Detection of Ventricular Fibrillation by Sequential Hypothesis Testing of Binary Sequences

J Pardey

Huntleigh Healthcare, Cardiology Products Division, United Kingdom

Abstract

A method is presented for the detection of ventricular fibrillation using binary sequences derived from the surface electrocardiogram. The binary sequences are used to obtain threshold crossing interval and Lempel-Ziv complexity measurements which together form the inputs to a neural network classifier. It is shown that the method outperforms the sequential hypothesis testing of either measurement on the MIT, AHA and CU databases.

1. Introduction

The detection of ventricular fibrillation (VF) by sequential hypothesis testing of measurements taken from the surface electrocardiogram (ECG) has been reported in several papers [1-7]. Two time-domain measurements which demonstrate some utility for this purpose are the threshold crossing interval (TCI) and the normalised complexity of the ECG as quantified using the Lempel-Ziv algorithm. Common to both these measurements is an initial conversion of the ECG to a binary sequence, and the published results for sequential hypothesis testing of either measurement show good performance on small, well-defined training and tests sets. However, the ANSI/AAMI standards EC38:1998 and EC57:1998 [8,9] require VF detection algorithms to be tested on three standard databases, namely the MIT arrhythmia database, the AHA database for evaluation of ventricular arrhythmia detectors, and the Creighton University (CU) ventricular tachyarrhythmia database. This paper presents results for sequential hypothesis testing of TCI or normalised complexity measurements on these databases. The results reveal unacceptable numbers of false positive and false negative detections. To address this the VF detection problem is reformulated in terms of the well-known pattern recognition paradigm of feature extraction and classification. It is then shown that using both TCI and complexity measurements as input features to an artificial neural network (ANN) classifier outperforms the sequential hypothesis testing of either TCI or normalised complexity measurements.

2. Methods

To convert the ECG to a binary sequence the ECG is first divided into fixed-length segments of n samples, $x_1,x_2,\ldots,x_n$, and the DC offset is subtracted from the samples in each segment:

$$\underline{x}_i = x_i - \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (1)$$

Each zero-mean sample, $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n$, is then compared against an amplitude threshold, T_d, to generate a binary sequence, $s_1,s_2,\ldots,s_n$:

$$s_i = \begin{cases} 0 & \text{if } \underline{x}_i < T_d \\ 1 & \text{otherwise} \end{cases} \qquad (2)$$

For a segment length of one second the threshold crossing interval (TCI) is calculated as follows:

$$\text{TCI} = \frac{1000}{N - 1 + t_2/(t_1+t_2) + t_3/(t_3+t_4)} \text{ ms} \qquad (3)$$

where N is the number of pulses in $s_1,s_2,\ldots,s_n$; t_1 is the time from the last threshold crossing in the previous segment to the start of $s_1,s_2,\ldots,s_n$; t_2 is the time from the start of $s_1,s_2,\ldots,s_n$ to the first threshold crossing in $s_1,s_2,\ldots,s_n$; t_3 is the time from the last threshold crossing in $s_1,s_2,\ldots,s_n$ to the end of $s_1,s_2,\ldots,s_n$; and t_4 is the time from the end of $s_1,s_2,\ldots,s_n$ to the first threshold crossing in the next segment. N and t_1–t_4 are illustrated in Figure 1.

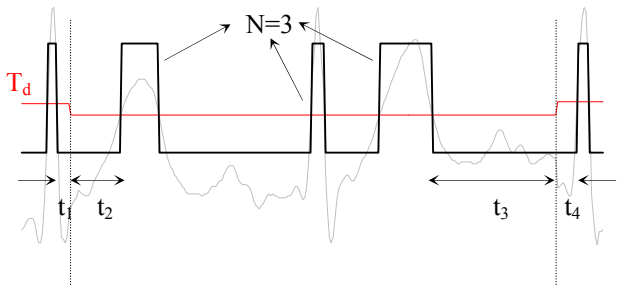

Figure 1. TCI calculation.

