Bayesian Networks Classifiers for Gene-Expression Data

Luis M. de Campos, Andrés Cano, Javier G. Castellano and Serafín Moral
Department of Computer Science and Artificial Intelligence. Granada University.
Granada, Spain
Email: {*lci,acu,fjgc,smc*}*@decsai.ugr.es*

Abstract—In this work, we study the application of Bayesian networks classifiers for gene expression data in three ways: first, we made an exhaustive state-of-art of Bayesian classifiers and Bayesian classifiers induced from microarray data. Second, we propose a preprocessing scheme for gene expression data, to induce Bayesian classifiers. Third, we evaluate different Bayesian classifiers for this kind of data, including the C-RPDAG classifier presented by the authors.

Keywords-Machine learning, supervised classification, Bayesian network, gene expression data, DNA microarray, Bayesian classifier, naïve Bayes, TAN, BAN, Selective naïve Bayes, C-RPDAG, Bayesian multinet.

I. INTRODUCTION

In a *supervised classification* problem we have a set of observations or cases, made up of a series of attributes or values that we observed and a variable that we want to predict. This is called variable to classify or, simply, class. A *classifier* is a function that maps an instance into a class label. Bayesian networks have been successfully applied to classification problems in many ways by inducing classifiers using different types of Bayesian network learning algorithms.

In the field of Bayesian classifiers, the first one which can be considered as a Bayesian network classifier, is the *naïve Bayes* (NB) classifier [1,2]. This simple classifier relies on two assumptions: (1) each attribute is conditionally independent from the other attributes given the class and (2) all the attributes have influence on the class. In the *directed acyclic graph* (DAG) representing the NB structure all the arcs are directed from the class to the attributes. The success of this classifier is mainly due to its simplicity, exhibiting a surprisingly competitive predictive accuracy [2,3]. At the other extreme in Bayesian classifiers, any algorithm for learning Bayesian networks can be used to induce a classifier [3,4,5].

The relative success of the NB classifier has motivated the development of many methods, which try to improve it. One way is, starting from the NB basic topology, complete it by adding *augmenting arcs* between attributes, for example, *tree-augmented naïve Bayesian network* (TAN) [3], where the attributes are linked together as a directed tree or the *Bayesian network augmented naïve Bayesian classifier* (BAN). In this case the NB structure is fixed and the search of the augmenting arcs is carried out using some algorithm

for learning Bayesian networks [4]. In the middle between TAN and BAN we have the *k-dependence Bayesian classifier* (KDB) [6], that is a Bayesian network whose associated DAG contains the structure of the naïve Bayes classifier and allows each attribute to have a maximum of k attribute nodes as parents.

There are also other variations of the NB model, this is the case of the *semi-naïve Bayesian classifier* [7] and the model proposed by [8]. They share the idea of merging several correlated attributes into new compound attributes. Other approximations remove correlated and irrelevant attributes using *feature subset selection* (FSS) to improve the naïve Bayes classifier: the *selective naïve Bayesian classifier* (SNB) [9,10] searches for a good subset of attributes, starting from an empty set and iteratively introducing the attribute that most improves the predictive accuracy, using cross-validation.

Out of the NB model and variations, the C-RPDAG classier is based on learning Bayesian networks with a local search and using a standard scoring function. The method carries out a simple local search in a space composed of a type of partially directed acyclic graphs (PDAGs), which combine two concepts of equivalence of DAGs: classification equivalence and independence equivalence. Using the BDeu score, this algorithm has proved more effective than other Bayesian network classifiers [11].

Besides that, Bayesian multinets are Bayesian network extensions where context-specific conditional independences can be represented, also we can use them for classification purposes [12]. In Bayesian multinets we have a distinguished variable and a Bayesian network for each value of this variable.

Gene expression is the process where a gene information is used in the synthesis of a functional gene product (e. g. proteins). DNA microarrays (also called DNA chips) [13] can be used to measure the expression levels of thousands of genes simultaneously in a single experiment. The study of the expressions of a huge number of genes in a single clinical test allows us to perform comparisons among tissues, pathological phases or the different responses to different biological conditions.

