

## MIT Open Access Articles

*Patient-Adaptive Ectopic Beat Classification using Active Learning*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** J. Wiens, J.V. Guttag. "Patient-Adaptive Ectopic Beat Classification using Active Learning" Proceedings of the 2010 Computing in Cardiology, IEEE. © Copyright 2010 IEEE

**As Published:** [http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5737921&contentType=Conference+Publications&sortType%3Dasc\\_p\\_Sequence%26filter%3DAND%28p\\_IS\\_Number%3A5737883%29%26rowsPerPage%3D50](http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5737921&contentType=Conference+Publications&sortType%3Dasc_p_Sequence%26filter%3DAND%28p_IS_Number%3A5737883%29%26rowsPerPage%3D50)

**Publisher:** Institute of Electrical and Electronics Engineers (IEEE)

**Persistent URL:** <http://hdl.handle.net/1721.1/73888>

**Version:** Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

**Terms of Use:** Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



# Patient-Adaptive Ectopic Beat Classification using Active Learning

J Wiens, JV Guttag

Massachusetts Institute of Technology, Cambridge MA, USA

## Abstract

*A major challenge in applying machine learning techniques to the problem of heartbeat classification is dealing effectively with inter-patient differences in electrocardiograms (ECGs). Inter-patient differences create a need for patient-specific classifiers, since there is no a priori reason to assume that a classifier trained on data from one patient will yield useful results when applied to a different patient. Unfortunately, patient-specific classifiers come at a high cost, since they require a labeled training set. Using active learning, we show that one can drastically reduce the amount of patient-specific labeled training data required to build a highly accurate patient-specific binary heartbeat classifier for identifying ventricular ectopic beats. Tested on all 48 half-hour ECG recordings from the MIT-BIH Arrhythmia Database, our approach achieves an average sensitivity of 96.2% and specificity of 99.9%. The average number of beats needed to train each patient-specific classifier was less than 37 beats, approximately 30 seconds of data.*

## 1. Introduction

In 24 hours, a Holter monitor can record over 100,000 heartbeats from one patient. This has prompted researchers to use machine learning to generate algorithms that automatically analyze such data, e.g., [1] and [2].

A major challenge in applying such automated techniques is dealing effectively with inter-patient differences. The morphology and interpretation of ECG signals can vary depending on the patient. Since machine learning techniques aim to estimate the underlying system that produced the data, if the system changes between training and testing, this can cause unpredictable results [3]. More concretely, inter-patient differences mean there is no *a priori* reason to assume that a classifier trained on data from one patient or even a large set of patients will yield useful results when applied to a different patient.

Therefore research in heartbeat classification has focused on patient-specific or patient-adaptive techniques that train on labeled data from the test patient.

[1] and [2] each showed that labeling the first 5 min-

utes or the first 500 beats of the test record and including it in the training set boosted overall classification performance for distinguishing ventricular ectopic beats (VEBs) from non-VEBs. We hypothesize that *passively* selecting the training set all from the initial portion (or a randomly chosen portion) of a record increases the risk of over-fitting to the patient's physiological state at that particular time. Moreover, records containing over 500 beats probably contain many redundancies, and asking a cardiologist to label all 500 beats is labor intensive.

In order to use expert knowledge more efficiently, we propose using *active learning* to train patient-specific heartbeat classifiers. By carefully choosing examples from the entire record, we iteratively build a training set that requires fewer cardiologist labels and then use this training set to build a patient-specific classifier.

When applied to the ECG data from the MIT-BIH Arrhythmia Database [4], our active learning algorithm classifies VEBs with a mean sensitivity of 96.2% and specificity of 99.9%. On average the algorithm used only 30 seconds of data. By comparison, [2] presents a patient-adapting classifier that uses 11 hours of previously collected training data and 500 labeled beats from the test patient to achieve a mean sensitivity of 80.6% and specificity of 99.9%. When tested on the same data set, our method used an average of 35 labeled data points from each record, and achieved a mean sensitivity of 98.9% and specificity of 99.9%. Additionally, we test our algorithm on ECG data from another cohort of patients admitted with NSTEMI, using two different cardiologists to show the clinical utility of our method.

The remainder of this paper proceeds as follows: Section 2 provides a brief overview of our active learning algorithm (details are provided in [5]). Section 3 presents the experimental results of testing our algorithm on ECG data from two different data sets. Finally, Section 4 summarizes the findings and contributions of this paper and makes suggestions for future work in this area.

## 2. Methods

Our approach to active learning for heartbeat classification is briefly described in the following two subsections.

## 2.1. Query selection

Given a new test record of unlabeled heartbeats, or rather feature vectors representing each beat, the algorithm begins by clustering the data, and then queries the point closest to the centroid in each cluster. Once labeled these examples compose our initial training set, and we train a linear support vector machine (SVM). At each following iteration, we use the current SVM to determine the next query, similar to [6].

