Wrapper- and Ensemble-Based Feature Subset Selection Methods for Biomarker Discovery in Targeted Metabolomics

Holger Franken[1], Rainer Lehmann[2,3], Hans-Ulrich Häring[2,3],
Andreas Fritsche[2,3], Norbert Stefan[2,3], and Andreas Zell[1]

[1] Center for Bioinformatics (ZBIT), University of Tübingen,
D-72076 Tübingen, Germany
`holger.franken@uni-tuebingen.de`
[2] Division of Clinical Chemistry and Pathobiochemistry (Central Laboratory),
University Hospital Tübingen, D-72076 Tübingen, Germany
[3] Paul-Langerhans-Institute Tübingen,
Member of the German Centre for Diabetes Research (DZD),
Eberhard Karls University Tübingen, Tübingen, Germany

Abstract. The discovery of markers allowing for accurate classification of metabolically very similar proband groups constitutes a challenging problem. We apply several search heuristics combined with different classifier types to targeted metabolomics data to identify compound subsets that classify plasma samples of insulin sensitive and -resistant subjects, both suffering from non-alcoholic fatty liver disease. Additionally, we integrate these methods into an ensemble and screen selected subsets for common features. We investigate, which methods appear the most suitable for the task, and test feature subsets for robustness and reproducibility. Furthermore, we consider the predictive potential of different compound classes. We find that classifiers fail in discriminating the non-selected data accurately, but benefit considerably from feature subset selection. Especially, a Pareto-based multi-objective genetic algorithm detects highly discriminative subsets and outperforms widely used heuristics. When transferred to new data, feature sets assembled by the ensemble approach show greater robustness than those selected by single methods.

1 Introduction

1.1 Background

Non-alcoholic fatty liver disease (NAFLD) is associated with insulin resistance, but can also be detected in insulin sensitive subjects. Insulin resistant individuals with NAFLD have a very high risk of developing type 2 diabetes (T2D) at an early stage and therefore require a timely initiation of ongoing, preventive intervention and drug therapy. In contrast, it is currently assumed that insulin sensitive people with NAFLD have a low risk of developing T2D, i.e., metabolic

control at longer intervals is sufficient [17,21]. In this work, we examine different methods for the discovery of novel metabolite biomarkers to discriminate benign versus malign fatty liver in prediabetic subjects.

Biomarkers are defined as a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention [1]. For the discovery of novel biomarkers, the young discipline of metabolomics has received increased attention in recent years. It measures small molecules or metabolites contained in cells, tissues, or fluids involved in metabolism to reveal information about physiological processes. These processes may be influenced by both, genetic pre-disposition and environmental factors such as nutrition, exercise or medication [5]. Modern high-throughput techniques are capable of performing great numbers of measurements to produce datasets which stand up to statistical scrutiny. At the same time, the amounts of data generated are too voluminous to be interpreted by hand and therefore require dimensionality reduction. Thus, data mining and bioinformatics techniques are essential to identify and verify markers that are biochemically interpretable and biologically relevant. In contrast to projection- or compression-based methods for dimensionality reduction, like Principal Component Analysis (PCA) or the use of information theory, feature selection methods select a subset of variables instead of altering them. Thus, they preserve the original meaning of the variables and facilitate interpretation by a domain expert [19].

Targeted profiling schemes are used to quantitatively screen for known compounds, which depict relevant metabolic pathways of the investigated conditions. In such an approach, features reflect calculated concentrations of predefined metabolites. Data mining techniques are affected by factors such as noise, redundancy, and relevance in the experimental data. Feature selection is therefore focused on the process of identifying and removing as much irrelevant or redundant information as possible [15]. In this context, feature selection techniques can be organized into filter and wrapper methods.

Filter methods, in most cases, compute a feature relevance score and discard low-scoring features. The remaining subset of features serves as input to a classification algorithm. Filter techniques easily scale to very high-dimensional datasets, are computationally fast and are independent of the classification algorithm. A common disadvantage of filter methods is that most proposed techniques are univariate; i.e., each feature is considered separately, ignoring feature dependencies [19].