Inducing Bayesian networks from gene expression data has some difficulties [14], the main problem is the *curse of*

the dimensionality: the number of features (genes) characterizing these data is thousands or tens of thousands, while the number of samples is limited because it is an expensive test. In DNA microarrays we are working at molecular level, therefore this data is affected by noise, measurement errors and missing values. Besides, a gene expression is a continuous value.

The rest of the paper is organized as follows. Section II shows the use of Bayesian classifiers to analyse DNA microarray data. Later Section III contains the experimental results of the evaluation for the different Bayesian classifiers of some gene expression public databases. Finally, Section IV contains the concluding remarks.

II. BAYESIAN CLASSIFIERS FOR DNA MICROARRAY DATA

In the literature, when we are learning Bayesian classifiers from gene expression data, we can observe that most approaches perform a *feature subset selection* (FSS)[1] and afterwards they use the classifier.

In many works the NB classifier is used for comparison purposes. For example, Inza et al. [15] study four supervised classification methods (NB among others). This work is focused in gene selection to improve the classification accuracy. In an extension of that work, they compare some filter and wrapper FSS methods [16], using discretized and continuous data. They obtain better results with a wrapper FSS and discretized data. On the other hand in [17], they also perform a gene selection but using EDAs [18] with four different starting points. For the search, a selective naïve Bayes is used.

The NB is used in [19] by Ben-Dor et al. to compare different FSS methods. The same authors [20] use the NB classifier in a similar way, but in this case this classifier is used to validate classes found with non supervised methods.

Also with comparison purposes, the NB classifier is used by Marmor et al. in [21] and Osareh et al. in [22]. In the second work, the authors also study some FSS techniques.

Some authors work directly with continuous data instead of using discretization methods. That is the case of Moler et al. in [23] where the supervised NB and the non supervised NB are studied. At the beginning, the authors use a non supervised NB to find subclasses. This classifier uses a Gaussian approach to deal with continuous variables. Besides they found two measures based in the NB for ranking the genes.

In [24], Fan et al. begin with a feature selection and a feature transformation, using later a continuous NB. The transformation converts features in new variables, which are mutually independent given the class.

[1]In this specific case also called *gene selection*.

Also dealing with the continuous nature of the gene expression data, Cano et al. [25] use a Gaussian NB. They perform a two steps gene selection: first, a filter selection and, second, using a Selective NB. After that, they learn a Bayesian Network using the selected genes. In [26] the same authors also use a Gaussian Selective NB with two improvements: the preordering of the feature set and the application of an irrelevant feature elimination method.

As we have said, apart from the NB classifier, there are other classifiers that have been also used with gene expression data.

Armañanzas et al. [27] use a KDB classifier, with $k = 4$. Bootstrap resampling with 1000 iterations is used to add reliability and robustness to the overall process. In each iteration a FSS is performed, concretely the *correlation feature selection* [28]. After the bootstrap, they build an averaged classifier using the KDB induced previously. In [29] they improve their method with consensus gene selection.

The BAN classifier has been used by Bosin et al. [30] with DNA microarray data. They use the MDL score to rank the genes. After that, a gene selection is performed. Finally with the selected features they build a BAN and a NB classifiers.

A general Bayesian network is used in [31] by Hwang, Zhang et al. for comparison purposes. They discretize the data in two states and perform a FSS. The Bayesian network is learned with a score+search paradigm. In [32] the p-metric is used for gene selection. In [33] they use a clustering method and discretize the data in three states.

Bockhorst et al. [34] use a Bayesian network to identify operons, where the network structure is made by an expert. In a previous work [35] they use a NB.

Diao et al. [36] present a method that combines clustering, Bayesian networks learning and disease gene selection.

The use of additional knowledge can improve the Bayesian classifiers accuracy when we are working with a very few cases. For example, Gevaert et al. [37] use microarray data and clinical data. They build the Bayesian networks with and without the clinical data. In [38] they consider data from Pubmed which are used in the prior distribution of the Bayesian networks.

Hwang et al. [39] also use Bayesian networks. They build an average network using different features orderings to avoid overfitting.

Trying to represent time-series data with classifiers, Tucker et al. [40] introduce a hybrid approach based on Bayesian classifiers and dynamic Bayesian networks.

