Continuous Detection of Abnormal Heartbeats from ECG Using Online Outlier Detection

Yuhang Lin[1⋆], Byung Suk Lee[1], and Daniel Lustgarten[2]

[1] Department of Computer Science, University of Vermont, Burlington, VT, USA
[2] Department of Medicine, University of Vermont, Burlington, VT, USA
{Yuhang.Lin, Byung.Lee, Daniel.Lustgarten}@uvm.edu

Abstract. Detecting abnormal heartbeats from an electrocardiogram (ECG) signal is an important problem studied extensively and yet is a difficult problem that defies a viable working solution, especially on a mobile platform which requires computationally efficient and yet accurate detection mechanism. In this project, a prototype system has been built to test the feasibility and efficacy of detecting abnormal ECG segments from an ECG data stream targeting a mobile device, where data are arriving continuously and indefinitely and are processed online incrementally and efficiently without being stored in memory. The processing comprises three steps: (i) segmentation using R peak detection, (ii) feature extraction using discrete wavelet transform, and (iii) outlier detection using incremental online microclustering. Experiments conducted using real ambulatory ECG datasets showed satisfactory accuracy. In addition, comparing personalized detection (tuned separately for each patient's ECG datasets) and non-personalized detection (tuned aggregated over all patients' datasets) confirms a definite advantage of personalized detection for ECG.

Keywords: ECG · Anomaly Detection · Outlier Detection · Data Stream.

1 Introduction

There has been a significantly large body of work on automatically detecting abnormal segments from electrocardiogram (ECG) signal. Different methods have been used for different work with different objectives, and in this project the objective is real-time online incremental detection with a lightweight computational algorithm. Ideally, the computation overhead should be light enough to run on a mobile platform such as a smartphone. The method chosen with this objective in mind is online outlier detection based on microclustering. Additionally the following choices have been made to support the objective: (1) only one lead ECG is used as opposed to the full 12-lead ECG, and (2) clustering is performed on features extracted from ECG segments using a computationally efficient ($O(N)$) and resilient to errors [14]. Our goal in this paper is to examine the method in light of the objective.

⋆ Current affiliation: Department of Computer Science, North Carolina State University, Raleight, NC, USA, ylin34@ncsu.edu.

The project started on Android smartphone platform, but then migrated to a laptop platform and stayed there until now. Yet, with the real-time processing expectation in mind, the algorithm chosen worked incrementally over incoming ECG data stream with instantaneous processing, i.e., without having to store the data in memory. The processing was done in three steps: (1) segmentation of the ECG data, (2) feature extraction from the ECG segments, and (3) online outlier detection from the features. Segmentation relied on R peak detection equipped with false R peak removal. Feature extraction used Haar discrete wavelet transform. Outlier detection used incremental microclustering [22].

The outcome was evaluated in terms of the detection accuracy using MIT-BIH arrhythmia ECG datasets. When the parameters for outlier detection were tuned personalized to individual patients' datasets, the sensitivity, specificity, and accuracy on average were 83%, 88%, and 92%, respectively, and when aggregated over all patients' datasets using the average parameter values, they were 56%, 87%, and 82%, respectively. In addition, the accuracy was higher when there was a clearer majority between normal or abnormal segments, that is, when the ECG segments were skewed in their distribution of abnormality. These results demonstrated the feasibility and efficacy of the detection method employed and strongly indicated the need for *personalized* detection.

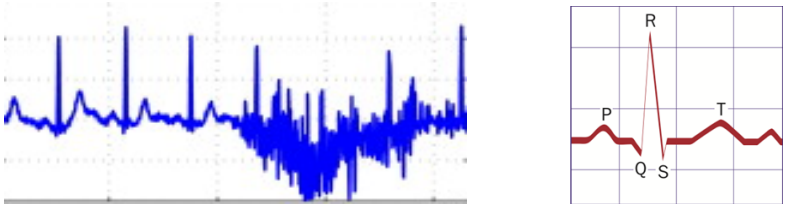
Main contributions of this paper can be summarized as follows: (i) to the best of our knowledge, this is the first project using online outlier detection mechanism to detect abnormal segments from an ECG signal; (ii) comprehensive evaluations on the accuracy of abnormal segment detection presents a new insight into the behavior of the online outlier detection mechanism and an empirical perspective on the merit of personalized anomaly detection as opposed to non-personalized.

Following this Introduction, Section 2 describes the ECG datasets used in the project, Section 3 discusses the steps of anomaly detection process, Section 4 reports the anomaly detection accuracy in the experiment results, Section 5 discusses related work, and Section 6 concludes the paper.

2 ECG Datasets

Electrocardiogram (ECG) is an electrical signal manifesting the heartbeat over time. It is a sequence of segments, one segment per heartbeat. Figure 1a shows a raw (i.e., unfiltered) ECG signal with noise in it. Figure 1b illustrates the composition of an ECG segment – each segment consists of a P wave, a QRS complex, and a T wave.

The ECG datasets used in this project were downloaded from MIT-BIH Arrhythmia Database, which contains 48 half-hour excerpts of two-channel ambulatory ECG recordings obtained from 47 subjects (i.e., patients) studied by the BIH Arrhythmia Laboratory from 1975 to 1979. All recordings were gathered using the sampling rate of 360 samples per second per channel. In this project, we used the ECG signal on the first channel (or lead) in each dataset, namely MLII, for all patients.