The complexity of $s_1,s_2,\ldots,s_n$ is denoted by c_n and calculated using the Lempel-Ziv algorithm [10]. c_n is usually expressed in its normalised form, C_n, where $0 \leq C_n \leq 1$, since C_n is largely independent of n for $n > 1000$:

$$C_n = c_n \log_2 n / n \quad (4)$$

The amplitude threshold, T_d, in (2) is re-calculated for each new segment, $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n$, according to whether TCI or C_n is being measured. If $\underline{x}_{max}$ and $\underline{x}_{min}$ are the maximum and minimum amplitude values in $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n$ then:

$$T_d = 0.2\underline{x}_{max} \quad (5)$$

for TCI [1,2,3,6] and:

$$T_d = \begin{cases} 0 & \text{if } P_c + N_c < 0.4\,n \\ 0.2\underline{x}_p & \text{otherwise} \end{cases} \quad (6)$$

for C_n [4,5,6,7], where P_c and N_c are the numbers of samples, $\underline{x}$, in $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n$ for which $0 < \underline{x} < 0.1\underline{x}_{max}$ and $0.1\underline{x}_{min} < \underline{x} < 0$ respectively, and $\underline{x}_p = \underline{x}_{max}$ if $P_c < N_c$ or $\underline{x}_p = \underline{x}_{min}$ otherwise. Using (5) and (6) on a given segment of ECG yields different binary sequences for TCI and C_n.

For TCI and/or C_n to be useful for VF detection there must be a measurable difference between the values obtained during VF compared to those obtained during non-VF. With this in mind the distributions of TCI and C_n values for various training sets of VF and non-VF data have been investigated in several papers [1-4,7]. The means, μ, and standard deviations, σ, of the Gaussian approximations to these distributions are reprinted in Table 1 for TCI and in Table 2 for C_n with $n \geq 1000$.

Table 1. Published values of μ and σ, in ms, for TCI.

n	μ_{VF}	σ_{VF}	$\mu_{non\text{-}VF}$	$\sigma_{non\text{-}VF}$	Ref.
200	105	6.5	220	16.5	[1]
250	158	16	350	75	[2]
250	180	47	264	156	[2]
n/k	210	62	280	62	[3]

Table 2. Published values of μ and σ for C_n.

n	μ_{VF}	σ_{VF}	$\mu_{non\text{-}VF}$	$\sigma_{non\text{-}VF}$	Ref.
1000	0.595356	0.043018	0.315968	0.077467	[4]
1200	0.589194	0.034976	0.321613	0.085398	[4]
1400	0.587551	0.033346	0.305053	0.076662	[4]
1600	0.587520	0.029779	0.299566	0.070999	[4]
1250	0.236900	0.036900	0.164100	0.027300	[7]

To classify segments of ECG as VF or non-VF via sequential hypothesis testing of TCI or C_n measurements, the sequential hypothesis testing algorithm must first be told what the distributions of TCI or C_n values during VF and non-VF look like. This is done by setting values of μ_{VF}, σ_{VF}, $\mu_{non\text{-}VF}$ and $\sigma_{non\text{-}VF}$ such as those in Tables 1 and 2 on the algorithm, and comparing these two distributions against the TCI or C_n measurements obtained from the segments of ECG to be classified. Specifically, if $X_1,X_2,\ldots,X_m$ is the series of TCI or C_n measurements obtained from m consecutive segments of ECG then sequential hypothesis testing proceeds by evaluating the following log-likelihood function, starting with m=1:

$$F_m = \frac{1}{\sigma^2_{VF}} \sum_{i=1}^{m} (X_i - \mu_{VF})^2 - \frac{1}{\sigma^2_{non\text{-}VF}} \sum_{i=1}^{m} (X_i - \mu_{non\text{-}VF})^2 \quad (7)$$

F_m is then compared against two decision thresholds:

$$F_m \leq 2\ln[ß/(1-\alpha)] + 2m\ln[\sigma_{non\text{-}VF} / \sigma_{VF}] \quad (8)$$

$$F_m \geq 2\ln[(1-ß)/\alpha] + 2m\ln[\sigma_{non\text{-}VF} / \sigma_{VF}] \quad (9)$$