III. EXPERIMENTS

In literature we have seen the importance of gene selection in order to deal with the gene expression data dimensionality. Besides most experts have chosen to

discretize the data instead of working with continuous approaches.

In this section, we are going to use some state-of-the-art Bayesian classifiers like the NB classifier, the TAN or the Selective NB, as well as other Bayesian classifiers like the C-RPDAG classifier and Bayesian Multinet. The reason to include a Bayesian multinet classifier in the experiments is to find *Cellular Context Mining* [41] and represent it with asymmetrical independences. The use of the C-RPDAG algorithm is motivated for its good results [11].

A. Data sets

We have selected nine data sets that have been widely used in the literature and which are based on different types of cancer. Those data can be obtained from its original authors and from *Kent Ridge Bio-medical Dataset* [42], where the data sets are stored in different formats.

Table I gives a brief description of the characteristics of each database, including the number of instances, genes (attributes) and states for the class variable.

Table I
GENE EXPRESSION DATA SETS FOR DIFFERENT TYPES OF CANCER USED IN OUR EXPERIMENTS.

Data Set	Instances	Genes	Classes
Breast-cancer	97	24481	2
CNS	60	7129	2
Colon-cancer	62	2000	2
DLBCL-MIT	77	7129	2
DLBCL-Stanford	47	4026	2
Leukemia ALL-AML	72	7129	2
Leukemia-MLL	72	12582	3
Lung-cancer	203	12600	5
Prostate-tumor	136	12600	2

The first data set (*Breast-cancer*) is about patients outcome prediction for breast cancer [43].

The *CNS* dataset is about patients outcome prediction for central nervous system embryonal tumor [44]. Only dataset C mentioned in the paper is used.

Alon et al. present in [45] a study from colon-cancer patients with tumor biopsies (*Colon-cancer*).

The data set from Alizadeh et al. [46] is about distinct types of diffuse large B-cell lymphoma (*DLBCL-Stanford* in table I).

Other data set about diffuse large B-cell lymphoma (DLBCL) is the one presented by Shipp et al. in [47] (*DLBCL-MIT* in table I). There are two kinds of classifications about diffuse large B-cell lymphoma addressed in the publication. First one is DLBCL versus Follicular Lymphoma morphology and the second problem is to predict the patient outcome. Only the first one was used.

The data set about Leukemia presented by Golub et al in [48] is about cancer classification (called *Leukemia ALL-AML*).

Also about leukemia, adding an additional type of cancer is the data set presented by Armstrong et al. in [49] (*Leukemia-MLL*).

About lung cancer is the data set presented by Bhattacharjee et al. in [50] (called *Lung-cancer*).

Finally, we have a prostate cancer data set from the work of Singh et al. [51] (*Prostate-tumor* in table I). In this work, two classification problems are presented: the first one is about tumor versus normal tissue and the second one is about prediction of clinical outcome. We have chosen the first one.

These data sets have been widely used given their public availability and the existence of previous results in the literature. In the table II we can observe the best results found for each data set in different papers [21,22,25,26,52,53,54,55].

Table II
BEST RESULTS FOUND IN THE LITERATURE.

Data set	Best accuracy
Breast-cancer	89,47
CNS	86,5
Colon-cancer	98,5
DLBCL-MIT	98,2
DLBCL-Stanford	97,6
Leukemia ALL-AML	98,8
Leukemia-MLL	100
Lung-cancer	95,2
Prostate-tumor	98,8

Some of these datasets had a train set and a test set, in those cases (*Breast-cancer*, *DLBCL-MIT*, *Leukemia ALL-AML*, *Leukemia-MLL* and *Prostate-tumor*) we have merged both sets because we are going to use cross validation in our experiments.

B. Preprocessing: Missing values imputation

In the different databases we have no missing data with the exception of the DLBCL-Stanford dataset. To work with this data without losing the cases with missing values, we have carried out a *local least squares imputation method* (*LLSImpute*). The software used for this imputation is implemented by the package *pcaMethods* [56].

