*Proceedings of the 2012 IEEE 12th International Conference on Bioinformatics & Bioengineering (BIBE), Larnaca, Cyprus, 11-13 November 2012*

# Influence of Algorithmic Parameters on Marker Selection in Genomic Datasets

T. Vigdideli, E. S. Bei, and M. Zervakis
Department of Electronic and Computer Engineering,
Technical University of Crete,
Chania Crete, Greece
tvigdideli@isc.tuc.gr, abei@isc.tuc.gr,
michalis@display.tuc.gr

D. Kafetzopoulos
Institute of Molecular Biology and Biotechnology
Foundation of Research and Technology Hellas (FORTH)
Heraklion Crete, Greece
kafetzo@imbb.forth.gr

*Abstract*— **The biological processes are widely studied by genome analysis leading to a large number of genes, thus making necessary the use of automated evaluation methods. In this study, we examine the influence of algorithmic parameters in the prediction power of a gene signature and in the selection process of the signature itself. We focus on one gene selection approach applied on a dataset of the budding yeast *Saccharomyces cerevisiae,* using quite different parameters and evaluate the influence on the selected signature. In particular, we adopt a recursive feature elimination process where at each step the prognostic power of the set of remaining genes is evaluated by five different classifiers, as well as by four classifier-fusions schemes. More specifically, we consider the logistic-sigmoid, kernel nearest centroid, kernel minimum squared error, kernel subspace, and support vector machines as classifiers with different parameters and/or kernel functions. We also study four fusion methods in order to reduce uncertainties related to the classifier evaluating the prognostic significance of genes. In all cases, the selection process is embedded into a cross validation scheme in order to enhance the confidence on the generalization of results. We consider the differences of signatures based on gene overlap and also the biological annotation of selected genes, using the MIPS FunCat architecture. We found out that a robust identification of a number of highly differential genes can offer "good" predictive power to the models. Furthermore, the classification accuracy achieved by mixtures of experts can be significantly better than the one of the individual classifiers. We also pointed out that different selection schemes result in a diverse size of gene signature, with differences in the selected genes. Nevertheless, when we annotate the genes of each signature we find that the same biological processes are invoked, with possibly small differences in the relative frequency of participation.**

***Keywords- classification; mixture of experts; gene selection; Saccharomyces cerevisiae; biological processes***

## I. INTRODUCTION

The interest on expression-profile datasets has drastically increased with the aim of facilitating diagnosis, prognosis, or therapy of diseases. The baker's yeast *Saccharomyces cerevisiae*, a widely used eukaryotic model organism, has also been proved a successful model for studies of human diseases and molecular pathways. Yeast studies have pioneered our current knowledge about main biological processes (BPs) such as cell cycle, regulation of gene expression, and metabolism in humans. They have promoted the analysis of disease genes and aberrant cellular pathways through molecular, biochemical and also bioinformatics techniques [1-2]. It is known that almost 50% of genes implicated in human diseases have homologs in yeast and vice versa, at least 31% of yeast proteins have homologs in humans [3], proving the key role of *S. cerevisiae* in the clarification of human gene function [1].

One of the main problems in the analysis of gene expression is to determine which genes are expressed differently in different tissues or in two phenotypically different conditions. There are two main approaches in this field, one searching for discriminant genes in a certain population and the other identifying groups of genes with high predictive power for the conditions studied. Finding the genes whose expression levels are associated with a particular disease is important for selecting the most appropriate therapy and for the prediction of its recurrence by scientists. Moreover, this would facilitate biologists, allowing the design of smaller microarray DNA, adjusted to a specific disease, recording the expression levels of only some tens of genes.

Besides statistical errors in the measurements of different datasets due to platform or procedural inconsistencies, different algorithms operating on the same dataset may deduce different signatures with minimal (or even no) overlap. Not surprisingly, even different parameters of the same algorithm may influence the results of marker evaluation. One of the issues with unstable performance of prediction algorithms is the "curse-of-dimensionality", which appears in genomic datasets where the number of genes is extremely larger than of available samples. Nevertheless, recent studies advocate properties to such datasets, which can be directly exploited to relieve uncertainties. Furthermore, from a different point of view, the enrichment of statistical techniques with relevant biological knowledge can shed light to the biological relationship of genes that otherwise appear different from their statistical evaluation. In addition, several meta-analysis studies reveal that two signatures that may show minimal overlap may still share significant biological overlap in the associated annotations, pathways or biological processes [4].

