

**Journal of Scientific Perspectives**

Volume 3, Issue 4, Year 2019, pp. 261-270

E - ISSN: 2587-3008

URL: <http://ratingacademy.com.tr/ojs/index.php/jsp>DOI: <https://doi.org/10.26900/jsp.3.028>

Research Article

**A SUPPORT VECTOR-BASED PREDICTIVE MODEL TO REVEAL THE RELATIONSHIPS AMONG ANTIBODY FEATURES AND THEIR EFFECTIVE FUNCTIONS AGAINST HIV****Ferdi SARAÇ \***

\* Suleyman Demirel University, TURKEY,

e-mail: [ferdisarac@sdu.edu.tr](mailto:ferdisarac@sdu.edu.tr)ORCID ID: <https://orcid.org/0000-0002-7080-1634>

Received: 24 September 2019; Accepted: 8 October 2019

**ABSTRACT**

Despite 4 decades' effort, an effective HIV-1 vaccine has not been produced owing to the inevitable antigenic diversity of the virus and millions of people around the world have lost their lives due to HIV. Increasing the knowledge of adaptive immune response to vaccination would ultimately lead to an effective HIV cure. Antibodies, which are responsible for protection and fighting against antigens, are vital parts of immune system response. In order to identify discriminative antibodies, which provide protection against HIV, and to disclose the associations between antibody features and their functional outcomes, computational methods, such as feature selection, regression and classification can be used to construct predictive models. Here we used our unsupervised K-Means Based Feature Selection (KBFS) method which is presented in our previous study, to identify functional antibodies that fight against HIV. The accuracy results for the proposed KBFS framework are compared with those presented in a recent study and are also compared with results from four different state-of-the-art unsupervised feature selection methods, namely MCFS, InFS, LapFS, and SPFS, along with the entire feature set. Then, support vector based systems are utilised to predict the associations between antibody features and their functional activities, namely gp120-specific antibody dependent cellular phagocytosis (ADCP), antibody dependent cellular cytotoxicity (ADCC) and cytokine release activities on RV144 vaccine recipients. Pearson Correlation Coefficient (PCC) metric is used to evaluate the prediction accuracy of the predictive models and to be consistent with the previous study. Our SVR based KBFS framework presented higher accuracy than the original study by improving prediction performance 16% for ADCP assay, 200% for the ADCC assay.

**Keywords:** Feature Selection, RV144, HIV

## **1. INTRODUCTION**

Antibodies are specialised Y-shaped glycoproteins (gp) that are produced by plasma cells to defend against intruders that cause infection. Antibodies are crucial for the immune system since they play a role in protecting against foreign substances or antigens. Antibodies consist of two antigen-binding fragments: fragment antigen-binding (Fab) and fragment crystallisable (Fc). Fab regions are the arms of the antibodies called immunoglobulin G (IgG) which are responsible for the identification of infected cells [1]. On the other hand, Fc regions stimulate the innate immune system to neutralise antigens. Antigens that exist in vaccines stimulate immune system response by instructing B-cells in order to produce antibodies which are responsible for protection. Vaccine-induced immunity effectors, or antibodies, are important defenders against antigens, including HIV viruses. Vaccination provides active protection since it trains the immune system to recognise antigens. Then, the immune system produces specific antibodies to fight against the antigens. The function of antibodies is to recognise and bind to antigens. This detection process begins when antibodies recognise a small region on the surface of an antigen called the epitope [2]. Vaccine-mediated antibodies are important defenders against intruders including Human Immunodeficiency Virus (HIV) [3]. HIV attacks and destroys the immune system; indeed, it causes depletion of CD4-positive lymphocytes. The RNA of HIV has only nine genes that contain the code necessary to produce structural enzymes [5]. HIV poses a number of immunological threats to the human immune system due to its extensive genetic diversity. Furthermore, HIV is capable of developing countermeasures to avoid the effect of antibodies. HIV can prevent itself from being detected by the immune system thanks to its reverse transcription ability. This ability enables HIV to mutate approximately  $3 \times 10^5$  per nucleotide base [6]. Therefore, producing an effective vaccine which can elicit antibodies to block HIV is vital to neutralise the virus [7]. Novel vaccine strategies are required to overcome the aforementioned challenges posed by HIV. Increasing the knowledge of associations between virus and immune system would ultimately result in producing an effective vaccine; an example is RV144. Functional antibodies are considered to be HIV inhibitors [8]. These inhibitory antibodies are capable of binding to virions, reducing their movement across mucus and mediating a variety of Fc receptor-mediated anti-HIV-1 activities, such as Antibody Dependent Cellular Cytotoxicity (ADCC) [9] [10]. ADCC-mediated antibodies can eradicate HIV infected CD4 cells [11] and block the transmission of HIV within 24 hours after viral entry [12].

