

Combining gene expression and interaction network data to improve kidney lesion score prediction

Davoud Moulavi¹, Mohsen Hajiloo¹, Jörg Sander¹, Philip Halloran², Russell Greiner¹

¹Department of Computing Science

²Department of Medicine and Medical Microbiology and Immunology

University of Alberta

Edmonton, Canada

Email: {moulavi, hajiloo, jsander, phil.halloran, rgreiner}@ualberta.ca}

Abstract— Doctors often use histopathology of needle biopsies (in the form of lesion score) to help diagnose kidney rejection, which is useful for identifying the appropriate treatment. As these lesion score are subjective and error prone, some researchers have tried to train a classifier to predict the rejection or non-rejection of a renal transplant, based on the gene expression microarrays of the patient’s renal biopsies. However the high dimensionality and intrinsic noisy nature of this data makes this task very challenging. The most common techniques for predicting lesion scores from microarrays just use a single regressor on a subset of genes selected by statistical feature selection methods. Due to the high dimensionality of microarray data, these models usually overfit. This paper presents a novel method for predicting lesion scores based on the majority vote of bagging regressors built on feature subsets selected by either statistical or biological feature selection approaches, including a model that uses interaction networks to select genes. Our experimental results show that focusing on genes that interact with many other genes (“Hub genes”) and also interact with statistically selected genes in interaction networks, provide significantly better results than other biological feature selection methods. These experimental results show that none of the statistical feature selection methods are significantly better than our Hub genes approach and that a simple fusion of Hub genes and the best statistical feature selected method can further increase the generalization power of the prediction model.

Index Terms— Kidney transplant, Lesion score, Microarray, Biological feature selection, Hub genes.

I. INTRODUCTION

Histopathology standards are used to assess needle biopsies, towards determining kidney allograft rejection. However this kind of diagnosis is very subjective, in that different physicians (or even the same physician on different days) may give different assessments of the same biopsy. This motivated a team of renal pathologists and transplant surgeons to develop the current standard in the Banff consensus system [1] for scoring the lesions of a biopsy. This system allows pathologist to assign each lesion a score in $\{0,1,2,3\}$, where 0 usually means that this aspect of the transplanted kidney is healthy and other scores suggest some problem in the kidney, where higher lesion scores means more severe lesion damage. Pathologists then use these lesion scores to predict rejection or

no rejection of a transplanted kidney.

Physicians can use these diagnoses to help treat patients – in particular, they can sometimes prevent the predicted rejection by giving anti-rejection therapies. Patients who reject their transplanted kidney may need to start the painful process of dialysis or to have another kidney transplant.

However, as this histopathology diagnosis is subjective, it is not a perfect gold standard. This suggests using some other approach to diagnose rejection, such as the molecular approach that is based on gene expression microarray data [2, 3]. The transcription of the genetic information contained within the DNA into messenger RNA (mRNA) molecules is called “gene expression”. These mRNAs are later translated into the proteins that perform most of the critical functions of cells. Gene expression is a complex process that allows a cell to respond to both environmental needs and to its own changing needs [4]. Microarrays, like the Affymetrix GeneChip probe arrays, enable scientists to study changes in expression levels of a large number of genes simultaneously. We can view a microarray study over a set of patients as a matrix whose rows each represent a specific patient and whose columns each represent a probe related to a specific gene; hence each cell of this matrix corresponds to the expression level of a specific probe for a specific patient. The result of a study might be a classifier (trained on matrix described above) that uses a patient’s microarray to predict some property of that patient. Previous studies suggest that we can use microarrays to predict lesion scores or rejection [2]. A molecular approach based on microarrays seems to promise fruitful insights into the kidney transplant process, but some challenges arise due to the nature of microarrays and the problem domain, including the large dimensionality (number of probes) that can cause overfitting, intrinsic noise of microarrays that can decrease reliability, and subjectivity in assigning lesion scores by pathologists. In this paper we address the challenge of the predicting the lesion scores by first reducing the dimensionality of the feature space by applying statistical feature selection methods and prior biological knowledge and by applying ensemble regression methods to learn the lesion scores. Section II gives a detailed

explanation of the proposed methods in this study. In particular, we consider seven feature selection methods (some based on prior biological knowledge and others on statistical approaches) to select the most important features, then use bagging models and data fusion to improve the final prediction of lesion scores. Section III includes the results of experiments and discussions, which show (1) that using the topological features of an interaction network are helpful for finding new genes relevant to our kidney problem and (2) fusing the results of the biologically inspired feature selection and statistical feature selection methods further improves the predictive accuracy. Section IV presents the conclusion. We close this section with a quick survey of related literature.

