

Automated ECG Profiling and Beat Classification

Miad Faezipour, Adnan Saeed and Mehrdad Nourani

Quality of Life Technology Laboratory
The University of Texas at Dallas, Richardson, TX 75080
{mxf042000, axs055200, nourani}@utdallas.edu

Abstract—Recent trends in clinical and telemedicine applications highly demand automation in (electrocardiogram) ECG signal processing and heart beat classification. A real-time patient-adaptive cardiac profiling scheme using repetition detection is proposed in this paper. We introduce a novel local ECG beat classifier to profile each patient's normal cardiac behavior. As ECG morphologies vary from person to person, and even for each person, it can vary depending on the person's physical condition, having such profile is essential for various diagnosis (e.g. arrhythmia) purposes, and can successfully raise an early warning flag for the abnormal cardiac behavior of any individual. Experimental results show that our technique follows the MIT/BIH arrhythmia database annotations with high accuracy.

Index Terms—ECG beat classification, packet processing, cardiac profile, repetition, hash functions.

I. INTRODUCTION

A. Background

The *Electrocardiogram* (ECG) signal is represented by the bioelectrical activity of the heart representing the cyclical contractions and relaxations of the human heart muscles [1]. Once the ECG signal has been picked up, effective pre-processing (denoising) is required for beat detection and feature extraction, as efficient identification of the ECG beats is necessary for arrhythmia analysis. The most important features include the information lying in the P, Q, R, S and T waves of the ECG signal [1]. ECG beats are then classified based on those features in order to detect different types of arrhythmias.

Several works have been done in the area of ECG beat detection. However, many of them are not suitable when high accuracy is needed. Pan and Tompkins developed a method using two adaptive thresholds and a search back algorithm [1] to detect the beats in a ECG signal that successfully performed 99.32% detection when tested on the MIT-BIH open-source arrhythmia database [2]. There is room for possible improvement in the existing algorithm with the development of the *Wavelet Transform* [3] [4] [5] and LabVIEW graphical programming language [6]. In one of our prior work, we identified the possible enhancements in the algorithm to reduce the complexity of the algorithm and to increase the SNR of the ECG signal before detection [7]. In this work, we achieved an overall beat detection accuracy of 99.51% on the MIT-BIH arrhythmia database.

After the preprocessing and beat detection stage, effective beat classification is required for correct analysis of different types of arrhythmia. The general trend is to develop automated systems to classify cardiac beats in real time to help simplify the diagnosis of heart diseases. For example, diagnosis of certain arrhythmia may take up to several hours when measured by hand. Even then, some vital information may be missed in between, due to the tedious manual procedure. Therefore, computer-based beat classification is essential and becoming the norm in clinical applications.

So far, many techniques such as maximum likelihood, (artificial) neural networks [8], and support vector machines [9] [10] have been introduced for ECG beat classification. Machine learning techniques learn from the samples of training data and map new data instances based on the information extracted from the annotated training data.

Support Vector Machines (SVM) have been around for quite some time and have grabbed much attention for ECG beat classification [9] [10] [11]. In this technique, training data sets with known classes are given to the SVM program. Based on the training data, SVM determines hyperplanes in the feature space for each class. The distances between the features of each instance and hyperplanes (classes) are computed and the hyperplane holding the minimum distance indicates that the instance is of that particular class. Each instance of data in SVM contains important features of the heart beat. For example, the instances of an SVM may contain QRS duration, RR interval, amplitude of P, Q, R, S and T points. In other words, each instance is a vector of certain features of a beat, often called *feature vector*. SVM methods provide good classification results using less training data sets and small feature vector sizes. In this paper, a novel repetition-based detection technique has been adopted to classify, or more effectively, *profile* a patient's cardiac behavior.

The artificial neural networks along with the mixture of experts (MOE) approach introduced in [8] is a technique that uses a local classifier in addition to a global classifier in order to classify ECG beats. Our profiling scheme, which is more or less particular to each individual, also falls in the same local-classifier category.

