

# Multi-Objective Genetic Programming for Visual Analytics

Ilknur Icke<sup>1</sup> and Andrew Rosenberg<sup>1,2</sup>

<sup>1</sup>The Graduate Center  
The City University of New York  
365 Fifth Avenue 10016  
New York, NY  
iicke@gc.cuny.edu

<sup>2</sup>Queens College  
The City University of New York  
65-30 Kissena Blvd. Flushing  
NY 11367-1575  
andrew@cs.qc.cuny.edu

**Abstract.** Visual analytics is a human-machine collaboration to data modeling where extraction of the most informative features plays an important role. Although feature extraction is a multi-objective task, the traditional algorithms either only consider one objective or aggregate the objectives into one scalar criterion to optimize. In this paper, we propose a Pareto-based multi-objective approach to feature extraction for visual analytics applied to data classification problems. We identify classifiability, visual interpretability and semantic interpretability as the three equally important objectives for feature extraction in classification problems and define various measures to quantify these objectives. Our results on a number of benchmark datasets show consistent improvement compared to three standard dimensionality reduction techniques. We also argue that exploration of the multiple Pareto-optimal models provide more insight about the classification problem as opposed to a single optimal solution.

## 1 Introduction

The amount of data poses both theoretical and practical challenges for machine learning. According to the *curse of dimensionality* [4] theorem, the number of samples needed for a classification task increases exponentially as the number of dimensions (variables, features) increases. Moreover, irrelevant and redundant features might hinder classifier performance. On the other hand, it is costly to collect, store and process data. In exploratory analysis settings, high dimensionality prevents the users from exploring the data visually. Feature extraction is a two-step process that seeks suitable data representations that would help us overcome these challenges. Feature construction step creates a set of new features based on the original features and feature selection is the process of finding the best features amongst them. Feature selection techniques are divided into filter, wrapper and embedded methods depending on the assessment method [10].

In this paper, we focus on feature extraction methods for visual analytics in classification problems. Visual analytics is a human-machine collaborative approach to data modeling where visualization provides means to understand the structure of the data such as revealing the clusters formed within the observed data and the relationships between the features. Various linear (such as

principal components analysis (PCA), multiple discriminants analysis (MDA), exploratory projection pursuit) and non-linear (such as multidimensional scaling (MDS), manifold learning, kernel PCA, evolutionary constructive induction) techniques have been proposed for dimensionality reduction and visualization. Traditionally, these algorithms aim to optimize a scalar objective and return a single optimal solution. However, recently it has been claimed that the analysis of multiple Pareto-optimal data models provides more insight about the classification problem as opposed to a single optimal solution [13].

Our algorithm is a novel adaptive approach to feature extraction for visual analytics that consists of Pareto-based multi-objective evolutionary constructive induction for feature construction and a hybrid filter/wrapper method for feature selection. Genetic programming approaches for feature selection/extraction in various classification tasks have been reported in the literature ([14, 17, 16, 9, 6]) where a single or scalar objective is optimized. A multi-objective method that optimizes accuracy of a simple classifier along with tree size as the complexity measure was described in [21]. Our algorithm differs from these methods in the sense that we aim to optimize human interpretability and discrimination power simultaneously and study visualization as a tool for interpretability. Our previous work studied a single objective (classifier accuracy) genetic programming method for data visualization ([12]) and a preliminary multi-objective implementation has been discussed in [11]. This paper further extends our work into a multi-objective setting with various measures for interpretability and also proposes ideas for model selection by analyzing the multiple Pareto-optimal solutions.

The remaining sections are organized as follows: section 2 introduces the MOG3P algorithm. Section 3 explains the optimization criteria used in a number of standard dimensionality reduction techniques versus the proposed criteria for the MOG3P algorithm. Experiment results on a number of benchmark datasets are reported in section 4. Section 5 presents the conclusions and future work.

## 2 The Multi-Objective Genetic Programming Projection Pursuit (MOG3P) Algorithm

We utilize the genetic programming (GP) framework in order to simultaneously evolve data transformation functions that would project the input data into a lower dimensional representation for visualization and classification (figure 1).

Each data transformation function is represented as an expression tree which is made up of a number of functions over the initial features and represents a 1D projection of the data. Each individual contains two such expression trees that evolve independently and generate a 2D view of the N-dimensional dataset. The algorithm is named Multi-Objective Genetic Programming Projection Pursuit (MOG3P) since it searches for *interesting* low dimensional projections of the dataset where the measure of interestingness consists of three equally important objectives. These objectives are: 1) classifiability: the generated data representation should increase the performance of the learning algorithm(s), 2) visual interpretability: easily identifiable class structures on 2D scatterplots, 3) semantic interpretability: the relationships between the original and evolved features should be easy to comprehend.