If (8) is satisfied then the VF hypothesis is accepted and the segment of ECG corresponding to $X_1,X_2,\ldots,X_m$ is classified as VF. Conversely if (9) is satisfied the non-VF hypothesis is accepted and the segment of ECG is classified as non-VF. However, if neither (8) nor (9) is satisfied then the test is inconclusive. In this case X_{m+1} is measured and $X_1,X_2,\ldots,X_m,X_{m+1}$ is used to evaluate F_{m+1} which in turn is compared against (8) and (9). This is repeated for F_{m+2}, F_{m+3}, etc., until one of the hypotheses is accepted or the number of steps, m, reaches a predefined upper limit (a lower limit can also be placed on m to avoid reaching a decision 'too hastily'). Either way, sequential hypothesis testing restarts on the next segment of ECG at m=1. The values of α and ß in (8) and (9) are error probabilities. Specifically, α is the probability of rejecting the VF hypothesis when it is true (false negative) and ß is the probability of rejecting the non-VF hypothesis when it is true (false positive). This offers a trade-off between the probability of error, α, ß, and the number of steps, m, required to reach a decision, with larger values of α and ß requiring fewer steps, m. The derivation of (7)–(9) is given elsewhere [1,3].

The sequential hypothesis testing procedure is illustrated in Figure 2. The first segment of ECG is used to obtain a measurement, X_1, which in turn is used to calculate F_1. However, it can be seen that F_1 lies between the two decision thresholds, so no decision is made and the next segment of ECG is used to obtain a second measurement, X_2. X_1 and X_2 are used to calculate F_2 but this also lies between the decision thresholds, so again no decision is made and the third segment of ECG is used to obtain another measurement, X_3. X_1, X_2 and X_3 are used to calculate F_3, which lies above the upper decision threshold, whereupon the segment of ECG corresponding to X_1, X_2 and X_3 is classified as non-VF.

Results for VF detection using sequential hypothesis testing of TCI or C_n measurements have been reported in several of the papers cited earlier [1-3,7]. These results look promising but should be regarded with caution for two reasons. First, the test sets used to generate these

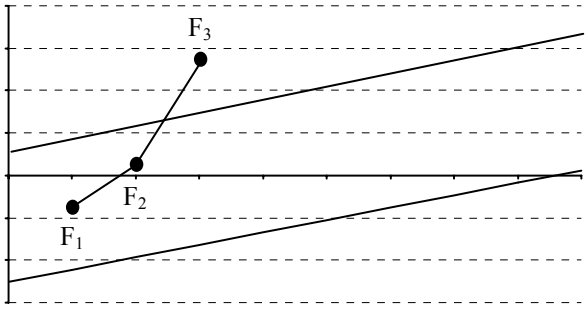

Figure 2. F_m versus m for sequential hypothesis testing.

results are typically small and consist of short, isolated segments of carefully selected VF and non-VF data. Such test sets do not reflect the clinical situation where the ECG is recorded continuously and for long periods of time, and is frequently corrupted by noise. Secondly, in most cases the test set is not independent of the training set since the data used to test the sequential hypothesis testing algorithm are also used to determine the values of μ_{VF}, σ_{VF}, $\mu_{non\text{-}VF}$ and $\sigma_{non\text{-}VF}$ used by the algorithm. For these reasons the results give little if any indication of the expected performance if the method were to be deployed clinically. Furthermore, the use of sequential hypothesis testing itself has two drawbacks. The first is the reliance on a single measurement, TCI or C_n, to discriminate VF from non-VF rather than multiple independent, or quasi-independent, measurements. The second is the reliance on *a priori* Gaussian approximations for VF and non-VF, and in particular the specific values of μ_{VF}, σ_{VF}, $\mu_{non\text{-}VF}$ and $\sigma_{non\text{-}VF}$ used for sequential hypothesis testing. It can be seen in Tables 1 and 2 that each of these values can differ significantly for different training sets, although the reasons for this are currently not understood.

To overcome these limitations a new method was developed that uses TCI and C_n measurements as inputs to a multilayer perceptron (MLP) artificial neural network. The use of neural networks for VF detection has been reported previously [11] but did not include an assessment against the MIT, AHA and CU databases.