C. Preprocessing: Discretization

One of the DNA microarrays problems that we have to deal with is that gene expressions take continuous values. As we have seen in section II, most authors use discretization methods to address this problem in Bayesian networks.

The advantage of the discretization process is stability versus random and systematic variations in the gene expressions. The disadvantage of this process is the information

Table III
PREDICTIVE ACCURACY FOR EACH BAYESIAN CLASSIFIER USING LEAVE-ONE-OUT.

Dataset	NB	TAN	BAN	SNB	UBN	C-RPDAG	BMN$_w$
Breast-cancer	100.0	98.97	100.0	98.97	100.0	100.0	98.97
CNS	96.67	98.33	98.33	96.67	100.0	100.0	96.67
Colon-cancer	98.39	98.39	97.77	98.39	100.0	100.0	98.39
DLBCL-MIT	100.0	100.0	100.0	100.0	100.0	100.0	100.0
DLBCL-Stanford	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Leukemia ALL-AML	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Leukemia-MLL	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Lung-cancer	98.52	97.54	98.52	96.55	98.52	98.52	97.54
Prostate-tumor	97.79	97.06	97.79	98.53	97.79	100.0	97.79
Average	99.04	98.92	99.17	98.79	99.59	99.84	98.82

loss[2].

To deal with the continuous nature of gene expression and Bayesian networks, the most used discretization method was proposed by Friedman et al. [61]. They discretize the expression levels into three states: under-expressed, baseline and over-expressed. We had used this discretization method at early stages of this study but the results were worse than expected.

We have chosen the entropy discretization of Fayyad and Irani [62]. This method has been widely used in machine learning. Using this approach most of the gene expression data have been discretizated in two states or, in some few cases, in three states. In some cases the expression values were discretized into a single state, which is equivalent to an irrelevant gene.

D. Preprocessing: Gene Selection

The final stage of the preprocessing step is gene selection. We need to address the curse of dimensionality for expression data and reduce the number of attributes to learn Bayesian classifiers in a reasonable time.

In this stage we have followed a similar method to the one presented by Armañanzas et al. in [29]: First, we have removed those genes discretized into a single state at the discretization stage. Second, we have chosen the most representative genes from high correlated groups (called prototype genes by Armañanzas et al.) where we have carried out a *correlation based filter*(CFS) [28] gene selection. And third, we have ranked the genes by a metric and we have eliminated all the features that do not achieve an adequate score. The ranking gene selection is motivated by the high number of variables in some problems after the CFS feature subset selection. For this sub-stage we are going

[2]Some researchers [57,58,59] have chosen the Bayesian network paradigm as the best method to induce genetic regulatory networks from gene expression data. The model used today for genetic regulatory networks is based on the Kauffam proposal [60] where each node is represented by two states: ON and OFF. Despite the simplicity of the model, it exhibits a good predictive behaviour and has been used as a model for real genetic network problems. Therefore we can think that a two or three states discretization is not a bad approximation.

to carry out the gene selection only in those problems with more than 100 features. The measure used to rank the genes was the *information gain* defined by Quinlan for the ID3 algorithm [63]. After the rank is made, only the top 100 genes are selected.

E. Experimental Results

The classifiers being considered in this study are naïve Bayes (NB), tree augmented naïve Bayes (TAN), Bayesian network augmented naïve Bayes (BAN), Selective naïve Bayes (SNB), a wrapper Bayesian multinet (BMN$_w$,) and two unrestricted Bayesian network learning algorithms (UBN and C-RPDAG). All the algorithms are implemented in the Elvira system [64].

Both BAN and UBN algorithms search locally in the space of augmented DAGs and unrestricted DAGs, respectively (with the classical operators of arc addition, arc deletion, and arc reversal), and are scoring-based. C-RPDAG, which is also scoring-based, searches locally in the C-RPDAG space using its specific operators. In these three methods, the selected scoring function is the same: BDeu [65], with the parameter representing the equivalent sample size set to 1 and an uniform structure prior. The starting point of the local search in UBN and C-RPDAG was the empty graph.

In the wrapper Bayesian multinet classifier [12] for each attribute X_i, we learn the classifier expanding this variable and determining its accuracy, which will be computed using 5-fold cross-validation. Afterwards the node X_j with the greatest accuracy is chosen to branch. The classifiers used at the leaves will be Naïve Bayes. The SNB classifier also follows a wrapper approach with 5-fold cross-validation.