B. Main Contribution

An efficient technique for classifying normal ECG beats is presented in this paper. We introduce a novel technique for profiling a patient's normal ECG. The main idea behind our approach is to consider ECG waveforms as packet streams and apply packet processing techniques, namely repetition-detection approaches [12] [13], to derive certain ECG patterns specific to a patient. Our approach is similar to the concept used in hardware-based string matching and repetition finding techniques used in worm detection in internet networking [13]. Counter units and hashing functions are used for the processing. We derive an adaptive ECG profile for any patient, since ECG morphologies pretty much vary from person to person, as well as condition to condition for the same person. Our technique clearly defines a normal region for a person, and can identify abnormal beats that fall outside this normal region. Our technique is a local beat classifier that can be designed on top of a global classifier for performance enhancement. To the best of our knowledge, this type of highly

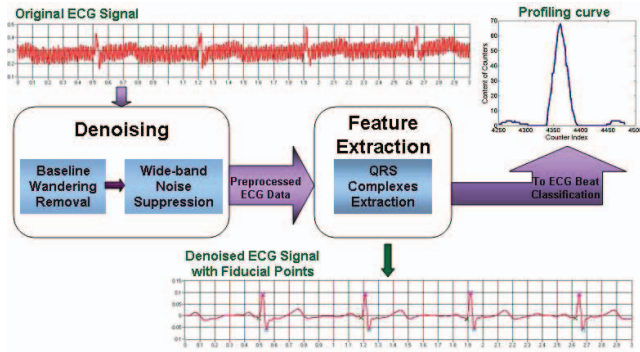


Figure 1. Overall view of cardiac signal processing.

efficient ECG behavior profiling has not been addressed in the literature.

II. PROFILING ECG WAVEFORM

The big picture of our signal ECG signal processes is shown in Figure 1. We first used an efficient method to accurately extract the QRS complexes of an ECG signal by means of wavelet analysis and adaptive thresholding [7]. Our beat classification is based on repetition profiling of the extracted features, as explained hereafter.

The main idea in ECG profiling is based on the underlying concept of pattern matching used in internet packet processing. We treat ECG signals as packets and apply similar methods used in packet processing such as identifying chunks of data packets that match certain particular strings. Similar idea is used in finding malicious internet data packets. All packets are processed and a stream that matches any string corresponding to malicious data is marked as suspicious. In essence, there is a certain pattern (or patterns) against which we compare the data packets.

In the case of ECG packets, there is no global pattern (with fixed waveform dimensions) for a normal ECG waveform, making the problem more complicated. The reason is that ECG is a physiological phenomenon particular to each individual, also depending on physical condition of the person. Therefore, we deploy an approach similar to repetition based pattern detection schemes to derive a normal ECG pattern for an individual.

• String Matching:

In general, string matching in networking applications is performed by comparing packet data bytes to a set of fixed signatures. In this process, data packets chunks are analyzed to check whether any portion(s) of the packet stream matches those signatures. When pre-defined fixed signatures are not available for comparison, repetition based pattern detection techniques can be used. The overall idea in repetition-based pattern detection is to identify strings that are frequently repeated a certain number of times [13]. In this technique, the input stream with total size of $|S|$ bytes is consistently broken into strings of L bytes. The boundaries of these strings with distance b are chosen consistently so that when a long string is repeated, the boundaries within that string are repeated at the same relative locations. To achieve this, the boundaries are decided based on the hash value of a certain sliding window of W -bytes. A boundary (border) is identified when

the hash value is a fixed $l = \log_2 L$ bit prime number. The data bits between two consecutive boundaries are examined for frequent occurrence. To achieve high performance real-time detection, without having to implement numerous counters, the *shared counters* technique is used. In this technique, multiple counters are used in which small storage units are designed to function in parallel. In the design, a multi-level hashing is used in two phases. The data bits between two consecutive boundaries are used to first generate a Phase I hash. This hash value is then used in a second phase hash to produce m independent hash values that index into m arrays. Assuming the width of each memory array to be n , each m hash memory would have 2^n locations in depth. Each array location that is indexed, is incremented. Effectively, m shared counters per signature during monitoring are used instead of one, adding to the accuracy of the conventional counting approach. If, say k out of m counters exceed a certain threshold THR value, the signature is sent to other processing units as a worm suspect for further investigation. More accurate per signature counting is performed to confirm the suspicion afterwards.

In our prior work [13], we approximated the distribution of the number of counters with content x to be a normal (bell-shaped curve) distribution:

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1)$$

where $\mu = N \cdot P = \frac{|S|}{L \times 2^n}$ and $\sigma = \sqrt{\mu}$. Based on the contents of the counters, a bell-shaped profiling curve would be provided, where anomalies would lie on the tail(s) of the curve [13].

• ECG Profiling:

In our approach, similar concept as above is used. We analyze the data in each heart beat and search for repetitions to extract an ECG pattern for any individual.