(a) Raw ECG signal. (b) ECG segment.

Fig. 1. ECG signal.

The ECG datasets are annotated with the codes denoting the normality or abnormality of segments at each R peak location. A complete list of those annotation codes can be found at the PhysioBank Annotation web site [19]. Table 1 shows the annotation labels used in this project. Database code is the R peak type annotated in the database, and AHA code is the type categorized by the American Heart Association (AHA). We used the AHA code in this project, and thus, the database codes 'V', 'F', '!', 'E','P', 'f','p','Q', '/', and –(empty) were considered abnormal. (The code '!' is a heart beat code but incorrectly listed as a non-beat code, and was corrected in this project.)

Table 1. List of ECG annotation codes used in this project.

Database Code	AHA Code	Description
N	Normal	Normal beats
L	Normal	Left bundle branch block beat
R	Normal	Right bundle branch block beat
A	Normal	Atrial premature beat
S	Normal	Supraventricular premature or ectopic beat (atrial or nodal)
!	Normal	Ventricular flutter wave
V	Abnormal	Premature ventricular contraction
j	Abnormal	Nodal (junctional) escape beat
F	Abnormal	Fusion of ventricular and normal beat
f	Abnormal	Fusion of paced and normal beat
E	Abnormal	Ventricular escape beat
Q	Abnormal	Unclassifiable beat
/	Abnormal	Paced beat
–	Abnormal	No annotation found within this segment

The downloaded ECG datasets have already been filtered through a bandpass filter to retain only the frequency range 0.1 Hz to 100 Hz and digitized at 360 Hz using hardware built at the MIT Biomedical Engineering Center and at the BIH Biomedical Engineering Laboratory [18]. Some ECG datasets still contained significant noise, which had adverse effect on the resulting accuracy, while most others were stable enough to be used without such an effect (see Section 4.2).

3 Approaches

This section discusses the specific approaches used in this project.

3.1 Overview

Figure 2 shows a high-level overview of the steps for detecting abnormal ECG segments from an incoming ECG signal. First, the signal is divided into consecutive segments (or heartbeats). Then, in the feature extraction step, each segment is transformed to a feature vector, which is mapped to a point in a feature vector space. The outlier detection algorithm then picks out those points farther than a threshold from other points. These outliers are considered abnormal ECG segments. In effect, the detection is done as an incremental unsupervised binary classification of each segment, i.e., either normal or abnormal. Each step is discussed in more detail in the rest of this section.

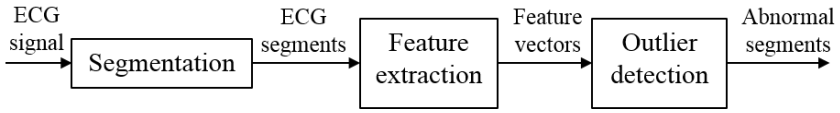


Fig. 2. Overview of the approach to continuous anomaly detection from ECG signal.

3.2 Segmentation

Segmentation comprises three steps: R peak detection, segment extraction, and false R peak removal.

R Peak Detection There are different algorithms used to detect R peaks from raw ECG data. In this project, Chen and Chen's *moving average based filtering algorithm* [6] was used for its good performance and low computation overhead. This algorithm performs three steps over a moving average of consecutive ECG samples: (1) linear high-pass filtering, (2) nonlinear low-pass filtering, and (3) decision making with adaptive threshold.

In the decision-making (step 3), an adaptive threshold T is updated in each moving window using the formula below

$$T = \alpha\gamma P + (1 - \alpha)T$$

where P is the local maximum newly detected in the waveform, α is the forgetting factor, and γ is the weighting factor to determine the contribution of peak values to threshold adjustment.

It is suggested in their algorithm that the moving average window size M can be 5 or 7 samples, α can be chosen from the range of 0.01 to 0.1, and γ can be 0.15 or 0.2. In this project, we set M to 5, α to a random number from 0.01 to 0.1 at each run, and γ to 0.17 (as its showed higher R peak location accuracy than 0.15 or 0.2).

Segment Extraction After detecting R peaks, the next step is to extract segments from the ECG data. We adopted the following formula, introduced in Veeravalli et al.'s work [23]:

$$P_{window} = QR_{max} + 0.2 * RR_{prev} + 0.1$$

$$T_{window} = 1.5 \times QTc_{max} \times \sqrt{RR_{prev}} - QR_{max}$$

where QR_{max} (= 0.08) is half of the maximum of QRS duration and QTc_{max} (= 0.42) is the maximum value for the QT coefficient in Bazet's formula [3] shown below.

$$QT_cB = \frac{QT}{\sqrt{RR}}$$

The extracted segment spans the P_{window} and the T_{window}.

Figure 3 shows an example ECG segment extracted using the formula. The yellow dot marks the R peak; on its left is the P window, and on its right is the T window.

False R Peak Removal While the adopted R peak detection and segment extraction algorithms worked adequately for most segments, there were quite a number of segments that contained two R peaks, where the first one was a true peak and the second one was a false peak. We, therefore, added one more step to remove the false second R peak from the segment. Specifically, if any extracted segment has two R peaks and if the second R peak is within the T_{window}, it is detected as a false R peak, and the end of the segment is cut 15 samples before it. This reduction length of 15 was chosen as a result of manually checking the results for different reduction lengths ranging from 0 to 25 at the increment of 5.