This paper proceeds with the algorithmic issues of the proposed methodology in Section II. It presents statistical

This research supported by "ONCOSEED" project funded by the NSRF 2007-13 of the Greek Ministry of Development, and by "YPERThEN" project, which is an INTERREG project funded by the EU and funds from Greece and Cyprus.

results and analyzes the biological consequences on a database of budding yeast in Section III. Finally, the paper concludes in Section IV.

## II. METHODOLOGY

We propose a new methodology for feature selection in order to derive the most important or predictive genes. At the base of this methodology, the recursive feature elimination algorithm based on linear neuron weights is firstly placed, in order to achieve recursive elimination of features. This block is succeeded by the estimation of the predictive power of "remaining genes" by means of different classifiers, as well as by mixtures of them. Multiple repetitions of this process provide an assignment of statistical significance to genes, based on their relative frequencies. The process of gene selection is embedded in a five-fold cross-validation scheme, in order to strengthen confidence in the results.

### *A. Recursive Feature Elimination Algorithm*

The recursive feature elimination based on linear neuron weights method (RFE-LNW) [5] is used to evaluate genes via their recursive elimination. In order to select the most differentially-expressed genes, this algorithm uses a variation of Fisher's coefficient and assigns higher weights to genes which differentiate more their expression in the two classes of interest and smaller or even zero weights to genes which are expressed in similar or the same way.

Particularly, weights are initially assigned to all genes randomly and are re-evaluated and potentially adapted from iteration to iteration, as in

$$w_i(t+1) = w_i(t) + \mu \cdot \text{sign}(e \cdot f'(u)) \cdot \text{sign}(g_i) \cdot f_2(g_i) \qquad (1)$$

where $\mu$ is the learning rate, e is the error function, $f'(u)$ is the derivative of a logistic-sigmoid function and $f_2(g_i)$ denotes the Fisher's coefficient for the $g_i$ gene. The genes are ranked according to the absolute values of weights and the genes with the smallest weight in absolute value are eliminated, considered as the less important. The classification accuracy of the surviving genes at each iteration is estimated by different learning and predictive models. The algorithm terminates when all genes have gradually been eliminated and then we can find out the smallest subset of genes with the highest classification accuracy for each model.

### *B. Classification Models*

The support vector machines [5, 6], as a state-of-art classification method, as well as other kernel-based classifiers [7], are used as learning and prediction models.

A support vector machine attempts to find the best separating hyperplane to distinguish between the two classes of interest, positive (+1) and negative (-1). The discrimination (decision) function is determined by the hyperplane that optimally separates the two classes, which is found when the boundaries of the two classes are as far as possible (maximal margin) from each other, and is placed in the middle of these boundaries. This hyperplane is defined in equation "(2)" only by instances (Support Vectors-SV) and not the whole data. Thus,

$$f(x) = \text{sgn}((\sum_{i=1}^{m} \lambda_i y_i k(x_i^{sv}, x)) + b) \qquad (2)$$

where m is the number of the training samples, $\lambda$ represents the vector of Lagrange multipliers, y is the label (either +1 or -1), k is a kernel function [5]. These instances or SVs are placed on the boundaries of the two classes and therefore define the boundaries (margins) of separation.

The kernel nearest centroid method uses training data to calculate the mass centers of the two classes, i.e. of the samples belonging to the positive and to the negative class and assigns the new test sample to the class whose centroid is closest (smallest quadratic distance) to the sample [7]. The kernel minimum squared error machine compares the kernel matrix between the new test sample and all the training samples, with the target vector of all training samples and assigns the new sample to either the positive or the negative class, according to the sign of this comparison [7]. The kernel subspace method computes the eigenvalues and the eigenvectors of the kernel matrices for the positive and the negative training samples. After that, it generates two discriminant functions by multiplying the largest eigenvalues and eigenvectors with the kernel matrix between the new test sample and all the training samples for each of the two classes separately. The new test sample is assigned to the class which is represented by the discriminant function with the largest value [7].