This work describes a system that learns an *ensemble classifier* that uses *microarray data* to better understand *kidneys*. As such, it relates to the many others studies that have used on microarray data to tackle kidney related problems.

Several previous works on biopsies have shown that the gene expression data has considerable potential for improving biopsy diagnoses [2, 3, 5]. Reeve et al. [2] conducted experiments to compare a histopathology-based approach and molecular-based approach for predicting kidney transplant rejection. They applied the predictive analysis of microarrays (PAM) method [6] to predict whether a kidney will be rejected. Their study showed that this molecular-based approach provided more insight for rejection prediction than histopathology based approach and was more accurate at the borderlines rejections. Sarwal et al. [5] used microarrays in a systematic study of gene expression patterns in biopsy samples from normal and dysfunctional renal allografts. They found consistent differences among the gene-expression patterns associated with acute rejection, nephrotoxic effects of drugs, chronic allograft nephropathy, and normal kidneys. Using microarrays, they identified molecular variations suggesting the existence of distinct molecular and prognostic variants of acute rejection, which could not previously be clearly defined on the basis of clinical or pathological criteria. Mueller et al. [3] performed a microarray study to assess donor kidney quality and the risk of delayed graft function (DGF) and found that the gene expression data reflects kidney quality and susceptibility to DGF better than available clinical and histopathological scoring systems. In addition to predicting kidney rejection [2], rejection subtype [5] and delayed graft function [3], microarrays can also be used to predict lesion scores, which is the goal of the current study.

Since microarrays typically have a large set of features but most studies involve only a small set of samples, feature selection techniques are usually applied to eliminate noisy and irrelevant features. Furthermore, feature selection increases the chance of producing more understandable results. Two broad categories of feature selection methods are: statistical feature selection methods [7] and biologically inspired feature selection methods [8, 9]. Statistical feature selection methods range from univariate filter approaches like shrunken centroid [6], t-test [18], ANOVA [18] and SNR [4] to multivariate filter

methods like mRMR [19] and wrapper approaches like genetic algorithms [20] and SVM-RFE [7]. Statistical feature selection methods select a group of statistically informative genes, but they do not necessarily provide any biological insight about the problem under study. On the other hand, many researchers have used biological selected genes for the prediction task. Using biological information for gene selection can also help in finding new pathway genes and complexes. Kennedy, Simoff, Skillicorn, and Catchpoole [9] applied an integrative method that used statistically selected genes and re-clustered these genes into groups of genes with similar biological functionality. These clusters allowed a biological interpretation and helped biologists to form new hypotheses. Muller et al. [10] used pathogenesis-based transcript sets (PBTs) that reflect major biologic events in allograft rejection. PBTs were correlated with histopathological lesion scores and were the highest in biopsies with apparent rejection that represent a measure of inflammatory disturbance in organ transplants.

There are some studies that use the idea of ensemble learning for combining their predictions. According to Polikar [13], ensemble learning is “the process by which multiple models are strategically generated and combined to solve a particular computational intelligence problem.” Ensemble learning is primarily used to improve the (classification, prediction, function approximation, etc.) performance of a system (involving a set of models), often by reducing the likelihood of an unfortunate selection of a poor model. The intuitive justification for ensemble learning is that typically no single approach or system will be uniformly superior to any other, and that the integration of several single approaches will enhance the performance of the final classifier (accuracy, reliability, comprehensibility) [14]. An ensemble learner will typically have overall better performance than the individual base learners when the base learners are accurate and diverse [23].

