# AF Classification from ECG Recording using Feature Ensemble and Sparse Coding

Bradley M Whitaker[1], Muhammad Rizwan[1], V Burak Aydemir[1],
James M Rehg[2], and David V Anderson[1]

[1] Electrical & Computer Engineering Department, Georgia Institute of Technology, Atlanta, GA, USA
[2] College of Computing, Georgia Institute of Technology, Atlanta, GA, USA

## Abstract

*Introduction: The aim of the Physionet/CinC Challenge 2017 is to automatically classify atrial fibrillation (AF) from a short single lead ECG recording. The Challenge provides 8,528 labeled ECG recordings; each recording was labeled as normal, AF, other, or noisy. In addition, the Challenge provides sample code which includes an R-peak detector and a simple classifier.*

*Algorithm: We use an ensemble of features extracted from the ECG signals to create a four-class support vector machine (SVM) classifier. Included in the feature set are statistics obtained from the ECG signal, its spectrum, and the RR-intervals. In addition, we learn a 32-element sparse coding dictionary on the sorted RR-intervals of the ECG signals. Using the dictionary, we calculate a sparse coefficient vector for each training sample and put these through a soft-margin linear SVM. The soft-margin scores are used as additional features in the final classifier.*

*Results: Our algorithm achieves cross-validated F1 scores of 0.874, 0.756, and 0.689 (for normal, AF, and other files, respectively), resulting in a final cross-validated challenge score of 0.773. The score when tested on a subset of the unknown data is 0.78 (with F1 scores of 0.88, 0.80, 0.65). The official challenge score was 0.77.*

*Conclusions: We developed an algorithm to classify ECG recordings as normal, AF, other, or noisy. Our results show that sparse coding is an effective way to define discriminating features from a list of sorted RR-intervals. In addition, these sparse codes complement more commonly used features in the classification task. Further work will attempt to increase the accuracy of the algorithm by exploring other features and classifiers while still using sparse coding as an unsupervised feature extractor.*

## 1. Introduction

This work describes the solution of our entry in the 2017 Physionet/CinC Challenge. The goal of the Challenge was to accurately classify ECG recordings. The Challenge details can be found at `https://physionet.org/challenge/2017/`. The database we used in conjunction with the challenge is described in detail in [1].

The novelty of our solution and the focus of this paper relates to using sparse coding as a tool for performing unsupervised feature extraction. Many researchers are investigating the use of sparse coding in classification tasks [2–9]. Our own previous work has found success in using sparse coding features alone or with other problem-specific features in a classification setting [10, 11].

## 2. Algorithm

### 2.1. ECG Preprocessing

When examining the ECG signal, we used two algorithms to extract the times and voltage values of the fiducial points of the ECG beat. The first algorithm was provided by the Challenge organizers and is based on the Pan and Tompkins (P&T) method of QRS detection [12, 13]. The second algorithm was an ECG delimiter based on the work of [14] that identified all parts of the PQRST complex.

For the R-peak detection part, we closely followed the algorithm presented in [14] with minor variations. Firstly, we used six second segments to perform both R-peak detection and delineation. Secondly, for each segment, we checked if any of the RR intervals within the segment was greater than 2.6 seconds. If it was, we declared that segment as noisy and ignored it during further processing.

There are also minor differences from [14] in our implementation of the PQST detection. For the Q detection, after the phasor transform of the windowed signal, we searched for a local minimum. If there was no local minimum, the start time of the window was declared as the Q point. The S point is defined as the global maximum of the absolute value of the seek window after the gamma. For P and T points, we analyze the local maxima of the seek windows. The local maximum with the greatest mag-

 ISSN: 2325-887X DOI:10.22489/CinC.2017.174-192

nitude before the R-peak is chosen as the P point. The local maximum with the greatest magnitude after the R-peak is chosen as the T point.

## 2.2. Feature Ensemble

After processing the ECG signal with both R-peak detection algorithms as well as the in-house PQRST algorithm, we calculated several statistics and other measures to use as features in the classifier. These features are described in Table 1.

## 2.3. Sparse Coding

We used the RR-interval vectors obtained using our in-house R-peak detector in our sparse coding dictionary learning algorithm. In order for the linear dictionary to extract meaningful features, we first sorted the RR-intervals and interpolated the sorted list to have a length of 50. The purpose of this was to allow the 'short' and 'long' RR-intervals from each signal to line up correctly. After sorting and interpolating the RR-intervals, we applied sparse coding as an unsupervised feature extractor.