In order to select the most appropriate kernel for each classifier we apply several kernels and compute the kernel-target alignment [8] measure for all the pairs of kernel-classifier. According to this measure, we finally propose the use of the following schemes: support vector machines with linear and quadratic kernel, kernel nearest centroid classifier with second degree polynomial kernel, kernel minimum squared error machine with Gaussian kernel and kernel subspace method with third degree polynomial kernel.

Furthermore, we present four methods of combining classifiers. The first mixture-of-experts approach implements a fusion of the above five individual classifiers in a fuzzy weight combination [9]. The five experts are trained over the training set and after that, they perform classification of the test samples. Then we implement five gate functions taking into consideration the average values and the variances of these individual decisions. Finally, the weighted distance from class mean gate functions and the decisions are combined into a weighted averaged form as in "(3)", where $x_i$ is the i-th test sample, $g_k()$ and $f_k()$ are the gate and decision functions

$$f(x_i) = \sum_{k=1}^{K} g_k(x_i) \cdot f_k(x_i) \qquad (3)$$

for the k-th classifier and f(.) is the decision function of the fusion scheme.

The second method [10] is a fusion of two logistic-sigmoid classifiers trained on the basis of gradient descent learning. More specifically, the two experts along with two softmax activation gate functions [11] assign different weights to the genes of the training samples. After the generation of the mixture's decision for each sample, the weights are updated based on the error of the mixture's decision, following a gradient descent learning algorithm. The training process is repeated until the weights converge. During the testing process, the two experts and the two softmax activation gate functions use their fixed trained weights to produce the mixture's decision, in the form of "(3)".

The third mixture of experts is a special blend of the two schemes above. It trains the two logistic-sigmoid experts similar the second mixture, with the difference that the weights are updated on the basis of their own classification error rather than the mixture's error.

Finally, the fourth method [10] uses a different form of expert mixture, as it implements classifier selection rather than classifier fusion. The training set is divided into two clusters with the Self-Organizing Map (SOM) algorithm. Then, two local experts are trained, each one on the basis of one cluster, and perform classification of the test samples based on the logic of the nearest neighbor distances. For each test sample we implement two weighted distance from class samples gate functions reflecting its distance from the training samples of the each cluster. The mixture's decision is a linear combination of the experts' decisions and the gate functions, similar to "(3)".

### *C. Gene Selection Strategy*

The proposed strategy for feature selection aims at selecting the most significant genes and consists of three levels. The first level addresses the full dataset and aims at deriving and initial set of many characteristic genes. The second level performs a more detailed selection of genes starting from the subset of genes obtained from the first level. Each of the previous level evaluates the appropriate number of genes to be selected in a fivefold cross validation circle, and then proceeds with the actual selection of specific genes by implementing twenty cycles of fivefold cross validation. The third level evaluates the predictive power of the significant genes selected from the second level, through ten cycles of fivefold cross validation. These steps are briefly discussed in the following.

First level of gene selection strategy: This level consists of two steps, a data evaluation and a gene selection one. At the first step of whole-data evaluation ($1^{st}$ sublevel), an assessment of the classification performance of the genes of the entire dataset is performed. Firstly, within a fivefold cross validation scheme, we apply the RFE-LNW algorithm for the recursive elimination of all genes. In each repetition of this algorithm, the less significant genes are removed from both the training and the test dataset and the predictive power of the surviving genes is assessed by the individual classifiers and mixtures of experts. When all genes have been gradually eliminated for all five iterations, our models have been trained on several possible training sets and have been tested for their classification accuracy. We are now able to compute for each model the average performance achieved for all stages of elimination and identify the smallest number of genes for which the model achieves the highest average accuracy.