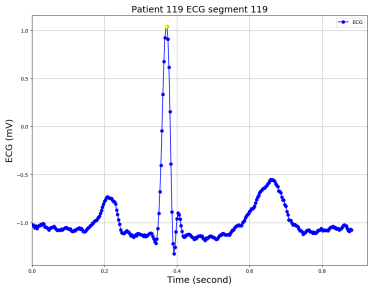


Fig. 3. A sample ECG segment.

3.3 Feature Extraction

Discrete wavelet transform (DWT) was used to transform each ECG segment to a feature vector. This step is, in effect, reducing the dimensionality of an ECG segment of approximately 300 samples to a feature vector of 32 coefficients. Daubechies wavelets and Haar wavelets were compared, and Haar was chosen for its better signal restoration ability and faster speed.

Haar wavelet transform takes a pair of consecutive numbers from the input sequence, calculates their pairwise average and puts the result in the first half, and calculates their pairwise difference and puts the result in the second half. Then, by taking the first half, which contains only the pairwise averages, we can approximate the original signal, and repeating this process, we can reduce the number of coefficients to half each time. We continued repeating until we had 32 coefficients in the first half, and then extracted the first 32 coefficients as the feature vector. In this project, it typically took 3 iterations to finish the process.

Symmetric padding: Given the recursive two-way division performed by Haar wavelet transform, the input length (i.e., number of data elements) should be a power of 2. Since an ECG segment length (i.e., number of ECG samples in it) is not a power of 2 for most segments, a certain number of data elements should be added to make it a power of 2. There are several different ways to do it [21],

and we chose the *symmetric padding*, which mirrors the data to increase the length to the nearest next power of 2. For example, if the segment has n samples $x_1x_2...x_{n-1}x_n$, it is mirrored on both sides to $...x_2x_1|x_1x_2...x_{n-1}x_n|x_nx_{n-1}...$ symmetrically until the resultant length is a power of 2.

3.4 Outlier Detection

The output from the feature extraction step is a continuous stream of feature vectors of 32 coefficients. Each feature vector is mapped to a point in a 32-dimensional feature space. In feature extraction, ECG segments that have similar shapes are mapped to points at similar coordinates in the feature space. Thus, normal segments, which have similar shapes, are mapped to similar coordinates and form a cluster. Abnormal segments, on the other hand, are mapped to "outliers", i.e., points far from other points in a cluster. So, outlier detection is an effective mechanism to identify a point farther off from others, and such outliers translate to abnormal ECG segments.

We adopted an outlier detection method called "*Microcluster-based Continuous Outlier Detection (MCOD)*" [14, 8]. This method is one of popularly used distance-based outlier detection algorithms [22], and works well as long as there is a majority between normal or abnormal segments. The algorithm requires the following three parameters to detect outliers.

- r: maximum allowed radius from a point
- k: minimum number of points required within the radius
- w: size (i.e., number of points) of a moving window

Microcluster-Based Continuous Outlier Detection (MCOD) MCOD is a distance-based outlier detection (DBOD) algorithm over a data stream, enhanced from Continuous Outlier Detection (COD) [14]. Figure 4 illustrates the distance-based outlier detection approach. The point a is not an outlier because there are 3 points b, c, and e within distance r from a. In contrast, the point e is an outlier because there is only one (i.e., less than 3) point b within the distance r from a [12, 13].

COD is computationally efficient in handling two cases – a new point entering the window and an old point leaving the window. In the former case, it checks if any existing outliers should become inliers after the addition of the new point. In the latter case, it checks if any existing inliers should become outliers after the removal of the old point. To handle these two cases, COD supports a range query to find points within the distance r and uses an event-based queue to check if an inlier becomes an outlier because of a point removed from the window.

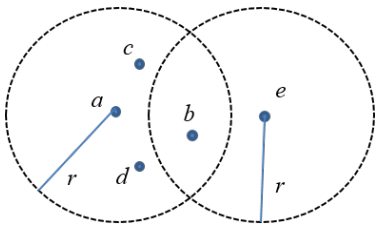


Fig. 4. Distance-based outlier detection when $k = 3$ (source: [8]).

Executing a range query can be expensive, especially when the dimensionality of points is high. MCOD can greatly reduce the number of range queries performed, thereby improving the overall performance of the algorithm. A microcluster can be thought of as a small sphere in the data space. The algorithm requires a microcluster to be of radius $r/2$ and contain more than k points at all time. Any point that belongs to a microcluster is never an outlier because there always exist more than k points within the range r in the same microcluster. In contrast, any point that does not belong to any microcluster is very likely to be an outlier. Every time a new point arrives in a stream, if the window is not full, then no point is removed. Otherwise, the oldest point is removed and, if it belongs to a microcluster, the number of members in that microcluster is reduced by one and, if the resulting number drops below $k+1$, the microcluster is removed and for all its members, their lists of nearest microcluster centers within a distance of $3r/2$ is updated.

MCOD may label an ECG segment as an outlier when it enters a window and then later change it to inlier. In this project, an ECG segment is considered an outlier only if it is labeled as an outlier throughout from the time it enters the sliding window till the time it exits the window.

3.5 Complexity Analysis

The outlier detection algorithm MCOD can tell if an ECG segment is normal or abnormal only after the feature point mapped from the segment passes through the window completely. Thus, the complexity can be expressed in terms of the segment size s and the window size w. The three-step approach – comprising segmentation, feature extraction, and outlier detection – requires $O(s + wk)$ memory space, where s is the largest ECG segment size, and takes $O(sw)$ run time, as discussed below.

