Multiclass Cancer Diagnosis Using Tumor Gene Expression profiles

MultiClass Cancer Use Case

Background

- ❖ Cancers are classified by the type of tissue in which the cancer originates or/and by the location in the body where the cancer first developed
- ❖ There is a wide spectrum in cancer morphology
- ❖ Many tumors are atypical or lack morphologic features
- ❖ Molecular diagnostics could offer the promise of precise human cancer classification

Goal

❖ Use JADBio's automated machine learning to develop an accurate diagnostic test based on tumor gene expression profiles by Ramaswamy et al.[1], to classify multiple cancers.

[1]Ramaswamy et al., "Multiclass Cancer Diagnosis Using Tumor Gene Expression Signatures.' PNAS December 18, 2001 98 (26) 15149-15154

Data

Data from Ramaswamy et al., in which multi-classification was used for diagnosing common adult malignancies

Samples:
190 tissue specimens
14 common tumor types

Potential predictors:
16063 genes and expressed sequence tags

Target:
Cancer type

Metrics	RC_D60715_at Numerical	RC_D60774_f_at Numerical	RC_D60794_i_at Numerical	RC_D60794_f_at Numerical	class Categorical
min	-521	-361	-3567	-620	
max	127	1754	0.000	60	
nanmean	-73.647	20.037	-1198.979	-100.795	
nanstd	105.577	229.624	678.772	103.709	
median	-44.500	-10	-1068.500	-62.500	
missing	0	0	0	0	0
mode					Leukemia

Ramaswamy,S., et al. Multiclass cancer diagnosis using tumor gene expression signatures. Proc. Natl. Acad. Sci., 98, 15149 LP – 15154, 2001. Data available at: https://www.openml.org/d/1106

Running JADBio
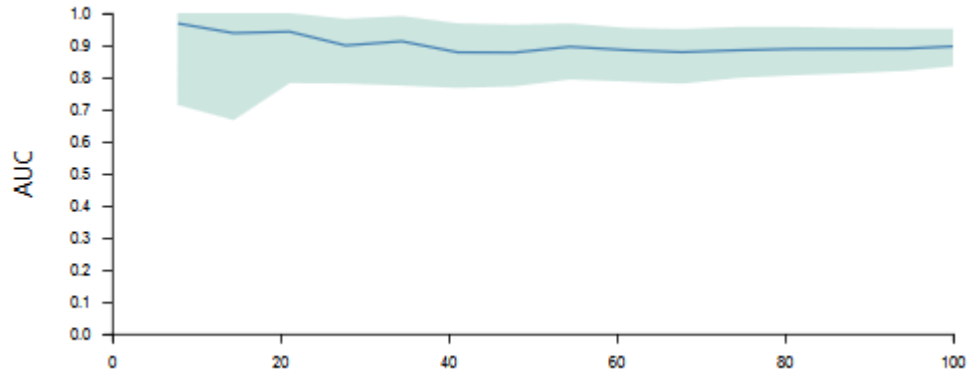
Preliminary analysis settings with optimized knowledge discovery and without enforced interpretability

During the Analysis, JADBio

1. Applies AI rules
2. Selects predictive features
3. Trains classification models
4. Optimizes model selection
5. Estimates model's performance
6. Creates plots

What you get:

✓ Optimal predictive model
✓ Estimates of model's performance
✓ Prioritized biomarkers

Models

When the Best Performing Model is not also interpretable (easily described in an equation or graphic) JADBio provides two models.

Best Performing Model: Classification Random Forests:
   AUC=0.896 [0.835, 0.951]