Thus, the first step of data evaluation derives an initial estimate of the minimum number of genes (say A) for each model, achieving the highest average classification performance. This number of A genes is used as a stopping criterion of the RFE-LNW algorithm for each iterative cycle (fold) in the next step of gene selection, which attempts an initial selection of predictive genes from the whole set of genes. We perform several runs of the feature selection algorithm, each stopping at A genes, and record these specific genes. After the implementation of twenty cycles of fivefold cross validation, we gather all recorded genes and compute their frequencies of occurrence. The genes with highest frequencies according to a threshold (knee of the graph) are selected. Thus, at the end of 100 CV repetitions, we plot the genes from the recorded list with their corresponding frequencies and we only keep the genes with the highest frequencies (say C) according to the graph's knee.

At the end of the first level of gene selection, we have a first selection of C most significant genes, which are retained from the dataset. These C specific genes are evaluated for their classification performance in the second level of data evaluation.

Second level of gene selection strategy: This level also consists of two steps, a data evaluation and a gene selection one, which are implemented starting from a number of C genes for all data samples. At the step of data evaluation, an overall assessment of the classification performance of the C genes is performed in the same manner as in the first level of data evaluation. At the end of this step we derive the minimum number of genes (say D) for each model, achieving the highest average classification performance. Then, the second step of gene selection performs a reduction of genes from C to D most significant genes with the RFE-LNW algorithm being applied multiple times.

The selection process is exactly the same as in the first level of gene selection, with the only difference being that the stopping criterion of the RFE-LNW algorithm is the number of D genes. We perform twenty cycles of fivefold cross validation and for each iteration we record the final D genes. At the end of 100 CV repetitions, we perform the selection of the most significant genes (say F) based on their relative frequencies of occurrence. These genes, which compose the "gene signature", are evaluated for their classification performance in the final level of signature evaluation.

Third level of gene selection strategy: This final level performs an assessment and evaluation of the classification performance of the F selected genes from the second level of gene selection. The dataset is reduced again in order to include only these F genes. Then we apply 10 cycles of fivefold cross validation, so that the classification models are trained and tested for their generalization ability for all the F specific genes. In this final level, the RFE-LNW algorithm is not used since there is no need for feature elimination.

Our gene selection strategy do not perform evaluation of individual genes for their discrimination ability between the two classes (as in filter methods), but evaluation of groups of genes for their predictive power (in accordance to wrapper methods). In this scheme the evaluation of genes changes at each iteration step and for every data fold selected. The proposed use of a two-level gene selection scheme aims to restrain such variations in the ranking of genes. The first level derives a large number of potentially useful markers from the initial list of genes. Alternatively, the second level attempts a refinement of these markers by repeating the selection process on a more constrained basis, starting from a smaller list of genes.

## III. RESULTS

We test our method on the microarray dataset of Eisen et al. (1998) [12], which contains 2467 yeast (*Saccharomyces cerevisae*) genes. We used data from time series during the following processes: 1) the cell division cycle after synchronization by the centrifugal elutriation (positive class) and 2) environmental conditions such as temperature and reducing shocks (negative class) (14 time series, respectively).

The goal of gene selection is to extract a small set of informative genes that are expressed differently in the two above processes which constitute the two classes of interest in our work [13]. Informative genes are those useful to train a model which can generalize, i.e. correctly predict the class of new samples. Generally, in this work we used statistical tools in recursive feature elimination, in order to produce a sorted list of genes that contain in decreasing order the most frequent genes from the set of most differentially expressed genes. Successive groups of genes from the different selection levels are used for training and testing the models using the samples of the two classes of interest.

Initially, the classification accuracy of the models for the entire set of 2467 genes is used as reference of the predictive value of genes and gene selection. Starting from this initial number of genes, we proceed with recursive feature elimination up to a level that achieves high classification accuracy with only a few genes. Thus, we use the predictive accuracy as a measure of relevance of each group of genes. In the process of gene elimination, however, the relevance of a gene is determined by the degree of differential expression in the classes of interest and by the relative frequency of appearance of this gene. In essence, our gene selection scheme is based on the hypothesis that, if a number of highly differential genes can be robustly identified, then with large possibility this set of genes will express "good" predictive power [14]. In essence, the proposed gene selection scheme exploits and combines the discriminative power as well as the predictive power of the genes in the selected signature. Then, the proposed methodology provides a reliable method to assign statistical significance to genes and highlights only those that are the most relevant from the initial set of genes.