Space Complexity The segmentation step requires linear processing of the incoming ECG data samples, and memory buffer large enough to hold the largest ECG segment suffices to support this processing. So, the space complexity is $O(s)$. The feature extraction step processes each segment at a time, and for each segment the run time is proportional to the segment size. Therefore, it requires only the buffer space to hold the largest segment, hence $O(s)$. The outlier detection step requires $O(wk)$ space. Readers are referred to the MCOD paper [14] for a proof of this complexity.

Time Complexity Since the segmentation step processes ECG data samples linearly, it takes $O(sw)$ to generate w feature points in the window of the outlier detection step. The feature extraction step to generate the w feature points in the window is $O(sw)$ as well because for each segment DWT takes linear time with the segment size and each segment generates one feature point in the window. The outlier detection step takes $O(l \log w) + O(m)$, where l is the average number of times feature points are re-labeled as outliers as a result of an old point removed from the window of size w and m is the average number of feature points within the maximum allowed radius r (discussed in the MCOD paper [14]). While the theoretical worst case time complexity of this step could

be $O(w \log w)$, in practice it is near $O(\log w)$ because both $l \ll w$ and $m \ll w$ hold. Thus, the total run time complexity for all three steps is $O(sw)+O(\log w)$, which asymptotically equals $O(sw)$.

4 Evaluations

This section presents the setup, results, and analysis of the experiments performed to evaluate the accuracy of detecting abnormal segments from an ECG data stream.

4.1 Experiment Setup

Development platform: The main development platform was Windows 10 laptop with 2.6GHz dual core CPU, 8GB RAM, and 240GB SSD. In addition, a virtual Ubuntu Server with 2.2GHz single core CPU, 1GB RAM and 25GB SSD, running MySQL and PHPMyAdmin on DigitalOcean cloud server was used to train and test algorithms and store the experiments results.

Datasets: MIT-BIH Arrhythmia datasets described in Section 2 were used in the experiments. (Due to space cosntraint, tables in this section show results from 20 randomly selected datasets. Results from all 48 datasets are available at `https://github.com/yuhang-lin/ECGAD_extended_result`.) Each ECG dataset was divided into training and testing datasets with 60%-40% split. Table 2 shows the number and ratio of abnormal segments in training dataset and testing dataset, respectively, for each dataset. Different patients show different abnormal segment ratios (i.e., ratio of abnormal segments over all segments), and the ratio varies widely. Notably, the patent 122 is a healthy patient with no heartbeat anomaly, and the patients 222 and 231 are in good shape as well, with only a few abnormal heartbeats. In contrast, the patient 217 is in a very poor shape, with approximately 90% heartbeats abnormal.

Performance measures We used the traditional performance measures – sensitivity, specificity, and accuracy. In light of detecting abnormal segments, (i) true positive means detecting an abnormal segment as abnormal, (ii) true negative means detecting a normal segment as normal, (iii) false positive means detecting a normal segment as abnormal, and (iv) false negative means detecting an abnormal as normal.

Outlier detection parameter tuning: The set of parameter values that maximizes the anomaly detection performance was found using a random search iterated 1000 times for each ECG dataset. 1000 iterations is more than enough, and it gives 99.996% probability of achieving near optimum within 1% of the true optimum. (A random search of n iterations has $1-(1-\epsilon)^n$ probability of finding parameter values achieving near-optimum within the error ϵ from the true optimum [7].) The ranges of each MCOD parameter used in the experiments are [0.1, 3.0) for the radius r, [2,80] for k, and [25,100] for window size w. After the training process of parameter tuning through random search, we picked the set of parameter values that maximized the accuracy, subject to the constraints of minimum 80% required of both sensitivity and specificity. (In case none met

Table 2. Ratios of abnormal segments in training and testing.

Patient Number	Training Dataset			Testing Dataset		
	Number of Abnormal Segments	Total Number of Segments	Ratio of Abnormal Segments	Number of Abnormal Segments	Total Number of Segments	Ratio of Abnormal Segments
106	224	1212	18.48%	296	808	36.63%
114	58	1138	5.10%	4	759	0.53%
116	73	1438	5.08%	38	959	3.96%
118	19	1383	1.37%	19	922	2.06%
119	232	1192	19.46%	213	796	26.76%
122	0	1485	0.00%	0	990	0.00%
200	536	1688	31.75%	466	1126	41.39%
201	135	1173	11.51%	65	782	8.31%
202	21	1281	1.64%	2	855	0.23%
205	39	1591	2.45%	41	1061	3.86%
207	307	1444	21.26%	426	963	44.24%
208	927	1760	52.67%	433	1174	36.88%
213	370	1950	18.97%	212	1300	16.31%
217	1125	1324	84.97%	838	883	94.90%
219	39	1291	3.02%	27	862	3.13%
221	281	1452	19.35%	116	969	11.97%
222	3	1500	0.20%	17	1001	1.70%
228	332	1387	23.94%	237	925	25.62%
231	2	942	0.21%	0	628	0.00%
233	497	1841	27.00%	342	1228	27.85%

(Ratios of all 48 datasets are available at `https://github.com/yuhang-lin/ECGAD_extended_result/blob/master/abnormal_segment_ratio.md`.)

the constraint, the lower bound was lowered progressively until one was found.) When more than one set of parameter values gave the same accuracy, then the one that had the smaller window size was picked because a smaller window size can output the outlier quicker and can be more robust when the input data is smaller.

