# Optimal ECG Lead System for Automatic Myocardial Ischemia Detection

Misha Glazunov[1], Alfonso Aranda[2], Carlo Galuzzi[3],

[1] TU Delft, Delft, the Netherlands
[2] Bakken Research Center, Medtronic, Maastricht, The Netherlands.
[3] Swinburne University of Technology, Melbourne, Australia

## Abstract

*In this paper, we study features that are not commonly used in clinical practice but may play a role in the automatic detection of acute myocardial infarction (AMI) using a reduced 3-lead ECG system: fragmented QRS in the time domain and intra-QRS in the time-frequency domain. Using chaos theory we reconstruct attractors from ECG and devise geometrical features and two main dynamical invariants: the correlation dimension and the Lyapunov exponent. For validation, we use the Physionet STAFF III dataset. We perform automatic classification using the gradient boosting machine and identify the optimal 3-lead ECG system, achieving promising results: the area under the ROC curve (AUC) is 0.91. It improves the results obtained with the baseline features such as ST-segment elevation and T-wave inversion: AUC 0.85. Finally, we combine new parameters with the baseline features and enhance the final model with previously introduced pseudo-vectorcardiography parameters. The results account for all regions of the heart ischemia: anterior, inferior, and posterior. The proposed automatic algorithm allows the easiest way to determine the first signs of AMI in a patient's ECG based on the input from the minimal number of leads.*

## 1. Introduction

The possibility to automatically diagnose ischemia using ECG with a minimal lead system and with high accuracy using machine learning has been shown in [1]. The goal of this paper is to look for new parameters that may help to improve the current algorithm and to identify the best lead system for every region in the heart where AMI is located. In this paper we address the following questions:

- Is it possible to improve AMI detection by introducing new parameters for classification? What are these parameters and how they can improve the current algorithm?
- Is it possible to improve AMI detection based on a different lead subset of 12 leads?
- Is there a single set of parameters fit for all AMI locations?

## 2. Methodology

**Dataset** This work is based on PhysioNet [2] STAFF III dataset [3] which contains data collected from patients receiving elective percutaneous transluminal coronary angioplasty (PCTA). The dataset registered both minutes of baseline (5 minutes at rest) and minutes during the surgery including the moment of balloon inflation. We use it as a basis for our algorithm to diagnose AMI since the moment of balloon inflation simulates conditions of complete coronary occlusion. The dataset preprocessing and annotation and the selection of the recording are described in [1]. Two recordings per patient were used: a baseline recording preceding intervention and at rest, and a recording selected in the interval between catheter balloon inflation and deflation during PTCA. The dataset contains a non-uniform distribution of the registered samples, most of the samples have inflation in the inferior region of the heart, while the posterior region represents only a small fraction: anterior 34%, inferior 45%, posterior 21%.

**ECG preprocessing** For fQRS and intra-QRS high-frequency band, we extract part of the signal that is annotated as the most relevant for distinguishing between baseline and inflation. The duration of this part is on average equal to 10 sec. Since both QRS methods depend on the high frequencies we use bandpass filtering in a range from 1 to 180 Hz. For attractors, we extract a third of the signal centered at the middle of the ECG. The signal is denoised with band-pass filtering in a range from 1 to 30 Hz.

### 2.1. Feature Engineering

#### 2.1.1. QRS features

**Fragmented QRS in time domain:**

Our algorithm identifies the number of spikes and the specific category of the fQRS based on classification suggested in [4], namely: RSR', rSr', rSR, notched S, notched

R, and fragmented QRS. Fragmented QRS is QRS with one or more spikes within the QRS complex [5]. Spikes are defined as local maxima within fiducials that mark the start and the end of each QRS complex.

**Intra-QRS HF-band in time-frequency domain:**

Intra-QRS power spectrum has a region in the high-frequency band that may have diagnostic significance for ischemia [6]. We use a wavelet transform to calculate the power spectrum of each QRS, with the Morlet wavelet as a kernel. To tackle the trade-off between time and frequency precision we use a range of cycles from 2 to 10 that are logarithmically incrementing for higher frequencies.

### 2.1.2. Attractors

Takens proved that it is possible to reconstruct the topological characteristics of the phase space from the time series of a single variable [7]. The method is based on the delay coordinates and it yields an attractor homeomorphic to the original one. The region obtained in the reconstructed space represents a dynamical attractor with a possibly fractal dimension.