After the initial pass, we re-cluster the data that lie on or within the margin of the current SVM. Next, for each cluster, the point closest to the SVM’s decision boundary is queried and added to the current training set. The SVM is updated based on this new training set and the data are again re-clustered. The algorithm stops either when we obtain a classifier that has no unlabeled data on or within its margin, or earlier depending on the stopping criterion. This iterative approach greatly reduces the number of queries needed, but increases the risk of converging to a solution that is not globally optimal. To reduce the chance of this occurring, we use two different forms of clustering, as discussed in the next section.

## 2.2. Clustering

We use hierarchical clustering to identify the most representative samples from each class, similar to [7], [8] and [9].

In hierarchical agglomerative clustering each data point is initially assigned its own cluster. Next, clusters with the smallest inter-cluster distance are combined to form new clusters. This is repeated until the desired number of clusters is reached. The distance between two clusters is defined by a linkage criterion based on a function of the Euclidean pairwise distances of the observations in each cluster. As described in [5], we investigated different linkage criteria, and chose two complementary criteria to reduce the risk of getting stuck in a locally optimal solution.

At each iteration of the algorithm, the data are clustered using both linkage criteria independently. The number of clusters here is not fixed, but is allowed to change depending on the input and the number of iterations. Initially, the maximum number of clusters,  $k$ , for each linkage criterion is set to two. The maximum number of clusters is incremented at each iteration. Depending on the input, this results in  $k$  to  $2k$  different clusters at each iteration.

## 3. Experimental results

To test the utility of our approach to active learning for heartbeat classification, we applied our algorithm to ECG data from two different sources, the MIT-BIH Arrhythmia Database at Physionet.org [4] and data from another co-

hort of patients admitted with NSTEMI. We tested our algorithm on a total of 52 different half-hour records. To get the data in a format that the algorithm could use, each record was first pre-processed and segmented as described in [5]. Once segmented each beat was transformed into a feature vector consisting of the 67 heartbeat features listed in Table 1.

Table 1. Heartbeat features used in experiments.

Index	Features
[1, ..., 60]	• Wavelet coefficients from the last 5 levels of a 6 level wavelet decomposition using a Daubechies 2 wavelet
[61, ..., 63]	• Energy in different portions of the beat
[64, ..., 66]	• RR-intervals
[67]	• Morphological distance between the current beat and the median beat of a record, calculated using the dynamic time warping approach described in [10].

We implemented our algorithm in MATLAB, and used  $SVM_{light}$  [11] to train the linear SVM at each iteration. We held the cost parameter of the linear SVM constant at  $C = 100$  throughout all experiments. For all experiments we used a stopping criterion based on the second derivative of the margin, described in [5].

## 3.1. Experiments on the MIT-BIH arrhythmia database

The MIT-BIH Arrhythmia Database consists of 48 annotated half-hour ECG records. For each record we used active learning to train a record-specific linear classifier to distinguish VEBs from non-VEBs.

We compared the performance of each classifier trained using active learning to the performance of a record-specific linear SVM classifier trained on all of the data. The performance of the SVMs trained using all of the data gives us a sense of the difficulty of each record. When evaluating the performance of the classifiers, we test on all of the data in order to maintain a consistent evaluation set. Each classifier assigns a positive label to what it believes are VEBs, and a negative label to everything else. The accuracy and sensitivity of each classifier is calculated as follows:

$$Accuracy = \frac{tp + tn}{tp + fn + tn + fp} \quad (1)$$

$$Sensitivity = \frac{tp}{tp + fn} \quad (2)$$

The results are shown in Table 2. Querying an average of approximately 37 beats yields an average accuracy and

Table 2. Our SVM active learning algorithm achieves close to the same performance as a classifier trained on all of the data, but uses over 98% less data.