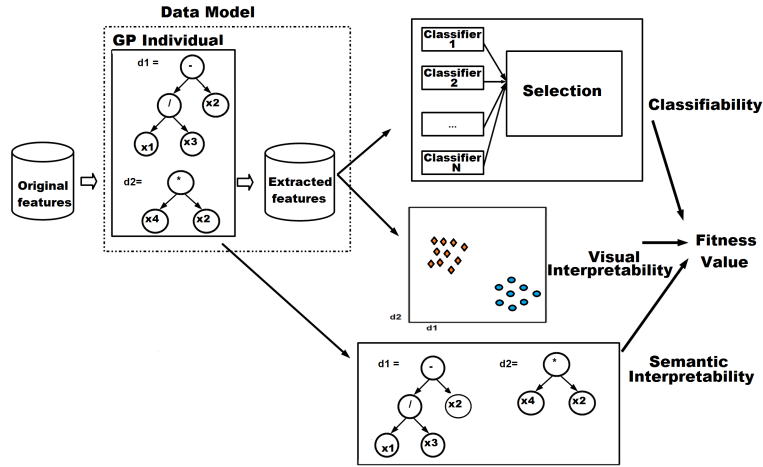


Fig. 1. Components of the MOG3P algorithm

### 3 Feature Extraction Criteria

In this section, we briefly cover the criteria utilized by the three standard dimensionality reduction methods and our MOG3P algorithm. Each standard method presented here optimizes a single criterion whereas the MOG3P utilizes multiple objectives simultaneously.

#### 3.1 Standard Methods

The following three methods are amongst the most widely used dimensionality reduction techniques [5] for data visualization:

**Principal Components Analysis (PCA).** PCA finds a lower dimensional representation such that maximum variation in the data is still retained. More specifically, 2-dimensional PCA representation can be generated by projecting the data onto the two eigenvectors of the data covariance matrix corresponding to the largest two eigenvalues. Each new feature is a linear combination of the original features.

**Multidimensional Scaling (MDS).** MDS algorithm searches for a mapping of the input data onto a  $k$ -dimensional space that preserves the inter-point distances. Given a matrix of inter-point distances between the original data points  $(\mathbf{x}_i, \mathbf{x}_j)$  where  $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$ , a variant of this algorithm named Sammon's mapping minimizes:  $\sum_{i \neq j} \frac{(d_{ij} - \|\mathbf{z}_i - \mathbf{z}_j\|)^2}{d_{ij}}$  where  $\mathbf{z}_i, \mathbf{z}_j$  are the new feature vectors in  $k$ -dimensional space. This algorithm does not explicitly define functions that transform the original features into the new features.

**Multiple Discriminants Analysis (MDA).** For a  $c$ -category classification problem, MDA finds a  $c - 1$  dimensional representation of the data such that the separation between the categories is maximized. This is achieved by maximizing the ratio of between-class scatter matrix to within-class scatter matrix. Similar to PCA, each new feature is a linear combination of the original features.

### 3.2 MOG3P Multi-Objective Criteria

The MOG3P algorithm aims to optimize the discriminative power and interpretability of the generated features for data classification problems. We examine the discriminative power under the classifiability criterion where the goal is to find a data representation that will increase the performance of the learning algorithm(s). For interpretability, we define two separate criteria.

**Classifiability.** We aim to optimize discriminative power of the constructed features by maximizing the classification accuracy. The classifiability criterion can be computed in a variety of ways, including measuring the accuracy of a single classifier, and aggregating the performance of an ensemble of classifiers. In this paper, we measure classifiability by the accuracy of a single classifier selected at random from an ensemble. This way, we aim to ensure that the features will not be biased towards one classification paradigm.

**Visual Interpretability.** For a data classification task, our goal is to find visualizations (scatterplots) of the dataset such that class structures are easily identifiable. Humans can easily judge the quality of a visualization regardless of the shape of the class boundaries (linear or nonlinear) or the shapes of clusters formed by members of the same class. Since there are a large number of candidate visualizations, it is not feasible to utilize human judgement. A number of numerical measures have been proposed to assess the quality of groupings automatically. Although the ultimate automated quality measure is not known, recent user studies in visual analytics literature have shown that measures of separation and compactness reflect human perception to some extent ([18, 19]). In this paper, we utilize the following measures used in exploratory projection pursuit ([8], [15]):