By applying this proposed methodology for nine different predictive schemes (individual classifiers and mixtures), we generate nine gene signatures at the end of the second level of gene selection, one for each tested model. These signatures are compared in terms of their accuracies and their differences/ similarities based on their potential overlap are discussed. We also compare the improvement in classification accuracy of each model from the initial set of 2467 genes to the selected group of genes in the corresponding gene signature. Finally, the signatures are compared in terms of the biological description and the biological significance of the selected genes.

### A. Statistical Results

Figure 1 presents the mean classification accuracy of all models for the dataset involving the initial 2467 genes and the final F selected genes. The numbers of the F selected genes are 44, 13, 13, 13, 11, for the kernel nearest centroid classifier with 2$^{nd}$ degree polynomial kernel, kernel minimum squared error machine with Gaussian kernel, kernel subspace method with 3$^{rd}$ degree polynomial kernel, SVM quadratic, Svm linear, and 9, 13, 13, 13 for the first, second, third and fourth mixture-of-experts, respectively. From these results, we conclude the following points:

- Each model results in a different number of selected genes, but not in completely different gene signatures; apart from the differences in the selected genes, there are some genes in common.
- The classification accuracy achieved by the models through the genes that constitute the signatures, is significantly better than that achieved through the initial set of genes with 2467 genes, as shown from figure 1. The only exception is the kernel nearest centroid classifier that seems to reduce its accuracy for a small number of genes.
- The mixture of experts approach can achieve higher accuracy, especially if the combination type is classifier selection, such as the fourth mixture. This mixture forms its decision based on the complementarity of the individual decisions, which have been obtained from the expert's training on different subsets of the feature space, separated by SOM.
- Last but not least, we observe that the performance of a single classifier depends not only on the form of its decision function but also on its kernel, since its performance varies using different kernels. Adapting a kernel [8] suitable to improve alignment with the labeled training samples, greatly enhances the alignment with the test samples, resulting in improved classification accuracy. Thus, the choice of kernel is a crucial issue for each classifier as it influences its classification accuracy.

In addition to these results, we computed the similarity percentage between the gene signatures of the models in pairs, with reference to the smaller set. With this test we attempt to clarify if comparing two sets of different size, all elements of the smaller set are included in the larger set; in this case the similarity percentage reaches 100%.

We found that the rates of gene similarity are high in many cases, as shown in tables I and II. However, there are also combinations of large disimilarities. The signatures of the individual classifiers have fairly common genes among