4.2 Experiment Results and Analysis

The results are presented in two different scenarios. One is *personalized*, where the parameters are tuned for individual patients as discussed in Section 4.1. The other is *aggregated*, where the average of the individual optimal parameter values are used as generic parameter values for all patients.

Personalized Results Table 3 summarizes the accuracy, sensitivity, and specificity obtained for each patient's ECG dataset when the outlier detection parameters were optimized for each dataset separately. This case reflects personalizing the anomaly detection for individual patients.

Overall, the performance using personalized parameters is good. 37 out of 48 datasets achieved accuracy higher than 90%. Note from Table 2 that ECG

Table 3. Best testing performance on each patient using personalized parameters (sensitivity N/A for zero abnormality ratio).

Patient #	Sensitivity	Specificity	Accuracy	Optimal Parameter Values		
				Window	K	Radius
106	0.98	0.97	0.98	97	45	2.19
114	0.67	0.95	0.95	80	14	0.92
116	0.97	0.97	0.97	87	13	3.05
118	0.74	0.94	0.93	65	4	1.65
119	1.00	0.89	0.92	31	21	2.58
122	N/A	1.00	1.00	28	3	2.53
200	0.92	0.94	0.93	74	40	2.10
201	0.94	0.96	0.95	73	36	1.42
202	1.00	0.93	0.93	85	5	0.91
205	1.00	1.00	1.00	98	65	1.33
207	0.28	0.97	0.65	59	25	1.90
208	0.96	0.58	0.72	75	26	1.67
213	0.81	0.87	0.86	80	17	1.73
217	1.00	0.00	0.94	99	80	1.08
219	0.68	0.97	0.97	60	5	3.04
221	0.96	1.00	0.99	44	19	2.12
222	0.06	1.00	0.98	25	4	2.07
228	0.60	0.91	0.83	74	33	2.09
231	N/A	1.00	1.00	25	14	2.87
233	0.97	0.94	0.95	27	10	3.07

Average from 48 datasets: sensitivity 0.83, specificity 0.88, accuracy 0.92.
(Results for all 48 datasets are available at `https://github.com/yuhang-lin/ECGAD_extended_result/blob/master/personalized_result.md`.)

datasets of the patients 122 and 231 have no abnormal segments in the testing data, so the sensitivity for them is not applicable (N/A).

For the ECG datasets of patients 207, 208, 213, and 228, the accuracy was lower than 90%, as low as 65% for the patient 207. There are a few reasons we believe can explain these lower accuracies.The first reason is the noise in the filtered dataset. The ECG datasets are from ambulatory devices, which cause significant noises such as baseline drifts, motion artifacts, and powerline noise. Although filtered, some datasets still show significant noise. (Figure 5 illustrates typical noisy segments from dataset 207.) Further removing noise from pre-filtered data would require sophisticated signal processing, and was beyond the scope of the project. The second reason is the change of statistics between training dataset and testing dataset during the performance evaluation. This in part can be reflected by the different abnormal segment ratio between training and testing as shown in Table 2. In this project, the performance tuning is not adaptive to such a change (called "concept drift") and, therefore, the algorithm may not be able to react to unexpected changes by adjusting the tuned parameter values.

Abnormal segment ratio and accuracy measures: Figure 6 shows the trend of accuracy, sensitivity, and specificity for datasets sorted by the abnormal segment

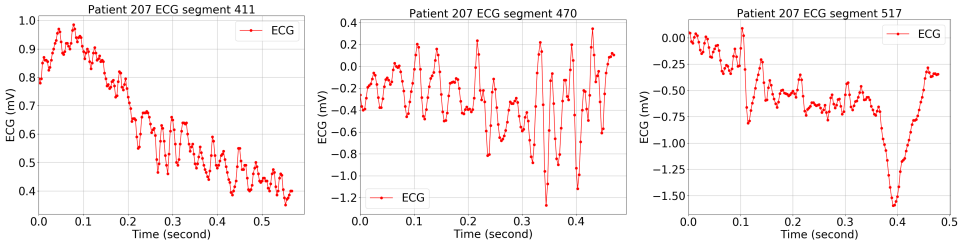


Fig. 5. Example noisy segments from the ECG of the patient 207. (The ECG dataset of patient 207 in particular shows the lowest accuracy overall among all datasets. Looking into the dataset in detail, we found that segments in this dataset have several different types of normal segments, such as 1457 left bundle branch block beats (L), 86 right bundle branch block beats (R), and 107 atrial premature beats (A), as well as several different types of abnormal segments such as 105 premature ventricular contractions (V), 472 Ventricular flutter waves(!) and 105 Ventricular escape beats (E) (see Table 1 for different annotation codes of ECG segments). Indeed, this dataset is mentioned as "an extremely difficult record" [20] in PhysioBank.)

ratio in the testing dataset. Note that the distribution of abnormality ratios in the datasets is skewed to approximately 45% or lower and approximately 90% or higher. The achieved accuracy is in a fairly consistent range across the two skewed ranges of abnormality ratio, which indicates robustness of the employed outlier detection mechanism to the ratio. The sensitivity shows a similar trend, but it drops very low when the ratio is near zero ($< 1\%$). It makes sense because lower ratio means fewer abnormal segments (i.e., true positives) and, hence, lower statistical significance. In contrast, the specificity drops very low when the ratio is near 1 ($> 90\%$). It makes sense because higher ratio means fewer normal segments (i.e., true negatives) and, hence, lower statistical significance.

