset in the rest of the paper).

VF rhythms include ventricular flutter, VF, and VT. As there is no VT annotation in CUDB, the VT segments in CUDB were reannotated by an experienced cardiologist. Non-VF rhythms include normal sinus rhythm, atrial fibrillation, ventricular bigeminy, first degree heart block, high-grade ventricular ectopic activity, nodal rhythm, paced rhythm, sinus bradycardia, supraventricular tachyarrhythmia, and ventricular escape rhythm. Note that segments labeled as asystole in which signals did not exceed 150 μV peak-to-peak and segments labeled as noise were excluded from the study dataset.

VF and Non-VF rhythms were marked segment-by-segment according to the length of an analysis window ranging from 1 to 10 s. The window segments are not overlapping. The 67 records used in the study dataset were split into training and test set using the odd and even record order in the database, e.g., the first record was put into the training set, while the second one went into the test set. Table I shows the details of the training and test set. The total segments of VF and Non-VF with different window length are shown in Table II.

B. ECG Signal Preprocessing Filtration

The sampling frequency of the ECG data was 250 Hz. The applied signal preprocessing included: 1) a high-pass filter with 1-Hz cutoff frequency to suppress residual baseline wander; 2) a second-order 30 Hz Butterworth low-pass filter to reduce high-frequency noise; and 3) a notch filter to eliminate power line interference.

C. VF Metrics

14 VF metrics were extracted based on a review of recent published documents as follows:

1) Complexity Measure (Complexity) [5]: A binary signal calculation is applied to a specific window of the ECG data (e.g., 5 s). A 0/1 binary string is generated by comparing the ECG data to a suitably selected threshold. The mean value x_m of the data points in the selected window is calculated and subtracted from each signal sample. The positive peak value V_p and the negative peak value V_n of the x_i data thus obtained are taken. The number of x_i in the interval $[0 < x_i < 0.1V_p]$ is denoted P_c and the x_i in the interval $[0.1V_n < x_i < 0]$ by N_c. If $(P_c + N_c) < 0.4n$, where n is the number of samples in the selected window, the threshold is selected to be $T_d = 0$; else if $P_c < N_c$, then $T_d = 0.2V_p$, otherwise $T_d = 0.2V_n$. The ECG data are converted into a string (*Binary_s*): if $x_i < T_d, s_i = 0$, or else $s_i = 1$.

A complexity measure *c(n)* is then computed using the following procedure: If S and Q represent two strings and *SQ* is their concatenation. *SQp* is the string *SQ* without its last element. At the beginning, *c(n)* $= 1$, $S = s_1, Q = s_2$, *SQp* $= s_1$. After a number of operations, $S = s_1, s_2, \ldots, s_r$ and $Q = s_{r+1}$. If Q is a substring of *SQp*, S does not change and Q becomes $Q = s_{r+1}, s_{r+2}, \ldots$, etc. until obtaining Q, which is not a substring of *SQp*. S is renewed to be S combined with Q $(S = s_1, s_2, \ldots, s_r, s_{r+1}, \ldots, s_{r+i})$, $Q = s_{r+i+1}$, and *c(n)* $=$ *c(n)* $+ 1$. The aforementioned procedures are repeated until Q is the last character. The normalized *C(n)* is computed as

$$C(n) = c(n)/b(n) \tag{1}$$

where *b(n)* gives the asymptotic behavior of *c(n)* for a random string, such that

$$b(n) \equiv n/\ln(n). \tag{2}$$

2) VF-Filter Leakage Measure (Leakage) [6]: The VF-filter technique corresponds to a narrow bandstop filter applied to the signal with central frequency equivalent to the mean signal frequency. Its output is the VF-filter leakage. The VF signal is considered to be of quasi-sinusoidal waveform. The mean period of a fixed length of data is obtained as

$$T = 2\pi \sum_{i=1}^{m} |V_i| \left(\sum_{i=1}^{m} |V_i - V_{i-1}| \right)^{-1} \tag{3}$$

where V_i are the signal samples and m is the number of data points.