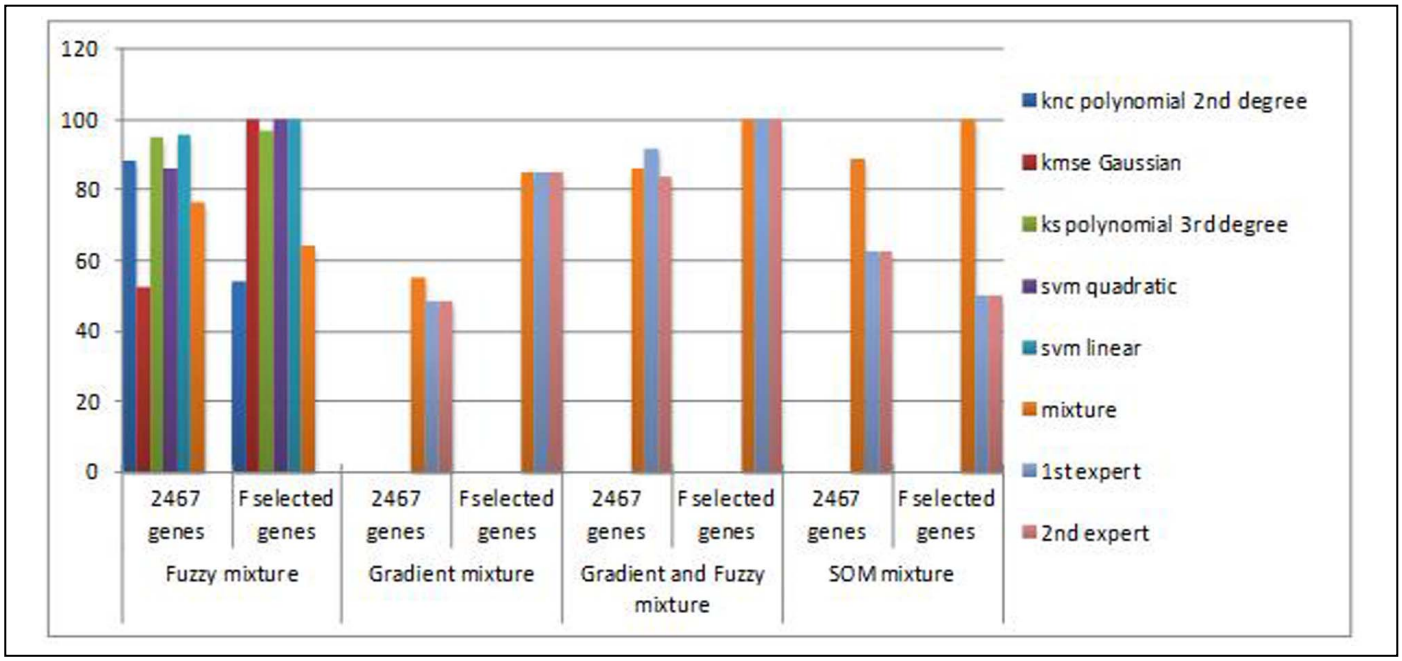


Figure 1. The average accuracy of the individual classifiers and the mixtures of experts, for the initial 2467 genes and the F selected genes.

themselves, as well as with their mixture. The classifier with the larger variation is the kernel nearest centroid with 2nd degree polynomial kernel, which is reasonable considering that its gene signature consists of many more genes than the other models. Finally, the similarity percentages of the signatures from different mixtures are high, indicating good stability of the mixure approach in the gene selection process.

### *B. Biological Results*

Genes that may look different from their statistical evaluation may have strong biological relationship. Thus, two or more gene signatures with small overlap between them, it is possible to share significant biological overlap in relevant descriptions or in related biological processes [13].

Towards this direction, we use the FunCat classification scheme [15], which is a hierarchically structured, organism independent, flexible and extendable system that allows the functional description of proteins from each organism, such as prokaryotic genomes, eukaryotic monocytes genomes, plants and animals. Taking into consideration the extremely broad and diverse spectrum of known protein functions, FunCat consists of 28 main functional categories that cover general fields like transcription and metabolism. It presents a hierarchical structure that resembles a tree, using up to six levels of increasing specialization; the second version used in this work includes 1307 functional categories of eukaryotic genomes of *Saccharomyces cerevisiae*. The FunCat descriptions were used to scan the significantly enriched categories in all different expressed genes. Considering each gene signature with this system, we determined a number of interesting biological processes involving the signature genes.

These processes are shown in figures 2 and 3. We observe that the 11, 13, 13, 13, 44 and 9 genes of the five individual classifiers and their mixtures, respectively, may not be the same but they all participate in the same 12 BPs of the same level (out of 28 main branches of FunCat), as illustrated in figure 2. The same happens with the gene signatures of the four mixtures, the genes of which share 10 BPs of the same level (out of 28 main branches of FunCat), as shown in fig. 3.

