A Game Theoretic Predictive Modeling Approach to False Alarm Reduction in Intensive Care Units (ICUs)

Fatemeh Afghah[1], Abolfazl Razi[1], Reza Soroushmehr[2,3], Somayeh Molaei[2,3], Hamid Ghanbari[4], and Kayvan Najarian[2,3,5]

[1] Electrical Engineering and Computer Science Department,
Northern Arizona University, Flagstaff, AZ 86011
{fatemeh.afghah, abolfazl.razi}@nau.edu,
[2] Department of Emergency Medicine
[3] Michigan Center for Integrative Research in Critical Care: M-CIRCC,
[4] Internal Medicine,
[5] Computational Medicine and Bioinformatics Department,
University of Michigan, Ann Arbor, MI 48109
{ssoroush,smolaei,ghhamid, kayvan}@med.umich.edu

Abstract.

Keywords: False alarm, feature selection, coalition game theory, classification

1 Introduction

In order to monitor a patient and also for the sake of diagnostic, prognostic and treatment, many monitoring and therapeutic devices are utilized in intensive care units (ICUs). These devices are also used to measure vital signs, support or replace impaired or failing organs and administer medications to patients [?]. Each of these devices might generate optic/acoustic alarms due to patient's physiologic condition, patient movement, motion artifact, malfunction of individual sensors and imperfections in the patientequipment contact [?]. Many of the alarms (80% to 99% [?])could be false and/or clinically insignificant which are not related to patient condition. These alarms could compromise quality and safety of care resulting to many problems such as cry–wolf effect and alarm fatigue among care–givers as well as the possibility of missing a real event due to care–givers insensitivity to these unreliable alarms known as cry–wolf effect.

Dealing with false alarms is widely considered the number one hazard imposed by the medical technology and an important concern in ICUs [?]. Many approaches have been utilized to decrease the number of false alarms such as adding short delay [?], improving the quality of signals, improvements in sensor technology and utilizing advanced multi-parameter models using [?,?]. Cvach has reviewed some research and non-research publications from 2000 to 2011 [?] and Imhof et al. gives an overview on different aspects of false alarm problem.

Using a machine learning approach, Li and Clifford have designed a framework for false alarm reduction on arrhythmia patients. They extracted 114 features from ECG, ABP, and PPG and used a genetic algorithm to select a subset of these features. Using a relevance vector machine (RVM) as a classifier, false alarm suppression was reported to be 86.4%, 100% and 27.8% respectively for asystole, extreme extreme bradycardia and extreme tachycardia. An automated method for false arrhythmia suppression was proposed in [?] that is based on quality assessment of normal and abnormal rhythms of ECG signals. In this method an ECG signal is downsampled to 125Hz and then QRS detection algorithms are applied. After that baseline wander is filtered and different signal quality indexes (SQI) are calculated and used in a SVM classifier where the obtained accuracy, sensitivity and specificity are respectively in 0.990, 0.985, and 0.994. Different approaches including KNN, Naive Bayes, Decision Tree, SVM and multi–layer Perceptron have been tested on a database from MIMIC II for alarms classification where features have been extracted from age, sex, CVP, SpO2, ABP, ECG and PAP [?]. The suppression rate for true alarm detection is between 2.33% and 17.73% for 5 alarms and false alarm suppression rate is between 71.73% to 99.23%. Charbonnier and Gentil have proposed a trend extraction that tracks the changes in signals using a fuzzy decision approach[?] and could filter 81% of the false alarm without filtering any true alarms where they tested their method on a small number of examples.

The above models considers a number of features/parameters extracted from multiple continuously-measured physiological signals, such as electrocardiogram (ECG) and arterial bold pressure (ABP) to create more reliable alarms. The major problem faced by these multi-parameter approaches is the presence of many parameters / features that individually have low impact on the model performance, and as such they might not be included in the model, while when coupled with other such parameters could significantly improve the performance of the accuracy and specificity of the alarm detection algorithms. Besides statistical evidence to this observation, the fact that physicians, by visual interpretation of the patterns in all patient signals, can very often correctly decide on the validity of the alarms caused by individual machines / monitors, suggests that when a suitable combination of all data/features are included in a model, false alarms can be reduced significantly [?].