Vaccination is a provider of active immunity since it stimulates the immune system to produce antibodies which fight against a virus. Interestingly, specific antibodies provide protection against specific antigens [13]. Moreover, the amounts of antibodies that are produced by the immune system are statistically related to the protection given, since antibodies will be needed for the subsequent attacks from antigens [14]. The functional characteristics of antibodies are also crucial for HIV protection; therefore, the identification of specific antibodies that mediate effector functions to neutralise HIV is essential for producing an effective HIV vaccine. Antibodies can also collaborate with other functions to provide prevention against viruses [15] [16]. Therefore, the identification of discriminative antibody features is crucial in producing prevention against HIV.

## **2. METHODS AND MATERIAL**

In this section methods and materials which are used in this study is presented.

### **2.1. The Dataset**

The RV144 data set provided in [1] is utilised in this study to model their antibody feature function relationships and to disclose HIV-specific antibodies. This data set contains

100 plasma samples (20 of them are placebo and 80 of them are vaccine injected) obtained from the individuals who participated in the RV144 vaccine trial at week 26.

Three different cell-mediated assays are used in this thesis: Antibody Dependent Cellular Phagocytosis (ADCP), Antibody Dependent Cellular Cytotoxicity (ADCC), and Natural Killer Cell Cytokine release. ADCC antibodies are capable of identifying infected cells, and these antibodies are involved in the binding of epitopes of HIV-infected cells [18]. Cytotoxicity activities are mediated by Natural Killer (NK) cells which can directly kill virally infected cells by adhering to them. Cytokine release activities includes the NK cell surface expression of CD107a and the quantitative detection of MIP-1- $\beta$  and IFN- $\gamma$  levels produced inside the cell. An anti- gen microsphere based liquid array is applied to determine antibodies (gp41, gp140, p24, gp120, and V1V2) and subclasses (IgG1, IgG2, IgG3, IgG4).

HIV-1 vaccine injection has been shown to be less effective due to the intrinsic variability of the virus. The identification of distinctive antibodies which correlate with protection against HIV-1 infection, along with increasing the knowledge of associations between immune mechanisms and HIV, would ultimately result in the development an effective vaccine against HIV.

In this study, three different cell-mediated assays, ADCC, ADCP, and Cytokine release, are used. The purposes of exploiting the RV144 data set are: (i) to differentiate functional antibodies; (ii) to identify the relationships between the human immune system and the HIV virus; and (iii) to test the effectiveness of the DFSFR framework for the given data set. The aim of this study is, therefore, to predict functional relationships between antibody features and their functional activities in RV144 vaccine recipients. Each data sample has twenty antibody features that consist of features related to IgG subclass and antigen specificity.

## **2.2. Feature Selection Methods**

Four different feature selection methods along with the entire feature set are used to identify discriminative antibody features from RV144 dataset.

### **2.2.1. K-Means Based Feature Selection (KBFS)**

KBFS is proposed in our previous study [19] and applied to a high dimensional age dataset. Here we adapted this method to disclose the relationship among antibody features and HIV.

At the first stage of KBFS, the data is transposed in order that samples become features and features become samples. The reason for this is to ensure that instead of samples the features are clustered in k-means algorithm. Next, based on user-specified integer (k), the data will be divided into k- clusters. The K-means clustering algorithm sorts the features based on their distance from the centroids in each cluster, and it usually exploits Euclidean or squared Euclidean distance measure. On the other hand, KBFS identifies three centroids based on their distance to the center of each cluster. The purposes of this is to minimise randomisation error and to deal with outliers.

Euclidean distance which is the most commonly used distance metric is exploited to calculate the distances among centroids and features. Euclidean distance can be calculated from:

$$J = \sum_{j=1}^K \sum_{i=1}^n ||x_i - C_j||^2$$

where  $x_i$ s  $i = 1, 2, \dots, n$  are a set of features to be partitioned to  $K$  clusters and  $C_j$ s  $j = 1, 2, \dots, K$  are the centroid points.