Importanly, the observed BPs shown in figures 2 and 3 reflect the expected cellular processes according to the examined data. For example, the overrepresented process in MIPS FunCat "cell cycle and DNA processing" refer to the events during the cell-division cycle, while "transcription" supports previous experimental findings about the high

TABLE I. THE SIMILARITY PERCENTAGES AMONG THE 5 INDIVIDUAL CLASSIFIERS AND THEIR MIXTURE'S

| *Similarity percentage %* | **SVM Linear** | **SVM Quadratic** | **KS Polynomial 3rd degree** | **KMSE Gaussian** | **KNC Polynomial 2nd degree** |
|---|---|---|---|---|---|
| **SVM Linear** | | 100% | 100% | 100% | 54% |
| **SVM Quadratic** | | | 84,62% | 84,62% | 61,54% |
| **KS Polynomial 3rd degree** | | | | 100% | 53,85% |
| **KMSE Gaussian** | | | | | 53,85% |
| **Fuzzy MoE** | 77% | 77% | 88,88% | 88,88% | 44,44% |

TABLE II. THE SIMILARITY PERCENTAGES AMONG THE 4 MIXTURES OF EXPERTS

| *Similarity percentage %* | **Fuzzy MoE** | **Som MoE** | **Gradient MoE** | **Gradient and fuzzy MoE** |
|---|---|---|---|---|
| **Fuzzy MoE** | | 88,88% | 88,88% | 77,77% |
| **Som MoE** | | | 92,31% | 84,62% |
| **Gradient MoE** | | | | 84,62% |

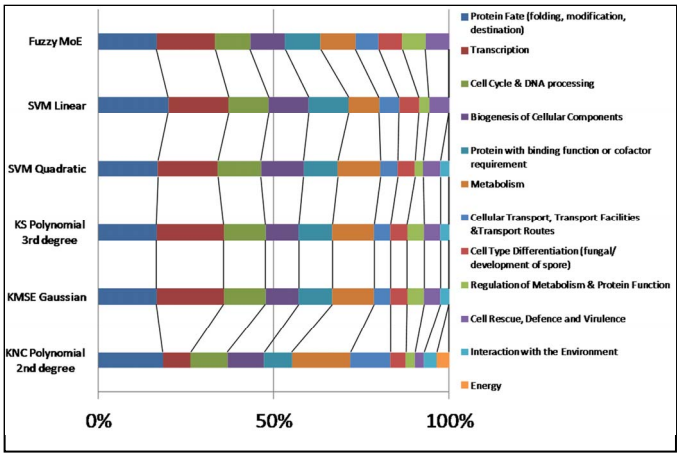


Figure 2. Comparative results of 12 main BPs for the 5 individual classifiers and the fuzzy mixture of them.

periodical transcriptional activity during the *S. cerevisae* cell cycle that can function as a cell cycle oscillator independently or in collaboration with the CDK oscillator [16].

These results confirm the need for exploration of biological effects related to the genotype and the biological support of similarity measures and/or distance functions used in statistical approaches, as to reduce uncertainty in decision making that include genomic data.

## IV. CONCLUSION

The selection of a subset of important genes significantly increases the accuracy of the classification models by reducing the uncertainty introduced by their parameters. Different algorithms that operate on the same dataset can lead to different signatures with little or no overlap, but also the different parameters of the algorithm itself can affect the results of marker evaluation. This study attempts to reveal and emphasize aspects of such parameter influence in the quality of the signature. We followed a fixed approach for the enhancement of genes that are expressed differently in two classes of interest. Thus, we tested the same way that weights are assigned to genes and the same criterion for gene selection for recursive elimination by the RFE-LNW algorithm. The factors that differentiate our study in the selection of most significant genes and leads to the formation of 9 different gene signatures are: 1) the different algorithmic parameters (different classification models) introduced for evaluation of subsets of genes and 2) the way in which we choose the final number of genes (different accuracy thresholds used at the first level of evaluation for each model). As a result, we derived certain differences in the selected gene signatures. Nevertheless, by considering the biological significance of these signatures, we conclude that there is significantly increased overlap in terms of biological processes, which necessitates the exploration of biological effects.

Concluding, we emphasize the need for further testing of our findings, towards identifying and validating such implications of gene selection and classification in different organisms, including human.