- *LDA Index* ( $I_{LDA}$ ) is based on Fisher’s discriminant analysis.  $I_{LDA} = \frac{|W|}{|W+B|}$   $B = \sum_{i=1}^k n_i (\bar{V}_i - \bar{V}_{..})(\bar{V}_i - \bar{V}_{..})'$  and  $W = \sum_{i=1}^k \sum_{j=1}^{n_i} (V_{ij} - \bar{V}_i)(V_{ij} - \bar{V}_i)'$  where  $V_{ij}$  are data points,  $\bar{V}_i$  and  $\bar{V}_{..}$  are group and dataset centroids,  $n_i$  is the number of points in group  $i$ ,  $B$  is the between-group sum of squares and  $W$  is the within-group sum of squares. Smaller values of  $I_{LDA}$  indicate more compact and better separated class structures.
- *C Index* ( $I_C$ ) is a cluster validation index:  $I_C = \frac{SD - SD_{min}}{SD_{max} - SD_{min}}$  where  $SD$  is the total sum, for all classes, of pairwise distances between samples of the same class (total  $p$  distances),  $SD_{min}$  and  $SD_{max}$  are the sums of  $p$  smallest/largest pairwise distances across the whole dataset. Smaller values of  $I_C$  indicate more compact and better separated class structures.
- *Davies-Bouldin Index* ( $I_{DB}$ ) is a measure of compactness and well separation:  $I_{DB} = \frac{1}{n} \sum_{i=1}^n \min_{i \neq j} \left\{ \frac{\delta(X_i, X_j)}{\Delta(X_i) + \Delta(X_j)} \right\}$  where  $\Delta(X_i)$  is intra-class distance for class  $i$  and  $\delta(X_i, X_j)$  is inter-class distance for classes  $i$  and  $j$ . Smaller values of  $I_{DB}$  indicate more compact and better separated class structures.
- *Dunn’s Index* ( $I_{Dunn}$ ) is a measure of compactness and well separation:  $I_{Dunn} = \min_{1 \leq i \leq n} \left\{ \min_{1 \leq j \leq n, j \neq i} \left\{ \frac{\delta(X_i, X_j)}{\max_{1 \leq k \leq n} \{\Delta(X_k)\}} \right\} \right\}$  where  $\delta, \Delta$  are defined as above. Smaller values of  $I_{Dunn}$  indicate more compact and better separated class structures.

**Semantic Interpretability.** The transformation functions (GP expressions) represent relationships between the original and the extracted features. Complex relationships are harder to comprehend and generalize. The most common measure of complexity used in previous research is the expression size without considering the nature of the data transformation functions. For example, a nonlinear relationship can be considered more semantically complex than a linear relationship between the features. The use of order of non-linearity as a complexity measure has been proposed in [20]. In this paper, we experiment with three measures for complexity: 1) expression size ( $I_{TS}$ ), 2) expressional complexity ( $I_{EC}$ ) which is the total number of nodes in all subtrees (from [20]), 3) a weighted version of the expressional complexity ( $I_{WEC}$ ) where each arithmetic operator is assigned a weight related to its linearity/non-linearity. An expression is considered more complex if the nonlinear operators are closer to the root.

## 4 Experiments

In this paper, we report results on a number of well-known datasets (table 1) from the data mining literature. We first create 2D representations of each dataset using principal components analysis (PCA), multidimensional scaling (MDS) and multiple discriminants analysis (MDA).

**Table 1.** Datasets

Name	#features	# samples	# classes
Bupa Liver Disorders (BUPA) [2]	6	345	positive:145, negative:200
Wisconsin Diagnostic Breast Cancer (WDBC) [2]	30	569	malignant:212, benign:357
Crabs [3]	5	200	4 (50 each)
Pima Indians Diabetes (PIMA) [2]	7	786	diabetic:286, non-diabetic:500

We report the 10-fold cross-validation performance of each classifier on these 2D representations as well as the original dataset. PCA and MDA construct new features based on linear transformations on the set of all original features and they can not uncover non-linear relationships. MDS does not construct an explicit mapping between the constructed and original features. In MOG3P, we utilize the ECJ toolkit( [7]) and experiment with WEKA( [1]) implementations of the following classifiers: Naive Bayes, Logistic Regression, SMO (support vector machine), RBF Network, IBk (k-Nearest Neighbors), Simple Cart and J48 (decision tree). Classifiability and visualization objectives can be used together or one at a time. Semantic interpretability objective is mandatory (table 2).

We implement a nested 10-fold cross validation scheme in order to assess generalization of the extracted features to unseen data. Each dataset is divided into 10-folds and 10 training and test set pairs are created. Each run uses one training set and only the optimal individuals are tested using the corresponding test set after each run. This process is repeated 10 times resulting in a total of 100 runs. In this paper, the quality of the two new features generated by each individual are compared to PCA, MDS and MDA methods in terms of the mean test set classification accuracy of the above mentioned classifiers.

Figures 2, 4, 6, 8 show visualizations of each dataset using the standard dimensionality reduction techniques. The visualizations reveal that the criteria optimized by the PCA and MDS algorithms (section 3) do not always generate data representations that facilitate classification. This effect is clearly evident on the Crabs dataset.

