cLP: Linear Programming with Biological Constraints and its Application in Classification Problems

Manli Zhou[1,2], Youxi Luo[1,3], Guoqin Mai[1], and Fengfeng Zhou[1,*]
[1] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, 518055, P.R. China
[2] University of Chinese Academy of Sciences, Beijing, 100049, China
[3] School of Science, Hubei University of Technology, Wuhan, Hubei, 430068, China
* Corresponding author: Fengfeng Zhou, phone: +86-755-86392200; fax: +86-755-86392299;
e-mail: FengfengZhou@gmail.com, or ff.zhou@siat.ac.cn. Web site: http://www.healthinformaticslab.org/ffzhou/.

***Abstract*—Feature selection represents a major challenge in the biomedical data mining problem, and numerous algorithms have been proposed to select an optimal subset of features with the best classification performance. However, the existing algorithms do not take into account the vast amount of biomedical knowledge from the literature and experienced researchers. This work proposes a novel feature selection algorithm, cLP, with the optimized binary classification accuracy. The proposed algorithm incorporates the biomedical knowledge as constraints in the linear programming based optimization model. The experimental data shows that cLP outperforms the other feature selection algorithms, and its constrained version performs similarly well with the unconstrained version. Although theoretically constraints will reduce the classification model performance, our data shows that the constrained cLP sometimes even outperforms the unconstrained version. This suggests that besides the benefit of including biomedical knowledge in the model, the constrained cLP may also achieve better classification performance.**

Keywords—constrained linear programming; feature selection; biological constraints.

I. INTRODUCTION

Modern biotechnologies have produced huge amount of biomedical knowledge and insights. Various tedious low-throughput screening technologies detected the phenotype-associated genes or other genomic functional elements by knocking out or knocking down one of these elements at a time, e.g. RNAi [1] or CRISPR/Cas9 [2, 3]. Genome-wide association studies and other high-throughput technologies were also widely used to screen the samples for millions of features at a time [4]. All these knowledge facilitates the frequent update of the biomedical textbooks, *e.g.* the 11^{th} version of the Lewin's GENES [5].

Biomedical data is being generated at an accelerated speed. Two years ago, Illumina's HiSeq 2000 machine may produce 200 Gbp genomic data with 100 bps per read within a week. The recently released X10 version of HiSeq achieves 1,800 Gbp total data production with 150 bps per read, and the running time of a full run is less than 3 days [6]. Such data amount leads to millions of features for a single sample [7], and causes major challenges for feature selection algorithms.

The majority of feature selection algorithms do not consider the existing biomedical knowledge in the model training. Taking a binary classification problem as an example, a feature selection algorithm usually tries to find a subset of features so that the intra-class pair-wise distance is significantly smaller than that of the inter-class distance. The ranking-based algorithm proposes a measurement of how discriminative each feature is between the two classes of samples, and some widely used ranking algorithms include t-test and Wilcoxon-test. And it still remains to be resolved how the existing biomedical knowledge is incorporated into the training of the feature selection model.

This work proposes a novel linear programming model with the biomedical knowledge based constraints. The comparative study shows that the proposed algorithm outperforms the existing feature selection algorithms in the measurements of sensitivity, specificity and accuracy. The experimental data also supports that the constrained linear programming model performs similarly well with the unconstrained version, and better on some datasets. Biological insights are also discussed about the features selected by the constrained model.

II. MATERIAL AND METHODS

A. Problem Model and Datasets

This study investigates the binary classification performance using the features selected by the feature selection algorithms. Firstly, a binary classification problem is defined to have two sample sets, *i.e.* the Positive and Negative sets usually consisting of the disease samples and controls, respectively. The optimization goal is to find a classification model that may accurately separate the two sample sets [8].

Four widely used datasets are used in this study to compare the classification performances based on the

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB13040400), Shenzhen Peacock Plan (KQCX20130628112914301), Shenzhen Research Grant (ZDSY20120617113021359), China 973 program (2010CB732606), the MOE Humanities Social Sciences Fund (No.13YJC790105) and Doctoral Research Fund of HBUT (No. BSQD13050). Computing resources were partly provided by the Dawning supercomputing clusters at SIAT CAS.

