# Blood Loss Severity Prediction using Game Theoretic Based Feature Selection

Abolfazl Razi[1], Fatemeh Afghah[2], Ashwin Belle [3], Kevin Ward [3] and Kayvan Najarian[4],

***Abstract*— Detection of hypovolemia in the early stages of hemorrhage is an important but unsolved problem in medicine. Many preventable deaths amongst critically injured patients happen due to delayed treatment of uncontrolled hemorrhage. Using a database of physiological signals collected during simulated hemorrhage on human subjects, our research applies a variety of signal processing techniques to extract a multitude of features that enable the prediction of the severity of hemorrhage. In this study, a novel feature selection method based on coalition game theory has been proposed which helps identify the most valuable features and thereby reduce the size of the feature space. This reduction in feature space not only improves the efficiency, but also improves the prediction accuracy and reliability of the developed model. This feature selection algorithm is independent of the underlying classification method and can be combined with any classification method based on the employed data. The proposed feature selection method significantly enhances the prediction accuracy by optimally selecting the features compared to the state of the art.**

## I. Introduction

Trauma induced hemorrhage has been a major factor in preventable deaths, both in civilian and combat settings [1], [2]. External or internal Hemorrhage when untreated can quickly lead to hypovolemia followed by hemorrhagic shock. Factors such as the types of treatment and how quickly it is administered can hinge the overall chances for survival of trauma patients. In combat settings nearly $20\%$ of the trauma patients die even before reaching a treatment facility, of which nearly $50\%$ of the deaths are caused by hemorrhage [3], [4]. Hence rapid assessment of the severity of hemorrhage as well as accurate triage decision and treatment is crucial for the survival of such patients. However, assessment of hypovolemia and the severity of blood loss in a patient can be a challenging problem, especially considering the time critical nature of hemorrhage. Therefore the ability to detect changes in blood volume and to be able to predict the severity of blood loss can be very important in providing early and successful intervention. There is a need for such a system since available physiological signals from the patient such as heart rate, oxygen saturation, arterial blood pressure and arterial hemoglobin does not reveal any early signs of hemorrhage until the onset of cardiovascular decompensation [5], by which time it could be too late. In our previous work we successfully developed a variety of signal processing and machine learning based systems which extracted features from heart rate variability (HRV) [6], morphology of electrocardiogram (ECG) [7], and from other physiological signals [8]. These systems utilized the variety of features and bio-markers that were extracted from these signals to predict the severity of hemorrhage. However, as the size of the features space being extracted grew larger, it became obvious that the size and quality of this feature set can highly affect the efficiency and performance of machine learning algorithms and its predictive accuracy. Hence, the following study proposes a novel game theoretic based feature selection approach, which improves the efficiency and accuracy of hemorrhagic shock prediction.

Using large feature sets imposes some restrictions during storage, search and classification steps. The possibility of redundant information as well non-informative features tends to cause a whole host of problems during the classification stage such as inefficiency, over fitting, reduction in accuracy etc. To overcome these issues and improve classification outcomes, it is important to reduce the feature space to a more concise and relevant set of features. Different feature selection algorithms have been studied in the literature [9]–[11] to reduce the data dimensions and recognize the irrelevant and redundant features. The irrelevant features do not have any useful information related to the target, while the redundant features do not provide any more information than the features, which have been already selected. These mechanisms improve the prediction algorithm performance and also reduce the required storage. In general, these methods are divided into three categories of embedded, filtering, and wrapper methods.

In embedded methods, the feature selection is not performed implicitly, rather the classification model is such that the contribution of irrelevant features become limited. For instance, in a fully Baysian RVM method, the irrelevant features does not contribute to the classification due to the vanishing corresponding coefficients in the model [12], [13]. However, these method are not desired in applications that the explicit list of active features are desired. For instance, in a genomic data analysis, providing the list of contributing genes is an essential requirement. The wrapper methods are based on the learning algorithms, where the classifier is retrained for any new data sets. Although this method results in a good performance, intensive required computations and the risk of over fitting limit the application of this approach in large datasets. Moreover, these methods are sensitive to the classification method and should be performed for each new classification method [14], [15]. Filtering methods, on the other hand, do not use a learning mechanism for feature

[1]A. Razi is with the Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708 abolfazl.razi@duke.edu

[2]F. Afghah is with the department of Electrical and Computer Engineering, North Carolina A&T State University, Greensboro, NC 27410 fafghah@ncat.edu

[3]A. Belle and K. Ward are with the Department of Emergency Medicine and Michigan Center for Integrative Research in Critical Care, University of Michigan, Ann Arbor, MI 48109 {bellea, keward}@umich.edu

[4]K. Najarian is with Department of Computational Medicine and Bioinformatics, and Michigan Center for Integrative Research in Critical Care, University of Michigan, Ann Arbor, MI 48109 kayvan@med.umich.edu

selection. In theses methods, the features are ranked based on a measurement metric related to the target such as distance, mutual information [16] or Pearson correlation.