Once the graphical structure of each algorithm has been determined, the parameter learning must be done. In all cases, we have used the same smoothed estimator, based on the Laplace correction [66], trying to prevent issues of unreliability and overfitting of the maximum likelihood estimator in data sets with few cases.

The results of this comparative study are displayed in Table III. For each classifier and dataset we show the predictive accuracy (the percentage of successful predictions on a test set different from the training set. More is better.). This

accuracy was estimated using leave-one-out cross-validation. We also have computed the non-parametric Friedman test [67] for comparing multiple classifiers [68]. In the table IV we can observe the average rankings for each classifier (less is better).

Table IV
AVERAGE RANKINGS OF THE BAYESIAN CLASSIFIERS USING FRIEDMAN'S TEST

Classifier	Ranking
NB	4.0
TAN	4.7222
BAN	4.0
SNB	4.6111
UBN	3.1667
C-RPDAG	2.7778
BMN_w	4.7222

We observe that better results than previous works (table II) have been achieved even with the NB classifier. These results suggest that our preprocessing stage does not suffer of information loss in the discretization stage nor in eliminating relevant genes in the genes selection stage on the chosen datasets.

Looking at tables III and IV, we see that the C-RPDAG classifier obtains the best results for all the problems, closely followed by the UBN approach.

Surprisingly, however, the NB has a better behaviour than the TAN approach, we think that it is due to the gene selection performed, where we obtain high relevant and uncorrelated genes.

The worst results come from the Bayesian multinet. We believe they are due to the data partitioning for each branch. We think that the SNB suffers the same problem in the learning step. Besides we must take into account that the SNB classifier makes another gene selection, that seems detrimental.

IV. CONCLUSIONS

When we use Bayesian classifiers to analyse DNA microarray data we must face some important problems inherent in this type of data. The two most important are the curse of dimensionality and the continuous nature of gene expression.

This paper has performed a survey of the distinct solutions given in previous works for Bayesian classifiers induced from expression data.

To address the problem of continuous data we have chosen a discretization method. In the experimental results, we can see that the discretization information loss is meaningless.

To deal with the dimensionality of the data, we have proposed a three steps gene selection: first, irrelevant genes are eliminated (in the discretization stage); second, we select uncorrelated genes (with the CFS features subset selection); and third, we eliminate irrelevant genes with a FSS ranking method.

We have carried out a study using various state-of-the-art Bayesian classifiers. The best results come from the C-RPDAG and the UBN approaches. The excellent results shown by the C-RPDAG classifier are the outcome of two reasons: one, it is more expressive than a classical Bayesian classifier (NB or TAN are structurally limited); two, these networks combine the ideas of classification equivalence and independence equivalence to produce a more reduced, focused and robust search space.

REFERENCES

[1] R. Duda and P. Hart, *Pattern classification and scene analysis*. John Wiley and Sons, New York, 1973.

[2] P. Langley, W. Iba, and K. Thompson, "An analysis of Bayesian classifiers," in *National Conference on Artificial Intelligence*, 1992, pp. 223–228.

[3] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian networks classifiers," *Machine Learning*, vol. 29, pp. 131–163, 1997.

[4] J. Cheng and R. Greiner, "Comparing Bayesian network classifiers," in *UAI '99: Proc. of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, 1999, pp. 101–108.

[5] ——, "Learning Bayesian belief network classifiers: Algorithms and system," in *AI '01: Proc. of the 14th Biennial Conference of the Canadian Society on Computational Studies of Intelligence*. Springer-Verlag, 2001, pp. 141–151.

[6] M. Sahami, "Learning limited dependence Bayesian classifiers," *Second International Conference on Knowledge Discovery in Databases*, pp. 335–338, 1996.

[7] I. Kononenko, "Semi-naive Bayesian classifier." in *European Working Session on Learning on Machine Learning*, 1991, pp. 206–219.

[8] M. J. Pazzani, "Searching for dependencies in Bayesian classifiers," in *Artificial Intelligence and Statistics IV*, ser. Lecture Notes in Statistics. Springer-Verlag, 1997.