Two main considerations while applying the method are the identification of the time delay $\tau$ and the embedding dimension $m$.

**Hyper-parameters:**

1. **Identifying time delay** $\tau$. There are several techniques suggested for this purpose in [8, 9]. We chose the autocorrelation function estimation with several lags for its simplicity. Significant changes in the correlogram suggest the potential values for our time delay. As an indicator of such significant change, we are using the first nadir in $r_k$. The value estimated in such a way turns out to be close on average to $\frac{1}{3}$ of a single heartbeat that may roughly correspond to the three main waves in the ECG: P-wave, QRS, and T-wave. So even though autocorrelation is based on the assumption of the weak stationarity of the time series, the obtained time lag does have physiological meaning.
2. **Identifying embedding dimension** $m$. First, we used an approach from [10]: we reconstructed an attractor in a 3-dimensional space, projected it to a 2-dimensional plane for preliminary analysis and simple feature extraction. In addition we calculated a fractal dimension. We chose to compute the correlation dimension due to the more effective numerical computations than the estimation of the dimension using the box-counting method. According to [11], the number of measurements required to determine the embedding dimension $D$ should be of order $10^D$. It means that in our case a proper estimation may be obtained only with a maximum of 5 dimensions due to the limit of the available data. So we calculated $D_{corr}$ values per patient and chose the threshold integer value that is greater than the maximum for the embedding dimension $m$.
3. **Correlation dimension**. Correlation dimension is a valuable qualitative parameter of the attractor, and may also be used to estimate the embedding dimension. For numerical $C(r)$ estimation we used the Grassberger-Procaccia algorithm [12], and then ran the estimation for the embedding dimensions with $step = 3$ from $m = 1$ to $m = 10$ in order to improve the performance.

**Simple geometric features:**

We started our analysis with $m = 3$ since it can be easily visualized. For feature extraction, we projected the reconstructed attractor on a plane perpendicular to the vector (1, 1, 1). As shown in [13], this particular projection removes baseline variation. We plotted a two-dimensional projection for all patients per lead and identified simple geometrical parameters that tend to variate between two subpopulations of the dataset: symmetry, an attractor size, and a density of the plotted trajectories.

**Lyapunov exponent:**

In our experiments with simple geometric features, the most promising results were achieved with density. From the point of view of the reconstructed phase space - density of the projected portrait reflects the level of the divergence of the trajectories of the reconstructed attractor: some of the trajectories are close to each other while others diverge. In chaos theory, there is a frequently used measure of the rate of the divergence of the points located on infinitesimally close trajectories during their evolution: the Lyapunov exponent. It is one of the major indicators of the chaotic behavior in the dynamic system. The main criterion is that the largest Lyapunov exponent $\lambda_1$ should be positive. For the Lyapunov exponent calculation, we used Wolf's algorithm [11].

## 3. Experiments and Results

### 3.1. Statistical analysis

We apply statistical analysis to check if there is a significant difference in parameters acquired on the previous step between two groups: baseline and inflation.

First, we use the Shapiro-Wilk test to check if the underlying random variable is normally distributed. After that, if it is indeed the case, we run a paired Student t-test with the mean and standard deviation estimated from the observations. The null hypothesis is that means of two groups are equal. However, if the test for normality fails then we apply a non-parametric Wilcoxon signed-rank test to see if there is a difference between populations. All parameters with $p < 0.05$ are considered to be statistically significant and subsequently used for classification.

From the point of view of parameters, the attractors contain most of the statistically significant values for all leads. In comparison, wavelets and fragmented QRS demonstrate

less significance, with fragmented QRS being more significant than wavelets.

From the point of view of leads, v2, v3, v5, v6 and III have the most significant parameters.

## 3.2. Machine learning

For classification, we use the gradient boosting classifier that demonstrated the best results in [1]. Since we are evaluating a binary classification we rely on the standard ROC-curve with sensitivity and (1-specificity) on the axes, and the area under the curve (AUC) as our model performance metric.

The dataset is split into a training part with 70% of the samples and a validation part with 30% of the samples. Since all parameters are statistically significant for at least one of the leads, we chose to do an exhaustive search for the best subset of 3 leads with all of our new features. To that end, we run training and validation of our classification model $\binom{9}{3} = 84$ times.