The narrow bandstop filter is simulated by combining the ECG data with a copy of the data shifted by half a period. The

VF-filter leakage is computed as

$$\text{Leakage} = \sum_{i=1}^{m} |V_i + V_{i-(T/2)}| \left[\sum_{i=1}^{m} (|V_i| + |V_{i-(T/2)}|) \right]^{-1}. \quad (4)$$

3) Spectral Analysis [7]: Each data segment is multiplied by a Hamming window and transformed in the frequency domain by fast Fourier transform (FFT). Four spectrum parameters, the first spectral moment normalized (FSMN) and $A1$, $A2$, $A3$, are obtained from

$$\text{FSMN} = \frac{1}{F} \frac{\sum \text{A}_i f_i}{\sum \text{A}_i} \quad (5)$$

where F is the frequency of the component with the greatest amplitude (called the peak frequency) in the range 0.5–9 Hz; f_i is the ith frequency in the FFT between 0 and 100 Hz; A_i is the corresponding amplitude; $A1$ is the sum of amplitudes between 0.5 Hz and F/2, divided by the sum of amplitudes between 0.5 Hz and 20 F; $A2$ is the sum of amplitudes between 0.7 and 1.4 F divided by the sum of amplitudes between 0.5 Hz and 20 F; $A3$ is the sum of amplitudes in 0.6 Hz bands around the second to eighth harmonics (2–8 F), divided by the sum of amplitudes in the span of 0.5 Hz to 20 F.

4) Time Delay Algorithm (Timedelay) [18]: Based on phase space reconstruction, the signal *s(t)* is plotted in a diagram in the following way: $x(t)$ is plotted on the x-axis, and $x(t+\tau)$ on the y-axis, τ being a proper time constant ($\tau = 0.5$ s for VF detection).

ECG data are down-sampled to a frequency of 50 Hz. Phase space plots $[x(t), x(t+\tau)]$ are plotted on a 40×40 grid, where the 40×40 grid stretches from the minimum to the maximum of the investigated ECG signal. The area of the plot filled by the curve is counted and the time delay is defined by

$$\text{Time delay} = \frac{\text{number of visited boxes}}{\text{number of all boxes}}. \quad (6)$$

5) Bandpass Filter and Auxiliary Counts [11]: An integer coefficient recursive digital filter with central frequency at 14.6 Hz and bandwidth from 13 to 16.5 Hz (–3dB) is applied on the ECG data. With a sampling frequency of 250 Hz, the filter is designed by

$$\text{FS}_i = \frac{14\text{FS}_{i-1} - 7\text{FS}_{i-2} + (S_i - S_{i-2})/2}{8} \quad (7)$$

where S_i is a signal sample with index i and FS_i is the filtered signal sample with index i.

Three auxiliary parameters are calculated from the absolute values of the digital integer-coefficient filter output (FS), named *Count1*, *Count2*, and *Count3*. Each parameter represents the number of signal samples with amplitude values within a certain amplitude range, calculated for a specific window.

1) *Count1*—Range: 0.5*max(|FS|) to max(|FS|);
2) *Count2*—Range: mean(|FS|) to max(|FS|);
3) *Count3*—Range: mean(|FS|) – MD to mean(|FS|) + MD, where max (|FS|), mean (|FS|), and mean deviation (MD) are computed for every 1 s time interval.

6) Covariance Calculation (CovarBin) [1]: Measures the variance of the corresponding binary signal (*Binary_s*) of ECG.

TABLE III
CLASSIFICATION PERFORMANCE OF EACH INDIVIDUAL METRIC (WINDOW SIZE = 5 S)

Metrics	Training set (%)					Test set (%)				
	Ac	Se	Sp	AUC	Ac_B	Ac	Se	Sp	AUC	Ac_B
Complexity	51.08	84.18	39.83	67.62	62.00	54.27	92.13	44.77	74.70	68.45
Leakage	93.38	94.41	93.04	98.16	93.72	95.35	86.03	97.69	98.47	91.86
FSMN	60.10	76.06	54.68	68.19	61.08	61.90	77.33	58.03	72.79	67.89
A1	47.17	71.98	38.73	58.29	55.36	45.32	55.38	42.80	43.68	49.09
A2	87.56	85.75	88.17	93.65	87.43	89.76	74.13	93.68	91.49	82.35
A3	71.73	65.08	73.99	74.66	69.54	73.92	70.29	74.83	80.03	72.56
Timedelay	90.31	98.69	87.46	97.81	93.07	93.93	93.52	94.03	97.56	93.78
Count1	94.28	95.48	93.87	98.49	94.68	94.25	84.30	96.74	95.09	90.52
Count2	**96.33**	97.77	95.85	99.57	96.81	**97.05**	91.42	98.47	99.21	94.94
Count3	96.27	97.53	95.84	99.55	96.68	97.01	89.04	99.01	99.20	94.03
CovarBin	79.28	98.66	72.70	93.01	85.68	83.25	94.80	80.35	94.25	87.58
FreqBin	49.84	90.90	35.88	67.67	63.39	54.06	93.22	44.23	76.70	68.73
AreaBin	82.14	85.87	80.87	93.02	83.37	89.91	93.45	89.02	94.26	91.24
Kurtosis	88.54	98.72	85.08	96.98	91.90	92.61	92.17	92.72	96.80	92.44

7) Frequency Calculation (FreqBin) [1]: Calculated by counting the number of binary signal (*Binary_s*) transitions between "0"and "1" and dividing it to the window length (e.g., 5 s), thus, obtaining the transitions for 1 s.