The MLP was trained so that TCI and C_n measurements presented to its inputs propagate through the MLP to provide at its output an estimate of the probability that the segment of ECG from which these measurements are taken is VF. This probability is denoted by $p(VF|TCI,C_n)$ where $0 \le p(VF|TCI,C_n) \le 1$; hence if $p(VF|TCI,C_n) > 0.5$ the segment of ECG is classifed as VF, otherwise it is classifed as non-VF. The method also uses an improved ECG to binary sequence converter that takes into account the polarity of the ECG and generates a single binary sequence from which both TCI and C_n can be measured:

$$T_d = \begin{cases} 0.2\underline{x}_{max} & \text{if } \underline{x}_{max} > \underline{x}_{min} \\ 0.2\underline{x}_{min} & \text{otherwise} \end{cases} \quad (10)$$

$$s_i = \begin{cases} 1 & \text{if } \underline{x}_i \ge T_d \ge 0 \text{ or } \underline{x}_i < T_d < 0 \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

The use of (10) and (11) is in preference to (5) and (6) since the latter yield different values of T_d, and hence different binary sequences, for TCI and C_n. Finally, a caveat arising from the scarcity of VF data is that the MLP's training set had to be taken from the AHA and CU databases; therefore the test set is not entirely independent since 3.5% of it is also in the training set.

The MLP was tested on the MIT, AHA and CU databases and its performance compared with sequential hypothesis testing of TCI or C_n measurements on the same databases. All methods were tested in the following way. The ECG was resampled to 250Hz and bandpass filtered using the Medilog ADAPT analysis algorithm [12,13]. VF detection was then performed on each channel separately using TCI and/or C_n measurements in conjunction with (a) sequential hypothesis testing or (b) the MLP classifier. For each method the results from all channels were combined using a logical AND function. This ensures that VF is signalled only where there is temporal overlap across all channels of the VF detections on each individual channel; elsewhere the ECG is classified as non-VF. The duration of an episode of VF extends from the time at which VF is first detected on all channels to the subsequent time at which non-VF is first detected on any channel. Finally, episodes of VF that did not meet a minimum duration requirement of 3 seconds were deleted, and those that did not meet a minimum separation requirement of 3 seconds were merged into a single episode. The values of α and ß used for sequential hypothesis testing were 0.00024 and 0.00089 respectively [1,4]. If no decision was reached after m=10 the segment of ECG corresponding to $X_1,X_2,...,X_{10}$ was assumed to be the same as the segment preceding it and classified as VF or non-VF accordingly. Since the ECGs were resampled to 250Hz only those values of μ_{VF}, σ_{VF}, $\mu_{non\text{-}VF}$ and $\sigma_{non\text{-}VF}$ in Tables 1 and 2 obtained using a value of *n* that is an integer multiple of 250 were selected. The following five methods were therefore compared:

A : Sequential hypothesis testing of TCI measurements using the values in row 2 of Table 1.
B : Sequential hypothesis testing of TCI measurements using the values in row 3 of Table 1.
C : Sequential hypothesis testing of C_{1000} measurements using the values in row 1 of Table 2.
D : Sequential hypothesis testing of C_{1250} measurements using the values in row 5 of Table 2.
E : Sequential hypothesis testing of C_{1250} measurements using the values in row 5 of Table 2.

For all methods both TCI and C_n were measured at one second intervals, which means that adjacent C_n measurements correspond to segments of ECG that overlap by 3

seconds for n=1000 or 4 seconds for n=1250 (adjacent TCI measurements do not overlap).

3. Results

The AHA database comprises 80 two-channel ECGs of 35 minutes duration each. It contains 10 episodes of VF (excluding a one-second episode in record 8206) ranging in duration from 67 seconds to 23 minutes 36 seconds. All 10 episodes of VF occur in records 8201–8210. The MIT database comprises 48 two-channel ECGs of 30 minutes duration each. It contains no episodes of VF but does contain 6 episodes of ventricular flutter in record 207. The CU database comprises 35 single-channel ECGs of 8¾ minutes duration each. It contains 47 episodes of VF ranging in duration from 12 seconds to 5 minutes 26 seconds. 'Hands off' testing was performed for each of the methods A–E in the previous section according to ANSI/AAMI EC38:1998 and EC57:1998 [8,9] and the results are presented in Table 3. The columns in this table list the numbers of true positive (TP), false negative (FN) and false positive (FP) VF detections along with the corresponding percentages for sensitivity (E Se) and positive predictivity (E +P). The results reveal that using both TCI and C_n measurements as inputs to an MLP classifier outperforms the sequential hypothesis testing of either measurement separately.