While evaluating the model using only our new parameters, we observed that the overall model performance demonstrated comparable results with an average $AUC = 0.81$. Nevertheless, we noticed that if at least two of the leads are located on opposite sides the accuracy increases. Moreover, AUC values change steadily per ischemia region as the subset of leads moves from v1 to III, i.e., the best results for the anterior region reveal subsets of leads near v1, however, the posterior region is better detected with subsets of leads near lead III.

Based on these results, we choose lead systems that have a total $AUC >= 90$. There are two such systems:

- leads v1, v4, III with $AUC = 0.90$
- leads v1, II, III with $AUC = 0.91$

As it can be seen in Figures 1 and 2, the lead system v1, v4, III has a lower AUC for the posterior region than the lead system v1, II, III, however, the latter is losing in anterior and inferior regions.

Systems with leads v1, v4, III and with leads v1, II, III both have Density v1-lead and Diameter III-lead among their Top 4 most significant relative influences. For leads v1, v4, III, another two parameters in Top 4 are Density v4-lead and Diameter v4-lead. For leads vI, vII, III - Symmetry II-lead and Density II-lead. Since the total area under the ROC-curve is larger than for the parameters identified in the previous work, we combined old parameters with the new ones for those two particular lead systems. Our models that combine all the parameters produce the following results:

- leads v1, v4, III with $AUC = 0.90$
- leads v1, II, III with $AUC = 0.92$

As shown in Figures 3 and 4 the AUC increased for all regions for lead system v1, II, III. As for the second system,

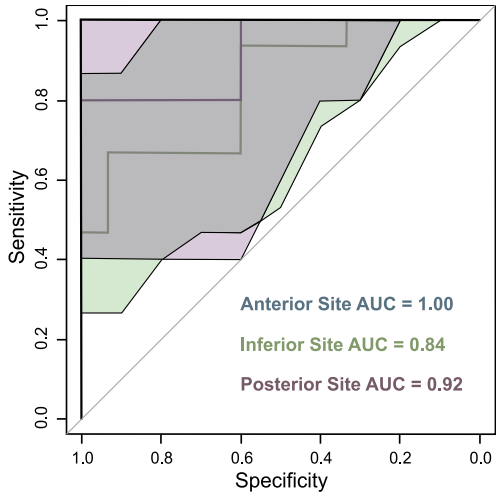


Figure 1. AUC for leads v1, v4 and III by region of ischemia (new features only)

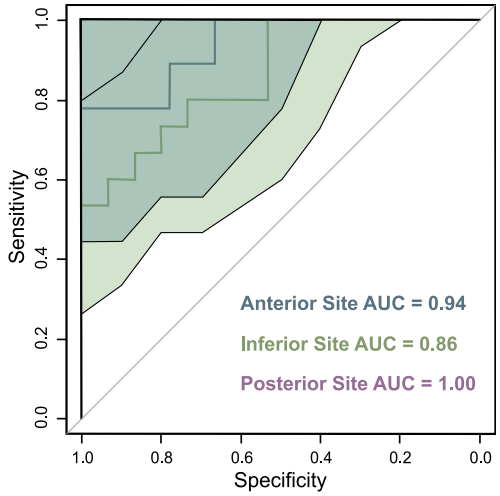


Figure 2. AUC for leads v1, II and III by region of ischemia (new features only)

there is a significant improvement in the inferior region and a very slight decrease for the anterior.

As it was mentioned before, since the latter system has two leads close to the side where the lead III is located, its result is better for the posterior region. However, the posterior region has fewer samples in the dataset. Also taking into consideration the better performance in general for sparse leads, we consider the lead system v1, v4, III as the most optimal one.

New features of attractors such as density, symmetry, diameter, and Lyapunov exponent together with parameters of the previous work such as ST-segment elevation have the highest relative influence for both systems: with leads v1, v4, III and with leads v1, II, III. ST-elevation for leads v4 and III is still the most important. Density for v1 is

another significant parameter for both systems.