**Table 2.** MOG3P Settings

Population size/Generations	400 / 100
Crossover/Mutation operators	One-point subtree crossover/subtree mutation
Crossover/Reproduction/Mutation Probability	0.9 / 0.05 / 0.05
Multi objective selection algorithm	Strength Pareto Evolutionary Algorithm (SPEA2 [22])
Archive Size	100
Function set symbols	{+, -, *, <i>protected</i> /, <i>min</i> , <i>max</i> , <i>power</i> , <i>log</i> }
Ephemeral Random Constants (ERC)	[-1, 1]
Terminal symbols	Variables in the dataset, ERC
Classifiability Objective (C)	maximize hit rate (minimize misclassification rate) of a random classifier per individual $\{I_{Random}\}$
Visualization objective (V)	minimize $\{I_{LDA}, I_C, I_{DB}, I_{Dunn}\}$
Semantic objective (S)	minimize $\{I_{TS}, I_{EC}, I_{WEC}\}$
Cross validation	10 times 10-fold cross validation (total 100 runs)

Classifier performance on these new feature representations are compared to the performance on the original dataset (tables 3, 4, 5, 6). For each dataset, the result for a dimensionality reduction method is italicized if it shows significant improvement over the original features. In this paper, all significance tests are reported based on a two-tailed t-test at  $\alpha = 0.05$ . As it is evident from the results, dimensionality reduction using the standard methods does not guarantee significantly better classifier performance compared to using the original features. On the other hand, for all four datasets, MOG3P consistently finds better data representations using any of the four visualization criteria in conjunction with the classifiability and semantic interpretability objectives.

We also compare the effects of different combinations of classifiability and visualization objectives for each dataset. Figure 10 shows the MOG3P results using the tree size ( $I_{TS}$ ) as the semantic interpretability measure. The most significant finding here is that, using any visualization objective along with the classifiability objective consistently outperforms the cases where either of these objectives were used alone. The other two complexity measures  $I_{EC}$  and  $I_{WEC}$  demonstrate similar results to  $I_{TS}$ , therefore they are not included here.

## 5 Conclusions

We present a visual analytics approach to data modeling that seeks to simultaneously optimize the human interpretability and the discriminative power using genetic programming. We introduce visual interpretability as an explicit objective for exploratory data analysis in classification problems. Different measures of interpretability and discriminative power can easily be incorporated into the algorithm in a multi-objective manner without forcing the user to make a-priori decisions on relative importance of these measures. Our experiments show that using visualization objectives along with the discrimination objectives, MOG3P finds better data representations for classification problems compared to the three standard dimensionality reduction techniques or wrapper techniques used alone. Moreover, the MOG3P algorithm is a data model mining tool providing the users with multiple optimal models aiming to help them discover the set of most informative features or select a classification algorithm by examining classifier performance across multiple models. Model selection can be performed either by choosing one best model or an ensemble of good models. Our ongoing work is concentrating on interactive and automated methods to analyze the pareto optimal solutions for model selection on higher dimensional problems.

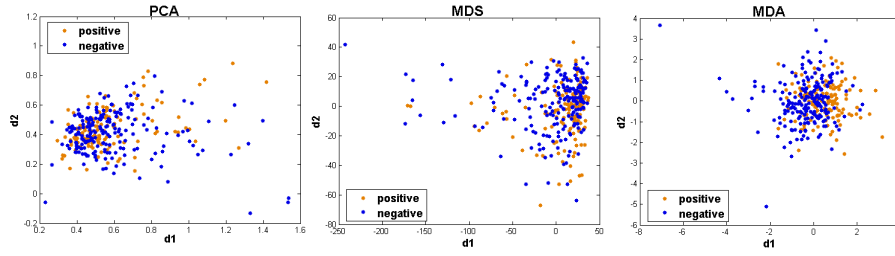


Fig. 2. PCA ( $I_{TS:46}$ ), MDS and MDA ( $I_{TS:46}$ ) visualizations of BUPA dataset

Table 3. Results on BUPA dataset

Classifier	PCA (2D)	MDS (2D)	MDA (2D)	All features	MOG3P (2D)	MOG3P (2D)	MOG3P (2D)	MOG3P (2D)
					$I_{Random}$ $I_{LDA,ITS}$	$I_{Random}$ $I_{C,ITS}$	$I_{Random}$ $I_{DB,ITS}$	$I_{Random}$ $I_{Dunn,ITS}$
N. Bayes	54.20	59.13	64.35	55.36	77.06	76.72	76.98	77.41
Logistic R.	57.97	58.84	68.12	68.12	77.3	77.50	76.98	77.82
SMO	57.97	57.97	59.42	58.26	74.04	73.30	75.3	75.85
RBF	61.45	59.13	64.06	64.35	77.85	78.05	77.44	77.41
kNN	55.65	57.68	60.29	62.9	78.16	78.8	77.87	77.52
CART	58.55	59.71	68.7	67.54	77.39	77.25	77.12	76.4
J48	57.97	56.82	69.28	68.7	76.95	76.97	76.66	76.08
Avg	57.68	58.47	64.89	63.60	<b>76.96</b>	<b>76.94</b>	<b>76.91</b>	<b>76.93</b>
(std)	(2.29)	(1.01)	(4.0)	(5.15)	<b>(5.63)</b>	<b>(5.71)</b>	<b>(5.45)</b>	<b>(5.53)</b>