In our ECG profiling system design, the data chunks we have chosen to use for frequent occurrence analysis is the time and amplitude information that lie in each QRS complex time-frame plus the time difference of the current and previous R points. To be consistent on the length of these data-chunks, we chose a fixed time-frame of 150ms centered at the R point, which corresponds to the largest QRS complex empirically [1]. Since multiple metrics are used here to quantify certain features of a heart beat, a (hash) function such as the summation of these quantities can be used to compact each instance (useful information between the borders of analysis) in a single string of data. The hash result would be the signature of Phase 1. In Phase 2, multiple other (hash) functions can be applied to the Phase 1 signature (e.g. mathematical functions such as mod, etc.) and multiple counters can be used to identify frequent occurrence of these hash results in parallel. We have chosen summation as our primary hash function, since as we will see further in the design, a linear function is used to maintain the range properties. Proper normalization is also required to result in an integer hash value. Essentially, when an R point is detected in the signal, the time and amplitude of all samples corresponding to 150ms of the ECG signal centered at the R point, and the current RR distance are all summed up to form one string (hash value). This string is then further analyzed for frequent occurrence.

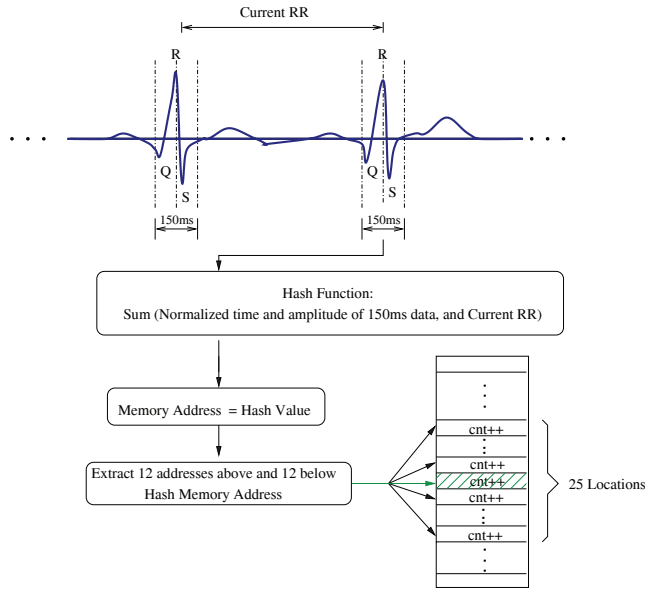


Figure 2. Principles of our ECG profiling system.

Similar to the second phase of the repetition based pattern matching detection technique [13], each hash value is also an index to a presumable memory/counter location. That is, the counter that the hash value points to is incremented. However, since ECG is a physiological signal, each heart beat pattern would not exactly repeat with the same pattern as it had appeared in the previous beat(s). Therefore, hash values would never reproduce the same result, and thus, hash memory contents would never be incremented if solely the above technique is used in our ECG profiling system. This is unlike internet patterns where we search for repeated strings that exactly resemble one another in byte representation. To deal with this issue, we use a valid range where hash values would represent more or less a similar pattern, and increment counters that fall within that range. Pan and Tompkins defined an empirical range of 92 to 116 percent of the average of RR intervals, and claim that beats that fall within this range indicate a normal range that do not require a search back algorithm [1]. Based on this fact, we chose 25, the difference between 92 to 116 as the number of memory locations we index and increment their contents. Essentially, for each hash generation, the content of 25 locations, 12 locations above, and 12 locations beneath the actual hash index would be incremented. Figure 2 shows the principles of our methodology after feature extraction.

In our design, we have used one primary hash function (summation with normalization) and report the results. As discussed in repetition-based pattern detection techniques, multiple hash functions can be used in parallel to form the shared-counter approach. Other linear hash functions that can be interpreted into ranges properly can also be used. The shared-counter approach yields better results in terms of false positives (FP) and false negatives (FN) [13].

• **Parameter Setting:**

Note carefully that to show the true face of this application, we expressed all parameters in actual time measurements

(second and millisecond). These metrics can be normalized to integer values by multiplying them with sampling frequency ($f_s = 360\text{samples/second}$) in our experiment.

- $L \approx RR_{interval}$: which is often in the range of 0.7 to 1.2 seconds representing the time gap expected between two consecutive heart beats.
- $W = 150\text{ ms}$: the size of the sliding window around QRS region. Considering the fact that we use two bytes for amplitudes and time index, W will be equivalent to $2 \times 0.150\text{sec} \cdot \text{samples/sec} \approx 108\text{ bytes}$.
- $b = RR_{interval}$: which indicates that the borders (boundaries) of processing are one heart beat apart.
- $m = 1$: showing only one memory unit in Phase II.
- $n = 10$: address bits of the memory arrays in Phase II with $d \approx 16$ bits wide.
- $k = 1$: indicating our policy in this particular experimentation to evaluate the system's performance using only one memory unit and one hash function (summation with normalized values) in the second phase.
- **THR**: is the threshold value for counters. Since we chose the hash function to be the average content of counters for the best performance, the THR will be any value fallen out of normal range. In other words, the abnormal heart beats are those that do not fit in the bell-shape normal behavior of counter values.