[9] P. Langely and S. Sage, "Induction of selective Bayesian classifiers," *Proceeding of the Tenth Conference on Uncertainty in Artificial Intelligence*, pp. 399–406, 1998.

[10] G. H. John, R. Kohavi, and K. Pfleger, "Irrelevant features and the subset selection problem," in *International Conference on Machine Learning*, 1994, pp. 121–129.

[11] S. Acid, L. M. de Campos, and J. G. Castellano, "Learning Bayesian network classifiers: Searching in a space of partially directed acyclic graphs," *Machine Learning*, vol. 59, no. 3, pp. 213–235, 2005.

[12] A. Cano, J. G. Castellano, A. R. Masegosa, and S. Moral, "Methods to determine the branching attribute in bayesian multinets classifiers," in *Symbolic and Quantitative Approaches to Reasoning with Uncertainty ECSQARU2005*, vol. 3571, 2005, pp. 932–943.

[13] M. Schena, D. Shalon, R. W. Davis, and P. O. Brown, "Quantitative monitoring of gene expression patterns with a complementary DNA microarray." *Science*, vol. 270, no. 5235, pp. 467–70, 1995.

[14] P. Spirtes *et al.*, "Constructing Bayesian network models of gene expression networks from microarray data," in *Proc. of the Atlantic Symposium on Computational Biology, Genome Information Systems and Technology*, 2000.

[15] I. Inza, B. Sierra, R. Blanco, and P. Larrañaga, "Gene selection by sequential search wrapper approaches in microarray cancer class prediction." *Journal of Intelligent and Fuzzy Systems*, vol. 12, no. 1, pp. 25–34, 2002.

[16] I. Inza, P. Larrañaga, R. Blanco, and A. J. Cerrolaza, "Filter versus wrapper gene selection approaches in DNA microarray domains." *Artificial Intelligence in Medicine*, vol. 31, no. 2, pp. 91–103, 2004.

[17] R. Blanco, P. Larrañaga, I. Inza, and B. Sierra, "Selection of highly accurate genes for cancer classification by estimation of distribution algorithms," in *Proc. of the Workshop Bayesian Models in Medicine, held within AIME 2001, Artificial Intelligence in Medicine*, 2001, pp. 29–34.

[18] P. Larrañaga and J. A. Lozano, *Estimation of distribution algorithms: A new tool for evolutionary computation*, ser. Genetic Algorithms and Evolutionary Computation. Kluwer Academic Publishers, 2001.

[19] A. Ben-Dor, N. Friedman, and Z. Yakhini, "Scoring genes for relevance," School of Computer Science & Engineering, Hebrew University, Jerusalem, Tech. Rep. 2000-38, 2000.

[20] ——, "Class discovery in gene expression data," in *RECOMB '01: Proc. of the fifth annual international conference on Computational biology*. ACM Press, 2001, pp. 31–38.

[21] M. Mramor, G. Leban, J. Demsar, and B. Zupan, "Visualization-based cancer microarray data classification analysis," *Bioinformatics*, 2007.

[22] A. Osareh and B. Shadgar, "Classification and diagnostic prediction of cancers using gene microarray data analysis," *Journal of Applied Sciences*, vol. 9, no. 3, pp. 459–468, 2009.

[23] E. Moler, M. Chow, and I. Mian, "Analysis of molecular profile data using generative and discriminative methods." *Physiological Genomics*, vol. 4, no. 2, pp. 109–126, 2000.

[24] L. Fan, K.-L. Poha, and P. Zhoub, "A sequential feature extraction approach for naïve bayes classification of microarray data," *Expert Systems with Applications*, vol. 36, no. 6, pp. 9919–9923, 2009.

[25] A. Cano, J. G. Castellano, A. R. Masegosa, and S. Moral, "Application of a selective Gaussian naive bayes model for diffuse large-B-cell lymphoma classification." in *Proc. of the 2nd European Workshop on Probabilistic Graphical Models(PGM'04)*, 2004, pp. 33–40.