An important disadvantage of the majority of filter-based methods is that the features are evaluated separately and the possible correlation between them is neglected. This fact degrades the feature selection/classification algorithms performance [17]. The most recently proposed methods consider the inter-feature dependence. However, the correlation based methods such as LDA [18], [19] do not capture the whole relation among features. The information theoretic methods [20], [21] maximize the mutual information between the features and the target class, while keeping the pairwise mutual information among features at a minimum level. These methods capture the inter-feature dependence and provide higher performances. However, obtaining inter-features mutual information is challenging and computationally expensive [22]. Furthermore, in majority of these methods, only pairwise features dependence among features are considered. Therefore, the impact of features, when considered within a bigger group is overlooked and these features may be discarded.

In this work, we utilize a *Coalition Game Theory* based feature selection method to address the aforementioned issues in feature selection. This game-theoretic based method accounts for the relevance between all potentially effective combination of the features to improve the classification algorithm performance. This enables the feature selection algorithm to recognize the features that despite their weak individual contribution to the classier, have an accountable impact when grouped with other features. Moreover, the number of searches using the proposed iterative method is significantly below the extensive search methods, while providing the equivalent performance.

The rest of this paper is organized as follows. An introduction to Coalition game theory is provided in section II. The utilized data of this study is introduced in section III. In section IV, a brief overview of the feature extraction method is provided. The utilized game theory based feature selection method is defined in section V. Numerical analysis results are presented in section VI, followed by conclusion in section VII.

## II. Introduction to Coalition Game

In coalition 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 between the members. Let $N$ be the the number of players and $P = \{P_1, ..., P_N\}$ be the set of players. The strategy of player $i$, $P_i$ is denoted by $X_i$. $S$ denotes a coalition set, $S \subseteq P$. In general, for a $N$-player game there exists $2^N$ possible coalitions of any size. The empty coalitions is shown by $\phi$, while grand coalition refers to the coalition of all players. Characteristic function $v(S)$ represents the total payoff can be gained by the members of coalition $S$. The characteristic of an empty coalition is zero, $v(\phi) = 0$. If $S_i$ and $S_j$, $(S_i, S_j \subseteq P)$ are two disjoint coalitions, the characteristic function of their union has super-additivity property, meaning that $v(S_i \cup S_j) \geq v(S_i) + v(S_j)$.

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 between the members. Let vector $\mathbf{x} = (x_1, ..., x_N), \mathbf{x} \in \mathbf{R}^N$ represent the amount of payoff that different members gained in a coalition. The allocation of coalition's payoff between the members must be efficient such that $\sum_{i=1}^{N} x_i = v(P)$ and also it needs to satisfy the individual rationality condition meaning that each player is getting paid better in a coalition than being on his own,$\forall i \in P, x_i \geq v(P_i)$.

## III. Data Analysis

The physiological signal database used in this study was collected at the U.S. Army Institute of Surgical Research (USAISR) in San Antonio, TX. The study was conducted on human volunteers and was approved by the internal review board of the research institution. In this study the subjects were subjected to a procedure called the lower body negative pressure (LBNP) which simulates hemorrhage [23]. The protocol begins with an initial 5-minute rest period wherein the data collected is used as baseline or normal. After which subjects are then exposed to successive levels of stepwise increasing negative pressure to the lower body, performed using a LBNP chamber, thereby simulating central hypovolemia in the subjects. Multiple physiological signals were collected during the protocol at $500$ samples per seconds, of which the most significant signals i.e. ECG, impedance and arterial blood pressure (ABP) are used in this study. The data collection ends when the subject faints. In total $176$ subjects data were used for this study. The time series data from each is labeled into four distinct stages of the severity of hemorrhage based on the levels of induced LBNP. The four stages are normal, mild, moderate and severe.

Various signal processing techniques are applied on these signals to extract a multitude of features from these time series waveforms. Features extracted from all the signals are then combined to form a very large feature set for each subject. The feature set is then reduced using a novel game theory based feature evaluation technique. Finally, the reduced feature set is used for classification of hemorrhagic severity using machine learning methods.