In wrapper methods, a search procedure generates and evaluates various subsets of features. A classification algorithm is trained and tested to evaluate specific subsets. To search the space of all feature subsets, the search algorithm is *wrapped* around the classification model. This problem is NP-hard, as the space of feature subsets grows exponentially with the number of features. Heuristic methods are therefore used to search for an optimal subset. Advantages of wrapper approaches include the interaction between feature subset search and

model selection and the ability to take feature dependencies into account. These methods have demonstrated their utility in various studies [19,12].

But the challenge in the search for biomarkers is not only to identify highly discriminative feature sets from a single data set. Subsets that are robust in the sense that they yield good results in the classification of new data sets, are particularly interesting. We assume that features that often contribute to highly predictive subsets independent of the method used for their selection can form a particularly robust pattern. Thus we applied a voting scheme (see Sec. 2.3) that integrates the single selection procedures into an ensemble.

1.2   State of the Art

Large numbers of features and limited sample sizes are typical drawbacks in the search for biomarkers in human studies. These give rise to hardly classifiable data sets that necessitate dimensionality reduction. Therefore, multivariate approaches such as PCA or partial-least-squares discriminant analysis (PLS-DA) are common. These techniques represent the extracted information as a set of new variables called components [7]. As these variables are not part of the original data set such methods lack in interpretability. Machine-learning algorithms are a more recent class of multivariate analysis techniques. Besides their desirable characteristic of preserving the original variables, they have demonstrated superior predictive accuracy than PLS-DA and PCA [14].

Several heuristic search algorithms have previously been applied as wrappers. Genetic algorithms have been shown to be a good choice for finding small feature subsets with high discriminatory power [23]. Furthermore, a modified form of SVMs was applied, and the unified maximum separability analysis (UMSA) algorithm was introduced for proteomic profiling [8,11]. Ressom et. al. proposed particle swarm optimization (PSO) for biomarker selection in mass spectrometric profiles of peptides and proteins [18].

In this work, we apply several search heuristics combined with various types of classifiers. The goal is to find feature subsets in different targeted metabolomics data sets that are able to classify plasma samples of insulin sensitive and insulin resistant subjects, both suffering from non-alcoholic fatty liver disease. This data set holds additional challenges as the group of sample donors was particularly singled out to be very homogeneous.

2   Methods

2.1   Plasma Samples

Plasma samples of 40 adults with NAFLD (20 insulin sensitive and 20 -resistant subjects) were analysed by a targeted metabolomics approach. All individuals were intensively phenotyped as part of the Tübingen Lifestyle Intervention Program (TULIP) and considered healthy according to physical examination and routine laboratory tests. Plasma samples were taken before and after a nine

month lifestyle intervention including dietary counseling and increased physical activity. In this work, *baseline* and *follow-up* denote samples and data acquired before and after the lifestyle intervention, respectively.

2.2   Data Acquisition

Biocrates (Innsbruck, Austria) measured the concentrations of 247 compounds in EDTA-plasma by targeted IDQ. This targeted metabolomics analytical platform combines flow injection, liquid chromatography, and gas chromatography mass spectrometric approaches. The applied instruments were an API 4000 QTrap tandem MS, a 7890 GC, and a 5795 MSD (Agilent, Waldbronn, Germany).

We considered a measured concentration to be reliable if it exceeds the noise level by at least a factor of three. To ensure validity we only considered metabolites that contained at least 70 % reliable measurements. This restriction led to an exclusion of 69 metabolites from the data set, leaving 178 compounds for the data analyses (21 amino acids, 21 acylcarnitines, 5 bile acids, 37 free fatty acids, 15 sphingomyelins, 70 phosphatidylcholines, 9 lysophosphatidylcholines).

2.3   Feature Subset Selection

To perform feature selection we implemented a modular Java[TM] software environment. This environment integrates the optimization framework EvA2 [9] and classification algorithms implemented in WEKA [4].

EvA2 is a comprehensive metaheuristic optimization framework with emphasis on Evolutionary Algorithms (EA) implemented in Java. It integrates several derivative-free, preferably population-based, optimization methods. From these methods, we applied the following algorithms as wrappers:

– a standard Genetic Algorithm (GA) [6]
– a Multi Objective Genetic Algorithm (MOGA) [3]
– the Cluster-Based Niching EA (CBN-EA) [22]
– Population-Based Incremental Learning (PBIL) [10]
– Hill Climbing (HC) [16]
– Monte Carlo Search (MCS)

The MOGA applies a Pareto-based ranking scheme to simultaneously optimize two objectives: the first is to maximize the classification performance, the second is to minimize the number of features contained in the subset.