[26] ——, "Selective gaussian naive bayes model for diffuse large-B-cell lymphoma classification: Some improvements in preprocessing and variable elimination," in *Symbolic and Quantitative Approaches to Reasoning with Uncertainty ECSQARU2005*, vol. 3571, 2005, pp. 908–920.

[27] R. Armañanzas, I. Inza, and P. Larrañaga, "Detecting reliable gene interactions by a hierarchy of Bayesian network classifiers," *Computer Methods and Programs in Biomedicine*, vol. 91, no. 2, pp. 110–121, 2008.

[28] M. A. Hall and L. A. Smith, "Feature subset selection: a correlation based filter approach," in *International Conference on Neural Information Processing and Intelligent Information Systems*. Springer, 1997, pp. 855–858.

[29] R. Armañanzas *et al.*, "Bayesian classifiers with consensus gene selection: A case study in the systemic lupus erythematosus," in *Progress in Industrial Mathematics at ECMI 2006*, 2007, vol. 12, pp. 560–565.

[30] A. Bosin, N. Dessì, D. Liberati, and B. Pes, "Learning Bayesian classifiers from gene-expression microarray data," in *Fuzzy Logic and Applications, 6th International Workshop, WILF 2005*, vol. 3849, 2005, pp. 297–304.

[31] K. B. Hwang *et al.*, "Applying machine learning techniques to analysis of gene expression data: Cancer diagnosis," in *Methods of Microarray Data Analysis (Proc. of CAMDA'00),*, 2002, pp. 167–182.

[32] B. T. Zhang and K. B. Hwang, "Bayesian network classifiers for gene expression analysis," in *A Practical Approach to Microarray Data Analysis*, 2003, pp. 150–165.

[33] J. H. Chang, K. B. Hwang, and B. T. Zhang, "Analysis of gene expression profiles and drug activity patterns by clustering and Bayesian network learning," in *Methods of Microarray Data Analysis II (Proc. of CAMDA'01),*, 2002, pp. 169–184.

[34] J. Bockhorst *et al.*, "A Bayesian network approach to operon prediction." *Bioinformatics*, vol. 19, no. 10, pp. 1227–35, 2003.

[35] M. Craven *et al.*, "A probabilistic learning approach to whole-genome operon prediction." in *Proc. of the 8th International Conference on Intelligent Systems for Molecular Biology*, vol. 8, 2000, pp. 116–27.

[36] Q. Diao *et al.*, "Disease gene explorer: Display disease gene dependency by combining Bayesian networks with clustering," in *3rd International IEEE Computer Society Computational Systems Bioinformatics Conference*, 2004, pp. 574–575.

[37] O. Gevaert *et al.*, "Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks," *Bioinformatics*, vol. 22, no. 14, pp. e184–e190, 2006.

[38] O. Gevaert, S. Van Vooren, and B. de Moor, "Integration of microarray and textual data improves the prognosis prediction of breast, lung and ovarian cancer patients." *Pacific Symposium on Biocomputing*, pp. 279–290, 2008.

[39] K. B. Hwang and B. T. Zhang, "Bayesian model averaging of Bayesian network classifiers over multiple node-orders: application to sparse datasets," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 35, no. 6, pp. 1302–1310, 2005.

[40] A. Tucker, V. Vinciotti, P. A. C. 't Hoen, and X. Liu, "Bayesian network classifiers for time-series microarray data," in *Advances in Intelligent Data Analysis VI, 6th International Symposium on Intelligent Data Analysis, IDA 2005, Proc.*, vol. 3646. Springer, 2005, pp. 475–485.

[41] I. Sen *et al.*, "Context-specific gene regulations in cancer gene expression data," in *Pacific Symposium on Biocomputing*, vol. 14, 2009, pp. 75–86.

[42] J. Li, H. Liu, and L. Wong, "Mean-entropy discretized features are effective for classifying high-dimensional biomedical data," in *Proc. of the 3nd ACM SIGKDD Workshop on Data Mining in Bioinformatics*, 2003, pp. 17–24.

[43] L. J. van 't Veer *et al.*, "Gene expression profiling predicts clinical outcome of breast cancer." *Nature*, vol. 415, no. 6871, pp. 530–536, 2002.