In this paper, we propose a coalition based game–theoretic predictive modeling approach to suppress the false alarm for five types of life threatening arrhythmias including asystole, extreme bradycardia, extreme tachycardia, ventricular tachycardia, and ventricular flutter/fibrillation. A main capability of the proposed method is finding discriminating combined / sub-sets of apparently low-impact features, which can have a significant impact on the specificity and accuracy of the alarm detection approaches.

Several data mining and feature reduction algorithms have been utilized in analysis of big data sets to improve the prediction accuracy and reliability through reducing the feature space to a more concise and relevant set of attributes [?,?,?,?,?]. However in the majority of these conventional methods, each of the features is evaluated separately, and as such, the possible correlation among them is neglected. Specifically, the existing methods either only account for the effect of individual features on the target or consider the inter-feature mutual information to obtain higher performance; however, it is often the case that a set of features that together have a considerable effect on the classifier, while each individual attribute in the set does not. Therefore, these features will most likely be filtered out that can result in significant degradation in the performance [**?**]. Here we propose a false alarm detection technique based on coalition game theoretical approach that accounts for intricate and intrinsic interrelation among all potentially effective combinations of the features to identify the most informative grouping. The unique advantage of this method is identifying the features, which despite their weak individual contribution to the classifier, have a quantifiable impact when grouped with other features.

Cooperative game theoretic approach has been recently utilized in feature selection algorithms [**?**,**?**,**?**]. In this paper, we develop a coalition game-theoretic based feature selection method that considers the interdependencies across the features and measures the contribution of features both individually and in group with others. This enables the feature selection algorithm to recognize the features that, despite their weak individual contribution to the classifier, have a quantifiable impact when grouped with other features.

In this paper, three main signals; ECG, ABP, and Spo2 (plethysmogram or PLETH) are used as the inputs of the alarm prediction model. In the first stage (i.e. signal analysis) wavelet coefficients of each signal at different levels of decomposition are calculated. Then, a number of statistical features such as mean, variance, median, kurtosis and entropy of the resulting wavelet coefficients are calculated for each level. The calculated coefficients along with the other parameters are used as features for our proposed Coalition game theoretic approach in which different combinations of features are considered for creating a predictive model that assesses the validity of the alarms.

The rest of this paper is organized as follows. Section 2 introduces the proposed signal analysis and feature extraction techniques. An introduction to coalition game theory followed by the description of the proposed game theoretic based feature selection method are presented in Section 3. The utilized data of this study is introduced in section **??**. Numerical analysis results are presented in Section 5, followed by conclusion in Section 6.

2 Signal Analysis and Feature Extraction

We applied a discrete wavelet transform (DWT) on the 1-D input signals, ECG II, PLETH and APB. The wavelet transform uses the decomposition (analysis) filters, fdf(a level 1 analysis filter), for the first level and the analysis filters, df (a analysis filters for level greater than 1), for subsequent levels. Supported wavelet transforms are the critically sampled DWT, double-density, dual-tree complex, and dual-tree double-density complex wavelet transform. The discrete wavelet transform (DWT) is an implementation of the wavelet transform using a discrete set of the wavelet scales and translations obeying some defined rules. DWT can decompose the signal into mutually orthogonal set of wavelets. The wavelet can be constructed from a scaling function which describes its scaling properties. The DWT of a signal is calculated by passing it through a series of filters. First the samples are passed through a low pass filter with impulse response resulting in a convolution of the two: $y_{high}[n] = (x * g)[n] = \sum_{k=-\infty}^{\infty} x[k]g[n-k]$.

The signal is also decomposed simultaneously using a high-pass filter. The outputs giving the detail coefficients (from the high-pass filter) and approximation coefficients (from the low-pass).

$$y_{low}[n] = (x * h)[n] = \sum_{k=-\infty}^{\infty} x[k]h[n-k] \tag{1}$$

The filter outputs are then subsampled by 2. The set of wavelets define a special filter bank, which can be used for signal component analysis and resulting wavelet transform coefficients can be further applied as signal features for its classification. Signal decomposition performed by a pyramidal algorithm is interpreting wavelets as pass-band filters.