## IV. Feature Extraction

For the feature extraction step, a windowing technique is used to parse through the entire waveforms. Non-overlapping windows of $20$ seconds are used for extracting a variety of statistical and domain based features. For the ABP and impedance signals statistical features such as mean, standard deviation, minimum, maximum, median, range, skewness and kurtosis are calculated for each window. Also using a peak detection technique breathing rate is extracted from the ABP signal [24]. For the ECG signal, in addition to these statistical features, the P-QRS-T components of the ECG waveform is detected [7], [24], [25] . From the detected QRS complex, a derivative signals such as heart rate and heart-rate variability (HRV) are computed. By applying discrete wavelet transform (DTW) on the HRV signal, features are extracted from each level of decomposition. In addition to the DTW features, more statistical as well as power spectral density based features are extracted from the HRV signal.

Further details on the HRV analysis can be found in one of our previous works [6]. Furthermore, several more features are extracted from the raw ECG signal by transforming it using Dual-Tree-Complex-Wavelet transform, in which in addition to all the statistical features, features such as complexity, mobility as well as information theoretic based KL distance are also extracted. So in all, 352 features are extracted from each window at each stage of the LBNP, thereby aggregating into a very large feature set for each subject. Thus a novel and efficient feature evaluation and reduction technique was developed to cope with the large feature space.

## V. Feature selection Using Game Theory

Cooperative game theory has been recently utilized in feature selection algorithms [26], [27]. A coalition game can be defined by the set of players and the characteristics function for every set $S \subseteq P$ as $(P, v(S))$. Following this framework, we model the features as game players, and the features can be classified in different coalitions, noting their impact on the classifier and their interdependency. Different possible grouping of the features are examined to recognize the optimal classification. Payoff of each coalition $S$, $v(S)$, measures the contribution for a coalition of selected features to the performance of the classifier (e.g. success rate in supervised learning).

If feature $i$ joins a coalition $S$, it may improve the classification capability of this coalition. This is called marginal importance and is defined as

$$\Delta_i(S) = v(S \cup \{i\}) - v(S) \tag{1}$$

In this approach, the Shapley value of player $i \in P$ is defined as the expected marginal importance of player $i$ to the set of players who precede this player.

$$\gamma_i(v) = \frac{1}{N!} \sum_{\pi \in \Pi} \Delta_i(S_i(\pi)), \tag{2}$$

where $\Pi$ is the set of all $N!$ permutations of $P$ and $S_i(\pi)$ is the set of features (players) preceding player $i$ in permutation $\pi$. The Shapley value proposes a fair solution of the coalition game, since it is efficient and the summation of the marginal importance of all players is equal to the characteristic function of the coalition $P$, $v(P)$.

Since the order of features in a coalition does not change the coalition power, the calculations in (2), can be further simplified by excluding the permutation of coalitions in the average:

$$\gamma_i(P, v) = \frac{1}{N!} \sum_{S \subseteq N/i} \Delta_i(S)|S|_i(N - |S| - 1))!, \tag{3}$$

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 (2, 3), the Shapely value solution accounts for all possible coalitions that can be formed by the players [28], thereby calculating this would become computationally intractable specially when the data set has a large number of features, as is the case in our study. 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. For instance, one may expect that only up to a few number of features may be related in our application. Therefore, we propose an algorithm based on utilizing the Multi-perturbation Shapley value measurement with coalition sizes up to $L$ rather than the original Shapely value. This factor is determined using an unbiased estimator based on Shapley value and has been utilized in analysis of neural systems [29] and gene multi-knockout studies [30].

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.

$$\gamma'_i(v) = \frac{1}{|\Pi_L|} \sum_{\pi \in \Pi_L} \Delta_i(S_i(\pi)), \tag{4}$$

where $\Pi_L$ denote 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 (which exponentially grows with $L$). 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 10. This is confirmed by simulation results in section VI. 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 [29]. Rather, we remove only $n_e$ features with the lowest Multi-perturbation Shapely value (if below Multi-perturbation Shapely threshold $\gamma_m$). 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 reached or ii) the classification accuracy of all remaining features fall below a threshold $T$. This proposed Coalition Game-theoretic Feature Selection (CFGS) algorithm is summarized as follows.