8) Area Calculation (AreaBin) [1]: Realized by summing the values of the binary signal (*Binary_s*) samples. *AreaBin* is equal to the maximum between the sum of the binary signal sample values and the sum of the inverted binary signal sample values.

9) Kurtosis (Kurtosis) [19]: Calculated as the fourth standardized moment of the ECG

$$\text{Kurtosis} = E\{(x - \mu_x)^4\}/\sigma^4 - 3 \quad (8)$$

where $E\{\}$ is the mathematical expectation operator, μ and σ are, respectively, the mean and standard deviation of the ECG segment x.

D. Feature Selection

The SVM classification performance of each individual metric is shown in Table III for both training and test sets for a 5 s window length. Se measures the proportion of VF segments that have been correctly identified as VF. Sp measures the proportion of non-VF segments that have been correctly identified as non-VF. Ac corresponds to the proportion of segments that have been correctly classified. Note that AUC is the area under receiver operating characteristic (ROC) curve, and Ac_B is the balanced Ac (see Section II-F for detail).

Since it is unlikely that all 14 parameters are useful (and in fact some may end up lowering the performance), a variable selection technique is required. Moreover, with a limited number of patterns from which to learn, it is important to keep the number of free parameters (a function of the number of features) which we need to learn as low as possible. A GA was therefore used to select the optimal subset of variables for VF classification [20]. By defining a "chromosome" to be a 14 element binary vector (each element representing one of the features with a "1"

indicating a feature is selected), the GA efficiently explores the space of variable combinations by "mutating" successful randomly chosen chromosomes. A population of 50 chromosomes was used with a 5% mutation rate, a 5% cloning rate, a 95% cull rate for crossover, and a 50-generation limit; the search space of possible variable combinations was rapidly explored. The fitness function that was minimized was the root mean squared error (rMSE) of a multivariate logistic regression. A bootstrapping procedure was performed by running the logistic regression and evaluating the rMSE on the training set with a 5 s window length. The GA selection was repeated 150 times and the selected variables were sorted by the frequency of selection. Variables with higher rank (more frequently chosen) were selected as the features to be used for classification.

E. Machine Learning Classification and Validation

During algorithm development, we used an SVM classifier (LIBSVM library) [21], [22] with a Gaussian radial basis function kernel defined by $K(x_n, x_m) = \exp(-\gamma \|x_n - x_m\|^2)$, where γ controls the width of the Gaussian kernel and plays a role in controlling the flexibility of the resulting classifier. x_n and x_m are two vectors expressed in the initial feature space. As the dataset is not balanced, a weight sector w is used on VF class data when training, where

$$w = \frac{\text{number of non-VF segments}}{\text{number of VF segments}}. \tag{9}$$

Every possible combination of selected features was fed into the SVM model as input to train the model using the training set with 5 s window length, beginning from the individual features, then pairs, triplets, and so on. The models were then evaluated on the test set. The feature combinations of each number of features with best performance were selected.

In the algorithm validation phase, an SVM was used to evaluate the classification performance of the algorithm. Each selected feature combination was validated using fivefold cross validation on the study dataset. The fivefold cross validation sorted the study dataset to fivefold by records rather than by the ECG segments. Then, fourfold of records were used for training and the last fold of the records was used for evaluation. This process was repeated five times as one integral procedure. The fivefold procedure was repeated randomly 50 times and the average of performance was used for evaluation. This approach was repeated on different window length (from 1 to 10 s).