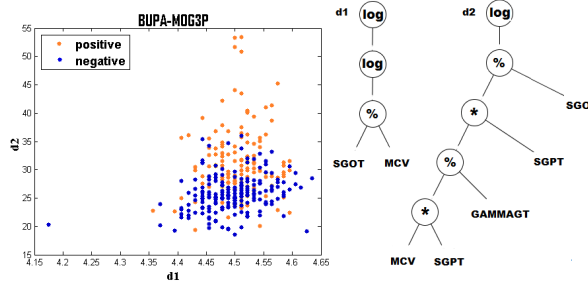


Fig. 3. One selected MOG3P model for the BUPA dataset ( $I_{TS:15}$ ). Variables: 6 numerical attributes related to liver disorders: mean corpuscular volume (MCV), alkaline phosphotase (ALPKHOS), alamine aminotransferase (SGPT), aspartate aminotransferase (SGOT), gamma-glutamyl transpeptidase (GAMMAGT) and the number of half-pint equivalents of alcoholic beverages drunk per day (DRINKS). This model uses 4 of the 6 original variables.

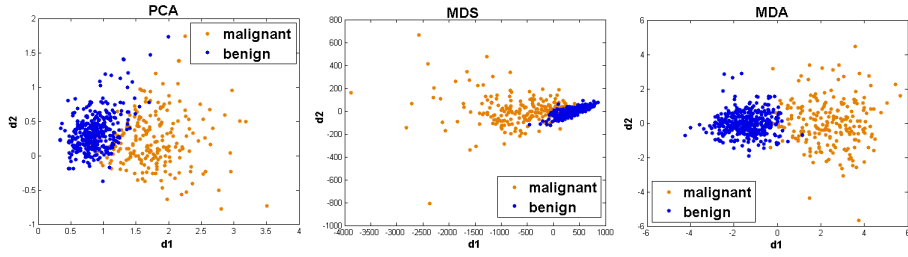


Fig. 4. PCA ( $I_{TS}:238$ ), MDS and MDA ( $I_{TS}:238$ ) visualizations of WDBC dataset

Table 4. Classification results on WDBC dataset

Classifier	PCA (2D)	MDS (2D)	MDA (2D)	All features	MOG3P (2D)	MOG3P (2D)	MOG3P (2D)	MOG3P (2D)
					$I_{Random}$ $I_{LDA,ITS}$	$I_{Random}$ $I_{C,ITS}$	$I_{Random}$ $I_{DB,ITS}$	$I_{Random}$ $I_{Dunn,ITS}$
N. Bayes	92.09	90.69	97.19	92.79	97.98	98.21	97.89	97.96
Logistic R.	94.38	93.15	97.72	94.73	97.89	98.28	98.3	98.37
SMO	93.32	87.52	96.84	98.07	97.61	97.87	97.54	97.66
RBF	94.73	92.79	98.24	93.5	98.21	98.28	98.08	98.21
IBk	91.21	90.86	96.66	94.73	98.17	98.40	98.3	98.3
CART	92.79	91.74	96.66	92.97	97.95	98.03	97.86	98.03
J48	92.97	92.44	97.19	93.67	97.98	98.14	97.89	98.01
Avg	93.07	91.31	97.21	94.35	<b>97.97</b>	<b>98.17</b>	<b>97.98</b>	<b>98.08</b>
(std)	(1.22)	(1.9)	(0.6)	(1.8)	(2.14)	(1.81)	(1.83)	(1.75)

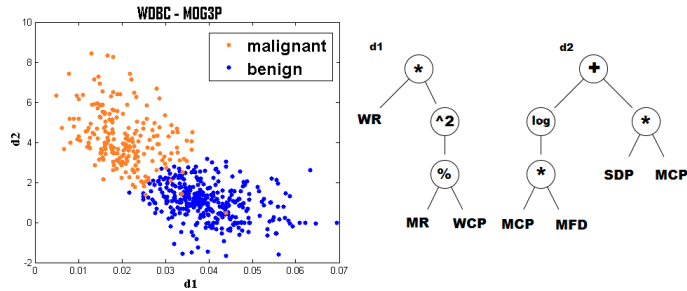


Fig. 5. One selected MOG3P model for the WDBC dataset ( $I_{TS}:14$ ). Variables (total of 30): the mean (M), standard deviation (SD) and worse (W) values of 10 measurements for breast cancer diagnosis: radius (R), texture (T), perimeter (P), area (A), smoothness (SM), compactness (CPT), concavity (C), concave points (CP), symmetry (SYM) and fractal dimension (FD). This model uses only 6 of the 30 original variables.