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

features selected by the investigated feature selection algorithms. The gene expression datasets of prostate cancer [9] and CNS cancer [10] are downloaded from the Broad Institute Genome Data Analysis Center (http://www.broadinstitute.org/cgi-bin/cancer/datasets.cgi). Two binary classification datasets on gastric cancer [11] are downloaded from the NCBI Gene Expression Omnibus (GEO) database with the ID GSE29272. The four datasets are denoted as Prostate, CNS, Gastric1 and Gastric2, respectively.

B. Linear Programming Model with Constraints (cLP)

This study proposes a formula of biological knowledge-based constraint for the linear programming model, denoted as cLP, to exploit the merits of both biological knowledge and the determined optimal solution of linear programming. Linear programming problem is theoretically proven to be solvable within a polynomial time [12], and have been widely used in many optimization problems [13, 14].

For a given binary classification problem $S_{n\times p}$ with n samples and p features, the proposed algorithm cLP aims to optimally separate the positive samples from the negative ones. Let n_t and n_f be the numbers of positive (T) and negative (F) samples, respectively. And we have $n=n_t+n_f$. Each feature has its weight ω_i, where i=1, 2, …, p. The linear programming model cLP(fFS) is described in the following formula, where fFS is the subset of features that must be retained in the final selected feature subset.

$$\min \sum_{i=1}^{p} w_i + \lambda\left(\sum_{k=1}^{n_t} \xi_k + \sum_{k=1}^{n_f} \xi_k\right) \quad (1)$$

subject to :

$$\sum_{i=1}^{p} \left|S_{ki} - m_i^t\right| w_i - \xi_k < \sum_{i=1}^{p} \left|S_{ki} - m_i^f\right| w_i, \text{ for } S_k \in T, k \in \{1, 2, ..., n_t\}; \quad (2)$$

$$\sum_{i=1}^{p} \left|S_{ki} - m_i^f\right| w_i - \xi_k < \sum_{i=1}^{p} \left|S_{ki} - m_i^t\right| w_i, \text{ for } S_k \in F, k \in \{1, 2, ..., n_f\}; \quad (3)$$

$$0 \le w_i \le 1, \text{ for } i \in \{1, 2, ..., p\}; \quad (4)$$

$$\xi_k \ge 0, \text{ for } k \in \{1, 2, ..., n_t + n_f\}; \quad (5)$$

$$w_s \ge \theta, \text{ for } s \in fFS. \quad (6)$$

The average level of the i^{th} feature of the positive and negative samples are denoted as m^t_i and m^f_i, i=1, 2, …, p, respectively. The first term of the optimization goal in formula (1) tries to minimize the sum of the weights of all the features, where only the selected features have non-zero weights. The second term in the optimization function measures the relaxed form of error rates [13]. The formula (2), (3) and (5) make sure that the positive and negative samples are closer to their respective centroid than to the other one. Formula (4) normalizes the feature weights between 0 and 1. Formula (6) makes sure that the pre-defined features in fFS have weights no smaller than θ. λ is evaluated in the range [0, 0.5] increased by 0.001, until the weights are converged.

The aforementioned linear programming model is solved by the lpSolveAPI in the software R version 3.1.1.

C. Classification Performance Measurements

Comprehensive comparisons are conducted to evaluate how cLP(fFS) performs when compared with the other feature selection algorithms. The four other feature selection algorithms used in this study are Prediction Analysis for Microarrays (PAM) [15], Regularized Random Forest (RRF) [16], t-test (Ttest) [17] and Wilcoxon-test (Wtest) [18]. The parameters of PAM was optimized by the amount of shrinkage, which is associated with the final list of selected features. The tree node impurity of RRF is measured by the Gini index [19], and all the features with positive Gini index are selected for further investigation. cLP, PAM and RRF will automatically return the optimized subset of features. Ttest and Wtest return only a measurement for each feature, and an ordered list of features based on this measurement. The numbers of features by Ttest and Wtest will be the same to that of the algorithm cLP.