Sparse coding is a matrix factorization problem that tries to decompose a data matrix ($\mathbf{Y}$) into the product of a dictionary matrix ($\mathbf{D}$) and a sparse coefficient matrix ($\mathbf{X}$):

$$\mathbf{Y} = \mathbf{D}\mathbf{X}. \tag{1}$$

Fig. 1 gives a visual representation of Eq. 1. Each column of $\mathbf{Y}$ represents a data sample, which in our case is the sorted interpolated list of RR-intervals. Each column of the dictionary matrix, $\mathbf{D}$, can be thought of as a commonly-occurring feature learned from the training data. Because we sorted the RR-intervals prior to learning the dictionary, the first entry of each dictionary elements corresponds to the shortest RR-interval extracted from the ECG signal. Each column of $\mathbf{X}$ is a sparse vector that indicates which dictionary elements (features) are used to reconstruct the corresponding data vector.

Mathematically, performing this matrix decomposition corresponds to solving the following $\ell_0$-regularized lease squares problem [19, 20]:

$$\min_{\mathbf{D}\in\mathcal{C},\{\S_{\mathrm{m}}\}} \quad \frac{1}{\mathbf{M}} \sum_{\mathrm{m}=1}^{\mathrm{M}} \frac{1}{2} \left\| \boldsymbol{y}_{\mathrm{m}} - \mathbf{D}\mathbf{x}_{\mathrm{m}} \right\|_2^2 + \lambda \left\| \boldsymbol{x}_{\mathrm{m}} \right\|_0 . \tag{2}$$

In this equation, each $\boldsymbol{y}_{\mathrm{m}}$ corresponds to a column of $\mathbf{Y}$

Table 1. Features used in SVM Classifier

| Feature | Description |
|---|---|
| 6 R-Peak Statistics | Mean, standard deviation, and median of R-peak values calculated from both R-peak algorithms |
| 10 RR-Interval Statistics | Mean, standard deviation, mean, min, and max of RR-intervals calculated from both R-peak algorithms |
| 8 Irregularity Features | AFEvidence, PACEvidence, IrregularityEvidence, and OriginCount, as defined in [15], calculated from both R-peak algorithms |
| 6 Entropy Features | Approximate Entropy [16], Sample Entropy [17], and Coefficient of Sample Entropy [18] calculated from both R-peak algorithms |
| 2 Algorithm Agreement Statistics | Mean and standard deviation of the difference between the statistics mentioned above calculated from the two R-peak algorithms (15 each) |
| 5 ECG Statistics | Mean, standard deviation, skew, kurtosis, and median of ECG signal |
| 1 Heart Rate | Number of R-peaks (using P&T method) divided by duration of sample |
| 1 RR-Interval Interdecile Range | Interdecile (10%-90%) range of the in-house RR-intervals (after removing intervals larger than 2.6s) |
| 9 PQRST RMS Statistics | Mean, standard deviation, and median of RMS values of P, R, and T waves |
| 9 PQRST Interval Statistics | Mean, standard deviation, and median of PR, RT, and QS intervals |
| 15 PQRST Amplitude Statistics | Mean, standard deviation, and median of P, R, T, S, and RS normalized amplitudes |
| 1 Polarity Feature | Percent of the positive polarity R-peaks |
| 1 Noise Feature | Number of noisy six-second segments |
| 4 Sparse Code Soft-Margin SVM Scores | Explained in Section 2.3 |

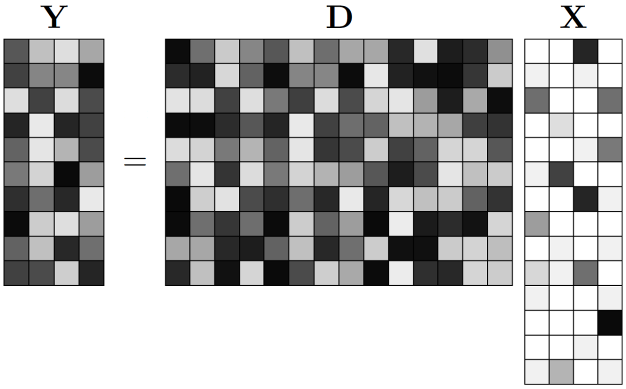


Figure 1. Visual representation of sparse coding. The feature matrix (left) is factored into the product of a dictionary matrix (center) and a sparse coefficient matrix (right).

and each $\boldsymbol{x}_{\mathrm{m}}$ corresponds to a column of $\mathbf{X}$. The dictionary matrix and the sparse coefficient vectors are jointly optimized. The dictionary is constrained to $\mathcal{C}$, the set of matrices whose columns have $\ell_2$-norm less than one, to prevent it from growing arbitrarily large. The $\lambda$ term is a tradeoff parameter between sparsity and fidelity.