## VI. Numerical Analysis Results

In this part, we present the numerical analysis to evaluate the performance of the proposed method. The game-theoretic

TABLE I

FEATURE SELECTION WITH THE PROPOSED CGFS METHOD COMPARED TO THE OTHER STATE OF THE ART METHODS. CLASSIFICATION/CLUSTERING ACCURACY OF DIFFERENT METHODS ARE SHOWN UTILIZING THE BEST 20 SELECTED FEATURES.

| | CLASSIFIER | | | |
|---|---|---|---|---|
| FEATURE SELECTION | NAIVEBAYES | RANDOM FOREST | K MEANS | SVM:LINEAR |
| NOFS | 0.5408 | 0.8616 | 0.7770 | 0.7712 |
| IBF | 0.5257 | 0.8371 | 0.7751 | 0.8062 |
| GR | 0.5293 | 0.8382 | 0.7771 | 0.8098 |
| CFS | 0.5678 | 0.8429 | 0.7771 | 0.7883 |
| CHI-SQ | 0.5337 | 0.8400 | 0.7685 | 0.8076 |
| RELEIF | 0.5023 | 0.7688 | 0.7724 | 0.7199 |
| CGFS | **0.6250** | **0.8625** | **0.7775** | **0.8875** |

**Algorithm 1** Proposed Coalition game-theoretic based feature selection (CGFS) algorithm

1) set parameters ($N, L, n_e, n_m, T, \gamma_m$, MaxIter)
2) **for** iteration 1 to MaxIter
3) randomly assign features to groups of size L
4) **for** group g=1 to number of groups
5) calculate $\gamma_i'$ for all feature inside group $g$
6) **end for**
7) remove up to $n_e$ features with $\gamma_i' < \gamma_m$
8) calculate classification performance ($Acc$)
9) exit if $Acc < T$
10) exit if number of remaining features $< n_m$
11) **end for**

based feature selection method is performed on our data set and the effect of the selected features in the clustering method is studied. As stated before, the impact of each feature on the classification performance is quantized by Multi-perturbation Shapely value through averaging over all coalitions up to size $L$, after a randomly developed groups. The parameters are set to $N = 352, L = 4, n_e = 1, n_m = 5, T = 0.4, \gamma_m = 0, \text{MaxIter} = 250$, unless specified otherwise. The proposed feature selection is general and the Multi-perturbation Shapely value of features at each round can be calculated as the accuracy of different classification/clustering algorithms. In this experiment, we utilized Naive Bayes (NB), Support Vector Machine (SVM) with linear, Random Forest as well as $K$ means clustering. Furthermore, the performance of these classifier on the extracted features from CGFS algorithm is compared with the performance the same classifier on the obtained features from various other feature selection techniques.

The results of this comparison using the most relevant 20 features are provided in Table 1. In this table, (NoFS) refers to classification including all features. We also examined several standard filter based methods using Weka ver. 3.6 package. In this experiment the following filter based feature selections are used: i) Information based feature selection (IBF), which maximizes the mutual information between selected features and the labels; ii) Gain ratio (GR) is another information based method that minimize the conditional entropy of class given the selected features; iii) The Correlation-based Feature Subset Selection (CFS) that selects a subset of features with the highest correlation with the labels and the lowest correlation with one another; iii) Chi-square method simply which chooses a subset of features by evaluating the chi-squared statistic with respect to the class; and iv) The Relief method evaluates the worth of a feature by sampling an instance and considering the value of the given feature for the nearest instance of the same and different classes.

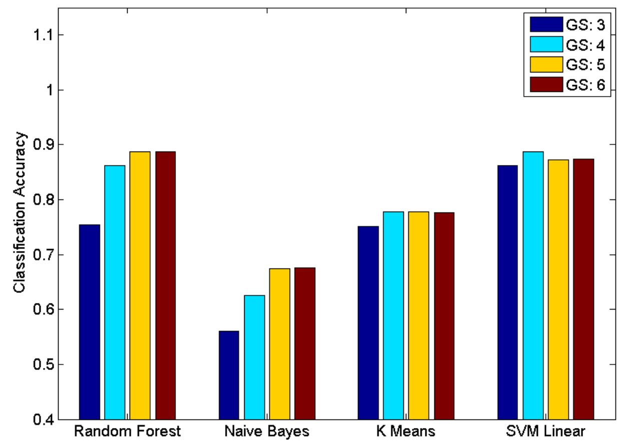


Fig. 1. Impact of coalition size in the performance of classification based on the selected features.