In KBFS, three centroid points are exploited for each cluster and features are ranked based on their absolute distance values to those centroids. A feature with the lowest distance to the any of three centroid points in a cluster is considered as the most important one. In KBFS, distance measure is calculated by:

$$J_{i1} = \sum_{j=1}^K \sum_{i=1}^n ||x_i - C_{j1}||^2$$

$$J_{i2} = \sum_{j=1}^K \sum_{i=1}^n ||x_i - C_{j2}||^2$$

$$J_{i3} = \sum_{j=1}^K \sum_{i=1}^n ||x_i - C_{j3}||^2$$

The weight of a feature is then calculated by:

$$WX_i = \frac{1}{\min (J_{i1}, J_{i2}, J_{i3})}$$

The K-means method randomly initialises the centroids and this might profoundly affect the clustering results. Therefore, the process of KBFS is repeated 100 times to minimise the randomisation error. At the end, the mean of the distances between the centroids and the features are calculated in order to rank features. Therefore,

$$\frac{1}{WX_i} = \frac{1}{p} \sum_{t=1}^p \min \left( \sum_{j=1}^K \sum_{i=1}^n ||x_i - C_{j1}||^2, \sum_{j=1}^K \sum_{i=1}^n ||x_i - C_{j2}||^2, \sum_{j=1}^K \sum_{i=1}^n ||x_i - C_{j3}||^2 \right)$$

where  $p = 1, 2, \dots, 100$ ,  $C$  represents clusters,  $x_i$ s are features where  $i = 1, 2, \dots, n$ ,  $K$  is the number of clusters,  $WX_i$  is the weight of  $i$ -th feature and  $C_j$ s are centroids.

### **2.2.2 Multi-Cluster Feature Selection (MCFS)**

MCFS is a clustering based feature selection method that utilises sparse learning technique to select relevant features. MCFS uses spectral analysis to measure the correlation between different features.

### **2.2.3 Laplacian Score Feature Selection (LapFS)**

LapFS is a graph based unsupervised feature selection method that selects features which are important to preserve the data manifold structure. LapFS utilises pairwise similarities between features that are measured by using the heat kernel.

### **2.2.4 Spectral Feature Selection (SPFS)**

SPFS can be defined as an extension of LapFS, however, SPFS works for both supervised and unsupervised scenarios. SPFS selects features which are consistent with the graph structure of the data; however, unlike LapFS, SPFS also independently evaluates features.

### **2.2.5 Infinite Feature Selection (InFS)**

In InFS, each node feature represented with a node in a graph and features are selected according to their centrality score.

### **2.3. Support Vector Regression**

Support Vector Machine (SVM) is a powerful statistical supervised learning model which has been exploited for both regression and classification tasks [24] [25]. In classification scenario, SVM realises a discriminative hyperline which achieves the largest margin between two classes [25]. The hyperline can be formulated as:

$$f(x) = w * x + b$$

where  $w$  is normal to hyperline and  $\frac{b}{\|w\|^2}$  is the perpendicular distance from the perpendicular distance from hyperline to the origin. The regression version of SVM is called Support Vector Regression (SVR).

SVR constructs a model function which indicates relationship between features and the target variable. In SVR,  $\epsilon$ -intensive loss function is used so that only features out of  $\epsilon$  tolerance are penalised given by:

$$C \sum_i^n \zeta^p$$

where  $p$  is a positive integer, and  $\zeta$  is the orthogonal distance away from the  $\epsilon$ -region. In this study, SVR is implemented with radial basis kernel using the LIBSVM library [26].

### **2.4. Performance Evaluation Metrics**

In this study, Pearson Correlation Coefficient metric is used to evaluate the effectiveness of predictive models and to compare our results with previous study [1].

PCC is an evaluation metric that is exploited to assess the performance of predictive models. The PCC evaluates the strength of the relationship between two variables. It can be calculated as:

$$PCC = \frac{\{n\Sigma\{x_i y_i\} - \Sigma\{x_i\}\Sigma\{y_i\}\}}{\sqrt{\Sigma x_i^2 - (\Sigma x_i)^2} \sqrt{\Sigma y_i^2 - (\Sigma y_i)^2}}$$

where  $x$  and  $y$  are vales the two quantitative variables and PCC indicates the linear association between them. A value of PCC that is equal to 1 indicates a perfect linear correlation.