F. Performance Measurements

We used the Se, Sp, Ac, and AUC to evaluate the performance of the algorithm. Since the dataset is unbalanced, a weight w is used to generate balanced result (Ac_B) when Ac is calculated as follows:

$$\text{Ac}_B = \frac{w\text{TP} + \text{TN}}{w\text{TP} + w\text{FN} + \text{FP} + \text{TN}}. \tag{10}$$

TABLE IV
ORDERED RANKING OF SELECTED VARIABLES BY THE GA OVER 150 RUNS

Rank	Variable name	#Times selected
1	*Leakage*	150
2	*Count2*	150
3	*CovarBin*	150
4	*FreqBin*	150
5	*AreaBin*	150
6	*Kurtosis*	150
7	*Complexity*	147
8	*A3*	123
9	*Count1*	117
10	*A1*	62
11	*FSMN*	46
12	*A2*	4
13	*Timedelay*	0
14	*Count3*	0

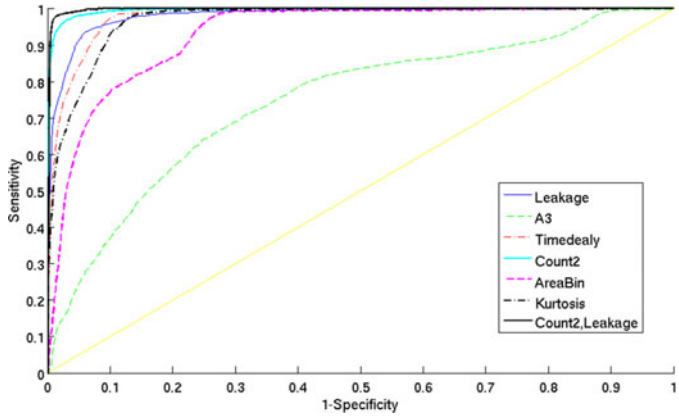


Fig. 1. ROC curves of two selected feature combinations (*Count2* and *Leakage*) and some individual features, window size = 5 s, on training set.

III. RESULTS

A. Feature Selection Result

Table IV shows the rank of variables sorted by the times of selection through 150 repeats of GA. There are six variables which were selected each of the runs and another three variables were selected in more than 50% of the GA runs. We selected these nine variables, including *Leakage*, *Count2*, *CovarBin*, *FreqBin*, *AreaBin*, *Kurtosis*, *Complexity*, *A3*, and *Count1*.

B. Best Performances of Feature Combinations

Table V shows the best performances of feature combinations and an SVM using the training and test dataset with 5 s window length. Fig. 1 compares the ROC curve of two selected feature combinations (Count2 and Leakage) with some individual features.

C. Validation Results

Using the feature combinations acquired from the last step, fivefold cross validation was repeated randomly 50 times using an SVM classifier on the study dataset with different window lengths and the average performance is shown in Table VI.

The best VF classification Ac rate on the evaluation fold by fivefold cross validation was 96.4% when two metrics, Count2

TABLE V
BEST PERFORMANCES OF FEATURE COMBINATIONS (WINDOW SIZE = 5 S)

# of features	Feature combinations	Training (%)					Test (%)				
		Ac	Se	Sp	AUC	Ac_B	Ac	Se	Sp	AUC	Ac_B
1	*Count2*	96.33	97.77	95.85	99.57	96.81	97.05	91.42	98.47	99.21	94.94
2	*Count2*, *Leakage*	98.11	98.39	98.01	99.79	98.20	97.52	89.04	99.65	99.67	94.35
3	*Count2*, *Leakage*, *A3*	**98.23**	98.48	98.15	99.79	98.32	97.55	89.23	99.64	99.68	94.44
4	*Count2*, *Leakage*, *A3*, *FreqBin*	98.16	98.48	98.05	99.80	98.27	97.52	89.16	99.61	99.70	94.38
5	*Count2*, *Leakage*, *A3*, *FreqBin*, *Complexity*	97.97	98.57	97.77	99.80	98.17	97.54	89.38	99.58	99.71	94.48
6	*Count2*, *Leakage*, *A3*, *FreqBin*, *Complexity*, *AreaBin*	97.62	98.54	97.31	99.78	97.93	97.67	90.02	99.58	99.70	94.80
7	*Count2*, *Leakage*, *A3*, *FreqBin*, *AreaBin*, *CovarBin, Kurtosis*	97.38	98.54	96.99	99.77	97.77	**97.70**	90.17	99.59	99.69	94.88
8	*Count2*, *Leakage*, *A3*, *FreqBin*, *Complexity*, *AreaBin*, *CovarBin, Kurtosis*	97.34	98.60	96.92	99.76	97.76	97.69	90.28	99.55	99.69	94.92
9	All 9 features	97.38	98.66	96.95	99.76	97.80	97.46	89.08	99.57	99.64	94.32

TABLE VI
FIVEFOLD VALIDATION RESULTS OF FEATURE COMBINATIONS (WINDOW SIZE = 5 S)