The selected features are used to train a classification model using two classification algorithms, *i.e.* Support Vector Machine (SVM) [20] and Naive Bayes (NBayes) [21]. The Radial Basis kernel is chosen for the SVM model. 5-fold cross validation (5FCV) strategy is used to calculate the classification performance measurements, *i.e.* sensitivity (*Sn*), specificity (*Sp*) and accuracy (*Ac*), as similar in [22-24]. *Sn*, *Sp* and *Ac* are defined to the percentages of corrected predicted positive samples, negative samples and all the samples, respectively.

This study focuses on the hypothesis that fixing biological knowledge in the classification modeling only leads to minor decrease in the classification performance, and performs similarly well to or better than the other feature selection algorithms. The default parameter values of data mining algorithms are usually optimized by the existing software for general purposes. So all the other parameters use the default values provided by the software R version 3.1.1.

III. Results and Discussion

A. cLP(fFS) versus cLP({})

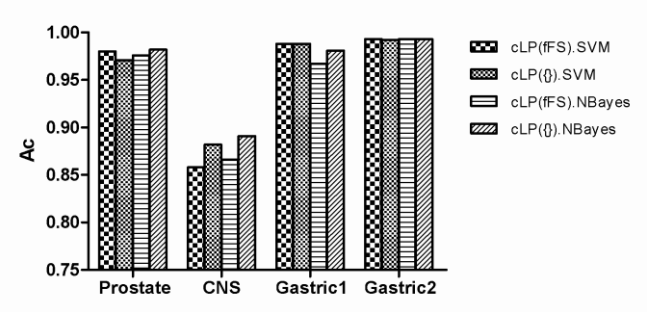
Firstly, a comparison is carried out to investigate whether fixing a few features in the model will significantly decrease the overall binary classification performance. By fixing a few features, the solution sub-space is smaller than the original one and theoretically the new optimal solution cannot outperform the optimal solution in the original solution space. For a given subset of features selected by cLP(fFS), the binary classification performance is calculated for the two classification algorithms, *i.e.* SVM and NBayes, on the aforementioned four datasets. All the four datasets are gene expression profiles produced by the microarray technology on cancer samples and their controls. The gene P53 is well known for its driving role in cancer development, verified by many experimental studies [25]. But its expression alone cannot classify whether an individual has a specific type of cancer. So it's intriguing to investigate whether P53 together with a few other chaperons may generate a good cancer classifier. The expression level of each probeset in a microarray platform is defined as a feature in the gene expression profile, and different microarray platform has different set of probesets. Table 1

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

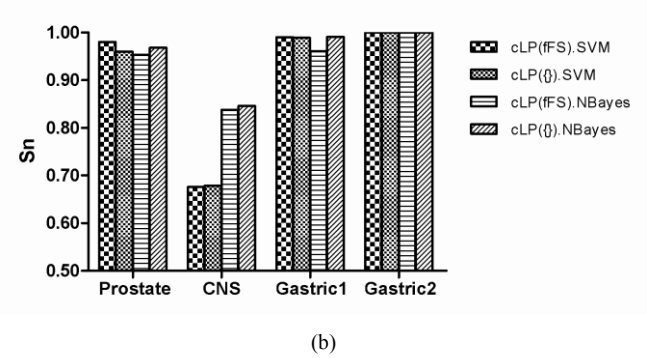
gives the probesets for *p*53 on each microarray platform of the dataset.