The minimization program presented in Eq. 2 is NP-hard [21], but there are common methods to approximate it and come up with workable solutions. In our work we employ convex relaxation by replacing the $\ell_0$-“norm” with its closest convex norm, the $\ell_1$-norm. We then alternate solving for $\mathbf{D}$ and $\mathbf{X}$ while keeping the other fixed. This method, known as the Alternating Minimization Algorithm, is outlined in Alg. 1 [22,23].

**Algorithm 1** Alternating Minimization.

**Require:** Signals $\{\boldsymbol{y}_{\mathrm{m}} \in \mathbb{R}^{\mathrm{N}}\}_{\mathrm{m}=1,\ldots,\mathrm{M}}$, initial dictionary $\mathbf{D_0} \in \mathcal{C}$, regularization term $\lambda$, number of iterations K

1: Initialize $\mathbf{D} \leftarrow \mathbf{D_0}$
2: **for** $\mathrm{k} = 1, \ldots, \mathrm{K}$ **do**
3:     **for** several $\mathrm{m} \in \{1, \ldots, \mathrm{M}\}$ (*in parallel*) **do**
4:         Calculate coefficient vectors:
5:         $\boldsymbol{x}_{\mathrm{m}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{y}_{\mathrm{m}} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_1$
6:     **end for**
7:     Update dictionary:
8:     $\mathbf{D} = \arg\min_{\mathbf{D} \in \mathcal{C}} \frac{\mathbf{1}}{\mathrm{M}} \sum_{\mathrm{m}=\mathbf{1}}^{\mathrm{M}} \frac{\mathbf{1}}{\mathbf{2}} \|\mathbf{y}_{\mathrm{m}} - \mathbf{D}\mathbf{x}_{\mathrm{m}}\|_{\mathbf{2}}^{\mathbf{2}}$
9: **end for**
10: **return** $\mathbf{D}$

In our implementation of Alg. 1, we update the dictionary using gradient descent, following the method reported in [8]. Line 5 of Alg. 1 is a well-studied problem known as ‘basis pursuit denoising’ [24]. We solve it using the software package `l1_ls`, developed by Koh, et. al. [25].

The intuition behind using sparse coding as a feature extraction tool is that each column of the learned coefficient matrix defines how much of each dictionary element (feature) is needed to reconstruct the respective column of the data matrix. Ideally, the trained dictionary will have some elements that correspond to normal heart features and other elements that correspond to abnormal heart features. For example, the sorted list of RR-intervals taken from a healthy heart is fairly flat, while the sorted list corresponding to AF would have a non-negligible slope.

After learning a dictionary on the data matrix, we calculated the sparse feature vector corresponding to each ECG file. We trained a cross-validated, linear, four-class, soft-margin SVM on the sparse vectors. The four scores from the soft-margin SVM are a measure of how likely the ECG belongs to each of the four classes. We used these four scores as features in our final classifier.

## 2.4. Classification

After extracting the 78 features from each ECG file, we used LIBSVM to train a 10-fold cross-validated RBF-kernel SVM that classified between normal, AF, other, and noisy files [26]. We used the modified cuckoo search algorithm to select the SVM learning parameters [27]. We searched for parameters that maximized the score while minimizing the range of the F1 scores corresponding to normal, AF, and other files.

# 3. Results

The ten-fold cross-validated score for the final SVM, which included the ECG-based features as well as the sparse coding scores, was 0.773. The F1 scores for normal, AF, other, and noisy files were 0.874, 0.756, 0.689, and 0.454, respectively. When tested on a subset of the unknown challenge data, the algorithm achieved a score of 0.78. The F1 scores for normal, AF, and other files were 0.88, 0.80, and 0.65, respectively. The F1 score for noisy files was not provided. The final official challenge score was 0.77.

# 4. Conclusion

The work in this paper shows that sparse coding can be used to augment a feature set in a classification setting. We combined an ensemble of ECG-based features with sparse coding soft-margin scores to produce our final classifier. Future work could explore incorporating sparse coding and our feature ensemble into different challenge solutions for more accurate AF detection from ECG readings.

## Acknowledgments

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1650044. This work was also supported by NIH award 1R01HL130619-01A1.


## References


[1] Clifford GD, Liu C, Moody B, Silva I, Li Q, Johnson A, Mark RG. AF classification from a short single lead ECG recording: The PhysioNet Computing in Cardiology Challenge 2017. In 2017 Computing in Cardiology Conference (CinC), volume 44. 2017; In Press.

[2] Huang K, Aviyente S. Sparse representation for signal classification. In Advances in neural information processing systems. 2006; 609–616.

[3] Mairal J, Leordeanu M, Bach F, Hebert M, Ponce J. Discriminative sparse image models for class-specific edge detection and image interpretation. In Forsyth D, Torr P, Zisserman A (eds.), Computer Vision ECCV 2008, volume 5304 of Lecture Notes in Computer Science. Springer Berlin Heidelberg. ISBN 978-3-540-88689-1, 2008; 43–56.