## **3. RESULTS AND DISCUSSION**

The RV144 data set provided in [1] is used in this study to model the antibody feature-function relationship. This data set contains 100 plasma samples (20 of which are placebo and 80 are vaccine-injected) obtained from the individuals participating in the RV144 vaccine trial at week 26. Three different cell-mediated assays are used: Antibody Dependent Cellular Phagocytosis; Antibody Dependent Cellular Cytotoxicity; and Natural Killer Cell Cytokine Release activities. The accuracy results for the proposed KBFS framework are compared with those presented in a previous study [1], and are also compared with results from four different

state-of-the-art unsupervised feature selection methods, namely MCFS, InFS, LapFS, and SPFS, along with the entire feature set. In this study, the PCC metric is used to be able to perform a consistent comparison with the previous study [1].

The SVR-based predictive models for the regression tasks are constructed using feature selection methods (filtered feature set). Their performance is then evaluated using a five-fold cross validation method. The RV144 data set is divided into two sets of samples. Four out of five samples, with a total of 64 samples, are utilised for training and the rest (16 samples) for testing purposes. This process is repeated 200 times by randomly creating subsets of the samples for the five-fold cross validation in order to avoid a bias towards and to assess the effect of randomisation in the cross validation. At the end, the mean performance and its corresponding standard deviation (std) values are obtained for each of the predictive models.

The prediction performance of unsupervised feature selection methods on three cell-mediated assays are summarised in Tables 1-3. Table 1 shows the PCC results of predictive models for Natural Killer Cell Cytokine release activities. The predictive models aim to estimate the level of cytokine release in order to understand its functionality for protection. The results suggest that KBFS outperforms state-of-the-art methods with 0.52 PCC using 16 features. SPEC yields the second-best result yielding 0.51 PCC, with 16 antibody features. Other methods produce average results.

The prediction results of unsupervised predictive models for ADCC activities are presented in Table 2. KBFS again produces the best result yielding 0.42 PCC using only 10 antibody feature. InFS produces the second-best result with 0.40 PCC utilising 14 antibody features. Other methods produce average results.

**Table 1.** Comparison of Unsupervised Feature Selection Methods for the Antibody Features and Natural Killer Cell Cytokine Release Activity Relationship

Metrics	PCC
KBFS (16)	0.52 ± 0.17
MCFS (16)	0.49 ± 0.17
Laplacian (16)	0.49 ± 0.18
SPEC (16)	0.51 ± 0.17
InFS (18)	0.49 ± 0.17
Baseline (20)	0.49 ± 0.17

**Table 2.** Comparison of Unsupervised Feature Selection Methods for the Antibody Features and Cellular Cytotoxic Activity Relationship

Metrics	PCC
KBFS(11)	0.43 ± 0.19
MCFS (18)	0.39 ± 0.18
Laplacian (12)	0.39 ± 0.18
SPEC (18)	0.41 ± 0.18
InFS (14)	0.40 ± 0.17
Baseline(20)	0.38 ± 0.18

**Table 3.** Comparison of Unsupervised Feature Selection Methods for the Antibody Features and Cellular Phagocytosis Activity Relationship

Metrics	PCC
KBFS(12)	0.65 ± 0.17
MCFS (17)	0.65 ± 0.14
Laplacian (3)	0.65 ± 0.15
SPEC (18)	0.61 ± 0.14
InFS (18)	0.64 ± 0.15
Baseline(20)	0.61 ± 0.15

Table 3 presents the prediction results of USFSMs for ADCP activities. As can be clearly seen in the table, KBFS filtered predictive model outperforms the predictive models implemented with the complete feature set, InFS and SPEC. On the other hand, KBFS, Laplacian Score and MCFS produce the same PCC results with 12, 3 and 17 antibody features respectively.

The prediction results of the proposed method are also compared with those of the previous study [1] where the same data set by using the same cross validation method is utilised (5-fold with 200 replicates) and comparison results are shown in Tables 4-6. The results appear to suggest that KBFS has a better quantitative accuracy than the predictive models constructed using Lars, GP and SVR as presented in the previous study for ADCC and ADCP assays, at 0.43 and 0.65 PCC respectively. In particular, the proposed approach yields as much as 1.16x and 3x better outcomes than the results of SVR for the ADCP and ADCC assays respectively. KBFS has slightly lower quantitative performance as compared to the predictive model for the Cytokine assay constructed using SVR as presented in the previous study. However, it still has better quantitative performance than the Lars and GP predictive models for the Cytokine assay.

Overall, the proposed KBFS method generally achieves the best performance on all cell-mediated assays, which thereby verifies that it is able to select informative antibody features.