TABLE I. **SUMMARY OF FEATURES TO BE FIXED IN THE SELECTED FEATURE SUBSETS FOR THE FOUR DATASETS.** *COLUMN "NUMBER" IS THE NUMBER OF FEATURES FOR P53 IN EACH DATASE. COLUMN "FFS" GIVES THE IDS OF ALL SUCH FEATURES*

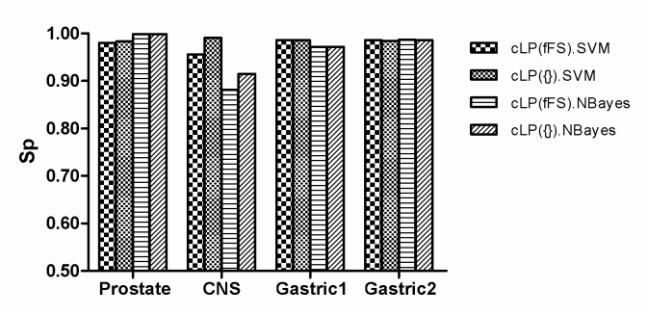
Dataset	Number	fFS
Prostate	3	1974_s_at, 31618_at, 1939_at
CNS	1	M22898_at
Gastric1	2	211300_s_at, 201746_at
Gastric2	2	211300_s_at, 201746_at



(a)

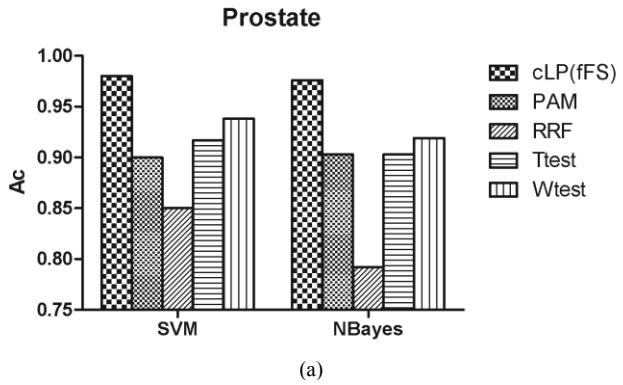


(b)

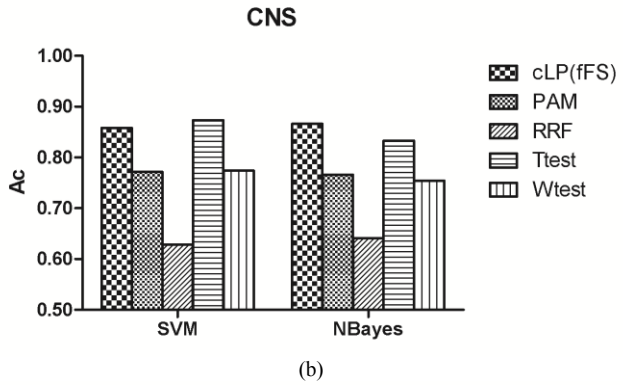


(c)

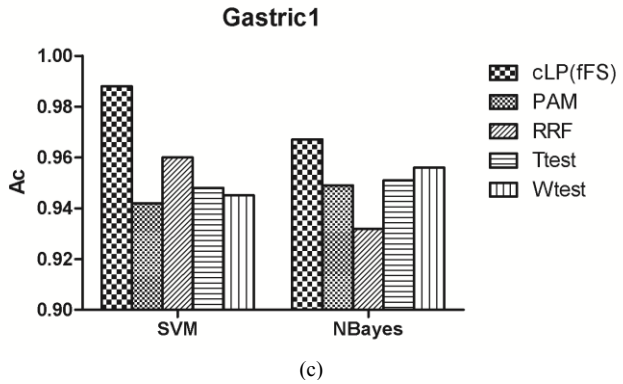
Fig. 1. ***Histogram plots of the binary classification performance of the selected features.*** *Each dataset is evaluated by the two feature selection algorithms cLP(fFS) and cLP({}), and the two classification algorithms SVM and NBayes. The classification performance is measured by (a) overall accuracy Ac,(b) sensitivity Sn and (c) specificity Sp, averaged over the 30 runs of the 5FCV strategy.*