[44] S. L. Pomeroy *et al.*, "Prediction of central nervous system embryonal tumour outcome based on gene expression," *Nature*, vol. 415, pp. 436 – 42, 2002.

[45] U. Alon *et al.*, "Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays," *Proc. of the National Academy of Sciences of the United States of America*, vol. 96, no. 12, pp. 6745–6750, 1999.

[46] A. Alizadeh *et al.*, "Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling," *Nature*, vol. 403, pp. 503–511, 2000.

[47] M. A. Shipp *et al.*, "Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning," *Nature Medicine*, vol. 8, pp. 68 – 74, 2002.

[48] T. R. Golub *et al.*, "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring," *Science*, vol. 286, pp. 531–537, 1999.

[49] S. A. Armstrong *et al.*, "MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia," *Nature Genetics*, vol. 30, pp. 41 – 7, 2002.

[50] A. Bhattacharjee *et al.*, "Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses," *Proc. of the National Academy of Sciences of the United States of America*, vol. 98, pp. 13 790 – 5, 2001.

[51] D. Singh *et al.*, "Gene expression correlates of clinical prostate cancer behavior," *Cancer Cell*, vol. 1, pp. 203 – 9, 2002.

[52] R. Blanco, "Learning Bayesian networks from data with factorisation and classification purposes. applications in biomedicine." Ph.D. dissertation, University of the Basque Country, Spain, 2005.

[53] A. R. Masegosa, "Model of supervised classification: Applications to genomics and information retrieval." Ph.D. dissertation, Granada University, 2009.

[54] A. Cano, J. G. Castellano, A. Masegosa, and S. Moral, "Aplicación de un modelo naïve bayes gaussiano con selección de variables al análisis de datos de expresión genética," in *Actas de las VI Jornadas de Transferencia Tecnológica de Inteligencia Artificial (TTIA'2005).*, 2005, pp. 81–88.

[55] H.-H. Chang and M. Ramoni, "Transcriptional network classifiers," *BMC Bioinformatics*, vol. 10, no. 9, p. S1, 2009.

[56] W. Stacklies *et al.*, "pcaMethods—a bioconductor package providing PCA methods for incomplete data," *Bioinformatics*, vol. 23, no. 9, pp. 1164–1167, 2007.

[57] E. Wit and J. Mcclure, *Statistics for microarrays*. John Wiley & Sons, 2004.

[58] K. Sachs *et al.*, "Causal protein-signaling networks derived from multiparameter single-cell data," *Science*, vol. 308, no. 5721, pp. 523–529, 2005.

[59] M. Bansal, V. Belcastro, A. Ambesi-Impiombato, and D. di Bernardo, "How to infer gene networks from expression profiles," *Molecular Systems Biology*, vol. 3, 2007.

[60] S. A. Kauffman, "Metabolic stability and epigenesis in randomly constructed genetic nets," *Journal of Theoretical Biology*, vol. 22, no. 3, pp. 437–467, 1969.

[61] N. Friedman, M. Linial, I. Nachman, and D. Pe'er, "Using Bayesian networks to analyze expression data." *Journal of Computational Biology*, vol. 7, no. 3-4, pp. 601–620, 2000.

[62] U. Fayyad and K. Irani, "Multi-valued interval discretization of continuous-valued attributes for classification learning," in *Proceeding of the 13th International joint Conference on Artificial Inteligence*, 1993, pp. 1022–1027.

[63] J. R. Quinlan, *C4.5: Programs for machine learning.* Morgan Kaufmann Publishers Inc., 1993.

[64] Elvira Consortium, "Elvira: An environment for probabilistic graphical models," in *Procs. of the 1st European Workshop on Probabilistics Graphical Models*, 2002, pp. 222–230.

[65] D. Heckerman, D. Geiger, and D. M. Chickering, "Learning Bayesian networks: The combination of knowledge and statistical data." *Machine Learning*, vol. 20, no. 3, pp. 197–243, 1995.

[66] I. J. Good, *The estimation of probabilities*. MIT Press, 1965.

[67] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *Journal of the American Statistical Association*, no. 32, pp. 675–701, 1937.

[68] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.