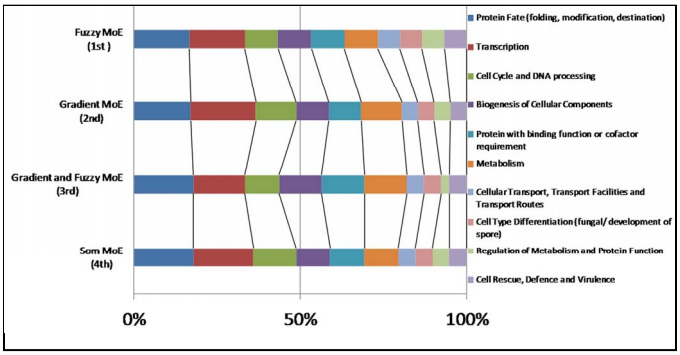


Figure 3. Comparative results of 10 main BPs for the 4 mixtures of experts.

## REFERENCES


[1] B. Suter, D. Auerbach, and I. Stagljar, "Yeast-based functional genomics and proteomics technologies: the first 15 years and beyond," Bio.Techniques, vol. 40(5), pp. 625-644, May 2006.

[2] P. C. Stirling, M. S. Bloom, T. Solanki-Patil, S. Smith, P. Sipahimalani, Z. Li, et al, "The complete spectrum of yeast chromosome instability genes identifies candidate CIN cancer genes and functional roles for ASTRA complex components," PloS. Genet, vol. 7(4), Apr 2011.

[3] L. H. Hartwell, "Yeast and cancer," Biosci. Rep, vol. 24(4-5), pp. 523-544, Aug-Oct 2004.

[4] J-Z. Xu, and C-W. Wong, "Hunting for robust gene signature from cancer profiling data: Sources of variability, different interpretations, and recent methodological developments," Cancer. Letters, vol. 296, pp. 9–16, 2010.

[5] M. Zervakis, and M. E. Blazadonakis, "Wrapper filtering criteria via linear neuron and kernel approaches," Computers. Biol. Med, vol. 38, pp. 894-912, 2008.

[6] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene Selection for Cancer Classification using Support Vector Machines," Mach. Learn, vol. 46, pp. 1-3, March 2002.

[7] S. Niijima, and S. Kuhara, "Gene subset selection in kernel-included feature space," Pattern. Recogn. Letters, vol. 27, pp. 1884-1892, 2006.

[8] N. Cristianini, J. Kandola, A. Elisseeff, and J. Shawe-Taylor. "On Kernel-Target Alignment," in Innovations in Machine Learning: Theory and Application's, Springer Verlag, 2006.

[9] J. Leski. "A fuzzy if-then rule-based nonlinear classifier," Int. J. Appl. Math. Comput. Sci, vol. 13, pp. 215-223, 2003.

[10] B. Tang, M. I. Heywood, and M. Shepherd, "Input partitioning to mixture of experts," Proceed. Intern. Joint. Conf. Neural. Networks, vol. 1, pp. 227-232, 2002.

[11] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G.E. Hinton, "Adaptive mixtures of local experts," Neural. computation, vol. 3(1), pp. 79–87, 1991.

[12] M. B. Eisen, P. T . Spellman, P. O. Brown, and D. Botstein, "Cluster analysis and display of genome-wide expression patterns," Nat. Acad. of Sciences., vol. 95, pp. 14863–14868, 1998.

[13] R. Maglietta, A . D'Addabbo, A. Piepoli, F. Perri, S. Liuni, G. Pesole, et al, "Selection of relevant genes in cancer diagnosis based on their prediction accuracy," Artif. Intellig. Medicine, pp. 29-44, 2006.

[14] K. Dobbin, and R. Simon, "Sample size planning for developing classifiers using high-dimensional DNA microarray data," Biostatistics, vol. 8 (1), pp. 101–117, 2007

[15] A. Ruepp, A. Zollner, D. Maier, K. Albermann, J. Hani, M. Mokrejs, et al, "The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes," Nucleic Acids Research, vol. 32 (18), pp. 5539–5545, 2004.

[16] D. A. Orlando, C. Y. Lin, A. Bernard, J. Y. Wang, J. E. Socolar, E. S. Iversen, et al, "Global control of cell-cycle transcription by coupled CDK and network Oscillators," Nature, vol. 453(7197), pp. 944-947, Jun 2008.