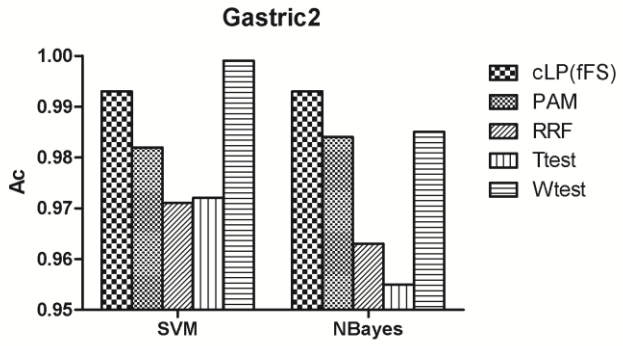
(a)



(b)



(c)



(d)

Fig. 2. ***Comparison of the binary classification accuracies Ac of the two classification algorithms SVM and Nbayes using the features selected by the five feature selection algorithms, i.e. cLP(fFS), PAM, RRF, Ttest and Wtest.*** *The averaged value Ac is calculated over 30 runs of the 5FCV strategy over the datasets (a) Prostate, (b) CNS, (c) Gastric1 and (d) Gastric2.*

The two versions of cLP algorithm, *i.e.* cLP(fFS) and cLP({}), perform similarly on all the four datasets, as in Fig. 1. The features fixed in our algorithm appear in the final set of selected features for the classification modeling, and the experimental participants/people will be predicted as having the disease or not. As expected, cLP({}) outperforms cLP(fFS) in most cases, with a minor increase in Ac. The maximal increase 2.3% in Ac is obtained when using NBayes on the dataset CNS, as in Fig 1 (a). The sensitivity is just increased by 0.8% in Fig. 1 (b), but the specificity has the largest increase of 3.4% compared with cLP(fFS) in Fig. 1 (c). The decrease of classification performance of cLP(fFS) never exceeds 3.0% for all the other datasets using both SVM and NBayes. Fig. 1 also shows that the fixation of two features for the dataset Gastric2 does not reduce and even slightly increases the classification performances.

B. cLP versus the other feature selection algorithms

The comparative study with the other feature selection algorithms is also conducted to investigate whether cLP(fFS) performs reasonably well. The overall accuracy averaged over 30 runs of the 5FCV strategy is plotted in Fig. 2. The pre-fixed feature subset fFS consists of the features of gene *p*53.

cLP(fFS) outperforms the other feature selection algorithms in most combinations of classification algorithms and datasets, as shown in Fig. 2. The features selected by cFS(fFS) achieve the top 2 accuracies among the 5 feature selection algorithms and 2 classification algorithms on the datasets Prostate and Gastric1. Ttest achieves 87.3% in Ac when using the classifier SVM, which is slightly better than the next two top-ranked classification models of cLP(fFS) combined with the classifiers SVM (85.8%) and NBayes (86.6%), respectively. Wtest works best with the classifer SVM with 99.9% in Ac, which is slightly better than the Ac 99.3% achieved by cLP(fFS) combined with both SVM and NBayes. Although there are two models (Ttest+SVM(CNS) and Wtest+SVM(Gastric2)) slightly better than the feature selection algorithm cLP(fFS), the experimentalists may be more confident in validating the cLP(fFS) models with the integration of their biological knowledge using the precious resources.

Our data shows that classification algorithms perform differently on individual datasets, *e.g.* RRF and the two ranking algorithms do not perform well on the dataset CNS. This may be due to that feature selection algorithms have different assumptions and optimization goals, and will work best on the datasets that fit these conditions. For example, Ttest assumes that the data follows a normal distribution, whereas Wtest has the assumption that the paired data are sampled randomly and independently. Fig. 2 suggests that the datasets Gastric1 and Gastric2 are easily separable by all the five feature selection algorithms, whereas the algorithms perform significantly different on the datasets Prostate and CNS. cLP(fFS) performs the best and the second on the datasets Prostate and CNS, respectively. And Ttest performs slightly better than cLP(fFS) on the dataset CNS. The data also suggests that it's important to evaluate how accurately the existing data mining algorithms perform on individual dataset.