Therefore, the three 1-D input signals for each patient is transformed into 18 vectors of wavelet coefficients. Including all wavelet coefficients as features to the classification setup is not efficient and may significantly decrease the generalization property of the trained model due to over-fitting. Therefore, we further reduce the number of features by extracting representative statistical and information-theoretic properties of the wavelet vectors as summarized in Table 1. In calculating information-theoretic properties (e.g. Entropy), we assume that the wavelet vector elements are derived from an unknown probability distribution.

Table 1: Statistical and Information-theoretic features of wavelet vectors.

No	Feature	No	Feature	No	Feature
1	mean	13	skewness	25	nS(10)
2	mode	14	harmonic mean	26	nS(100)
3	median	15	interquartile range	27	nS(1000)
4	max	16	Shannon Entropy	28	nS(10000)
5	min	17	Log Entropy	29	nS(25000)
6	range	18	nT(1)	30	nS(50000)
7	variance	19	nT(10)	31	nS(65000)
8	std (σ)	20	nT(100)	32	an_1
9	μ_3	21	nT(500)	33	an_2
10	μ_4	22	nT(1000)	34	an_3
11	coefficient of var	23	nT(5000)	35	an_5
12	kurtosis	24	nS(1)	36	an_10

In Table 1, features 1 to 10 are typical statical properties of the signal, where μ_n is the n^{th} standardized sample moment i.e. $\mu_n = \frac{\Sigma_{i=1}^{N}(X_i-\bar{X})^n}{N}$, where $X_1, \ldots, X_N$ are the N^{th} wavelet coefficients associated with each signal probe. *Kurtosis* and *skewness* define the shape of probability distributions, such that *kurthosis* measures the peakedness of distribution and is defined as the ratio of the forth standardized moment to the square of variance (i.e. $\kappa(X)\frac{E[(X-\mu)^4]}{\left(E[(X-\mu)^2]\right)^2} = \frac{\mu_4(X)}{\sigma_4(X)}$). Likewise, *skewness* is a measure of the symmetry of distribution around zero and is defined as $\lambda(X) = E\left[\left(\frac{X-\mu(X)}{\sigma(X)}\right)^3\right] = \frac{\mu^3(X)}{\sigma^3(X)}$. *Harmonic mean* is defined as $\frac{N}{\sum_{i=1}^{N} 1/X_i}$. *Interquartile range* is calculated based on the difference between the 75th and 25th percentiles. Shannon entropy is an information theoretic property of the square of coefficients approximated by their sample counterparts and calculated as $-\sum_{iu=1}^{N} X_i^2 \log_2 X_i^2$. Log energy is $\sum \log X + i^2$. $T(\alpha) = \sum_{i=1}^{N} 1(\|X_i\| > \alpha)$ counts the number of times that the value of wavelet coefficients exceed the threshold α, whereas $nS(\alpha) = N - 2\sum_{i=1}^{N} 1(X_i^2 \leq \alpha^2) + 2\alpha^2 \sum_{i=1}^{N} 1(X_i^2 > \alpha^2) + \sum_{i=1}^{N}[X_i^2 1(X_i^2 \leq \alpha^2)]$. Finally, $an_m(X) = \sum_{i=1}^{N} |X_i|^m$ is the m^{th} norm of the vector of the absolute values of the coefficients. Hereafter, features 1 to 15 are called as *Feature set 1:Statistical*, which includes statistical information and features 16 to 36 as *Feature set 2:Entropy*, which mainly includes information-theoretic and geometrical properties of the coefficients.

3 Feature Selection

Here we map the coalition game-theoretical methodology to the feature selection problem by considering the competing features as the game players, where the features can be classified in different coalitions by noting their impact on the classifier and also by their interdependency.