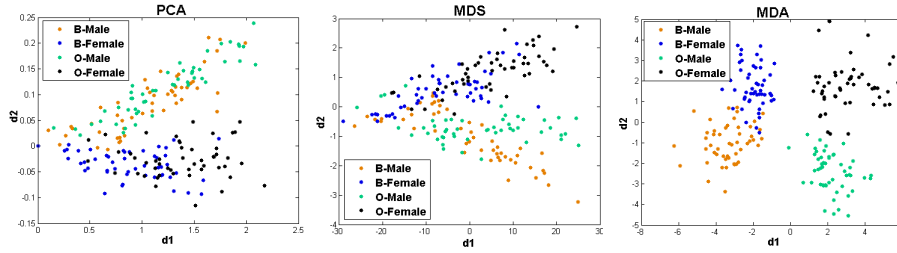


Fig. 6. PCA ( $I_{TS}:38$ ), MDS and MDA ( $I_{TS}:38$ ) visualizations of Crabs dataset

Table 5. Classification results on Crabs dataset

Classifier	PCA (2D)	MDS (2D)	MDA (2D)	All features	MOG3P (2D)	MOG3P (2D)	MOG3P (2D)	MOG3P (2D)
					$I_{LDA,ITS}$	$I_{C,ITS}$	$I_{DB,ITS}$	$I_{Dunno,ITS}$
N. Bayes	57.5	67	93.5	38	97.4	96.7	96.25	96.25
Logistic R.	59.5	63	94.5	96.5	98	96.95	96.6	96.25
SMO	54.5	59	94.5	63.5	97.15	96.45	96.15	95.55
RBF	67	69	96	49	97.65	96.75	96.7	96.25
IBk	57	67.5	93	89.5	97.5	97.35	96.8	96.65
CART	57.5	61	94	75.5	97.05	96.6	96.2	96.1
J48	56.5	59	92.5	73.5	97.35	97	96.25	96.45
Avg	58.5	63.64	94	69.36	97.44	96.83	96.42	96.24
(std)	(4.03)	(4.19)	(1.16)	(20.93)	(3.62)	(4.13)	(4.2)	(4.47)

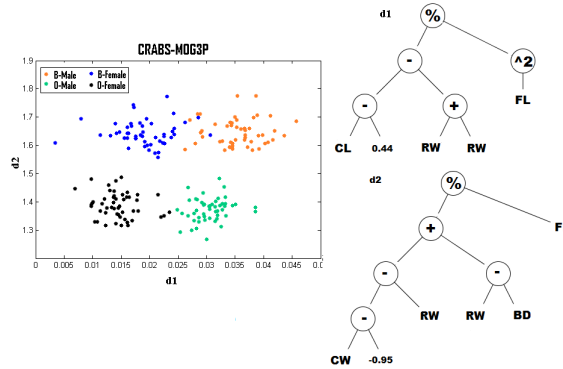


Fig. 7. One selected MOG3P model for the Crabs dataset ( $I_{TS}:21$ ). Variables: frontal lobe (FL), rear width (RW), carapace width (CW), claw length (CL) and body depth (BD). Feature d1 separates males from females, d2 separates the two species of crabs.

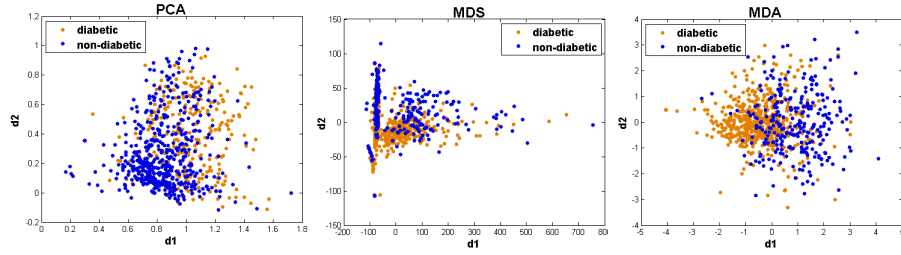


Fig. 8. PCA ( $I_{TS}:54$ ), MDS and MDA ( $I_{TS}:54$ ) visualizations of PIMA dataset