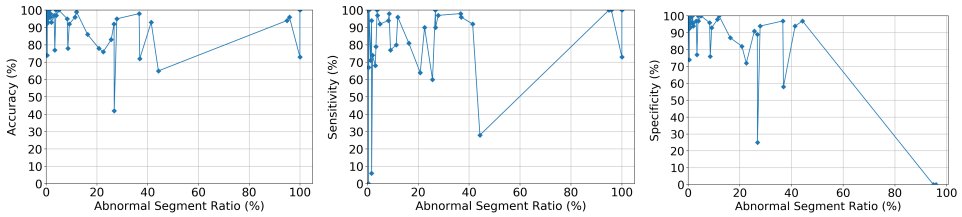


Fig. 6. Accuracy measures for different abnormal segment ratios.

Aggregate (i.e., Non-personalized) Results Based on the optimum parameter values determined in the personalized anomaly detection experiment (see Table 3), we calculated their mean values as the generic parameter values used commonly for all 48 patients, namely, non-personalized. The mean values are 62 for w, 17 for k, and 1.8 for r. Table 4 shows the resulting performances.

The accuracy achieved using the non-personalized approach is lower than that of the personalized approach for 31 out of 48 datasets, although 24 datasets still achieved accuracy higher than 90%. Overall there were significant degradation of accuracy. The datasets for patients 217 and 219 in particular sustained the

Table 4. Performance on each patient when not personalized (window $w = 62$, $k = 17$, and radius $r = 1.8$ for all patients' datasets; sensitivity N/A for zero abnormality ratio).

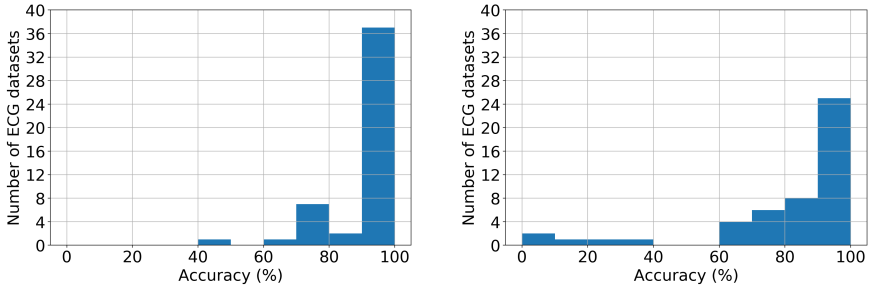
Patient #	Sensitivity	Specificity	Accuracy	Patient #	Sensitivity	Specificity	Accuracy
106	0.68	0.96	0.86	207	0.21	0.97	0.61
114	0.33	1.00	0.99	208	0.82	0.75	0.78
116	1.00	0.81	0.82	213	0.84	0.83	0.83
118	1.00	0.87	0.88	217	0.34	0.16	0.33
119	0.30	0.78	0.65	219	1.00	0.14	0.17
122	N/A	0.98	0.98	221	0.97	0.99	0.99
200	0.37	0.98	0.72	222	0.18	0.99	0.98
201	0.03	0.99	0.90	228	0.60	0.91	0.83
202	0.00	0.94	0.94	231	N/A	1.00	1.00
205	0.98	1.00	1.00	233	1.00	0.55	0.68

Average from 48 datasets: sensitivity 0.56, specificity 0.87, accuracy 0.82.
(Results for all 48 datasets are available at `https://github.com/yuhang-lin/ECGAD_extended_result/blob/master/nonpersonalized_result.md`.)

biggest degradation – from 94% to 33% for the patient 217 and from 97% to 17% for the patient 219.

4.3 More on Personalized versus Non-personalized

The histograms in Figure 7 show the number of ECG datasets in each 10% range of accuracy when the anomaly detection was personalized and not personalized. It is visually evident that personalized detection by far outperforms non-personalized detection. Numerically, the chi-squared distance of the personalized histogram from the non-personalizedis 10.2.



(a) Personalized detection. (b) Non-personalized detection.

Fig. 7. Number of ECG datasets in different accuracy ranges (bin size = 10%).

The fairly large difference in the accuracy performance is understood when the distribution of the optimal sets of outlier detection parameters (i.e., w, k, r) are examined, as shown in the scatter plot in Figure 8. It shows the MCOD parameters tuned personalized for each patient's ECG dataset (see Table 3) and also the aggregated mean values of them (i.e., $w = 62, k = 17, r = 1.8$) used in the non-personalized case. The parameter values tuned for different datasets are widely spread in the parameter space, as indicated by their standard deviations 24, 21, and 0.85 for w, k, and r, respectively. These observations confirm that

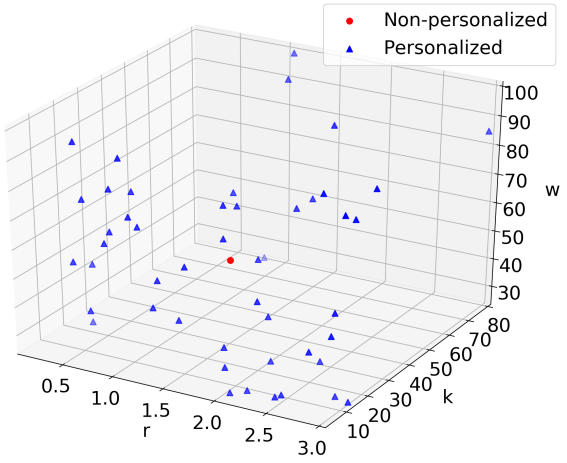


Fig. 8. Scatter plot of personalized MCOD parameters.