The results in this Table 1, demonstrate that the proposed method outperforms the competing methods by a significant margin, due to considerable impact of coalition among features as an important but overlooked characteristic of the real data. The superiority of the proposed CGFS algorithm comparing to CFS is due to the fact that the correlation based methods do not capture nonlinear dependence among the features. The GR and Chi-square methods treat the features separately and IBF only accounts for the pairwise inter-feature relations. Our proposed method even shows a better accuracy comparing to NoFS, since it eliminates the irrelevant features.

To eliminate the potential high complexity of the proposed coalition based algorithm, we propose to split the features into smaller groups at each iteration and then calculate the Shapely value for each feature inside the group. The optimal group size is a data-dependent parameter and is in the range of 1 to 10 in our case. The preliminary results show that the range of 4 to 6 captures all the inter-feature dependencies, while significantly reducing the computational costs. The feature grouping was performed randomly at each round of the algorithm; hence all features have the chance of visiting the dependent features. The results enable us to obtain high accuracy using about 20 features across almost all clustering/classification methods, which shows the quality of the final set of features. Moreover, the majority of surviving 20 features are information-theoretic based features that are the crux of our proposed feature extraction/optimization tasks, as discussed above.

As mentioned earlier, an important parameter of the proposed method is the group size. Since, the search complexity is highly dependent on the group size, finding the lowest possible group size that captures the whole inter-feature relation is very crucial that accelerates the feature selection method. This impact is depicted in Figure 1. As a general trend, increasing group size enhances the classification success rate, since considers the inter-feature relations in a bigger groups. However, this effect saturates and hence there is no point to use larger group sizes. The optimal group size is totally dependent on the nature of data, which is about 4 to 6 for our data, regardless of the employed classification/clustering method.

## VII. Conclusion

In this paper, we study a game theoretic feature-selection technique to extract the most effective features in prediction of the blood loss severity. In the proposed method, the problem of feature selection is modeled as a coalition game, where the features are considered as the game players and the marginal contribution of each player reflects the average contribution of the corresponding feature in the prediction process considering in conjunction with the other potential feature subsets. The results demonstrate that the utilized game theoretic method provides a superior classification/clustering performance, while significantly reducing the number of searches compared to the other state-of-the methods.

## References


[1] P. Teixeira, K. Inaba, P. Hadjizacharia, A. Brown, R. T. Browder, T. T. Noguchi, and D. Demetriades, "Preventable or potentially preventable mortality at a mature trauma center," *J Trauma*, vol. 63, pp. 1338–1346, 2007. [Online]. Available: http://www.nejm.org/doi/full/10.1056/NEJMoa0804748

[2] J. F. Kelly, A. E. Ritenour, D. F. McLaughlin, K. A. Bagg, A. N. Apodaca, C. T. Mallak, L. Pearse, M. M. Lawnick, H. R. Champion, C. E. Wade, and J. B. Holcomb, "Injury severity and causes of death from operation iraqi freedom and operation enduring freedom: 2003-2004 versus 2006," *J Trauma*, vol. 64, pp. 521–526, 2008.

[3] M. H. Weil, L. Becker, T. Budinger, K. Kern, G. Nichol, I. Shechter, R. Traystman, H. Wiedemann, R. Wise, M. Weisfeldt, and G. Sopko, "Post resuscitative and initial utility in life saving efforts (pulse): A workshop executive summary," in *Resuscitation*, vol. 50, 2001, pp. 23–35.

[4] H. B. Alam, D. Burris, and J. A. DaCorta, "Hemorrhage control in the battlefield: Role of new hemostatic agents," *Military Medicine*, vol. 170, pp. 63–69, 2005.

[5] V. Convertino, K. Ryan, C. Rickards, J. Salinas, J. McManus, W. Cooke, and J. Holcomb, "Physiological and medical monitoring for en route care of combat casualties," *Journal of Trauma*, vol. 64, pp. 342–353, 2008.

[6] S.-Y. Ji, A. Belle, K. R. Ward, K. L. Ryan, C. A. Rickards, V. A. Convertino, and K. Najarian, "Heart rate variability analysis during central hypovolemia using wavelet transformation," *Journal of Clinical Monitoring and Computing*, pp. 1–14, 2013.

[7] A. A. R. Bsoul, S. Ji, K. Ward, K. Ryan, C. Rickard, V. Convertino, and K. Najarian, "Prediction of severity of blood volume loss using features based on p, t and qrs waves," in *2009 American Heart Association (AHA) Resuscitation Science Symposium*, 2009.