From WEKA, we applied different classifier types:

– k-nearest-neighbor (kNN), an instance-based learner, which compares each new instance to existing ones using a distance metric
– K*NN, another instance-based method using an entropic distance measure [2]
– Naive Bayes, a classifier based on Bayes' formula for conditional probabilities, depending on the assumption that attributes are independent
– the J4.8 decision tree, a reimplementation of the C4.5 decision tree which has been shown to have a very good combination of error rate and speed [13]

- random forest, a metalearner which baggs ensembles of random trees
- linear SVM, a maximum-margin based classifier we chose due to its amenities concerning interpretability compared to nonlinear models [20]

While we are interested in features, that provide good classification performance, this is not the sole criterion of our search. We also want our selected features to be interpretable by a domain expert. Hence, we focused on classification methods that allow conclusions to be drawn about the involved features.

The modular structure of our software environment allows us to combine any of the aforementioned search strategies with an arbitrary classification method and thereby design a multitude of wrapper-based feature subset selection procedures. Each procedure performs a search for a good subset using the selected classifier to evaluate feature subsets.

For the evaluation of each feature subset, we performed a stratified nested cross validation consisting of an inner loop for model selection of optimal parameters and an outer loop for external validation. The optimal parameter combination was determined in the inner loop within a two-fold cross-validation. The performance of the selected parameters was then evaluated in the outer loop using three-fold cross-validation according to the area under the receiver operating characteristic curve (AUC). We carried out this validation scheme five times to avoid bias induced by the random number generator and computed the average AUC.

We combined each of the listed search strategies from EvA2 with each classifier to determine the feature subset that produces the best classification performance. Thereby we also assessed the question of the optimal wrapper-classifier combination for this task.

Additionally, we regarded the tested wrapper-classifier combinations as an ensemble and screened their selected subsets for common features. We implemented a voting scheme to extract features that were selected by many of the search procedures assuming that these form a particularly robust pattern. For this purpose, we compared all feature subsets selected by the individual procedures by their average AUC. Those with an AUC less than the median of all subsets were discarded. Thus, we only incorporate feature subsets that show a comparatively high predictive potential. Each of these subsets votes for the features it contains. We weight the votes by the average AUC of the voting subset and sum them into a voting score for every feature. Hence, votes from highly predictive subsets have a greater impact than others. From these voting scores we calculated the 90 % quantile as a lower bound. All features that received a voting score exceeding that bound were included in the consensus feature subset.

2.4   Experimental Design

From the measurements described in Sec. 2.2, we generated several different data sets. Firstly, we formed a data set from each of the listed compound classes individually. Secondly, we integrated all metabolites into one dataset to select subsets containing features from all groups under consideration. We performed
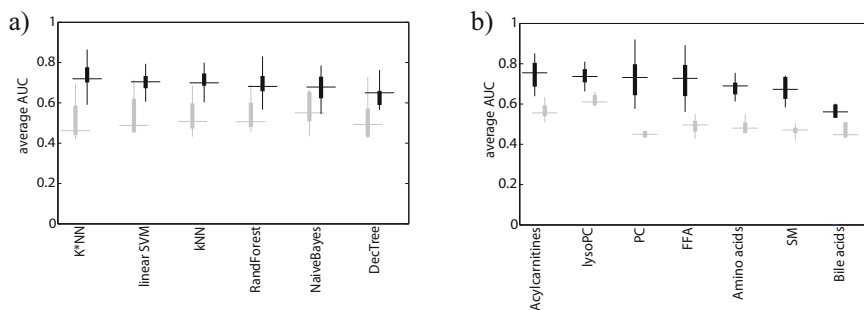
Fig. 1. Comparison of a) the classification performances of the applied wrappers and b) the predictive potential of the compound classes before (grey) and after (black) feature subset selection

this approach for the baseline as well as for the follow-up data. All data sets were mean-centered and scaled to unit variance. We applied each of the feature subset selection procedures described in Sec. 2.3 to these data sets to find subsets that accurately discriminate between insulin sensitive and insulin resistant subjects.