Table 3. Results for sequential hypothesis testing (SHT) v an MLP on the combined AHA, MIT and CU databases.

	TP	FN	FP	E Se	E +P
A: TCI → SHT	16	47	76	25	17
B: TCI → SHT	23	40	40	36	36
C: C_{1000} → SHT	0	63	0	0	–
D: C_{1250} → SHT	43	20	624	68	6
E: TCI+C_{1000} → MLP	**47**	**16**	**57**	**75**	**45**

4. Discussion and conclusions

A new VF detection algorithm has been presented that uses an MLP neural network to classify segments of ECG as VF or non-VF. The algorithm uses a new ECG to binary sequence converter that takes into account the polarity of the ECG and generates a single binary sequence from which both TCI and C_n can be measured. These TCI and C_n measurements form the inputs to the MLP. The results show an improvement on previous methods but still fall short of the accuracy required for clinical use. To address this two further improvements are being investigated. These are the use of frequency-domain measurements [14,15] as additional MLP inputs alongside the TCI and C_n time-domain measurements, and sequential hypothesis testing of the MLP output values, thereby using the MLP for data fusion rather than for classification. For this it will first need to be established that the MLP output produces values for σ_{VF}, $\mu_{non\text{-}VF}$ and $\sigma_{non\text{-}VF}$ that are independent of the MLP's training set.

References

[1] Thakor NV, Zhu YS, Pan KY. Ventricular tachycardia and fibrillation detection by a sequential hypothesis testing algorithm. IEEE Trans Biomed Eng 1990; 37: 837-43.

[2] Clayton RH, Murray A, Campbell RWF. Comparison of four techniques for recognition of ventricular fibrillation from the surface ECG. Med & Biol Eng & Comput 1993; 31:111-7.

[3] Chen S-W, Clarkson PM, Fan Q. A robust sequential detection algorithm for cardiac arrhythmia classification. IEEE Trans Biomed Eng 1996; 43:1120-5.

[4] Zhang X-S, Zhu Y-S, Thakor NV, Wang Z-Z. Detecting ventricular tachycardia and fibrillation by complexity measure. IEEE Trans Biomed Eng 1999; 46:548-55.

[5] Moraes JC, Blechner M, Vilani FN, Costa EV. Ventricular fibrillation detection using a leakage / complexity measure method. IEEE Computers in Cardiology 2002; 29:213-6.

[6] Amann A, Tratnig R, Unterkofler K. Reliability of old and new ventricular fibrillation detection algorithms for automated external defibrillators. Biomed Eng Online 2005; 4(60).

[7] Chen S-W. Complexity measure based sequential hypothesis testing for real time detection of lethal cardiac arrhythmias. EURASIP Journal on Advances in Signal Processing 2007.

[8] ANSI/AAMI EC38:1998 Ambulatory Electrocardiographs.

[9] ANSI/AAMI EC57:1998 Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms.

[10] Lempel A, Ziv J. On the complexity of finite sequences. IEEE Trans Inf Theory 1976; 22:75- 81.

[11] Clayton RH, Murray A, Campbell RWF. Recognition of ventricular fibrillation using neural networks. Med & Biol Eng & Comput 1994; 32:217-20.

[12] Pardey J, Jouravleva S. The next generation Holter revolution: from analyse–edit–print to analyse–print. IEEE Computers in Cardiology 2004; 31:373-6.

[13] Pardey J. On the use of artificial neural networks in a commercial Holter algorithm. IEEE Computers in Cardiology 2006; 33:485-8.

[14] Kuo S, Dillman R. Computer detection of ventricular fibrillation. IEEE Computers in Cardiology 1978; 347-9.

[15] Barro S, Ruiz R, Cabello D, Mira J. Algorithmic sequential decision making in the frequency domain for life threatening ventricular arrhythmias and imitative artifacts: a diagnostic system. J Biomed Eng 1989; 11:320-8.

Address for correspondence

Dr. J. Pardey.
Huntleigh Healthcare, Cardiology Products Division,
Lion House, Oriental Road, Woking, Surrey GU22 8AP, UK
E-mail: james.pardey@huntleigh-diagnostics.co.uk