**Table 4.** A Comparison of the Results with the Previous Study for the Antibody Features and Cellular Phagocytosis Activity Relationship

Regression	PCC
Lars [1]	0.61±0.15
GP [1]	0.53±0.16
SVR [1]	0.56±0.19
KBFS	0.65 ±0.17

**Table 5.** A Comparison of the Results with the Previous Study for the Antibody Features and Cellular Cytotoxic Activity Relationship

Regression	PCC
Lars [1]	0.42±0.18
GP [1]	0.24±0.21
SVR [1]	0.14±0.24
KBFS	0.43±0.19

**Table 6.** A Comparison of the Results with Previous Study for the Antibody Features and Natural Killer Cell Cytokine Release Activity Relationship

Regression	PCC
Lars [1]	0.51±0.21
GP [1]	0.46±0.24
SVR [1]	0.55± 0.15
KBFS	0.52±0.17

#### 4. CONCLUSION AND FUTURE WORK

In this study, a support vector-based (SV-based) predictive model is used to predict the associations among cell-mediated activities and multivariate antibody features in RV144 data set. Antibody features are filtered out using USFSMs in order to obtain attributes of high efficacy. This data set consists of 20 placebo samples and 80 vaccine injected samples, for a total of 100 data samples. Cross-validated SV-based predictive models presented higher accuracy than in the original study [1].

The RV144 data set is used to test the predictive capability of the proposed KBFS model for the given data set and to provide better generalisation and performance compared to a study conducted on the same data set [1].

The goal of the study is to disclose associations among antibody features and their effector functions. The effector functions can be described as actions of the immune system to fight against HIV. Therefore, the identification of specific antibody features involved in fighting against HIV is crucial in neutralising the virus.

Experimental results conducted on RV144 Vaccine data set suggest that the proposed KBFS method, outperforms the state-of-the-art USFSMs as well as the method used in the previous paper on the same data set. KBFS has a better quantitative accuracy performance than the predictive models constructed using Lars, GP and SVR presented in the data set paper for ADCC, and ADCP assays. KBFS has a little less quantitative performance as compared to predictive model for Cytokine assay constructed using SVR presented in the data set paper. However, it still has better quantitative performance than the Lars and GP predictive models for the Cytokine assay. By utilising KBFS framework, the number of features are reduced to 11 for ADCC assay, 12 for ADCP assay and 16 for Cytokine assay. However, in data set paper, the number of selected features are not indicated; instead, filtered set is mentioned without providing the number of selected features.

Experimental results conclude that the proposed unsupervised framework, KBFS, achieves generally the best performance on all assays, which thus verifies that it is able to reveal discriminative antibody features that provide protection against HIV.