[4] Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y. Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence Feb 2009;31(2):210–227. ISSN 0162-8828.

[5] Elad M, Figueiredo MA, Ma Y. On the role of sparse and redundant representations in image processing. Proceedings of the IEEE 2010;98(6):972–982.

[6] Kavukcuoglu K, Ranzato M, LeCun Y. Fast inference in sparse coding algorithms with applications to object recognition. arXiv preprint arXiv10103467 2010;.

[7] Wright J, Ma Y, Mairal J, Sapiro G, Huang T, Yan S. Sparse representation for computer vision and pattern recognition. Proceedings of the IEEE June 2010;98(6):1031–1044. ISSN 0018-9219.

[8] Charles AS, Olshausen BA, Rozell CJ. Learning sparse codes for hyperspectral imagery. Selected Topics in Signal Processing IEEE Journal of 2011;5(5):963–978. ISSN 1932-4553.

[9] Lee K, Hyung Z, Nam J. Acoustic scene classification using sparse feature learning and event-based pooling. In Applications of Signal Processing to Audio and Acoustics (WASPAA), 2013 IEEE Workshop on. IEEE, 2013; 1–4.

[10] Whitaker BM, Anderson DV. Heart sound classification via sparse coding. In 2016 Computing in Cardiology Conference (CinC). Sept 2016; 805–808.

[11] Whitaker BM, Suresha PB, Liu C, Clifford GD, Anderson DV. Combining sparse coding and time-domain features for heart sound classification. Physiological Measurement 2017;38(8):1701.

[12] Behar J, Oster J, Clifford GD. Combining and benchmarking methods of foetal ECG extraction without maternal or scalp electrode data. Physiological Measurement 2014; 35(8):1569.

[13] Pan J, Tompkins WJ. A real-time qrs detection algorithm. IEEE Transactions on Biomedical Engineering March 1985;BME-32(3):230–236. ISSN 0018-9294.

[14] Martinez A, Alcaraz R, Rieta JJ. A new method for automatic delineation of ECG fiducial points based on the Phasor Transform. 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBC10 2010;4586–4589. ISSN 0967-3334.

[15] Sarkar S, Ritscher D, Mehra R. A detector for a chronic implantable atrial tachyarrhythmia monitor. IEEE Transactions on Biomedical Engineering March 2008;55(3):1219–1224. ISSN 0018-9294.

[16] Pincus SM. Approximate entropy as a measure of system complexity. Proceedings of the National Academy of Sciences 1991;88(6):2297–2301.

[17] Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology Heart and Circulatory Physiology 2000;278(6):H2039–H2049. ISSN 0363-6135.

[18] Lake DE, Moorman JR. Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices. American Journal of Physiology Heart and Circulatory Physiology 2011;300(1):H319–H325. ISSN 0363-6135.

[19] Olshausen BA, Field DJ. Sparse coding with an overcomplete basis set: A strategy employed by v1? Vision research 1997;37(23):3311–3325. ISSN 0042-6989.

[20] Tosic I, Frossard P. Dictionary learning. Signal Processing Magazine IEEE March 2011;28(2):27–38. ISSN 1053-5888.

[21] Tillmann AM. On the computational intractability of exact and approximate dictionary learning. IEEE Signal Processing Letters Jan 2015;22(1):45–49. ISSN 1070-9908.

[22] Mairal J, Bach F, Ponce J. Sparse modeling for image and vision processing. Foundations and Trends in Computer Graphics and Vision 2014;8(2-3):85–283. ISSN 1572-2740.

[23] Lee H, Battle A, Raina R, Ng AY. Efficient sparse coding algorithms. In Schölkopf B, Platt J, Hoffman T (eds.), Advances in Neural Information Processing Systems 19. MIT Press, 2007; 801–808.

[24] Chen SS, Donoho DL, Saunders MA. Atomic decomposition by basis pursuit. SIAM journal on scientific computing 1998;20(1):33–61.

[25] Koh K, Kim SJ, Boyd S. An interior-point method for large-scale l1-regularized logistic regression. Journal of Machine learning research 2007;8(7):1519–1555. Software available at `https://stanford.edu/~boyd/l1_ls/`.

[26] Chang CC, Lin CJ. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2011;2:27:1–27:27. Software available at `www.csie.ntu.edu.tw/~cjlin/libsvm`.

[27] Walton S, Hassan O, Morgan K, Brown M. Modified cuckoo search: A new gradient free optimisation algorithm. Chaos Solitons Fractals 2011;44(9):710 – 718. ISSN 0960-0779.



Address for correspondence:
Bradley Whitaker (b.whitaker@gatech.edu)
85 Fifth Street NW Office 543, Atlanta, GA, 30308, USA