3.1 Coalition-based Game-theoretic Feature Selection

Cooperative game theory has been recently utilized in feature selection algorithms [**?**,**?**,**?**,**?**]. In these games, the players cooperate with each other by forming various sub-groups called *coalitions*. These games are defined based on exhaustive scenarios that players may form a group and how the total shared payoff is divided among the members. For a transferable utility coalition (TU-coalition) game with n players, let N denote the set of players, $N = \{1, 2, ..., n\}$. A coalition of players, S defines a sub-set of N, $S \subseteq N$. In general, for a n-player game there exists 2^n possible coalitions of any size. The empty coalition is denoted by ϕ, while grand coalition refers to the coalition of all players, N.

The n-player coalition game can be defined with the pair of (N, v), where $N = \{1, 2, ..., n\}$ is the set of players and the *characteristic function*, v is a real-valued function defined on the set of all coalitions, $v : 2^N \to \mathbb{R}$. For a coalition S, $S \subseteq N$, the characteristic function, $v(S)$ represents the total payoff can be gained by the members of this coalition. This function satisfies the following conditions,

- characteristic function of an empty coalition is zero, $v(\phi) = 0$, and
- if S_i and S_j, $(S_i, S_j \subseteq N)$ are two disjoint coalitions, the characteristic function of their union has super-additivity property, meaning that $v(S_i \cup S_j) \geq v(S_i) + v(S_j)$.

Here, we model the features as the players of the game, and the characteristic function of a coalition, v is measured by contribution of its members (features) to the performance of the classifier (e.g. success rate in supervised learning). Different possible grouping of the features are examined to recognize the optimal coalition. The contribution of feature i in classification accuracy when it joins a coalition S is defined by *marginal importance* as follows

$$\Delta_i(S) = v(S \cup \{i\}) - v(S) \tag{2}$$

A solution of a coalition game is determined by how the coalition of players can be formed and how the total payoff of a coalition is divided among the members. Let's define the value function, γ that assigns an n-tuple of real numbers, $\gamma(v) = (\gamma_1(v), \gamma_2(v), ..., \gamma_n(v))$ to each possible characteristic function, in which $\gamma_i(v)$ measures the value of player i in the game with characteristic function v. If the following axioms are satisfied, Shapley value can be utilized as a fair unique solution of the coalition game [**?**]. The Shapley axioms for $\gamma(v)$ are

- Efficiency (group rationality): $\sum_{i \in N} \gamma_i(v) = v(N)$, meaning that the summation of values for all players is equal to the value of grand coalition.
- Symmetry: If for players i and j, $i, j \in N$ and for every coalition S not containing i and j we have $v(S \cup \{i\}) = v(S \cup \{j\})$, then $\gamma_i(v) = \gamma_j(v)$.
- Dummy player: If for player i and for every coalition S not containing i, we have $v(S) = v(S \cup \{i\})$, then $\gamma_i(v) = 0$
- Additivity: For characteristic functions u and v, we have $\gamma(u + v) = \gamma(u) + \gamma(v)$, meaning that the value of two games played at the same time is equal to summation of their values if played at different times.

The Shapley value of player i is defined as the weighted mean of its marginal importance over all possible subsets of the players.

$$\gamma_i(v) = \frac{1}{n!} \sum_{\pi \in \Pi} \Delta_i(S_i(\pi)), \tag{3}$$

where Π is the set of all $n!$ permutations over N and $S_i(\pi)$ is the set of features (players) preceding player i in permutation π.

Since in feature selection, the order of features in a coalition does not change the value of coalition, the calculations in (3), can be further simplified by excluding the permutation of coalitions in the average:

$$\gamma_i(N, v) = \frac{1}{n!} \sum_{S \subseteq N/i} \Delta_i(S)|S|_i(n - |S| - 1))!, \tag{4}$$

where $S \subseteq N/i$ presents the coalitions that player i does not belong to. It is equivalent to the weighted average of coalitions, where the weight of each coalition is the number of its all possible permutations.

As shown in (3, 4), the Shapely value solution accounts for all possible coalitions that can be formed by the players [?]. Since in false alarm detection problem, the data set includes a large number of features, thereby calculating the Shapley value would be computationally intractable. Furthermore, considering the coalitions of a large number of features or all of them is practically unnecessary, since the maximum number of feature may interact with one another is much less than the total number of features. Therefore, we utilize the Multi-perturbation Shapley value measurement with coalition sizes up to L rather than the original Shapely value, which is determined using an unbiased estimator based on Shapley value [?] and [?].