C. cLP(fFS) selected features align with the existing cancer biomarkers

We investigate the biological functions of the 24 features selected by cLP(*p*53) on the dataset Gastric2. After excluding the two features/probesets of the gene *p*53, each of the other 22 features corresponds to a gene, and 10 of the 22 genes are known to be linked with the gastric cancer.

7 genes are known to be involved in the regulation of tumor development and metastasis. Cell proliferation is one of the major biological processes that are required by the tumor progression, and there are 5 such genes in the cLP(fFS) selected features, including the prostate stem cell antigen PSCA (205319_at) [26], the homeobox A10 (TF) HOXA10 (213150_at) [27], the sclerostin domain containing 1 SOSTDC1 (213456_at) [28], the S100 calcium binding protein A7 S100A7 (205916_at) [29], and the biglycan BGN (213905_x_at) [30, 31]. There are two tumor suppressor genes, *i.e.* the inhibin, beta A INHBA (210511_s_at) [32, 33] and the sulfatase 1 Sulf 1 (212344_at) [34].

The three other genes are also associated with the onset and development of gastric cancer. The interleukin 1 receptor antagonist IL1RN (216244_at) [35] and the procollagen-lysine - 2-oxoglutarate 5-dioxygenase 3 Plod3 (202185_at) are known to be gastric cancer biomarkers with their genomic alternations [35, 36]. The last gene, the period homolog 2 (Drosophila) PER2 (205251_at), encodes a Period family gene, shows up- and down- regulated expression levels in the gastric tumor tissues, when compared with the normal controls [37] and tumor-adjacent tissues [17], respectively.

Generally speaking, 10 out of the 22 genes are strongly associated with the tumor regulation and progression.

IV. CONCLUSIONS

This study explores the possibility of combining the merits of both biomedical knowledge and data mining algorithms, by adding constraints for the linear programming model. The proposed constrained linear programming algorithm, cLP(fFS), outperforms the other feature selection algorithms in almost all the cases. The embedded constraints will also facilitate the vast amount of biological knowledge from the literature and experienced biomedical researchers into the classification modeling. We believe that cLP(fFS) will greatly facilitate the training of biologically meaningful classification models and hypothesis-driven biomedical biomarker detections. A comprehensive comparison of how the parameters may be optimized and impact the classification performance will be investigated in the future work.

V. ACKNOWLEDGEMENTS

The constructive comments from the reviewers are appreciated.

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

REFERENCES

[1] R. Yan, A. Hallam, P. G. Stockley, and J. Boyes, "Oncogene dependency and the potential of targeted RNAi-based anti-cancer therapy," Biochem J, vol. 461, pp. 1-13, Jul 1 2014.

[2] W. Jiang, D. Bikard, D. Cox, F. Zhang, and L. A. Marraffini, "RNA-guided editing of bacterial genomes using CRISPR-Cas systems," Nat Biotechnol, vol. 31, pp. 233-9, Mar 2013.

[3] B. Wiedenheft, S. H. Sternberg, and J. A. Doudna, "RNA-guided genetic silencing systems in bacteria and archaea," Nature, vol. 482, pp. 331-8, Feb 16 2012.

[4] R. J. Klein, C. Zeiss, E. Y. Chew, J. Y. Tsai, R. S. Sackler, C. Haynes, et al., "Complement factor H polymorphism in age-related macular degeneration," Science, vol. 308, pp. 385-9, Apr 15 2005.

[5] J. E. Krebs, E. S. Goldstein, and S. T. Kilpatrick, Lewin's GENES XI: Jones & Bartlett Learning, 2012.

[6] Illumina. (2014, July 28, 2014). HiSeq X Ten System Performance Parameters. Available: http://www.illumina.com/systems/hiseq-x-sequencing-system/performance-specifications.ilmn

[7] Q. Cai, B. Zhang, H. Sung, S. K. Low, S. S. Kweon, W. Lu, et al., "Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1," Nat Genet, Jul 20 2014.