We analysed the results from different perspectives. The primary question is whether feature subset selection enhances classification performance at all. From this the question arises whether the applied procedures differ in their performance and which of them appears the most suitable for the task. Furthermore, we test our assumption that features that are selected by many different procedures form a particularly robust pattern. From the biological perspective we further consider the predicitve potential of the individual compound classes. We investigate whether these differ in their discriminative power and what impact the lifestyle intervention has on the different compound classes.

3   Results and Discussion

3.1   Benefit of Feature Subset Selection

First, we addressed the question whether the applied classification algorithms benefit from feature subset selection in terms of their classification accuracies. In Fig. 1a), grey boxes indicate the distribution of average AUCs each classifier achieved when applied to the complete data sets. Black boxes display the average AUCs when applied to the best selected feature subset, respectively. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually. Fig. 1b) shows the prediction accuracies based on the individual data sets. Again, the distributions of the classification results across all applied classifiers on the complete data sets are given by the grey boxes and black boxes illustrate the corresponding distributions of AUCs based on the best selected feature subsets from each data set. Both figures demonstrate that, in general, classification performances without
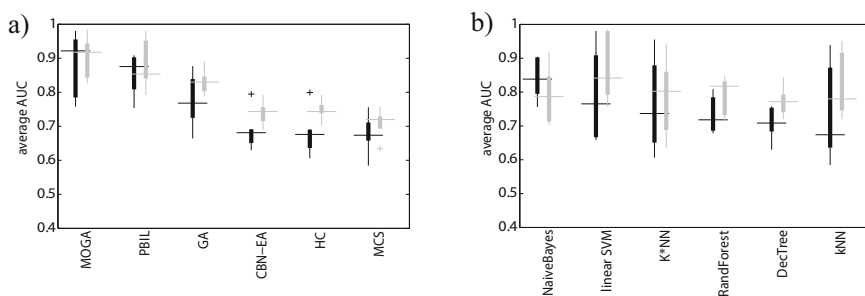
Fig. 2. Prediction accuracies achieved by the applied a) wrappers and b) classifiers on the baseline (black) and follow-up (grey) data sets

feature subset selection are close to an AUC of 0.5 and were significantly improved by the reduction of irrelevant information. All classifiers similarly benefit from the use of feature subset selection.

The individual compound classes differ in their gain in predictive potential. On the one hand, lysoPCs already classify the samples fairly well without feature subset selection and show a slight increase. On the other hand, phosphatidylcholines (PC), which show hardly any classification performance without feature subset selection, strongly benefit and even contain the most discriminative feature subsets afterwards.

3.2   Performance of Feature Subset Selection Procedures

We tested all feature subset selection procedures described in Sec. 2.3 on each individual compound class data set. The most discriminative subset selected from the baseline data consists of 10 phosphatidylcholines (PC). It yields an average AUC of 0.92 and was selected using MOGA and the K*NN classifier. Using the follow-up data sets, the best performing subset from a single compound class was again selected from the phosphatidylcholines. It also contains 10 metabolites and achieved an average AUC of 0.93. The employed classifier in this case was the linear SVM also wrapped by MOGA. The most discriminative subsets from the baseline and the follow-up data sets were selected from the same compound class, respectively, but they do not overlap with regard to individual compounds. Analogously, we analysed the data sets consisting of all considered metabolites. From the baseline data set, a subset of 19 compounds yields the best classification performance with an average AUC of 0.980. From the follow-up data, the most predicitve feature subset includes 23 metabolites and has an average AUC of 0.984. Both subsets were selected using MOGA as the wrapper and the linear SVM as the classifier. In this case, we found three compounds that were in both subsets.