The features that are used in our profiling scheme provide behavior analysis (classification) results for different types of arrhythmia versus the normal ECG behavior. Repetition-detection analysis for features such as the R-R interval and QRS duration provide classification results for the ventricular or atrial premature beat type. In other words, anomalies on the profiling curve indicate the presence of such premature beats. On the other hand, if the information lying in the ST segment of the signal is analyzed for frequent occurrence, abnormalities may indicate the presence of myocardial ischemia episodes leading to certain heart attacks [11]. Essentially, different ECG features or a combination of features, collectively, can classify the beats into normal vs. abnormal beats, where the abnormality type (class) depends on the features used.

III. EXPERIMENTAL RESULTS

We have applied our patient-adaptive profiling scheme on the first-minute time-frame of the data readings in the open-source MIT-BIH arrhythmia database [2]. In this experiment, we have used one hash function (summation with normalized values) in the first and second phase.

In general, abnormal ECG beats are reflected as distortions and humps on the tail of the bell-shaped curve. Highly distorted bell curve indicates the presence of too many irregular beats. Figure 3 shows the distribution of the counter contents for datasets 103, 100, 200, 112, 105 and 232, respectively. As can be seen, a bell-shaped curve represents the normal distribution of the content of the counters. The curve clearly depicts a normal region of the ECG beats. Any non-zero counter content that falls outside the normal region would indicate the existence of abnormal beats. Datasets 103 and 112 do not contain any abnormal beats, as the curves reflect this

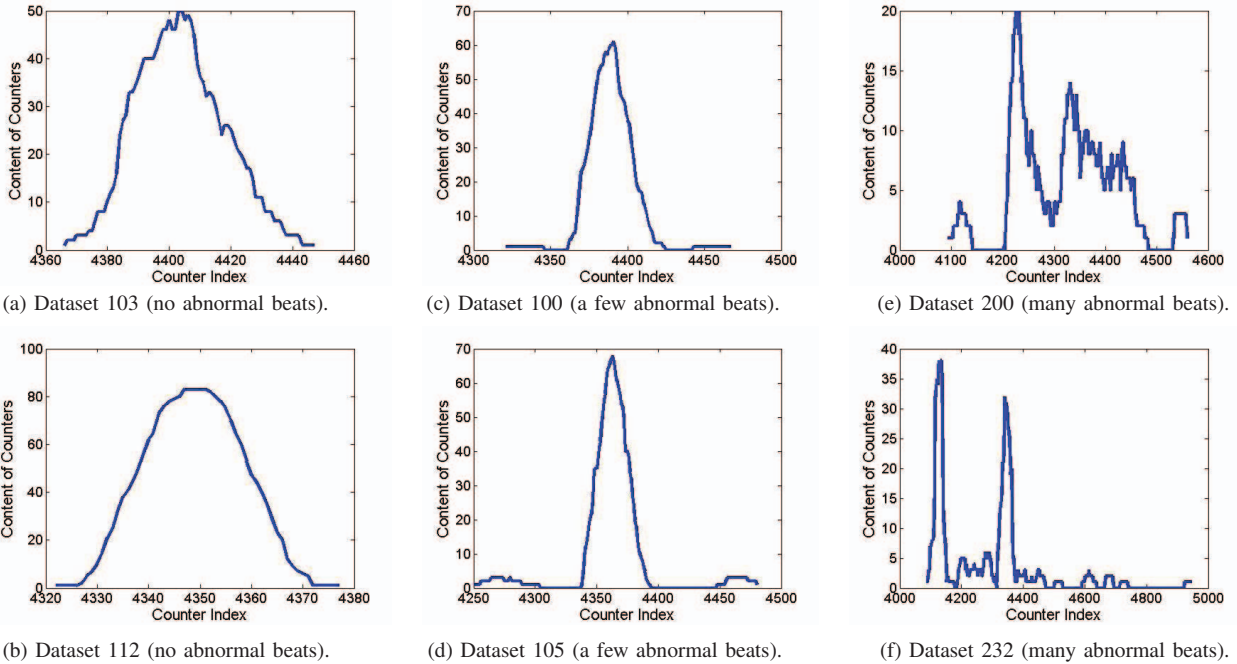


Figure 3. Distribution of counter contents for MIT-BIH arrhythmia database readings.