Ensemble models have been successfully applied in tasks involving microarrays, starting with Tan and Gilbert [14], who investigated the applicability of ensemble methods like bagging and boosting to some cancer diagnosis problems. Various ensemble learning methods have been applied to microarrays since then. A comparison of bagging and boosting and single C4.5 classifiers showed that ensemble methods (bagging and boosting) often perform better than a single decision tree in this classification task [14]. Furthermore, they showed that bagging classifiers often outperform boosting classifiers when dealing with microarray data [14]. Ensemble learners usually beat single learners since every learning algorithm employs a different search strategy to identify the true concept. If the number of the training samples is too small (which is often the case in microarray data), the individual learner can induce different classifiers that all produce sub-optimal performance. Thus, by averaging the different hypotheses, the combined classifier may produce a good approximation to the true concept. Moreover, to avoid local optima of the individual search strategies, an ensemble

classifier may provide a better approximation to the true concept by performing different initial searches and combining the outputs. Lastly, due to the limited amount of training data, an individual classifier may not represent the true hypothesis. Thus, through considering diverse base classifiers, it may be possible for the final classifier to approximately represent the true hypotheses. Various ensemble models have been applied to microarrays: Peng's ensemble SVM (enSVM) [15] builds SVM classifiers on selected subsets of genes and at performance time, combines these classifiers by a simple majority voting scheme. Each iteration of Dettling's method [16] trains a set of classifiers by applying a bagging algorithm, rather than train a single classifier on the current bootstrap sample. A bias-variance tradeoff study shows that bagged classifiers have considerably lower variance but the same expected bias as the single base classifier [16]. Studies that train individual decision trees on microarrays show that they usually overfit and so do not perform well on the test set [14]. These earlier results show that ensemble methods can be useful, especially for analyzing microarray data.

The remainder of this paper gives a detailed explanation of applied and proposed methods of this study, discusses the experiments and results and presents the conclusion.

II. PROPOSED METHOD

We divide the task of accurately predicting the lesion score into two main stages. In the first stage, we will try to reduce the dimensionality of the feature space. This is critical for learning a lesion score classifier that uses microarray data, due to the high dimensionality of microarray data. Here we consider seven different feature selection methods; three based on statistical approaches and the other four are based on prior biological knowledge (including on that finds and uses Hub genes).

In the second stage, we train bagging SMO [Sequential Minimal Optimization] regression models with the RBF [Radial Basis Function] kernel on only the selected subsets of genes. Then, we will fuse the predictions obtained on the Hub genes with the best prediction model of statistical feature selection methods by using an averaging method.

The following subsections first describe the statistical feature selection methods used in this paper and then present the biological feature selection methods.

A. Feature Selection Methods

- Statistical Feature selection

We used the following three statistical measures for selecting features:

A gene's variance is a simple statistical measure that simply measures how much the expression of that single gene varies over the patients. (Note it does not use the labels of the patients.) Here, we simply select the 500 genes with the highest variance.

The second statistical approach uses the squared t-test (SST) measure to select the input features of the model.

Intuitively, this method selects the genes that have better total discrimination between the four classes of lesion scores. We define the ability of a gene to discriminate between class 1 and 2 as a SST measure using the following equation:

$$STT(g;1,2) = \frac{(\mu_1(g) - \mu_2(g))^2}{\sigma_1(g)^2 + \sigma_2(g)^2} \quad (1)$$

where $\mu_i(g)$ is the average and $\sigma_i^2(g)$ is the variance of gene g over each class $i \in \{1,2\}$. This heuristic prefers genes that have better discrimination between class 1 and 2 i.e., if the difference of its mean between two classes is high in units of the variance of the classes. Here, we compute the sum $STT(g) = \sum_{i < j} STT(g; i, j)$ over the $C(4,2)$ pairs of classes as a measure for class discrimination of each gene. Then we choose the 500 genes that have the highest $STT(g)$ values.

The third statistical feature selection method just uses the 30 genes that Reeve et al. [2] found using the predictive analysis of microarray (PAM) method. Their research showed that these 30 genes were effective for predicting kidney rejection/no-rejection.

- Biologically inspired Feature selection

We also considered four different biologically inspired feature selection methods, which each try to select the genes that are biologically meaningful for predicting lesion scores.

The first biological feature subset involves the PBT (pathogenesis-based transcript) sets introduced by Muller et al. [10], which reflect the major biological events in allograft rejection. Each PBT is a set of genes. Here we took union over all 36 PBTs, corresponding to a total of 7280 probesets.

The second biological feature subset is the 169 probesets that appear in the "renal cell carcinoma pathway" specified in a KEGG database [24].

The third biological feature subset is the 57 probesets that are related to the "allograft rejection pathway" specified in a KEGG database.

The fourth biological feature selection method uses the interaction network of genes. Highly connected genes (called "Hub genes") in the gene networks are thought to play an important role in cells and in organizing the behavior of biological modules [11, 12]. Note that the expression values of these Hub genes might not be significantly different across the classes.