In Fig. 2a), we compare the performance of the applied optimization heuristics. For each heuristic, it shows the distributions of average AUCs across all classifiers achieved by the best feature subset selected from all metabolites. Black
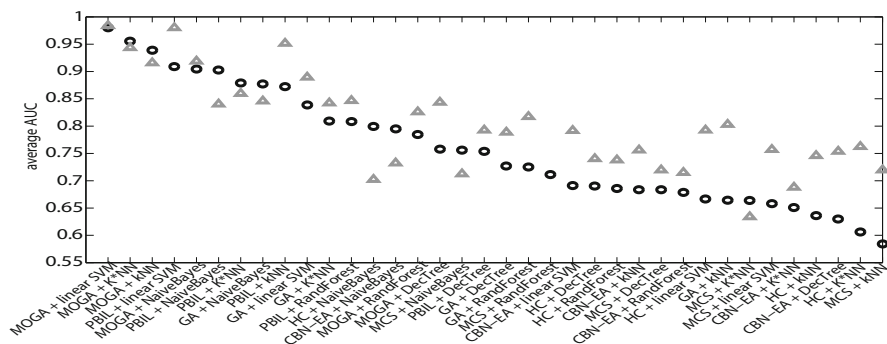
Fig. 3. Prediction accuracies achieved by the applied wrapper-classifier combinations on the baseline (black circles) and follow-up (grey triangles) datasets

boxes represent results obtained from the baseline data set and grey boxes depict results from the data acquired after lifestyle intervention. Entries are arranged in descending order of the median results per wrapper on the baseline data. MOGA and PBIL achieved the best results on both the baseline and the follow-up data sets. In both cases, they are followed by the GA. CBN-EA, HC and MCS achieve comparable results, which are considerably worse than those of MOGA and PBIL on both data sets.

Fig. 2b) displays a comparison of the classifiers applied to evaluate the tested feature subsets. For each classifier it shows the distribution of average AUCs across all wrappers achieved by the best feature subset selected from all metabolites. Again, black boxes depict results obtained from baseline data and grey boxes represent results from follow-up data. They are arranged in descending order of the median results per classifier on the baseline data. The results scatter strongly for most of the considered classifiers. The Naive Bayes classifier performs well on the baseline data, though the overall best feature set was selected using the linear SVM. Naive Bayes performs noticeably worse on the follow-up data sets. In this case the linear SVM ranks highest in terms of the median performance and overall best subset. The AUCs obtained by the J4.8 decision tree display the smallest variance, but no highly discriminative feature subsets could be selected using this classifier. The random forest shows a very similar behaviour, though its median performance is slightly better. However, a ranking of the performed classification algorithms is not revealing, due to the great variance in their results.

Fig. 3 gives an overview of the results achieved by each applied combination of classification algorithm and search heuristic. The presented average AUCs were obtained from the data sets including all considered compounds acquired before (black circles) and after lifestyle intervention (grey triangles). The data is sorted in descending order of the baseline results. The top three classification results were achieved using MOGA as the employed search heuristic. Among the top five classification results, MOGA can be found four times, followed by
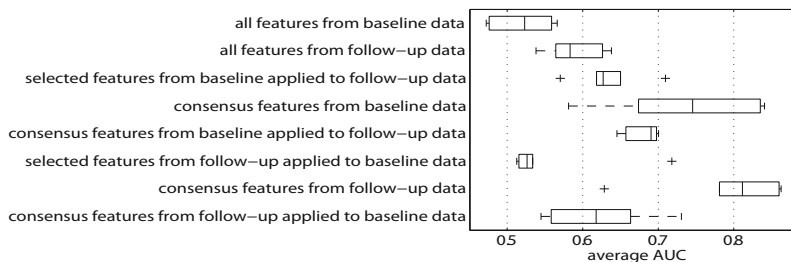
Fig. 4. Performance comparison of consensus feature sets with selected feature sets and complete data sets

PBIL, which appears four times among the top ten results. All feature sets selected using either of these heuristics are placed within the upper half of this arrangement. As previously mentioned, the results concerning the classifiers are not as clear. However, it can be observed that the achieved results of the applied combinations of wrappers and classifiers show significant correlation between baseline and follow-up data sets despite the scatter in the classifier performances (Spearman's $\rho = 0.73$, $p = 3.7e - 7$). Figures 2 and 3 show that, in the applied configurations, the choice of the search heuristic has a greater influence on the achieved results than the classifier applied for the evaluation of feature sets.

From the considered wrappers, MOGA delivered the most discriminative feature subsets and outperformed widely used heuristics such as the standard GA. This is an interesting finding because the MOGA has to deal with a tradeoff between the classifiaction accuracy and the size of the selected feature sets, whereas the GA is focused only on the accuracy. A possible explanation for this outcome might be that the said tradeoff prevents the MOGA from getting trapped in local optima and therefore it directs the search trajectory towards more generalizable solutions. This observation requires further investigation.