In our proposed algorithm, at each round, the features are randomly divided into groups of size L. Then, we calculate the corresponding Multi-perturbation Shapely value of feature i inside its group, $\gamma_i'(v)$ considering all possible coalitions of size $1 \leq l \leq L$. This is equivalent to randomly sampling from uniformly distributed feature i, $\gamma_i'(v)$ is calculated as follows.

$$\gamma_i'(v) = \frac{1}{|\Pi_L|} \sum_{\pi \in \Pi_L} \Delta_i(S_i(\pi)), \tag{5}$$

where Π_L denotes the sampled permutation on sub-groups of features of size L. There is an essential trade-off to set L in the proposed method. Large L values consider higher order relations, while increasing the complexity of finding Multi-perturbation Shapely value at each subgroup. We conjecture that the optimum value of L for our datasets taking into account various factors such as the nature of data, number of features, and the inter-feature dependence is in the range of 3 to 6. This is confirmed by simulation results in section 5. It is worth noting that in most feature selection algorithms, each feature is being considered separately or equivalently $L = 1$.

Since the size of subgroups and the role of each group at the classification for the normalized data is almost equal, at the end of each iteration, the n_e less effective features are removed from the list, regardless of the enclosing subgroup. In order to minimize the impact of individual grouping, at the end of each iteration, we do not remove all features with Multi-perturbation shapely value below threshold as in [?]. Rather, we remove only n_e features with the lowest Multi-perturbation Shapely value (if below Multi-perturbation Shapely threshold γ_m). We choose n_e a small number, because i) the complexity reduces linearly with n_e and ii) the features with lower Multi-perturbation Shapely value may have a higher impact, when belong to another group in the next iterations. After removing the less contributing features, we randomly permute the remained features and repeat regrouping. Therefore, over the long run, the features are most likely visit any other features, since $L \ll N$. We terminate the algorithm if one of the following two conditions are violated; i) the minimum number of features n_m is

Table 2: Alarms definition

Alarm Type	Definition
Asystole	There is no QRS for at least 4s
Extreme Bradycardia	Heart rate $<$ 40 bpm for 5 consecutive beats
Extreme Tachycardia	Heart rate higher $>$ 140 bpm for 17 consecutive beats
Ventricular Tachycardia	At least 5 ventricular beats with heart rate $>$ 100 bpm
Ventricular Flutter/Fibrillation	Fibrillatory, flutter, or oscillatory waveform for at least 4s

reached or ii) the classification accuracy of all remaining features fall below a threshold T.

4 Data Set Description??

The database used for this study, which is publicly available through Physionet [?], was produced by four hospitals in the USA and Europe , using monitors with different manufacturers, unit-specific protocols, software versions and unit types. The definition of the alarm is presented in Table 2[?]. The total number of records is 219 and for each alarm a label including 'true', 'false', or 'impossible to tell' has been assigned by expert annotators. Interference from pacemakers and other noise artifacts may be present in the ECG signals.

5 Numerical Analysis Results

Experimental results are provided in this section for the proposed alarm validation method as well as other state of the art explicit feature selection methods including Chi-square, Gain Ratio, Relief and Info Gain methods. The Chi-square method evaluates a subset of features by finding their corresponding chi-squared statistics with respect to the class. The Gain ratio (GR) is an information based method that minimizes the conditional entropy of class given the selected features. The Relief method is an iterative alogorithm that starts with an initial weights for features and then iteratively adjusts the weights by randomly choosing an instance from data and weighting each feature based on its corresponding distance between the selected data instance and the closest instances in different classes to highlight features with higher discriminative properties. The Information Gain Ratio maximizes the mutual information between the selected features and the class labels. The numerical results are obtained utilizing the proposed coalition-game theoretic method where the multi-perturbation Shapley value is calculated for coalitions' size up to 4, $L = 4$.