Table 6. Classification results on PIMA dataset

Classifier	PCA (2D)	MDS (2D)	MDA (2D)	All features	MOG3P (2D)		MOG3P (2D)	
					$I_{Random}$ $I_{LDA,ITS}$	$I_{Random}$ $I_{C,ITS}$	$I_{Random}$ $I_{DB,ITS}$	$I_{Random}$ $I_{Dunn,ITS}$
N. Bayes	71.22	74.22	76.43	76.30	82.03	81.90	81.68	81.86
Logistic R.	72.00	74.35	78.26	77.21	81.54	81.39	81.47	81.75
SMO	72.14	74.74	77.73	77.34	81.51	81.25	81.30	81.24
RBF	72.4	73.44	76.82	75.39	82.17	82.41	81.55	81.79
IBk	61.07	64.45	68.36	70.18	82.01	82.44	80.79	81.32
CART	68.49	73.7	75.39	75.13	81.79	81.96	81.15	81.41
J48	69.40	71.88	75.65	73.83	81.58	81.81	80.60	80.93
Avg	69.53	72.4	75.52	75.06	<b>81.81</b>	<b>81.87</b>	<b>81.22</b>	<b>81.47</b>
(std)	(4.01)	(3.62)	(3.32)	(2.48)	<b>(4.32)</b>	<b>(4.19)</b>	<b>(4.26)</b>	<b>(4.32)</b>

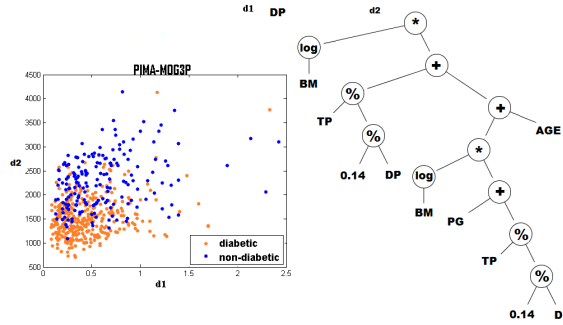


Fig. 9. One selected MOG3P model for PIMA dataset ( $I_{TS}:22$ ). Variables: times pre-grant (TP), plasma glucose concentration (PG), diastolic blood pressure (BP), triceps skin fold thickness (FT), body mass index (BM), diabetes pedigree function (DP), age (AGE). This model uses 5 out of the 7 original variables.

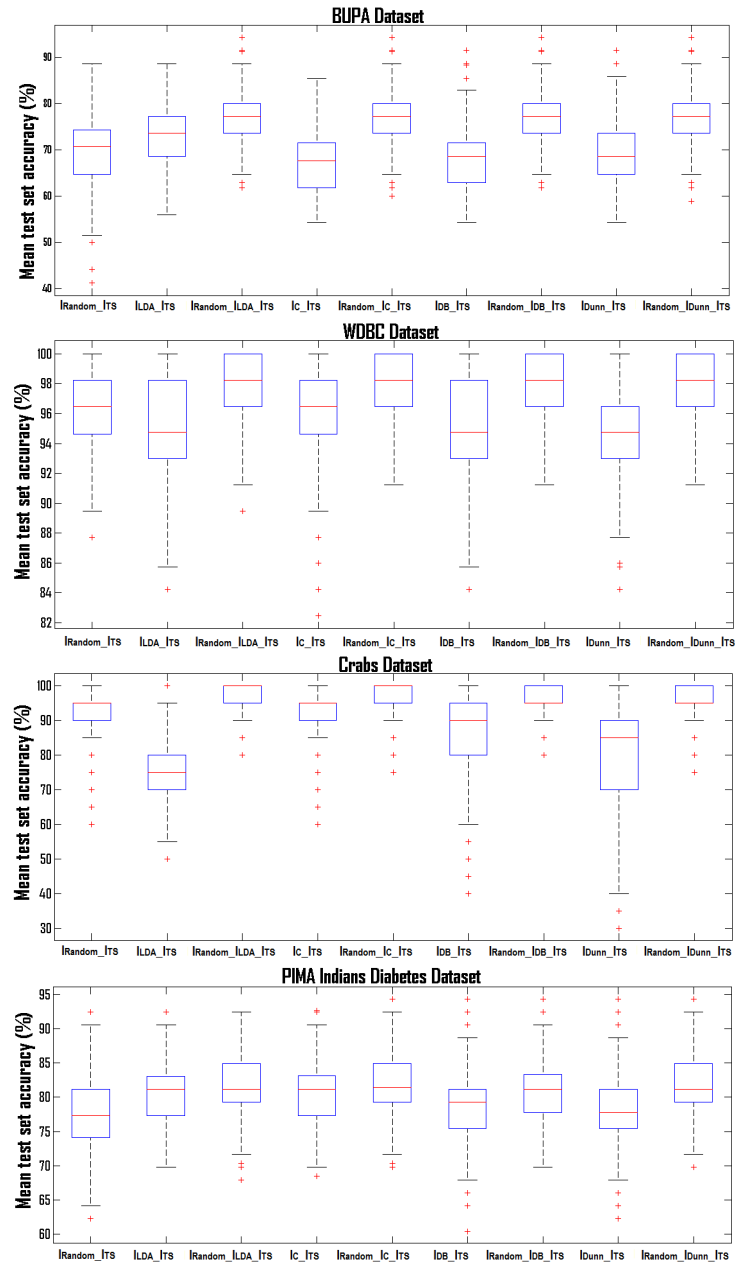


Fig. 10. Comparison of MOG3P objectives (semantic interpretability measure:  $I_{TS}$ ). Using classifiability and visual interpretability objectives together outperforms either objective alone.