[8] S. Ji, A. Bsoul, K. Ward, K. Ryan, C. Rickard, V. Convertino, and K. Najarian, "Incorporating physiological signals to blood loss prediction based on discrete wavelet transformation," in *2009 American Heart Association (AHA) Resuscitation Science Symposium*, 2009.

[9] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, March 2003. [Online]. Available: http://dl.acm.org/citation.cfm?id=944919.944968

[10] Y. Saeys, I. n. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, Sep. 2007.

[11] C. Lazar, J. Taminau, S. Meganck, D. Steenhoff, A. Coletta, C. Molter, V. de Schaetzen, R. Duque, H. Bersini, and A. Nowe, "A survey on filter techniques for feature selection in gene expression microarray analysis," *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, vol. 9, no. 4, pp. 1106–1119, 2012.

[12] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Sep. 2001. [Online]. Available: http://dx.doi.org/10.1162/15324430152748236

[13] Y. Mohsenzadeh, H. Sheikhzadeh, A. Reza, N. Bathaee, and M. Kalayeh, "The relevance sample-feature machine: A sparse bayesian learning approach to joint feature-sample selection," *Cybernetics, IEEE Transactions on*, vol. 43, no. 6, pp. 2241–2254, 2013.

[14] J. Huang, Y. Cai, and X. Xu, "A wrapper for feature selection based on mutual information," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 2, 2006, pp. 618–621.

[15] R. Kohavi and G. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, pp. 273–324, 1997.

[16] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 8, pp. 1226–1238, 2005.

[17] J. Fan, R. Samworth, and Y. Wu, "Ultrahigh dimensional feature selection: Beyond the linear model," *Journal of Machine Learning Research*, vol. 10, pp. 2013–2038, 2009.

[18] D. M. Witten and R. Tibshirani, "Penalized classification using fisher's linear discriminant," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 10 2011.

[19] M. T. W. Irina Gaynanova, James G. Booth, "Supervised classification using sparse fisher's lda," *arXiv:1301.4976*.

[20] Z. Nenadic, "Information discriminant analysis: Feature extraction with an information-theoretic objective," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 8, pp. 1394–1407, 2007.

[21] M. Chen, W. R. Carson, M. R. D. Rodrigues, L. Carin, and A. R. Calderbank, "Communications inspired linear discriminant analysis." in *ICML*. icml.cc / Omnipress, 2012.

[22] T. Suzuki, M. Sugiyama, and T. Tanaka, "Mutual information approximation via maximum likelihood estimation of density ratio," in *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, 2009, pp. 463–467.

[23] W. H. Cooke, K. L. Ryan, and V. A. Convertino, "Lower body negative pressure as a model to study progression to acute hemorrhagic shock in humans," *J. Appl. Physiol.*, vol. 96, pp. 1249–1261, 2004.

[24] Y. Luo, "The severity of stages estimation during hemorrhage using error correcting output codes method," *VCU digital archives, Virginia Commonwealth University, Richmond, Virginia*, 2012.

[25] A. Bsoul, S. Ji, K. Ward, and K. Najarian, "Detection of p, qrs, and t components of ecg using wavelet transformation," in *2009 IEEE/ICME International Conference on Complex Medical Engineering (CME2009)*, 2009.

[26] X. Sun, Y. Liu, J. Li, J. Zhu, H. Chen, and X. Liu, "Feature evaluation and selection with cooperative game theory," *Pattern Recogn.*, vol. 45, no. 8, pp. 2992–3002, Aug. 2012. [Online]. Available: http://dx.doi.org/10.1016/j.patcog.2012.02.001

[27] G. Cohen, S. Dror and G. Ruppin, "Feature selection via coalitional game theory," *Neural Computation*, vol. 19, no. 7, pp. 1939–1961, 2007.

[28] L. S. Shapley, "A value for $n$-person games," *H. W. Kuhn, and A. W. Tucker (Eds.), Contributions to the theory of games*, vol. 2, pp. 307–317, 1953.

[29] A. Keinan, B. Sandbank, C. Hilgetag, I. Meilijson, and E. Ruppin, "Axiomatic scalable neurocontroller analysis via the shapley value," *Artificial Life*, vol. 12, pp. 333–352, 2006.

[30] K. M. Kaufman, A. and E. Ruppin, "Multi-knockout genetic network analysis: The rad6 example," in *IEEE Computational Systems Bioinformatics Conference (CSB04)*, 2004, pp. 332–340.