# of features	Training					Evaluation				
	Ac (%)	Se (%)	Sp (%)	AUC (%)	Ac_B (%)	Ac (%)	Se (%)	Sp (%)	AUC (%)	Ac_B (%)
1	96.1±5.6	95.9±0.6	96.2±0.6	99.3±0.1	96.1±0.5	95.7±3.4	95.4±3.3	95.7±4.5	99.2±0.7	95.5±2.6
2	96.9±0.5	96.6±0.6	97.0±0.5	99.6±0.1	96.8±0.5	**96.3±3.4**	96.2±2.7	96.2±4.6	99.6±0.4	96.2±2.3
3	96.9±0.5	96.7±0.5	97.0±0.6	99.6±0.1	96.9±0.5	96.0±3.9	96.2±2.5	95.9±5.1	99.6±0.5	96.1±2.5
4	97.2±0.4	96.9±0.5	97.3±0.5	99.7±0.1	97.1±0.4	95.9±3.7	95.9±2.8	95.8±4.8	99.5±0.5	95.8±2.5
5	97.2±0.4	97.0±0.5	97.3±0.5	99.7±0.1	97.1±0.4	95.8±3.7	95.9±3.0	95.6±5.0	99.5±0.5	95.7±2.5
6	97.0±0.5	97.1±0.4	97.0±0.6	99.7±0.1	97.0±0.5	95.5±4.5	96.1±2.7	95.3±5.9	99.5±0.5	95.7±2.9
7	96.9±0.5	97.2±0.4	96.8±0.6	99.7±0.1	97.0±0.5	95.3±4.7	96.1±2.7	94.9±6.2	99.5±0.7	95.5±3.2
8	96.9±0.5	97.2±0.4	96.8±0.6	99.7±0.1	97.0±0.4	95.3±4.5	96.2±2.7	95.0±6.1	99.5±0.5	95.6±2.9
9	96.7±0.6	97.3±0.4	96.5±0.8	99.7±0.1	96.9±0.5	95.1±4.6	96.2±2.7	94.7±6.1	99.5±0.6	95.4±3.1

and Leakage, were used with 8 and 10 s window lengths, where the corresponding Ac on training folds was 97.0%. With a 5 s window length, the Ac on evaluation was as high as 96.3%. The Ac rate is lower when shorter windows are selected; 96.1% with 4 s, 95.8% with 3 s, 95.2% with 2 s, and 92.7% with a 1 s window. Balancing the data does not substantially change the results.

The combination of the two metrics Count2 and Leakage result in the best performance with each window length group, except 2 and 4 s window groups, where the three metric combinations of Count2, Leakage, and A3 yield the best performance. The classification Ac rate on the training folds increase when the numbers of selected metrics increase from 2 to 5. However, the Ac on the evaluation folds does not increase accordingly. The Se of VF detection increases with increasing number of features. However, the Sp decreases accordingly.

IV. DISCUSSION

The method proposed in this paper selected only two features on a relatively short window length. Thus, it provides real time operation of VF detection.

A bandpass filter with bandwidth from 13 to 16.5 Hz was used to trace the non-VF rhythm complexes when Count2 was being extracted. The filter output for non-VF rhythms shows obviously peaks of QRS; thus, a comparatively small number of signal samples locate in the ranges defined for Count2. However, the output for VF has no clearly defined peaks. It is associated with a higher number of samples in the range of Count2. The Leakage algorithm shifted the data by half a period and then combined them with the original data. The quasi-sinusoidal VF signal will, therefore, be canceled.

Fixed thresholds were used for VF classification in [6], [11], and [13]. A series of thresholds of Count1, Count2, and Count3 and complicated logical judgment was also employed in [11]. These thresholds were selected by retrospective analysis on the entire database without further evaluation. We use a machine learning algorithm to avoid the fixed threshold selection problem. The results show much more robustness on out-of-sample data. Cross validation was performed to mitigate the risk of overtraining.

The Count2 and Leakage metrics appear to provide complimentary information, since the former measures the half bandwidth of the signal centered on the dominant frequency, and the latter measures the power in the sidebands outside of the central frequencies.

Compared to the use of a 10 s window, the Ac using a 5 s window drops only by 0.1% even though there is a 50% shortening of the window size. Reducing the window size to 2 s can still provide a high performance with Ac of 95.2%, Se of 95.1%, Sp of 95.1%, and AUC of 0.992. A shorter window means quicker detection of such potentially lethal rhythms and so possibly faster treatment for the patient. Under zero-noise conditions, a shorter window should not matter too much, but because the episode of VF may be obscured partially by noise, Ac over shorter windows equates to a more accurate alarm under such conditions.