Here we use the cytoscape software [21] to build the interaction network from a set of genes. Cytoscape is free software that helps to obtain the interaction network for genes of interest.

These days many sources of interaction such as protein-protein interactions and protein-DNA interactions are highly available in the literature. Cytoscape uses text-mining techniques that can extract functional relationship between genes of interest. In this context two genes are linked if they are frequently mentioned in the same sentence. This connection may indicate a biochemical association, colocalization or coexpression

relationship [21]. However, these relationships are not always true but are useful in aggregate.

Here we started from the genes that Reeve et al. [2] found by using PAM method. Cytoscape used these genes to produce an interaction network that includes 8143 nodes and 37783 interactions between them. A “Hub gene” is defined as a gene that is connected to at least one of the original PAM genes and has more than 20 interactions. This produced a set of 90 Hub genes. We also include each of the original PAM genes that were not connected to any Hub gene in the output feature subset of Hub genes. This added 10 more genes, bringing the total number of genes in this final feature subset to 100 genes.

B. Learners

- *Support vector regression*

In this section we briefly describe the support vector regression (SVR) [17]. Let the set

$$D = \{ (x_1, y_1), \dots, (x_n, y_n) \}$$

define the whole training set, where each $x_i \in R^n$ is the input instance (a single patient) and $y_i \in R$ is the target value for that patient. In the ordinary linear regression model, we seek $w \in R^n$ and $b \in R$ such that

$$f(x_i) = w \cdot \phi(x_i) + b \quad (2)$$

minimizes the squared residual error over the data D, where $\phi(x)$ is the high dimensional feature space that is nonlinearly mapped from the input space x.

The SVR extends this linear regression models to obtain sparse solutions [17] by replacing the standard quadratic loss function by linear ϵ -insensitive loss function defined as following:

$$L(f(x) - y) = \begin{cases} 0 & |f(x) - y| < \epsilon \\ |f(x) - y| - \epsilon & \text{otherwise} \end{cases} \quad (3)$$

where ϵ is a precision parameter representing the radius of the tube located around the regression function.

The SVR method utilized the concept of kernel functions and kernel tricks to use effective dot product computations in the feature space. In the SVR formulation any function satisfying Mercer’s condition can be used as the kernel function [17]. The following are three commonly used kernel functions:

$$\text{Linear: } K(x_i, x_j) = x_i x_j \quad (4)$$

$$\text{Polynomial: } K(x_i, x_j) = (1 + x_i x_j)^p \quad p > 0 \quad (5)$$

Radial Basis Function (RBF):

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right) \quad (6)$$

In which, p and σ are adjustable kernel parameters.

- *Bagging*

Since microarray data are very noisy, we have used bagging ensemble method for learning predictors on each training data set obtained using different feature selection method. Bagging, a name derived from bootstrap aggregation, is one of the simplest and also most effective ensemble methods [22]. It was originally designed for classification and is often applied to decision tree models, but it can be used with any type of model for classification or regression.

The method uses multiple versions of a training set by using the bootstrap, i.e. producing a size-m training sample by sampling the original size-m sample with replacement. Each of these data sets is used to train a different model. The outputs of the models are combined by averaging (in the case of regression) or majority voting (in the case of classification) to create a single output.

- *Fusion Method*

For combining decision of different predictors trained on distinct training sets we can use different fusion methods. Here, we use averaging to combine the final decision of the different predictors trained on different datasets. The averaging method is one of simplest but most effective fusion methods, which is usually applied for the regression problems. In this method the output of the different predictors will be averaged as the final output of the learning task. It has been proven that using this simple method can produce a classifier whose variance dramatically less than any of the classifiers produced by a base predictor [13].

III. EXPERIMENTAL RESULTS

We analyzed microarray gene expression data (Affymetrix HG-U133) with 54675 probesets of renal transplant biopsies of 234 samples taken from 173 unique patients. The patients were biopsied between one week and thirty two years post-transplant at the University of Alberta hospital.