3.3 Ensemble Method

As described in Sec. 2.3, we assembled consensus feature sets from metabolites that were contained in many highly predictive subsets. From the baseline and follow-up data we obtained consensus sets that each contain 18 features. Seven features are in both of them. We tested the consensus sets for their predictive potential by using them as inputs for all of the applied classification algorithms. To measure their benefit, we compare the achieved classification results to those obtained using the data sets containing all metabolites. To evaluate the robustness of the consensus subsets, we applied the set that was assembled from the baseline analyses to the follow-up data and vice versa. Fig. 4 presents the results of these comparisons. In this figure, the term *selected features* refers to the best feature sets selected from the corresponding data as described in Sec. 3.2, *consensus features* denotes the consensus feature set as characterized in Sec. 2.3, respectively, and *all features* stands for the complete data sets.
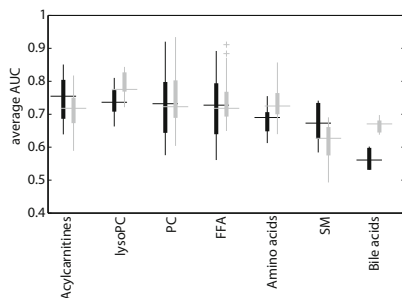
Fig. 5. Discriminative potential of compound classes at baseline (black) and follow-up (grey)

With regard to the benefit of the consensus feature sets, the results show that in both cases these sets have a significantly greater discriminative potential than the corresponding complete data sets. Furthermore, they also achieve significantly better results than the complete data sets, when applied to the data they were not derived from. In contrast, the *selected* feature sets did not perform considerably better than the complete data sets on the data, they were not selected from. The consensus feature sets do not quite attain the discriminative power of the best subsets selected by the single procedures. But when transferred from the baseline to the follow-up data sets or vice versa, the consensus sets prove to be more robust than those selected by single methods and reproducible results are obtained. This is also indicated by the finding, that the consensus feature sets obtained from baseline and follow-up data overlap in 7 out of 18 metabolites whereas the best subsets selected by single procedures, overlap by only 3 out of 19 and 23 compounds, respectively. The best subsets from the individual compound class data sets do not overlap at all.

3.4   Discriminative Potential of Compound Classes

Fig. 5 shows the distributions of classification performances that were achieved based on the individual compound class data sets (see Sec. 2.4) using all applied wrappers and classifiers. Black and grey boxes represent results from the baseline and the follow-up data, respectively. The entries are arranged in descending order by the median results from the baseline data sets. With the exception of the bile acids, no considerable differences exist in the results from the baseline data sets, with regard to the medians. In terms of the best selected feature subsets, however, the differences between the compound classes are more distinct. Highly discriminative subsets could be selected from the phosphatidylcholines (PC) and the free fatty acids (FFA) in the baseline and follow-up data. In general, the differences between the median results are greater at follow-up. LysoPCs, amino acids and bile acids show an increase in their discriminative power, whereas acylcarnitines and sphingomyelins decrease.

The selected feature subsets and actual metabolites will be analysed and discussed in more detail from the metabolic point of view in a separate publication.

4   Conclusions

The results demonstrate that highly discriminative subsets of interpretable features can be selected from otherwise hardly classifiable metabolomics data using a wrapper approach. Generally, in such an approach the choice of the wrapper has greater influence on the results than the applied classifier. It turned out that the J4.8 decision tree and random forests are not a good choice for the task, though. As a search strategy the Pareto-based MOGA proved to perform especially well and might be a useful tool in similar issues. With respect to transferability, however, the comparison and combination of several independently generated individual solutions revealed robust and reproducible subsets, which might become starting points for further investigation and test development for the early diagnosis of malign fatty liver in prediabetic subjects.