[8] M. Yang, B. Liu, M. Zhao, F. Li, G. Wang, and F. Zhou, "Normalizing electrocardiograms of both healthy persons and cardiovascular disease patients for biometric authentication," PLoS One, vol. 8, p. e71523, 2013.

[9] D. Singh, P. G. Febbo, K. Ross, D. G. Jackson, J. Manola, C. Ladd, et al., "Gene expression correlates of clinical prostate cancer behavior," Cancer Cell, vol. 1, pp. 203-9, Mar 2002.

[10] S. L. Pomeroy, P. Tamayo, M. Gaasenbeek, L. M. Sturla, M. Angelo, M. E. McLaughlin, et al., "Prediction of central nervous system embryonal tumour outcome based on gene expression," Nature, vol. 415, pp. 436-42, Jan 24 2002.

[11] G. Wang, N. Hu, H. H. Yang, L. Wang, H. Su, C. Wang, et al., "Comparison of global gene expression of gastric cardia and noncardia cancers from a high-risk population in china," PLoS One, vol. 8, p. e63826, 2013.

[12] N. Karmarkar, "A New Polynomial-Time Algorithm for Linear-Programming," Combinatorica, vol. 4, pp. 373-395, 1984.

[13] Y. Wang, Q. F. Wu, C. Chen, L. Y. Wu, X. Z. Yan, S. G. Yu, et al., "Revealing metabolite biomarkers for acupuncture treatment by linear programming based feature selection," BMC Syst Biol, vol. 6 Suppl 1, p. S15, Jul 16 2012.

[14] G. D. Guo and C. R. Dyer, "Simultaneous feature selection and classifier training via linear programming: A case study for face expression recognition," 2003 Ieee Computer Society Conference on Computer Vision and Pattern Recognition, Vol 1, Proceedings, pp. 346-352, 2003.

[15] R. Tibshirani, T. Hastie, B. Narasimhan, and G. Chu, "Diagnosis of multiple cancer types by shrunken centroids of gene expression," Proc Natl Acad Sci U S A, vol. 99, pp. 6567-72, May 14 2002.

[16] H. T. Deng and G. Runger, "Feature Selection via Regularized Trees," 2012 International Joint Conference on Neural Networks (Ijcnn), 2012.

[17] P. Baldi and A. D. Long, "A Bayesian framework for the analysis of microarray expression data: regularized t -test and statistical inferences of gene changes," Bioinformatics, vol. 17, pp. 509-19, Jun 2001.

[18] W. M. Liu, R. Mei, X. Di, T. B. Ryder, E. Hubbell, S. Dee, et al., "Analysis of high density expression microarrays with signed-rank call algorithms," Bioinformatics, vol. 18, pp. 1593-9, Dec 2002.

[19] L. Breiman, "Random Forests," Machine Learning, vol. 45, pp. 5-32, 2001/10/01 2001.

[20] V. N. Vapnik, The Nature of Statistical Learning Theory, 2 ed. New York: Springer-Verlag, 1999.

[21] H. Zhang, "Exploring conditions for the optimality of Naive bayes," International Journal of Pattern Recognition and Artificial Intelligence, vol. 19, pp. 183-198, Mar 2005.

[22] F. Zhou and Y. Xu, "cBar: a computer program to distinguish plasmid-derived from chromosome-derived sequence fragments in metagenomics data," Bioinformatics, vol. 26, pp. 2051-2, Aug 15 2010.

[23] K. Li, M. Yang, G. Sablok, J. Fan, and F. Zhou, "Screening features to improve the class prediction of acute myeloid leukemia and myelodysplastic syndrome," Gene, vol. 512, pp. 348-54, Jan 10 2013.

[24] G. Wang, F. Zhou, V. Olman, F. Li, and Y. Xu, "Prediction of pathogenicity islands in enterohemorrhagic Escherichia coli O157:H7 using genomic barcodes," FEBS Lett, vol. 584, pp. 194-8, Jan 4 2010.