Due to the high dimensionality of the microarray expression data we considered seven different feature selection methods (described in Section II) to reduce the dimensionality. As our goal is to compare the various feature selection methods, we fixed a single regressor for predicting the lesion scores. We used bagging SMO regression classifiers with RBF kernel on each feature subset. Intuitively, since microarray data are noisy we have to use predictive models that are robust to the noise. SMO regression model is one of the robust models due to the use of large margin concept. Also we used bagging ensemble of SMO regression because the ensemble methods such as bagging have the ability of reducing the variance error of our predictor. Even though each lesion has a discrete value $\{0, 1, 2, 3\}$, we view their values as real, and so treat this task as a regression task. We therefore evaluated each learner based on 10-fold cross validation RMSE (root mean squared error) of the regressor it produced. Note that our regressor produced real values, not necessarily one of the discrete values $\{0, 1, 2, 3\}$.

The first experiment compared the Hub Gene feature selection method with renowned statistical feature selection methods like PAM, High Variance, and STT. Here we trained bagging SMO regression classifier on the selected genes and used 10-fold cross validation as the evaluation method. Table I shows the average RMSE results, when repeating this experiment for 100 runs. The statistical methods that select a set of significantly expressed genes were comparable to Hub genes method (i.e. this difference is not significant, based on a paired t-test with $p < 0.05$.)

TABLE I

Comparison of RMSE for Bagging SMO regression classifier trained on features selected by Hub genes method and statistical feature selection methods

Feature Selection	RMSE
Hub genes	0.6706 ± 0.0481
PAM	0.6746 ± 0.0516
High Variance	0.6827 ± 0.0558
Squared T-Test (STT)	0.6616 ± 0.0483

The second experiment compares the prediction power of Hub genes versus other biologically selected genes like PBT genes, renal cell carcinoma pathway genes, allograft rejection pathway genes. Here we trained the bagging SMO regression classifier on each gene set and evaluated using 10-fold cross validation RMSE. Table II gives the average RMSE result over 100 runs. Hub genes are significantly better than any of the other biological feature selection methods (paired t-test, at $p < 0.05$).

TABLE II

Comparison of RMSE for Bagging SMO regression classifier trained on features selected by Hub genes method and other biological feature selection methods

Feature Selection	RMSE
Hub genes	0.6706 ± 0.0481
PBT	0.7612 ± 0.0534
Renal cell carcinoma	0.7340 ± 0.0395
Allograft rejection	0.6907 ± 0.0525

The third experiment asks if fusing a model trained on genes selected by a statistical method like STT and a model trained on genes selected by Hub genes method will produce better results. We have trained one bagging SMO regression on gene subsets selected by STT and another selected by Hub genes. We have also used fusion by averaging to combine the output of these two models. Table III shows the results of repeating these 10-fold cross validation experiments for 100 runs. Here combining the output of bagging regressor on Hub genes and the output of bagging regressor on STT genes by fusion, produces a model with statistically better performance (paired t-test, $p < 0.05$).

TABLE III

RMSE comparison of fusion and Bagging SMO regression classifier trained on genes selected by Hub genes method and Squared T-Test (STT) method.

Feature Selection	RMSE
Fusion by Averaging	0.6431 ± 0.0403
Hub genes	0.6706 ± 0.0481
Squared T-Test (STT)	0.6616 ± 0.0483

Figure I presents the performance of all of the models discussed here: bagging models built using biologically selected features like Hub genes, PBT genes, renal cell carcinoma pathway genes, allograft rejection pathway genes and features selected by statistical methods like PAM, high variance, and Squared T-Test (STT) as well as the performance of models formed by fusing output of bagged STT and bagged Hub genes. The results show that using Hub genes as an input feature selection method is more promising than other biological feature selection results and is comparable to statistical methods. Also it shows that the best performance obtained using the fusion model of the best statistical based and prior biological knowledge based feature selection methods.

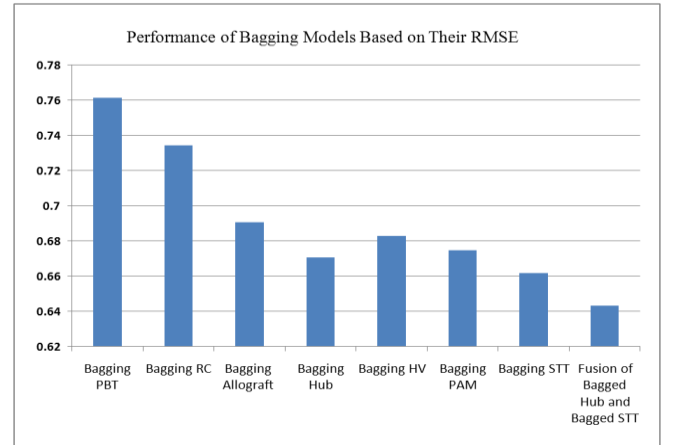


Figure I. Performance of Bagging Models Built over Statistically Selected Genes and Biologically Selected Genes and Fusion of Bagged Hub genes and Bagged STT Models.