Active vs. Complete Learning Results					
Rec.	# Beats	# VEB	Complete Sens. %	Active Sens. %	Active # Queries
100	2270	1	100.0	100.0	12
101	1861	0	-	-	10
102	2185	4	100.0	100.0	23
103	2082	0	-	-	9
104	2226	2	100.0	0.0	10
105	2563	41	100.0	100.0	25
106	2026	520	100.0	100.0	43
107	2135	59	100.0	100.0	17
108	1761	17	100.0	100.0	41
109	2530	38	100.0	100.0	45
111	2123	1	100.0	100.0	16
112	2537	0	-	-	8
113	1793	0	-	-	10
114	1878	43	100.0	100.0	42
115	1951	0	-	-	9
116	2410	109	100.0	100.0	33
117	1533	0	-	-	10
118	2276	16	100.0	100.0	27
119	1986	444	100.0	100.0	25
121	1861	1	100.0	100.0	12
122	2473	0	-	-	9
123	1516	3	100.0	100.0	16
124	1617	47	100.0	87.2	23
200	2599	826	100.0	100.0	134
201	1962	198	100.0	100.0	39
202	2134	19	100.0	100.0	36
203	2976	444	94.1	91.7	124
205	2654	71	100.0	98.6	19
207	2329	210	99.0	98.1	124
208	2950	992	99.7	99.4	86
209	3003	1	100.0	100.0	10
210	2645	195	99.5	96.9	105
212	2746	0	-	-	10
213	3248	220	95.5	89.5	62
214	2259	256	100.0	100.0	50
215	3361	164	100.0	100.0	42
217	2207	162	100.0	100.0	58
219	2153	64	100.0	98.4	37
220	2045	0	-	-	10
221	2426	396	100.0	99.5	17
222	2480	0	-	-	8
223	2603	473	99.8	100.0	99
228	2052	362	100.0	100.0	62
230	2254	1	100.0	100.0	12
231	1569	2	100.0	100.0	17
232	1779	0	-	-	8
233	3076	830	100.0	100.0	79
234	2752	3	100.0	100.0	20

sensitivity of 99.9% and 96.2%. Using all of the data to train a linear SVM classifier for each record, results in an average accuracy and sensitivity of 99.9% and 99.7% respectively. In exchange for a small drop in sensitivity, we are able to reduce the amount of labeled training data by over 98%.

In addition, for each record we looked at the performance of a classifier trained using our algorithm, versus the performance of a classifier trained on a fraction of each record. More precisely, each classifier is trained on the first  $x\%$  of each record. This passive selection of the training set has been proposed by other researchers, as discussed previously.

Figure 1 shows the amount of labeled data used when the training set is actively selected, and the corresponding mean sensitivity of the record-specific classifiers. Note that while our algorithm uses on average less than 1.6% of

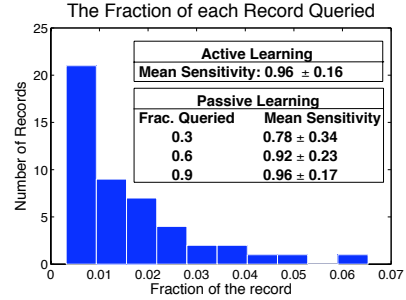


Figure 1. To achieve the same mean sensitivity as when the training set is actively selected, one would have to train on the first 90% of each record.

each record to train, passive learning requires 90% of each record to obtain the same mean sensitivity.

### 3.2. Experiments with cardiologists

The MIT-BIH Arrhythmia Database is the most widely used database of its kind. Because of this, many researchers are at risk for over-fitting to this database. This prompted us to test our algorithm on ECG data from a different source. For these experiments, we used data from another cohort of patients admitted with NSTEMI.

As a preliminary experiment, we looked at a small subset of this data. We considered 4 randomly chosen records, from a subset of patients who experienced at least one episode of ventricular tachycardia in the 7 day period following randomization. For each record, we consider the first half-hour. This totaled 8230 heartbeats.

We applied our active learning algorithm to each record once for each of two cardiologists. The cardiologists were presented with ECG plots of heartbeats, in context, like the one shown in Figure 2. Independently, the cardiologists were asked to label queries according to the following key: 1 = clearly non-PVC, 2 = ambiguous non-PVC, 3 = ambiguous PVC, 4 = clearly PVC. On average, the cardiologists were asked to label between 15-20 beats from each record.

This task differs slightly from the previous task of classifying beats as VEBs vs. non-VEBs. In this section we consider the task of detecting PVCs from non-PVCs because unlike for the MIT-BIH Arrhythmia Database, for this data there are no previously provided labels that can be treated as ground truth. To avoid having a cardiologist label 8230 heartbeats, we use existing PVC classification software from EP Ltd [12]. We applied this software to the same four records and assumed that as long as the two classifiers trained using our algorithm, agreed with each other and EP Ltd, then the beat was correctly classified. Out of a possible 8230 disagreements there were only 6.

For this experiment no changes were made to our algorithm; we simply asked the cardiologists to label beats

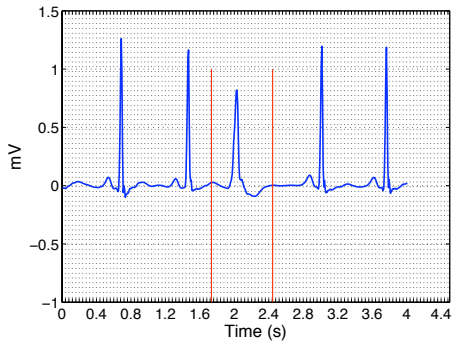


Figure 2. The classifiers trained using active learning both labeled the beat delineated by the red lines as a PVC, whereas a hand-coded classifier labeled it as a non-PVC.

according to the new task. This result shows the potential flexibility of active learning algorithms to adapt to new tasks.