[25] K. H. Khoo, C. S. Verma, and D. P. Lane, "Drugging the p53 pathway: understanding the route to clinical efficacy," Nat Rev Drug Discov, vol. 13, pp. 217-36, Mar 2014.

[26] Z. Zeng, X. Wu, F. Chen, J. Yu, L. Xue, Y. Hao, et al., "Polymorphisms in prostate stem cell antigen gene rs2294008 increase gastric cancer risk in Chinese," Mol Carcinog, vol. 50, pp. 353-8, May 2011.

[27] Y. Han, W. W. Tu, Y. G. Wen, D. P. Li, G. Q. Qiu, H. M. Tang, et al., "Identification and validation that up-expression of HOXA13 is a novel independent prognostic marker of a worse outcome in gastric cancer based on immunohistochemistry," Med Oncol, vol. 30, p. 564, 2013.

[28] G. Gopal, U. M. Raja, S. Shirley, K. R. Rajalekshmi, and T. Rajkumar, "SOSTDC1 down-regulation of expression involves CpG methylation and is a potential prognostic marker in gastric cancer," Cancer Genet, vol. 206, pp. 174-82, May 2013.

[29] J. Liu, X. Li, G. L. Dong, H. W. Zhang, D. L. Chen, J. J. Du, et al., "In silico analysis and verification of S100 gene expression in gastric cancer," BMC Cancer, vol. 8, p. 261, 2008.

[30] L. Hu, Y. T. Duan, J. F. Li, L. P. Su, M. Yan, Z. G. Zhu, et al., "Biglycan enhances gastric cancer invasion by activating FAK signaling pathway," Oncotarget, vol. 5, pp. 1885-96, Apr 15 2014.

[31] B. Wang, G. X. Li, S. G. Zhang, Q. Wang, Y. G. Wen, H. M. Tang, et al., "Biglycan expression correlates with aggressiveness and poor prognosis of gastric cancer," Exp Biol Med (Maywood), vol. 236, pp. 1247-53, Nov 2011.

[32] T. Oshima, K. Yoshihara, T. Aoyama, S. Hasegawa, T. Sato, N. Yamamoto, et al., "Relation of INHBA gene expression to outcomes in gastric cancer after curative surgery," Anticancer Res, vol. 34, pp. 2303-9, May 2014.

[33] Q. Wang, Y. G. Wen, D. P. Li, J. Xia, C. Z. Zhou, D. W. Yan, et al., "Upregulated INHBA expression is associated with poor survival in gastric cancer," Med Oncol, vol. 29, pp. 77-83, Mar 2012.

[34] H. Y. Ma, F. Zhang, J. Li, M. L. Mo, Z. Chen, L. Liu, et al., "HSulf-1 suppresses cell growth and down-regulates Hedgehog signaling in human gastric cancer cells," Oncol Lett, vol. 2, pp. 1291-1295, Nov 2011.

[35] B. S. He, Y. Q. Pan, Y. F. Xu, C. Zhu, L. L. Qu, and S. K. Wang, "Polymorphisms in interleukin-1B (IL-1B) and interleukin 1 receptor antagonist (IL-1RN) genes associate with gastric cancer risk in the Chinese population," Dig Dis Sci, vol. 56, pp. 2017-23, Jul 2011.

[36] L. Cheng, P. Wang, S. Yang, Y. Yang, Q. Zhang, W. Zhang, et al., "Identification of genes with a correlation between copy number and expression in gastric cancer," BMC Med Genomics, vol. 5, p. 14, 2012.

[37] M. L. Hu, K. T. Yeh, P. M. Lin, C. M. Hsu, H. H. Hsiao, Y. C. Liu, et al., "Deregulated expression of circadian clock genes in gastric cancer," BMC Gastroenterol, vol. 14, p. 67, 2014.

978-1-4799-7294-4/14/$31.00 ©2014 IEEE