The alarm typing rate for all feature selection methods are evaluated in combination with Bayes Net classification as a representative classifier. In all simulations, the 30 most informative features are selected to compare the performance of different feature selection techniques.. The comparison results in Fig. 1 suggest a considarable improvement for the proposed method in discarding the false alarms compared to the competitor methods. The alarm typing success rate for the proposed method is about 75% meaning that only 25% of alarms are deemed false, whereas the false alarm report rate for the best competitor method (Gain Ratio) is at least $100 - 68.88\% \approx 31\%$. The improvement is due to potential synergic impact of coalitions among features which is overlooked or not directly addressed in other methods. The proposed method outperforms the case of incorporating all wavelet coefficients (represented by None in Fig. 1) due to eliminating the irrelevant features. Another important observation in Fig. 1 is that the obtained success rate using feature set 1 (Statistical features) is slightly better than that of feature set 2 (Entropy-based features), meaning that feature set 1 provides more useful information in recognizing the true alarms. Interestingly, this is consistent among all feature selection methods. Although feature set 2 is relatively successful in identifying the true alarms, however adding it to the statistical features does not enhance alarm typing success rate suggesting that it does not bear additional information. It is notable that the promising rate of 75% is obtained using only 30 statistical features for any subject, which significantly reduces the risk of over-fitting compared to using all 18000 wavelet coefficients for each signal.

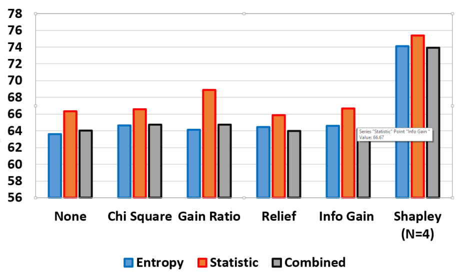
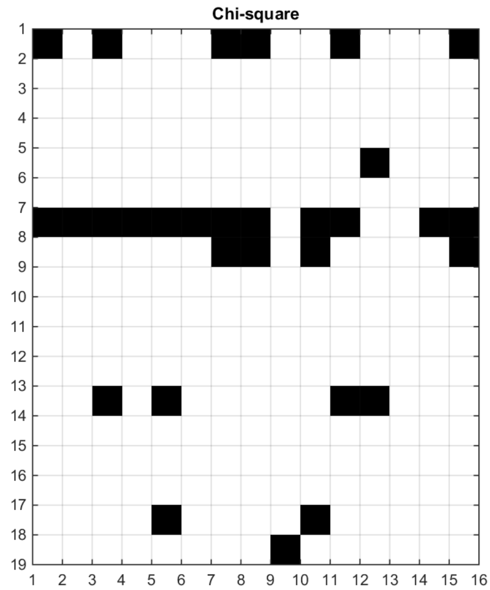


Fig. 1: True alarm recognition rate for the first 30 features using different feature selection methods with Bayes Net classification.

Fig. 8, presents the appearance of 30 most informative features in identifying alarm validity for all methods. The average over all methods is also depicted in Figure 8. This figure reveals that all statistical properties contribute almost equally to the false alarm recognition. However, there is a significant difference in the contribution of various signal source levels. The average appearance of features and signals are re-depicted in Fig. 11. It is clear from the results in Fig. 11 that the first level of ECG and $PLETH$ signals play significantly higher roles in the alarm validation. Indeed, the collective contribution of levels 2 to 6 are less than contribution of level 1 solely. However, all levels of signal $APB\ signal$ contribute almost equally for alarm recognition.