IV. CONCLUSION

This paper has examined several different types of feature selection methods to use when predicting the kidney lesion scores from gene expression data. We showed that a system that uses the topological features of an interaction network is helpful for finding new genes related to the kidney lesion prediction problem. Using Hub genes produces a system that is more accurate than ones that use other biological feature selection methods like PBT genes, allograft rejection pathway genes, and renal cell carcinoma pathway genes; used by itself, it is also competitive with statistical feature selection methods like applying squared t-test (STT), high variance, and the PAM method. Furthermore, fusing the models trained on Hub genes with another trained using the statistical feature selection method STT further improved the predictive accuracy. These

results show that prior biological knowledge extracted from the literature can help to produce better models for predicting kidney lesion scores from microarray data.

In the future, we plan to apply the idea of selecting Hub Genes to other problems. We will also explore other ways to use a gene's connectivity in an interaction network to select a set of genes. Finally, since the overall final goal of the kidney transplant rejection problem is to predict reject/no-reject, we will extend this model to this project prediction task.

ACKNOWLEDGMENT

This work was supported by funding and/or resources from NSERC, the Alberta Ingenuity Centre for Machine Learning, Genome Canada, Genome Alberta, the University of Alberta, the University of Alberta Hospital Foundation, Roche Molecular Systems, Hoffmann-La Roche Canada Ltd., Alberta Ministry of Advanced Education and Technology, the Roche Organ Transplantation Research Foundation, the Kidney Foundation of Canada, Stromedix, and Astells. Dr. Halloran also holds the Muttart Chair in clinical Immunology.

The authors gratefully acknowledge the other members of the ATAGC (Alberta Transplant Applied Genomics Centre) team for preparing the data used in this study and for providing important information and advice.