## 6 Acknowledgements

This research was supported in part by a grant of computer time from the City University of New York's High Performance Computing Research Center under NSF Grants CNS-0855217 and CNS - 0958379.

## References

1. <http://www.cs.waikato.ac.nz/~ml/weka/>.
2. [www.ics.uci.edu/~mllearn](http://www.ics.uci.edu/~mllearn).
3. [www.stats.ox.ac.uk/pub/PRNN](http://www.stats.ox.ac.uk/pub/PRNN).
4. R. Bellman. *Adaptive Control Processes*. Princeton Univ. Press, 1961.
5. R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley, 2000.
6. C. Estèbanez, R. Aler, and J.M. Valls. A method based on genetic programming for improving the quality of datasets in classification problems. *International Journal of Computer Science and Applications*, 4:69–80, 2007.
7. Sean Luke et al. Ecj: A java-based evolutionary computation research system. <http://cs.gmu.edu/~eclab/projects/ecj/>.
8. C. Garcia-Osorio and C. Fyfe. Comparing exploratory projection pursuit artificial neural networks. In *3rd International Workshop on Practical Applications of Agents and Multiagent Systems, IWPAAMS 2004*, pages pp. 201–210, 2004.
9. Hong Guo and Asoke K. Nandi. Breast cancer diagnosis using genetic programming generated feature. *Pattern Recognition*, 39(5):980–987, 2006.
10. Isabelle Guyon, Steve Gunn, Masoud Nikravesh, and Lofti Zadeh, editors. *Feature Extraction, Foundations and Applications*. Series Studies in Fuzziness and Soft Computing, Physica-Verlag, Springer, 2006.
11. Ilknur Icke and Andrew Rosenberg. Multi-objective genetic programming projection pursuit for exploratory data modeling. In *New York Academy of Sciences, 5th Annual Machine Learning Symposium*. 2010.
12. Ilknur Icke and Andrew Rosenberg. Symbolic regression for dimensionality reduction. In *Late Breaking Abstracts, GECCO 2010*, 2010.
13. Yaochu Jin and Bernhard Sendhoff. Pareto-based multiobjective machine learning: An overview and case studies. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, 38(3):397–415, 2008.
14. Krzysztof Krawiec. Genetic programming-based construction of features for machine learning and knowledge discovery tasks. *Genetic Programming and Evolvable Machines*, 3(4):329–343, 2002.
15. Eun-Kyung Lee, Dianne Cook, Sigbert Klinke, and Thomas Lumley. Projection pursuit for exploratory supervised classification. *Journal of Computational and Graphical Statistics*, 14(4):831–846, December 2005.
16. Mohammed Muharram and George D. Smith. Evolutionary constructive induction. *IEEE Trans. on Knowl. and Data Eng.*, 17(11):1518–1528, 2005.
17. Fernando E. B. Otero, Monique M. S. Silva, Alex A. Freitas, and Julio C. Nievola. Genetic programming for attribute construction in data mining. In *Genetic Programming, Proceedings of EuroGP2003, volume 2610 of LNCS*, pages 384–393, 2003.
18. Mike Sips, Boris Neubert, John P. Lewis, and Pat Hanrahan. Selecting good views of high-dimensional data using class consistency. In Hans-Christian Hege, Ingrid Hotz, and Tamara Munzner, editors, *EuroVis 2009 : Eurographics/ IEEE-VGTC Symposium on Visualization 2009*, volume 28 of *Computer Graphics Forum*, pages 831–838, Berlin, Germany, 2009. Blackwell.
19. Andrada Tatu, Peter Bak, Enrico Bertini, Daniel A. Keim, and Jörn Schneidewind. Visual quality metrics and human perception: an initial study on 2d projections of large multidimensional data. In *AVI-Proceedings of AVI Working Conference on Advanced Visual Interfaces*, pages 49–56, 2010.
20. Ekaterina Vladislavleva, Guido Smits, and Dick den Hertog. Order of nonlinearity as a complexity measure for models generated by symbolic regression via pareto genetic programming. *IEEE Trans. Evolutionary Computation*, 13(2):333–349, 2009.
21. Yang Zhang and Peter Rockett. A generic multi-dimensional feature extraction method using multiobjective genetic programming. *Evolutionary Computation*, 17(1):89–115, 2009.
22. Eckart Zitzler, Marco Laumanns, and Lothar Thiele. Spea2: Improving the strength pareto evolutionary algorithm. Technical report, ETH Zentrum, Gloriastrasse 35, CH-8092 Zurich, Switzerland, 2001.