6 Conclusion



0.3

Fig. 2: Chi-Square
0.3

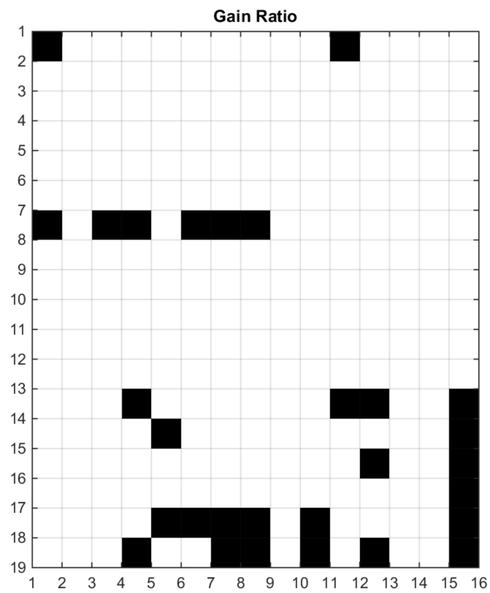
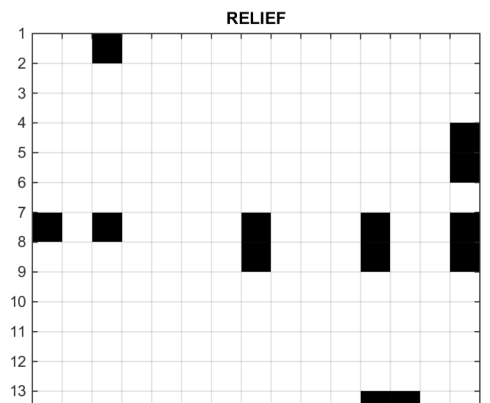


Fig. 3: Gain Ratio
0.3



0.45
(a)
t

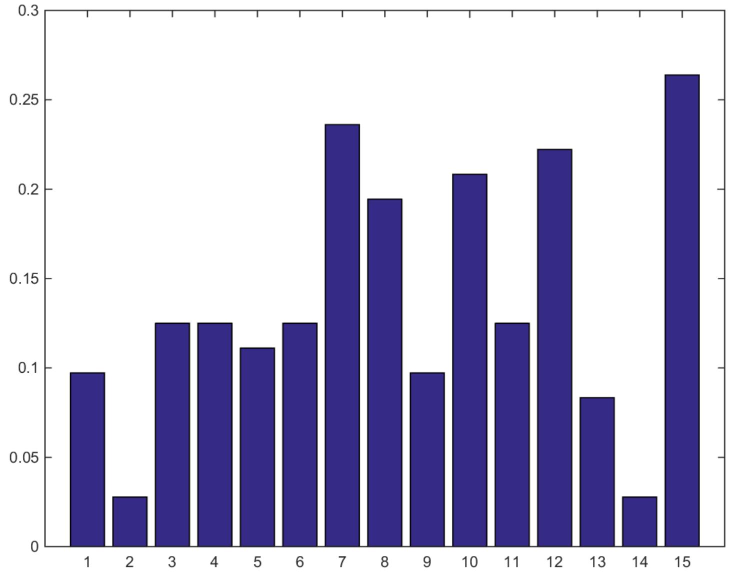


Fig. 9: Average Appearance of Features

0.45
(a)
t

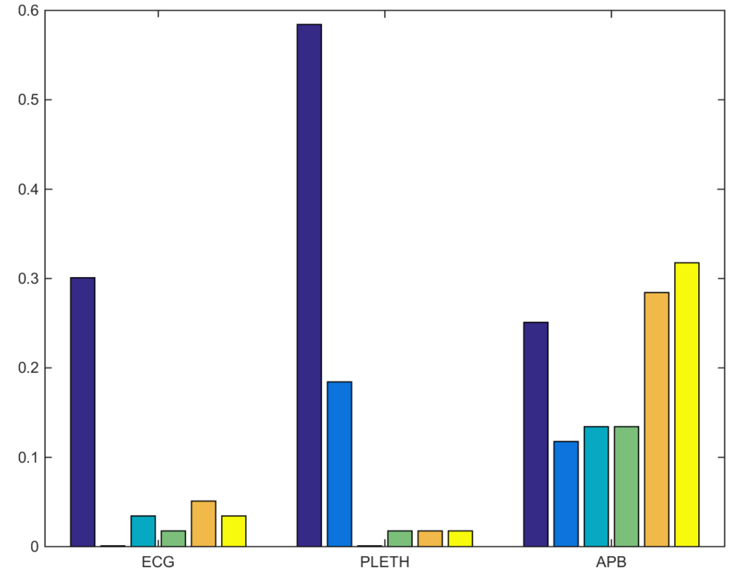


Fig. 10: Average Appearance of Signals

Fig. 11: Average Results for Feature Selection