ECG varies a lot for individual patients and, therefore, personalized detection is much desired.

5 Related Work

There is a large body of work done on anomaly detection from ECG. In this section, we discuss briefly what we believe are a representative sample reflecting the state of the art in three aspects of this project: (a) machine learning methods used for automatic detection of abnormal ECG segments, (b) feature extraction methods to reduce ECG segments to feature vectors, and (c) distance-based outlier detection from a data stream.

Machine Learning Methods: Various machine learning methods have been used for ECG classification, such as decision tree [4], support vector machine [24], artificial neural network [26], and their ensemble [15]. These methods, however, are geared for offline classifications and are not necessarily handling individual ECG segments separately. In contrast, some recent work are far more suitable for online real-time classification of ECG segments as done in this project. For example, Veeravalli et al. [23] used dynamic time warping (DTW) based similarity calculation, personalized to individual patients by obtaining the normal ECG segment through clustering (K-means). For another example, Chauhanv and Vig [5] used long short-term memory (LSTM) recurrent neural networks (RNN) as a predictive model trained with normal ECG segments to detect abnormal segments. Both methods have the advantage of working well with continuously arriving ECG segments. To the best of our knowledge, there is no prior work that examined using outlier detection based on online clustering, which was the goal in this project.

Feature Extraction Methods: As mentioned in Section 3.3, feature extraction in this project is for dimensionality reduction from an ECG segment to a feature

vector. There have been two different ways the extracted features are used. One way is as a synopsis of an ECG segment characterizing a certain anomaly [9]. For example, a P-R interval can be used to detect premature ventricular contraction, an R-R interval to detect premature atrial heartbeat, and the QRS duration to detect a ventricular premature complex [17]. Another way is as an input model of the segment to a subsequent machine learning algorithm. In this project, it is the discrete wavelet transform (DWT), chosen for its efficiency and resilience to noise. There are several others, such as principal component analysis (PCA) [16], rank correlation coefficient (RCC) [11], and B-splines [10]. They all extract dominant features that represent the ECG signal approximately but differ in the specific sense of dominance. Specifically, DWT selects the first 2^n, where $n = 5$ in our work, coefficients as the dominant features; PCA selects dominant linear components that, when linearly combined, approximates the input signal; RCC selects a subset of ECG data samples whose RCC values are the highest, where RCC is a measure of the correlation based on the ranks of data values; B-splines are used as bases that are linearly combined to fit ECG signal "curve" lines, and the resulting "knots" and parameters of the B-splines are used as the features.

Outlier Detection Methods: In addition to MCOD [8], the distance-based outlier detection method used in this project, there are other methods that can be used for outlier detection. For example, MOA supports the following ones that we believe are in the mainstream of online outlier detection over data streams: Exact-STORM and ApproxSTORM (STORM stands for "Stream Outlier Miner") [1], Abstract-C [25], COD [14] and MCOD, and additionally AnyOut [2]. All of these methods except AnyOut are distance-based methods developed progressively for improvements. (AnyOut is a method enabling the detection of an outlier "any time" the time expires, and is orthogonal to the detection mechanism (e.g., distance-based, density-based)). As shown in the comprehensive experiments conducted by Tran et al. [22], MCOD performs best among all distance-based methods.

6 Conclusion

The feasibility and accuracy of detecting abnormal segments from an ECG data stream using distance-based online outlier detection have been demonstrated in this project. Combined with features extracted using Haar discrete wavelet transform, the microcluster-based continuous outlier detection algorithm successfully detected abnormal ECG segments with higher than 90% accuracy for a majority of datasets. The accuracy performance compared between personalized and non-personalized anomaly detection scenarios showed that personalized showed by far higher accuracies.

There are several issues in the employed algorithms that still warrant further work. First, the outlier detection mechanism was implemented as a binary classifier to normal versus abnormal, without distinguishing among different types of abnormality. It is suggested that the mechanism is extended to a multi-class classifier that can label the anomaly type of abnormal segments. Second, the R peak detection algorithm used in this project has a significant room for improve-

ment so it will not result in multiple R peaks in the same segment as happened in this project. There are more advanced techniques, and one of them should be adopted for better results. Third, the online outlier detection algorithm used in this paper works well when the ECG data stream is stationary, and as a result, the accuracy performance was somewhat inadequate for some ECG datasets. It would be desired to enhance the algorithm to be adaptive to the change of the ECG segment statistic, such as abnormal segment ratio, to adjust the outlier detection parameters according to the change of statistic.

As mentioned in Introduction, the project initially started out on an Android smartphone platform. The project will continue to migrate the program codes of all steps into the Android platform. Then, a performance profile (i.e., the elapsed time of individual steps of the processing algorithm) will be built to assess the real-time "fitness' of the method used in this project.