REFERENCES

- [1] K. Solez, R. B. Colvin, L. C. Racusen, B. Sis, P. F. Halloran, P. E. Birk, P. M. Campbell, M. Cascalho, A. B. Collins, A. J. Demetris, C. B. Drachenberg, I. W. Gibson, P. C. Grimm, M. Haas, E. Lerut, H. Liapis, R. B. Mannon, P. B. Marcus, M. Mengel, M. J. Mihatsch, B. J. Nankivell, V. Nickleit, J. C. Papadimitriou, J. L. Platt, P. Randhawa, I. Roberts, L. Salinas-Madruga, D. R. Salomon, D. Seron, M. Sheaff and J. J. Weening "Banff 05 meeting report: differential diagnosis of chronic allograft injury and elimination of chronic allograft nephropathy," *American Journal of Transplantation*, vol. 7, no. 3, pp. 518–526, Feb. 2007.
- [2] J. Reeve, G. Einecke, M. Mengel, B. Sis, N. Kayser, B. Kaplan and P. F. Halloran, "Diagnosing Rejection in Renal Transplants: A Comparison of Molecular- and Histopathology-Based Approaches," *American Journal of Transplantation*, vol. 9, no. 8, pp. 1802–1810, Jun. 2009.
- [3] T. F. Muellera, J. Reeve, G. S. Jhangri, M. Mengel, Z. Jacaj, L. Cairo, M. Obeidat, G. Todd, R. Moore, K. S. Famulski, J. Cruz, D. Wishart, C. Meng, B. Sis, K. Solez, B. Kaplan and P. F. Halloran, "The Transcriptome of the Implant Biopsy Identifies Donor Kidneys at Increased Risk of Delayed Graft Function," *American Journal of Transplantation*, vol. 8, no. 1, pp. 78–85, Dec. 2007.
- [4] T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M.L. Loh, J.R. Downing, M.A. Caligiuri, C.D. Bloomfield, and E.S. Lander, "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring," *Science Magazine*, vol. 286, pp. 531-537, 1999.
- [5] M. Sarwal, Mei-Sze Chua, Neeraja Kambham, Szu-Chuan Hsieh, T. Satterwhite, M. Masek, and O. Salvatierra Jr., "Molecular heterogeneity in acute renal allograft rejection identified by DNA microarray profiling," *The New England Journal of Medicine*, vol. 349: 125–138, 2003.
- [6] R. Tibshirani, T. Hastie, B. Narasimhan, and G. Chu, "Diagnosis of multiple cancer types by shrunken centroids of gene expression " *PNAS*, vol. 99, no. 10, 6567-6572, May. 2002.
- [7] Y. Saeys, I. Inza, and P. Larranaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, 2007.
- [8] F. Rapaport, A. Zinovyev, M. Dutreix, E. Barillot, and J. P. Vert. "Classification of microarray data using gene networks," *BMC Bioinformatics*, 8(1):35+, Feb. 2007.
- [9] P. J. Kennedy, S. J. Simoff, D. B. Skillicorn, and D. R. Catchpoole, "Extracting and explaining biological knowledge in microarray data," In *PAKDD*, volume 3056 of *Lecture Notes in Computer Science*, pp. 699–703. Springer, 2004.
- [10] T. F. Mueller, G. Einecke, J. Reeve, B. Sis, M. Mengel, G. S. Jhangri, S. Bunnag, J. Cruz, D. Wishart, C. Meng, G. Broderick, B. Kaplan and P. F. Halloran, "Microarray analysis of rejection in human kidney transplants using pathogenesis-based transcript sets," *American Journal of Transplantation*, vol. 7: 2712–2722, Oct. 2009.
- [11] J.D.J. Han, N. Bertin, Tong Hao, D.S. Goldberg, G.F. Berriz, L.V. Zhang, D. Dupuy, A.J.M. Walhout, M.E. Cusick, F.P. Roth, M. Vidal, "Evidence for dynamically organized modularity in the yeast protein–protein interaction network," *Nature* 430, 88-93, 2004.
- [12] M.R. Carlson, B. Zhang, Z Fang, P.S Mischel, S. Horvath, S.F. Nelson, "Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks," *BMC Genomic*, vol. 7, Mar. 2006.
- [13] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits and Systems Magazine*, vol. 6, no.3, pp. 21-45, 2006
- [14] A.C. Tan and D. Gilbert, "Ensemble machine learning on gene expression data for cancer classification," *Applied Bioinformatics*, vol. 2, pp. 75-83, 2003.
- [15] Y. Peng, "A novel ensemble machine learning for robust microarray data classification," *Computers in Biology and Medicine*, vol. 36, pp. 553-573, 2006.
- [16] M. Dettling, "Overfitting for tumor classification with gene expression data," *Bioinformatics*, vol. 20, no. 18, pp. 3583-3593, 2004.
- [17] C. M. Bishop, "Pattern Recognition and Machine learning,"
- [18] P. Jafari, F. Azuaje, "An assessment of recently published gene expression data analyses: reporting experimental design and statistical factors," *BMC Med. Inform. Decis. Mak.*, 6:27, Mar 2006.
- [19] C. Ding, H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *Journal of Bioinformatics and Computational Biology*, vol. 3, no. 2, pp. 185-205, 2005.
- [20] T. Jirapech-Umpai, and S. Aitken, "Feature selection and classification for microarray data analysis: evolutionary methods for identifying predictive genes. *BMC Bioinformatics*, 6, 148.
- [21] M.S. Cline, M. Smoot, E. Cerami, A. Kuchinsky, N. Landys, C. Workman, R. Christmas, I. Avila-Campilo, M. Creech, B. Gross, K. Hanspers, R. Isserlin, R. Kelley, S. Killcoyne, S. Lotia, S. Maere, J. Morris, K. Ono, V. Pavlovic, A.R. Pico, A. Vailaya, P.L. Wang, A. Adler, B.R. Conklin, L. Hood, M. Kuiper, C. Sander, I. Schmulevich, B. Schwikowski, G.J. Warner, T. Ideker, and G.D. Bader. "Integration of biological networks and gene expression data using Cytoscape," *Nature Protocols*, vol. 2, pp.2366-2382, 2007.
- [22] T.G. Dietterich, "Ensemble methods in machine learning," In *Proceedings of the First International Workshop on Multiple Classifier Systems, MCS. Cagliari, Italy. LNCS,1857:1–15, Jun. 2000.*
- [23] L. I. Kuncheval and C. J. Whitaker, "Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy", *Journal of Machine Learning*, Springer, vol. 51, no. 2, May 2003.
- [24] Kyoto Encyclopedia of Genes and Genomes <http://www.genome.jp/kegg>