References

1. Atkinson, A., Colburn, W., DeGruttola, V., DeMets, D., Downing, G., Hoth, D., Oates, J., Peck, C., Schooley, R., Spilker, B., et al.: Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. Clinical Pharmacology & Therapeutics 69(3), 89–95 (2001)
2. Cleary, J., Trigg, L.: K*: An Instance-based Learner Using an Entropic Distance Measure. In: Proceedings of the 12th International Conference on Machine Learning (1995)
3. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation 6(2), 182–197 (2002)
4. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.: The WEKA data mining software: an update. ACM SIGKDD Explorations Newsletter 11(1), 10–18 (2009)
5. Harrigan, G., Goodacre, R.: Metabolic profiling: its role in biomarker discovery and gene function analysis. Springer, Netherlands (2003)
6. Holland, J.: Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor (1975)
7. Holmes, E., Nicholson, J., Nicholls, A., Lindon, J., Connor, S., Polley, S., Connelly, J.: The identification of novel biomarkers of renal toxicity using automatic data reduction techniques and PCA of proton NMR spectra of urine. Chemometrics and Intelligent Laboratory Systems 44(1), 245–255 (1998)

8. Koopmann, J., Zhang, Z., White, N., Rosenzweig, J., Fedarko, N., Jagannath, S., Canto, M., Yeo, C., Chan, D., Goggins, M.: Serum diagnosis of pancreatic adenocarcinoma using surface-enhanced laser desorption and ionization mass spectrometry. Clinical Cancer Research 10(3), 860 (2004)

9. Kronfeld, M., Planatscher, H., Zell, A.: The EvA2 optimization framework. In: Blum, C., Battiti, R. (eds.) LION 4. LNCS, vol. 6073, pp. 247–250. Springer, Heidelberg (2010)

10. Larranaga, P., Lozano, J.: Estimation of distribution algorithms: A new tool for evolutionary computation, vol. 2. Springer, Netherlands (2002)

11. Li, J., Zhang, Z., Rosenzweig, J., Wang, Y., Chan, D.: Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. Clinical Chemistry 48(8), 1296 (2002)

12. Li, T., Zhang, C., Ogihara, M.: A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression. Bioinformatics (Oxford, England) 20(15), 2429–2437 (2004)

13. Lim, T.: A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-Three Old and New Classification Algorithms. Machine Learning 40, 203–228 (2000)

14. Mahadevan, S., Shah, S.L., Marrie, T.J., Slupsky, C.M.: Analysis of metabolomic data using support vector machines. Analytical Chemistry 80(19), 7562–7570 (2008)

15. Masseglia, F., Poncelet, P., Teisseire, M.: Successes and new directions in data mining. Information Science Publishing (2008)

16. Mitchell, M., Holland, J., Forrest, S.: When will a genetic algorithm outperform hill climbing? Ann Arbor 1001, 48109

17. Petersen, K., Dufour, S., Befroy, D., Lehrke, M., Hendler, R., Shulman, G.: Reversal of Nonalcoholic Hepatic Steatosis, Hepatic Insulin Resistance, and Hyperglycemia by Moderate Weight Reduction in Patients With Type 2 Diabetes. Metabolism 54, 603–608 (2005)

18. Ressom, H.W., Varghese, R.S., Abdel-Hamid, M., Eissa, S.A.L., Saha, D., Goldman, L., Petricoin, E.F., Conrads, T.P., Veenstra, T.D., Loffredo, C.A., Goldman, R.: Analysis of mass spectral serum profiles for biomarker selection. Bioinformatics (Oxford, England) 21(21), 4039–4045 (2005)

19. Saeys, Y., Inza, I.N., Larrañaga, P.: A review of feature selection techniques in bioinformatics. Bioinformatics (Oxford, England) 23(19), 2507–2517 (2007)

20. Schölkopf, B., Smola, A.: Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond (Adaptive Computation and Machine Learning), 1st edn. The MIT Press (2001)

21. Stefan, N., Kantartzis, K., Häring, H.U.: Causes and metabolic consequences of Fatty liver. Endocrine Reviews 29(7), 939–960 (2008)

22. Streichert, F., Stein, G., Ulmer, H., Zell, A.: A clustering based niching EA for multimodal search spaces. In: Liardet, P., Collet, P., Fonlupt, C., Lutton, E., Schoenauer, M. (eds.) EA 2003. LNCS, vol. 2936, pp. 293–304. Springer, Heidelberg (2004)

23. Zou, W., Tolstikov, V.: Probing genetic algorithms for feature selection in comprehensive metabolic profiling approach. Rapid Communications in Mass Spectrometry 22(8), 1312–1324 (2008)