References

1. Angiulli, F., Fassetti, F.: Distance-based outlier queries in data streams: The novel task and algorithms. Data Mining and Knowledge Discovery **20**(2), 290–324 (March 2010)
2. Assent, I., Kranen, P., Baldauf, C., Seidl, T.: AnyOut: Anytime outlier detection on streaming data. In: Proceedings of the Conference on Database Systems for Advanced Applications. pp. 228–242 (2012)
3. Bazett, H.C.: An analysis of the time-relations of electrocardiograms. Heart **7**, 353–370 (1920)
4. Bensaid, A.M., Bouhouch, N., Bouhouch, R., Fellat, R., Amri, R.: Classification of ECG patterns using fuzzy rules derived from id3-induced decision trees. In: Proceedings of the Conference of the North American Fuzzy Information Processing Society. pp. 34–38 (August 1998)
5. Chauhan, S., Vig, L.: Anomaly detection in ECG time signals via deep long short-term memory networks. In: Proceedings of the IEEE International Conference on Data Science and Advanced Analytics. pp. 1–7 (October 2015)
6. Chen, H.C., Chen, S.W.: A moving average based filtering system with its application to real-time QRS detection. In: Computers in Cardiology. pp. 585–588 (September 2003)
7. Firebug: Practical hyperparameter optimization: Random vs. grid search. `https://stats.stackexchange.com/q/209409` (2016), [Online; accessed 18-April-2018]
8. Georgiadis, D., Kontaki, M., Gounaris, A., Papadopoulos, A.N., Tsichlas, K., Manolopoulos, Y.: Continuous outlier detection in data streams: An extensible framework and state-of-the-art algorithms. In: Proceedings of the ACM SIGMOD International Conference on Management of Data. pp. 1061–1064 (June 2013)
9. Houghton, A.R., Gray, D.: Making Sense of the ECG. CRC Press (2007)
10. Karczewicz, M., Gabbouj, M.: ECG data compression by spline approximation. Signal Processing **59**(1), 43 – 59 (1997)
11. Khare, S., Bhandari, A., Singh, S., Arora, A.: ECG arrhythmia classification using spearman rank correlation and support vector machine. In: Deep, K., Nagar, A., Pant, M., Bansal, J.C. (eds.) Proceedings of the International Conference on Soft

Computing for Problem Solving (SocProS 2011) December 20-22, 2011. pp. 591–598 (2012)

12. Knorr, E.M., Ng, R.T.: Algorithms for mining distance-based outliers in large datasets. In: Proceedings of the 24rd International Conference on Very Large Data Bases. pp. 392–403 (1998)
13. Knorr, E.M., Ng, R.T., Tucakov, V.: Distance-based outliers: Algorithms and applications. The VLDB Journal **8**(3-4), 237–253 (Feb 2000)
14. Kontaki, M., Gounaris, A., Papadopoulos, A.N., Tsichlas, K., Manolopoulos, Y.: Continuous monitoring of distance-based outliers over data streams. In: Proceedings of the IEEE International Conference on Data Engineering. pp. 135–146 (April 2011)
15. Macek, J.: Incremental learning of ensemble classifiers on ECG data. In: 18th IEEE Symposium on Computer-Based Medical Systems. pp. 315–320 (June 2005)
16. Monasterio, V., Laguna, P., Martinez, J.P.: Multilead analysis of t-wave alternans in the ECG using principal component analysis. IEEE Transactions on Biomedical Engineering **56**(7), 1880–1890 (July 2009)
17. Patel, A.M., Gakare, P.K., Cheeran, A.N.: Article: Real time ECG feature extraction and arrhythmia detection on a mobile platform. International Journal of Computer Applications **44**(23), 40–45 (April 2012)
18. PhysioNet: MIT-BIH Arrhythmia Database Directory (Introduction). `https://physionet.org/physiobank/database/html/mitdbdir/intro.htm` (2010), [Online; accessed 29-May-2018]
19. PhysioNet: PysioBank Annotation. `https://www.physionet.org/physiobank/annotations.shtml` (2016), [Online; accessed 25-May-2018]
20. PhysioNet: MIT-BIH Arrhythmia Database Directory (Records). `https://www.physionet.org/physiobank/database/html/mitdbdir/records.htm#207` (2018), [Online; accessed 14-June-2018]
21. PyWavelet: Signal extension modes PyWavelets Documentation. `http://pywavelets.readthedocs.io/en/latest/ref/signal-extension-modes.html` (2018), [Online; accessed 2-June-2018]
22. Tran, L., Fan, L., Shahabi, C.: Distance-based outlier detection in data streams. Proceedings of the VLDB Endowment **9**(12), 1089–1100 (August 2016)
23. Veeravalli, B., Deepu, C.J., Ngo, D.: Real-Time, Personalized Anomaly Detection in Streaming Data for Wearable Healthcare Devices, pp. 403–426. Springer (2017)
24. Venkatesan, C., Karthigaikumar, P., Paul, A., Satheeskumaran, S., Kumar, R.: ECG signal preprocessing and svm classifier-based abnormality detection in remote healthcare applications. IEEE Access **6**, 9767–9773 (2018)
25. Yang, D., Rundensteiner, E.A., Ward, M.O.: Neighbor-based pattern detection for windows over streaming data. In: Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology. pp. 529–540 (2009)
26. Yang, T.F., Devine, B., Macfarlane, P.W.: Artificial neural networks for the diagnosis of atrial fibrillation. Medical and Biological Engineering and Computing **32**(6), 615–619 (November 1994)