## REFERENCES

- [1] S. Sadanand, T. J. Suscovich, G. Alter, Broadly neutralizing antibodies against hiv: New insights to inform vaccine design, *Annual review of medicine* 67 (2016) 185–200.
- [2] A. S. Clem, et al., Fundamentals of vaccine immunology, *Journal of global infectious diseases* 3 (2011) 73.
- [3] G. E. Seabright, K. J. Doores, D. R. Burton, M. Crispin, Protein and glycan mimicry in hiv vaccine design, *Journal of molecular biology* (2019).
- [4] M. Rolland, P. T. Edlefsen, B. B. Larsen, S. Tovanabutra, E. Sanders- Buell, T. Hertz, C. Carrico, S. Menis, C. A. Magaret, H. Ahmed, et al., Increased hiv-1 vaccine efficacy against viruses with genetic signatures in env v2, *Nature* 490 (2012) 417–420.
- [5] H. S. Bagalb, Cellular and Molecular Biological Studies of a Retroviral Induced Lymphoma, Transmitted via Breast Milk in a Mouse Model, Ph.D. thesis, University of Toledo, 2008.
- [6] A. Perelson, P. Essunger, D. Ho, Dynamics of hiv-1 and cd4+ lympho- cytes in vivo., *AIDS* (London, England) 11 (1996) S17–24.
- [7] K. L. Williams, M. Stumpf, N. E. Naiman, S. Ding, M. Garrett, T. Go- billot, D. V´ ezina, K. Dusenbury, N. S. Ramadoss, R. Basom, et al., Identification of hiv gp41-specific antibodies that mediate killing of in- fected cells, *PLoS pathogens* 15 (2019) e1007572.
- [8] C. Nilsson, S. Aboud, M. Bakari, E. F. Lyamuya, M. L. Robb, M. A. Marovich, P. Earl, B. Moss, C. Ochsenbauer, B. Wahren, et al., Potent functional antibody responses elicited by hiv-1 dna priming and boosting with heterologous hiv-1 recombinant mva in healthy tanzanian adults, *PLoS one* 10 (2015) e0118486. Bagalb, H. S. (2008). *Cellular and Molecular Biological Studies of a Retroviral Induced Lymphoma, Transmitted via Breast Milk in a Mouse Model* (Doctoral dissertation, University of Toledo).
- [9] G. D. Tomaras, B. F. Haynes, Strategies for eliciting hiv-1 inhibitory antibodies, *Current Opinion in HIV and AIDS* 5 (2010) 421.
- [10] H. L. Robinson, Non-neutralizing antibodies in prevention of hiv infec- tion, *Expert opinion on biological therapy* 13 (2013) 197–207.
- [11] Y. Guan, M. Pazgier, M. M. Sajadi, R. Kamin-Lewis, S. Al-Darmarki, R. Flinko, E. Lovo, X. Wu, J. E. Robinson, M. S. Seaman, et al., Di- verse specificity and effector function among human antibodies to hiv-1 envelope glycoprotein epitopes exposed by cd4 binding, *Proceedings of the National Academy of Sciences* 110 (2013) E69–E78.
- [12] R. Ahmad, S. T. Sindhu, E. Toma, R. Morisset, J. Vincelette, J. Menezes, A. Ahmad, Evidence for a correlation between antibody- dependent cellular cytotoxicity-mediating anti-hiv-1 antibodies and prognostic predictors of hiv infection, *Journal of clinical immunology* 21 (2001) 227–233.
- [13] M. E. Ackerman, A.-S. Dugast, G. Alter, Emerging concepts on the role of innate immunity in the prevention and control of hiv infection, *Annual review of medicine* 63 (2012) 113–130.
- [14] F. Sarac, V. Uslan, H. Seker, A. Bouridane, Exploration of unsupervised feature selection methods in relation to the prediction of cytokine release effect correlated to antibody features in rv144 vaccines, in: *Bioinformatics and Bioengineering (BIBE), 2015 IEEE 15th International Confer- ence on, IEEE, 2015, pp. 1–4.*

- [15] M. E Ackerman, G. Alter, Opportunities to exploit non-neutralizing hiv-specific antibody activity, *Current HIV research* 11 (2013) 365–377.
- [16] S. A. Plotkin, Correlates of protection induced by vaccination, *Clinical and Vaccine Immunology* 17 (2010) 1055–1065.
- [17] I. Choi, A. W. Chung, T. J. Suscovich, S. Rerks-Ngarm, P. Pitisut-tithum, S. Nitayaphan, J. Kaewkungwal, R. J. O’Connell, D. Francis, M. L. Robb, et al., Machine learning methods enable predictive model- ing of antibody feature: function relationships in rv144 vaccinees, *PLoS computational biology* 11 (2015) e1004185.
- [18] W. S. Lee, M. S. Parsons, S. J. Kent, M. Lichtfuss, Can hiv-1-specific adcc assist the clearance of reactivated latently infected cells?, *Frontiers in immunology* 6 (2015).
- [19] F. Sarac, H. Seker, A. Bouridane, Exploration of unsupervised feature selection methods to predict chronological age of individuals by utilising cpg dinucleoties from whole blood, in: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2017.
- [20] D. Cai, C. Zhang, X. He, Unsupervised feature selection for multi- cluster data, in: *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2010, pp. 333– 342.
- [21] X. He, D. Cai, P. Niyogi, Laplacian score for feature selection, in: *Advances in Neural Information Processing Systems*, 2005, pp. 507–14.
- [22] Z. Zhao, H. Liu, Spectral feature selection for supervised and unsupervised learning, in: *Proceedings of the 24th international conference on Machine learning*, ACM, 2007, pp. 1151–1157.
- [23] G. Roffo, S. Melzi, M. Cristani, Infinite feature selection, in: *Proceed- ings of the IEEE International Conference on Computer Vision*, 2015, pp. 4202–4210.
- [24] H. Drucker, C. J. C. Burges, L. Kaufman, A. J. Smola, V. N. Vapnik, *Support Vector Regression Machines*, volume 9 of *Advances in Neural Information Processing Systems*, MIT Press, 1996.
- [25] V. N. Vapnik, An overview of statistical learning theory, *Neural Net- works, IEEE Transactions on* 10 (1999) 988–999.
- [26] C.-C. Chang, C.-J. Lin, LIBSVM: A library for support vector machines, *ACM Transactions on Intelligent Systems and Technology* 2 (2011) 1–27.