The standard deviation on the evaluation folds is five to ten times larger than that on the training folds after 50 repetitions

TABLE VII
PERFORMANCE COMPARISONS OF THE BEST PERFORMING METHODS CITED IN THE LITERATURE

Algorithms	Ac (%)	Se (%)	Sp (%)	Window size (s)	Database used	# subjects	# events
TCI, Thakor et al. [14]‡	N/A	98	75	8	AHA, CUDB, VFDB	65	161
ACF, Chen et al. [14]‡	N/A	78	32	8	AHA, CUDB, VFDB	65	161
VF-filter, Kuo and Dillman [14]‡	N/A	94	91	8	AHA, CUDB, VFDB	65	161
Spectrum analysis, Barro et al. [14]‡	N/A	79	93	8	AHA, CUDB, VFDB	65	161
Complexity measure, Zhang et al. [14]‡	N/A	66	75	8	AHA, CUDB, VFDB	65	161
Time delay, Amann et al. [18]‡	96.2	79.0	97.8	8	AHA, MIT-BIH, CUDB	123	333583
SCA, Amann et al. [15]‡	96.2	71.2	98.5	8	AHA, MIT-BIH, CUDB	123	333583
Discriminant analysis, Jekova [1]‡	N/A	94.1	93.8	10	AHA, CUDB, VFDB	67	N/A
Filter & Counts, Jekova and Krasteva [11]‡	94.7	94.4	95.9	10	AHA, CUDB, VFDB	67	12254
Imporved Filter & Counts, Fokkenrood, et al.[23]‡	98	97	98	6	MIT-BIH, CUDB, VFDB	105	N/A
Method proposed in this study (training)‡	**98.1**	**98.4**	**98.0**	**5**	AHA, CUDB, VFDB	67	20478
Method proposed in this study (validation)	**96.3±3.4**	**96.2±2.7**	**96.2±4.6**	**5**	AHA, CUDB, VFDB	67	20478
Method proposed in this study (validation)	**95.2±3.3**	**95.1±2.8**	**95.1±4.5**	**2**	AHA, CUDB, VFDB	67	51732

‡Note: all the other methods presented here except our proposed method report statistics on in-sample unbalanced training data. Our in-sample training results exceed all other methods. Out-of-sample results cause a drop of only 2% in performance. MIT-BIH is the MIT-BIH arrhythmia database.

of fivefold cross validation. This suggests that there are characteristic diversities of VF between patients, and a systemic validation is necessary for evaluating the performance of the algorithm.

Table VII shows the performance comparison of some of the cited methods. Note that the performance of our method is reported both on the in-sample training dataset and out-of-sample validation dataset; the others report results when using the entire dataset.

The study is limited by the data in the standard databases. These are often hand-picked to be of high quality, and so that the performance in the presence of severe noise could be examined. This is of course true of all previous studies. Furthermore, a more exhaustive set of features (beyond the current literature) were not explored, since it was assumed that the previous authors had already examined most of the promising approaches. The point of this paper was to illustrate the performance of previously published metrics (and combinations of them) in a statistically valid manner, with out-of-sample testing, which has not been previously attempted in the literature.

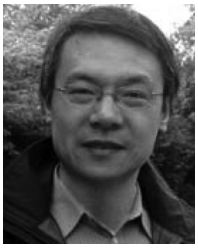
V. CONCLUSION

The method proposed in this paper performs almost equivalently on both the in-sample training and out sample validation dataset. All other methods reported in the literature have reported in-sample statistics, and therefore, a much lower performance for those techniques should be expected in reality. Such an issue was seen when we performed univariate tests on the metrics used by other authors. In our paper, we used a systematic cross-validation approach to reduce the possibility of overtraining, and we have demonstrated that we are able to produce a best-in-class VF detector across a diverse set of databases. In particular, we are able to produce high performances with smaller windows sizes than have previously been attempted. Moreover, our method is rapid to implement since it only requires two relatively simple features.