TABLE I

PERFORMANCE EVALUATION OF OUR ECG PROFILING SCHEME ON MIT-BIH ARRHYTHMIA DATABASE.

Dataset	MIT Abnormal Beats (%)	Our Abnormal Area (%)	Error (%)
mitdb103	0	0	0
mitdb112	0	0	0
mitdb100	1.35	1.35	0
mitdb105	4.18	4.819	0
mitdb200	39.08	41.02	1.94
mitdb232	31.58	30.52	0.06
All	68.12	61.68	6.44

fact. Datasets 100 and 105 contain very few abnormal beats, which is reflected on the curves as very few counter contents outside the normal region. Datasets 200 and 232 contain too many abnormal beats, as a normal region can be hardly defined on the curves. All these results match the MIT-BIH arrhythmia database annotations for beat classification.

To numerically verify our results, we computed the percentage of area under the curve for the abnormal region and compared it with the percentage of abnormal beats within the first minute time-frame of the MIT-BIH database. Table I compares the performance of our scheme with the MIT-BIH arrhythmia database beat annotations.

IV. CONCLUSION

We introduced a technique for profiling the normal ECG waveform for any individual by classifying the detected features of the beats by means of repetition-based packet processing techniques. An error of 6.44% was observed when our approach was tested against the MIT-BIH arrhythmia database. Our technique can efficiently be used for automatic ECG beat classification in early warning monitoring systems.

REFERENCES

- [1] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," in *IEEE Transactions on Biomedical Engineering*, vol. 32, no. 3, pp. 230-236, Mar. 1985.
- [2] MIT-BIH Arrhythmia Database, www.physionet.org/physiobank/database/mitdb.
- [3] C. Li, C. Zheng and C. F. Tai, "Detection of ECG Characteristic Points Using Wavelet Transforms," in *IEEE Transactions on Biomedical Engineering*, vol. 42, no. 1, pp. 21-28, Jan. 1995.
- [4] P. M. Agante and J. P. Marques de Sa, "ECG Noise Filtering Using Wavelets with Soft-Thresholding Methods," in *Computers in Cardiology*, pp. 535-538, Sep. 1999.
- [5] J. S. Sahambi, S. N. Tandon and R. K. P. Bhatt, "Using Wavelet Transforms for ECG Characterization: An On-line Digital Signal Processing System," in *IEEE Engineering in Medicine and Biology Magazine*, vol. 16, no. 1, pp. 77-83, Jan-Feb 1997.
- [6] National Instruments, "User Manuals for LabVIEW Toolset Version 8.5.1", 2008.
- [7] M. Faezipour, T. M. Tiwari, A. Saeed, M. Nourani and L. S. Tamil, "Wavelet-Based Denoising and Beat Detection of ECG Signal," in *Proceedings of the IEEE-NIH Life Science Systems and Applications Workshop, (LISSA'09)*, pp. 100-103, Apr. 2009.
- [8] Y. H. Hu, S. Palreddy and W. J. Tompkins, "A Patient Adaptive ECG Beat Classifier Using a Mixture of Experts Approach," in *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 9, pp. 891-900, Sep. 1997.
- [9] M. H. Song, J. Lee, S. P. Cho, K. J. Lee and S. K. Yoo, "Support Vector Machine Based Arrhythmia Classification Using Reduced Features," in *International Journal of Control, Automation, and Systems*, vol. 3, no. 4, pp. 571-579, Dec. 2005.
- [10] R. Besrou, Z. Lachiri and N. Ellouze, "ECG Beat Classifier Using Support Vector Machine," in *Proceedings of the 3rd IEEE International Conference on Information and Communication Technologies: From Theory to Applications (ICTTA'08)*, pp. 1-5, Apr. 2008.
- [11] M. Mohebbi and H. A. Moghadam, "An Algorithm for Automated Detection of Ischemic ECG Beats Using Support Vector Machines," in *Proceedings of the 15th IEEE Conference on Signal Processing and Communications Applications, (SIU 2007)*, pp. 1-4, Jun. 2007.
- [12] B. Madhusudan and J. Lockwood, "Design of a System for Real-Time Worm Detection," in *Proceedings of the 12th Annual IEEE Symposium on High Performance Interconnects (HOTI'04)*, pp. 77-83, Aug. 2004.
- [13] M. Faezipour, M. Nourani and R. Panigrahy, "A Real-Time Worm Outbreak Detection System Using Shared Counters," in *Proceedings of the 15th Annual IEEE Symposium of High Performance Interconnects (HOTI'07)*, pp. 65-72, Aug. 2007.