#### 4. Summary

In this paper, we showed how active learning can be successfully applied to the binary classification task of identifying ectopic beats. For each of the 48 half-hour records from the MIT-BIH Arrhythmia Database, we applied our algorithm to build a record-specific classifier for the binary classification task of separating VEBs from non-VEBs. Our experiments include even the most difficult records from the database (paced patients and patients with fusion beats), which are often excluded in the literature. Compared to classifiers using a complete training set, we achieved a negligible drop in mean specificity and a 3.5% drop in mean sensitivity in exchange for a reduction in the amount of training data of over 98%. We also compared the classification performance of our algorithm to a classifier trained on the first portion of each record. This passive selection of the training set is common in the literature. However, to achieve the same mean accuracy achieved by our method, one would have to train on the first 90% of each record.

Finally, we tested our algorithm using two cardiologists and data from a different cohort of patients admitted with NSTEMI. The method required only a small number of cardiologist-supplied labels and classified beats as PVCs vs. non-PVCs with a performance comparable to that of a hand-coded classifier designed specifically for classifying PVCs vs. non-PVCs. Active learning techniques are more flexible than hand-coded techniques since they can easily adapt to new tasks. These preliminary results are encouraging, but more work is needed to compare the performance of hand-coded techniques to active learning techniques on a larger scale.

In conclusion, when possible, active learning techniques

should be used in lieu of passive techniques, when training record-specific classifiers. In the context of classifying heartbeats, active learning can dramatically reduce the amount of effort required from a physician to produce accurately labeled heartbeats for previously unseen patients.

#### Acknowledgements

We would like to thank Benjamin Scirica, Collin Stultz, and Zeeshan Syed for sharing their expert knowledge in cardiology and for their participation in our active learning experiments.

#### References

- [1] Hu YH, Palreddy S, Tompkins W. A Patient-Adaptable ECG Beat Classifier Using a Mixture of Experts Approach. *Biomedical Engineering IEEE Transactions* on Sept. 1997; 44(9):891–900. ISSN 0018-9294.
- [2] de Chazal P, Reilly R. A Patient-Adapting Heartbeat Classifier Using ECG Morphology and Heartbeat Interval Features. *Biomedical Engineering IEEE Transactions* on Dec. 2006;53(12):2535–2543. ISSN 0018-9294.
- [3] Qu H, Gotman J. A Patient-Specific Algorithm for the Detection of Seizure Onset in Long-Term EEG Monitoring: Possible Use as a Warning Device. *IEEE Transactions On Biomedical Engineering* Feb 1997;44:115–122.
- [4] Mark R, Moody G. MIT-BIH Arrhythmia Database, 1997. URL <http://ecg.mit.edu/dbinfo.html>.
- [5] Wiens J. Machine Learning for Patient-Adaptive Ectopic Beat Classification. Master’s thesis, MIT, EECS, May 2010.
- [6] Tong S, Koller D. Support Vector Machine Active Learning with Applications to Text Classification. *Journal of Machine Learning Research* 2002;2:45–66. ISSN 1532-4435.
- [7] Dasgupta S, Hsu D. Hierarchical Sampling for Active Learning. In *ICML ’08: Proceedings of the 25th international conference on Machine learning*. New York, NY, USA: ACM. ISBN 978-1-60558-205-4, 2008; 208–215.
- [8] Xu Z, Yu K, Tresp V, Xu X, Wang J. Representative sampling for text classification using support vector machines. In *Proceedings of the twenty-fifth European Conference on Information Retrieval*. Springer, 2003; 393–407.
- [9] Nguyen HT, Smeulders A. Active learning using pre-clustering. In *Proceedings of the twenty-first international conference on Machine learning*. New York, NY, USA: ACM. ISBN 1-58113-828-5, 2004; 79.
- [10] Syed Z, Gutttag J, Stultz C. Clustering and Symbolic Analysis of Cardiovascular Signals: Discovery and Visualization of Medically Relevant Patterns in Long-term Data Using Limited Prior Knowledge. *EURASIP Journal on Advances in Signal Processing* 2007;2007:97–112.
- [11] Joachims T. *Making Large-scale Support Vector Machine Learning Practical*. Cambridge, MA, USA: MIT Press, 1999. ISBN 0-262-19416-3.
- [12] Hamilton P. Open Source ECG Analysis. In *Computers in Cardiology*, volume 29. 2002; 101–104.