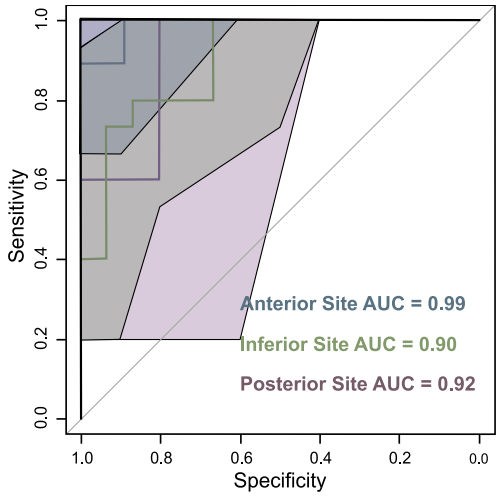


Figure 3. AUC for combined features: v1, v4, III

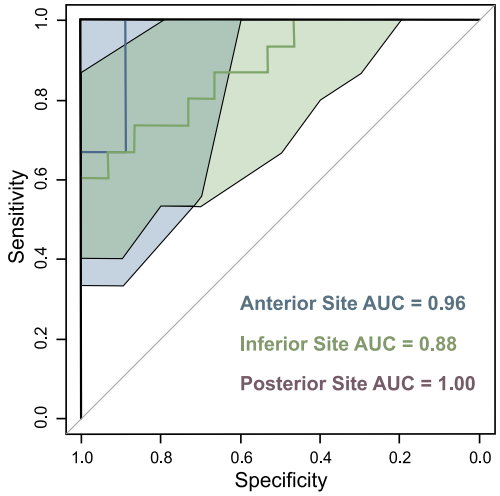


Figure 4. AUC for combined features: v1, II, III

## 4. Conclusion

We studied several parameters that may play a role in ischemia diagnostics. We looked at several characteristics of the QRS complex in both time and time-frequency domains, and also approached signal analysis from the point of view of non-linear dynamics and chaos theory.

For research questions, we managed to extract several new parameters statistically significant for MI detection. Also, we were able to identify the most optimal lead system with the minimum number of 3 leads that improve the diagnostic procedure for all heart regions in comparison with the previous work. In addition, we managed to find a single set of parameters and leads that have a high prediction accuracy.

Further research may include more parameters that represent dynamical invariants together with a set of fractal dimensions for a multi-fractal analysis.

## References

[1] Aranda A, Bonizzi P, Karel J, Peeters R. Myocardial ischemia diagnosis using a reduced lead system. volume 2018. 07 2018; 5302–5305.

[2] Goldberger A, Amaral L, Glass L, Hausdorff J, Ivanov P, Mark R, Mietus J, Moody G, Peng CK, Stanley H. Physiobank, physiotoolkit, and physionet : Components of a new research resource for complex physiologic signals. Circulation 07 2000;101:E215–20.

[3] Martinez J, Pahlm O, Ringborn M, Warren S, Laguna P, Sornmo L. The staff iii database: Ecgs recorded during acutely induced myocardial ischemia. 09 2017; .

[4] Take Y, Morita H. Fragmented qrs: What is the meaning? Indian pacing and electrophysiology journal 09 2012; 12:213–25.

[5] Das M, Khan B, Jacob S, Kumar A, Mahenthiran J. Significance of a fragmented qrs complex versus a q wave in patients with coronary artery disease. Circulation 05 2006; 113:2495–501.

[6] Gramatikov B, Iyer V. Intra-qrs spectral changes accompany st segment changes during episodes of myocardial ischemia. Journal of Electrocardiology 09 2014;48.

[7] Takens F. Detecting Strange Attractors in Turbulence. Lecture Notes in Mathematics, volume 898. ISBN 978-3-540-11171-9, 11 2006; 366–381.

[8] Bassingthwaighte J, Liebovitch L, West B. Fractal Physiology. 01 1994. ISBN 978-1-4614-7572-9.

[9] Theiler J. Estimating fractal dimension. J Opt Soc Am A 06 1990;7:1055–1073.

[10] Lyle J, Charlton P, Bonet-Luz E, Chaffey G, Christie M, Nandi M, Aston P. Beyond hrv: Analysis of ecg signals using attractor reconstruction. 03 2018; .

[11] Wolf A, Swift J, Swinney H, Vastano J. Determining lyapunov exponents from a time series. Physica D Nonlinear Phenomena 07 1985;16:285–317.

[12] Grassberger P, Procaccia I. Measuring strangeness strange attractors. Physica D Nonlinear Phenomena 10 1983; 9:189–208.

[13] Eades P. Symmetry finding algorithms. Machine Intelligence and Pattern Recognition 12 1988;6.

Address for correspondence:

Misha Glazunov
Hertogsingel 75, 6211NE, Maastricht, the Netherlands
michael.glazunoff@gmail.com