REFERENCES

[1] I. Jekova, "Shock advisory tool: Detection of life-threatening cardiac arrhythmias and shock success prediction by means of a common parameter set," *Biomed. Signal Process. Control*, vol. 2, pp. 25–33, Jan. 2007.
[2] N. V. Thakor, Y. S. Zhu, and K. Y. Pan, "Ventricular tachycardia and fibrillation detection by a sequential hypothesis testing algorithm," *IEEE Trans. Biomed. Eng.*, vol. 37, no. 9, pp. 837–843, Sep. 1990.
[3] S. Chen, N. V. Thakor, and M. M. Mower, "Ventricular fibrillation detection by a regression test on the autocorrelation function," *Med. Biol. Eng. Comput.*, vol. 25, pp. 241–249, May 1987.
[4] J. Millet-Roig, J. J. Rieta-Ibanez, E. Vilanova, A. Mocholi, and F. J. Chorro, "Time-frequency analysis of a single ECG to discriminate between ventricular tachycardia and ventricular fibrillation," *Comput. Cardiol.*, vol. 26, pp. 711–714, 1999.
[5] X. S. Zhang, Y. S. Zhu, N. V. Thakor, and Z. Z. Wang, "Detecting ventricular tachycardia and fibrillation by complexity measure," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 5, pp. 548–555, May 1999.
[6] S. Kuo and R. Dillman, "Computer detection of ventricular fibrillation," in *Proc. Comput. Cardiol.*, 1978, pp. 347–349.
[7] S. Barro, R. Ruiz, D. Cabello, and J. Mira, "Algorithmic sequential decision making in a frequency domain for life threatening ventricular arrhythmias and imitative artifacts: a diagnostic system," *J. Biomed. Eng.*, vol. 11, pp. 320–328, Jul. 1989.
[8] J. N. Watson, G. R. Clegg, M. Holzer, F. Sterz, and C. E. Robertson, "Evaluating arrhythmias in ECG signals using wavelet transforms," *IEEE Eng. Med. Biol. Mag.*, vol. 19, no. 5, pp. 104–109, Sep. 2000.
[9] R. H. Clayton, A. Murray, and R. W. F. Campbell, "Recognition of ventricular fibrillation using neural networks," *Med. Biol. Eng. Comput.*, vol. 32, pp. 217–220, Mar. 1994.
[10] F. Alonso-Atienza, E. Morgado, L. Fernandez-Martınez, A. Garcıa-Alberola, and J. L. Rojo-Alvarez, "Combination of ECG parameters with support vector machines for the detection of life-threatening arrhythmias," *Comput. Cardiol.*, vol. 39, pp. 385–388, 2012.
[11] I. Jekova and V. Krasteva, "Real time detection of ventricular fibrillation and tachycardia," *Physiol. Meas.*, vol. 25, pp. 1167–1178, Aug. 2004.
[12] V. Krasteva and I. Jekova, "Assessment of ECG frequency and morphology parameters for automatic classification of life-threatening cardiac arrhythmias," *Physiol. Meas.*, vol. 26, pp. 707–723, Jun. 2005.
[13] R. H. Clayton, A. Murray, and R. W. F. Campbell, "Comparison of four techniques for recognition of ventricular fibrillation from the surface ECG," *Med. Biol. Eng. Comput.*, vol. 31, pp. 111–117, Mar. 1993.
[14] I. Jekova, "Comparison of five algorithms for the detection of ventricular fibrillation from the surface ECG," *Physiol. Meas.*, vol. 21, pp. 429–439, Aug. 2000.
[15] A. Amann, R. Tratnig, and K. Unterkofler, "Reliability of old and new ventricular fibrillation detection algorithms for automated external defibrillators," *Biomed. Eng. OnLine.*, vol. 4, no. 60, 2005.
[16] J. Moraes, M. Blechner, F. Vilani, and E. Costa, "Ventricular fibrillation detection using a leakage/complexity measure method," *Comput. Cardiol.*, vol. 29, pp. 213–216, 2002.

[17] I. Jekova and P. Mitev, "Detection of ventricular fibrillation and tachycardia from the surface ECG by a set of parameters acquired from four methods," *Physiol. Meas.*, vol. 23, pp. 629–634, Aug. 2002.
[18] A. Amann, R. Tratnig, and K. Unterkofler, "Detecting ventricular fibrillation by time-delay methods," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 1, pp. 174–177, Jan. 2007.
[19] Q. Li, R. G. Mark, and G. D. Clifford, "Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter," *Physiol. Meas.*, vol. 29, pp. 15–32, Jan. 2008.
[20] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, MA, USA: Addison-Wesley, 1989.
[21] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 27, pp. 1–27, Apr. 2011.
[22] C. C. Chang and C. J. Lin, "LIBSVM—A library for support vector machines." [Online]. Available: http://www.csie.ntu.edu.tw/∼cjlin/libsvm/
[23] S. Fokkenrood, P. Leijdekkers, and V. Gay, "Ventricular tachycardia/fibrillation detection algorithm for 24/7 personal wireless heart monitoring," *Pervasive Comput. Quality Life Enhancement, Lecture Notes Comput. Sci.*, vol. 4541, pp. 110–120, 2007.