**Best Interpretable Model: Classification Decision Tree:**
AUC=0.679 [0.615, 0.751]

Signature

From the complete dataset, JADBio produces **Signatures** each of which includes a minimal-size subset of variables that are jointly predictive of the outcome of interest.
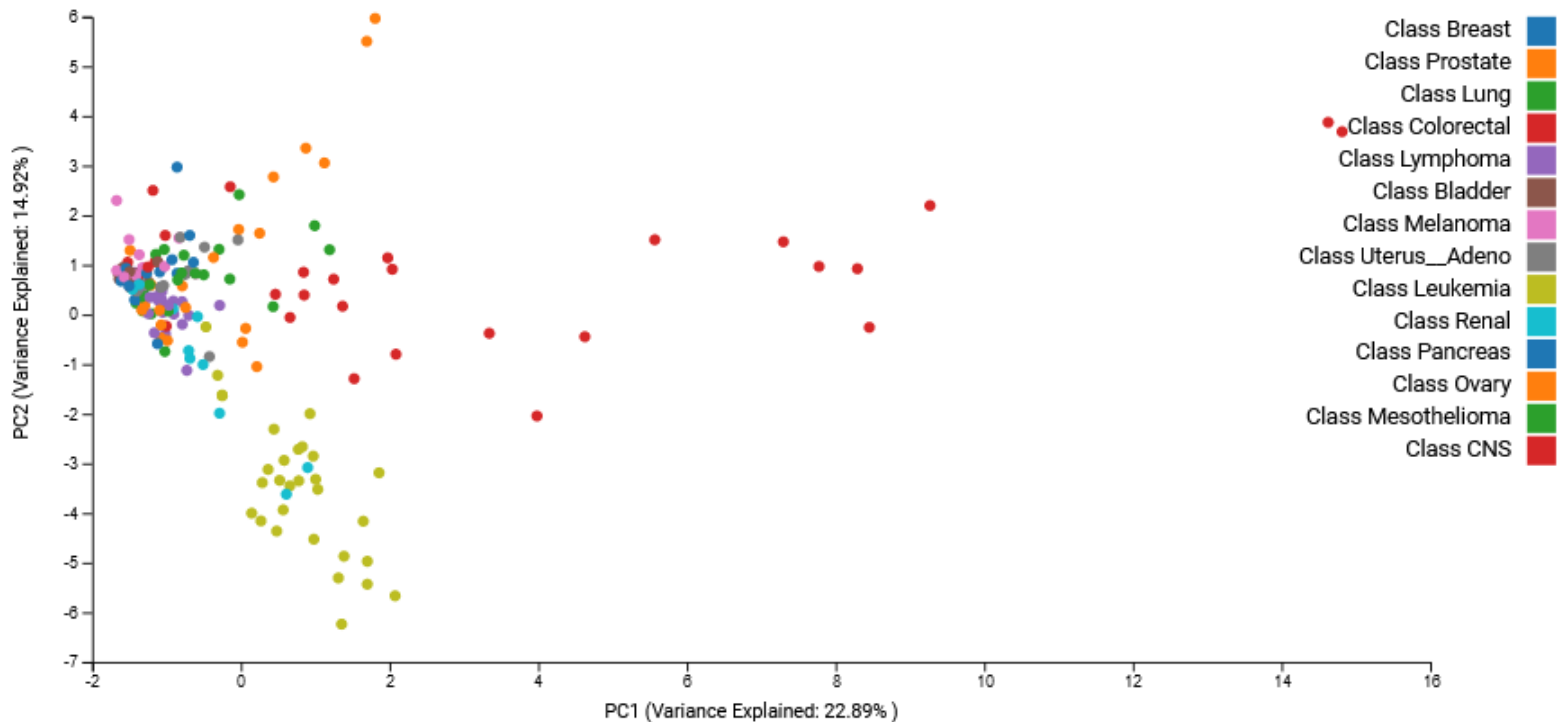
	Predictor 1	Predictor 2	Predictor 3	Predictor 4	Predictor 5	Predictor 6	Predictor 7
☐	X03635_at	X69699_at	RC_AA479910_at	X01703_at	RC_AA419609_at	U77180_at	AF001548_rn

25 Selected features
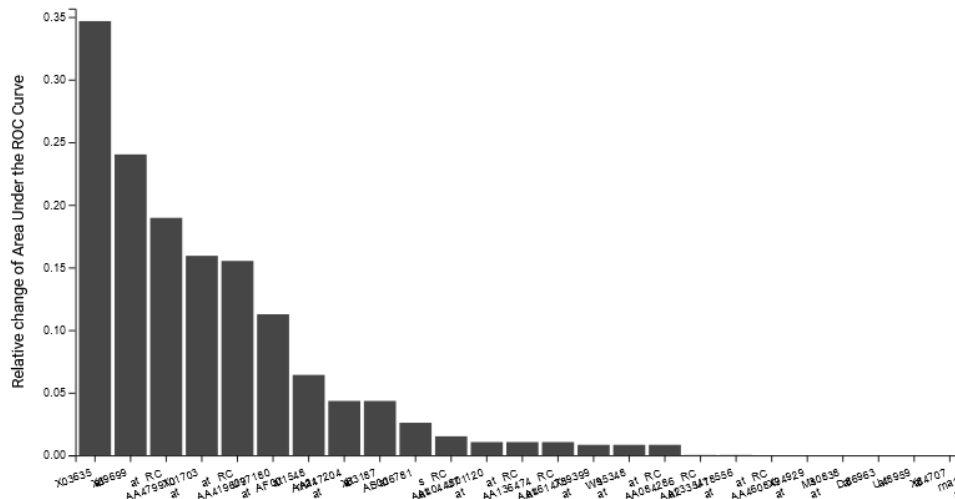
Analysis visualization

Supervised Principal Component Analysis (PCA)

Conclusion

❖ JADBio®, used gene expression profiles, provided by Ramaswamy et al., to produce interpretable and high-performance regression models for the prediction of different classes of cancer

❖ The process required no expert knowledge in advanced machine learning techniques.

❖ These results show JADBio can produce an accurate predictive model for classifying cancer types in an automated process.

www.jadbio.com

Thank you!

sales@jadbio.com

+30 2810 391257
+30 2810 391216

Science and Technology Park of Crete
N. Plastira 100, Vassilika Vouton GR-700 13, Heraklion, Greece , P.O.
Box 1447