Qiao Li received the B.Eng. degree in computer application from Shanghai Jiao Tong University, Shanghai, China, in 1991, the M.Med. degree in physiology from Shandong Medical University, Jinan, China, in 1994, and the Ph.D. degree in biomedical engineering from Shandong University, Jinan, China, in 2008.

He became a Research Assistant in 1994, a Lecturer in 1997, and an Associate Professor in 2001 at Institute of Biomedical Engineering, Shandong University. He took up a Visiting Scholar at MIT from 2006 to 2007, working with Dr. Clifford. In 2011, he took a sabbatical to join Dr. Clifford's research group at the University of Oxford. His research interest includes multidimensional biomedical signal processing, artifact and noise analysis, machine learning, and physiological database.

Dr. Li received the Martin Black Prize in 2009 with Dr. Clifford.

Cadathur Rajagopalan (M'85–SM'06) received the B.S. degree in electrical engineering from the University of Bangalore, Bengaluru, India, the M.S. degree in engineering (instrumentation) from the Indian Institute of Technology, Kharagpur, India, and the Ph.D. degree in control systems and biomedical engineering from the University of Lille, Lille, France.

He spent several years working at the Université de Lille and the Centre Hospitalier d'Arras on problems related to the analysis of physiological signals from the neonatal intensive care unit. He was then with the Department of Cardiology, Oklahoma University Health Sciences Center working on the analysis of signals from isolated myocytes, electrocardiograms, and high-resolution electrograms. He is currently a Biomedical Engineering Fellow in the advanced development division of Mindray DS USA (formerly Datascope), a medical device company. He has been involved in the design of several Food and Drug Administration-approved medical devices including the design of algorithms for the analysis of physiological signals. Some of these devices have won awards including the Medical Design Excellence Award. He is also a Principal Investigator, and part of the clinical and technical due diligence team of the Technology Research Institute, a global research group within Mindray. His current research interests include biomedical signal processing, electrocardiogram analysis, user interface design, and new ways to present physiological data to clinicians. He has several patents, represents Mindray at the Association for the Advancement of Medical Instrumentation, and is on the U.S. technical advisory group of the International Electrotechnical Commission.

Gari D. Clifford (M'02–SM'06) received the B.Sc. (Hons.) degree in physics from the University of Exeter, Devon, U.K., in 1992, the M.Sc. degree in mathematical and theoretical physics from the University of Southampton, Hampshire, U.K., in 1995, and the Ph.D. degree in biomedical engineering from the University of Oxford, Oxfordshire, U.K., in 2002.

In 2003, he became a Postdoctoral Researcher at MIT, transitioning to Research Scientist and then Principal Research Scientist. In 2009, he returned to the University of Oxford to take a faculty position as a Lecturer in Biomedical Engineering. He is also the Director of the Centre for Doctoral Training in Healthcare Innovation, Department of Engineering Science at Oxford's Institute of Biomedical Engineering, where he runs the Intelligent Patient Monitoring Group, focusing on low-cost healthcare technologies, biomedical signal processing, and machine learning applied to a range of chronic and acute healthcare problems. He is also a Fellow of Kellogg College where he directs the Centre for Affordable Healthcare Technology. He is also a James Martin Fellow at the George Institute for Global Health where he is the Acting Director of its Affordable Healthcare Technologies programme. Dr. Clifford is also a Lecturer in Medicine at Harvard University. He has licensed several patents and has been involved in the development of several Food and Drug Administration-approved medical devices and actively collaborates with industry. His research interests include mHealth, electronic medical records, machine learning, multidimensional biomedical signal processing, artifact and noise analysis, linear and nonlinear time-series analysis, and modeling of the electrocardiogram and the cardiovascular system.

Dr. Clifford is on the Editorial Board of IOP's Physiological Measurement and Biomedical Engineering OnLine, as well as the International Advisory Board for the NIH resource PhysioNet. His research group has won several awards including the Martin Black Prize, the Vodafone Foundation Prize, the Wireless mHealth prize, the Engineering World Health prize, and three citations in IOP Phys Measurement's "Select" for "significant impact on future research."