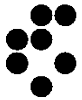
Evaluation Method for Feature Rankings and their Aggregations for Biomarker Discovery

I. Slavkov, B. Ženko and S. Džeroski



http://kt.ijs.si

Machine Learning in Systems Biology

MLSB09

September, 2009

Ljubljana, Slovenia

Outline

- Introduction
- Problem Definition
- Evaluation Method for Feature Rankings
- Experiments
- Results
- Conclusions

Biomarkers

- **Biomarkers** are biological parameters associated with the presence or status of a certain disease

- **Biomarker discovery** is the process of finding the biological parameters that have the strongest association with the presence or status of a disease

Feature Ranking vs. Feature Selection

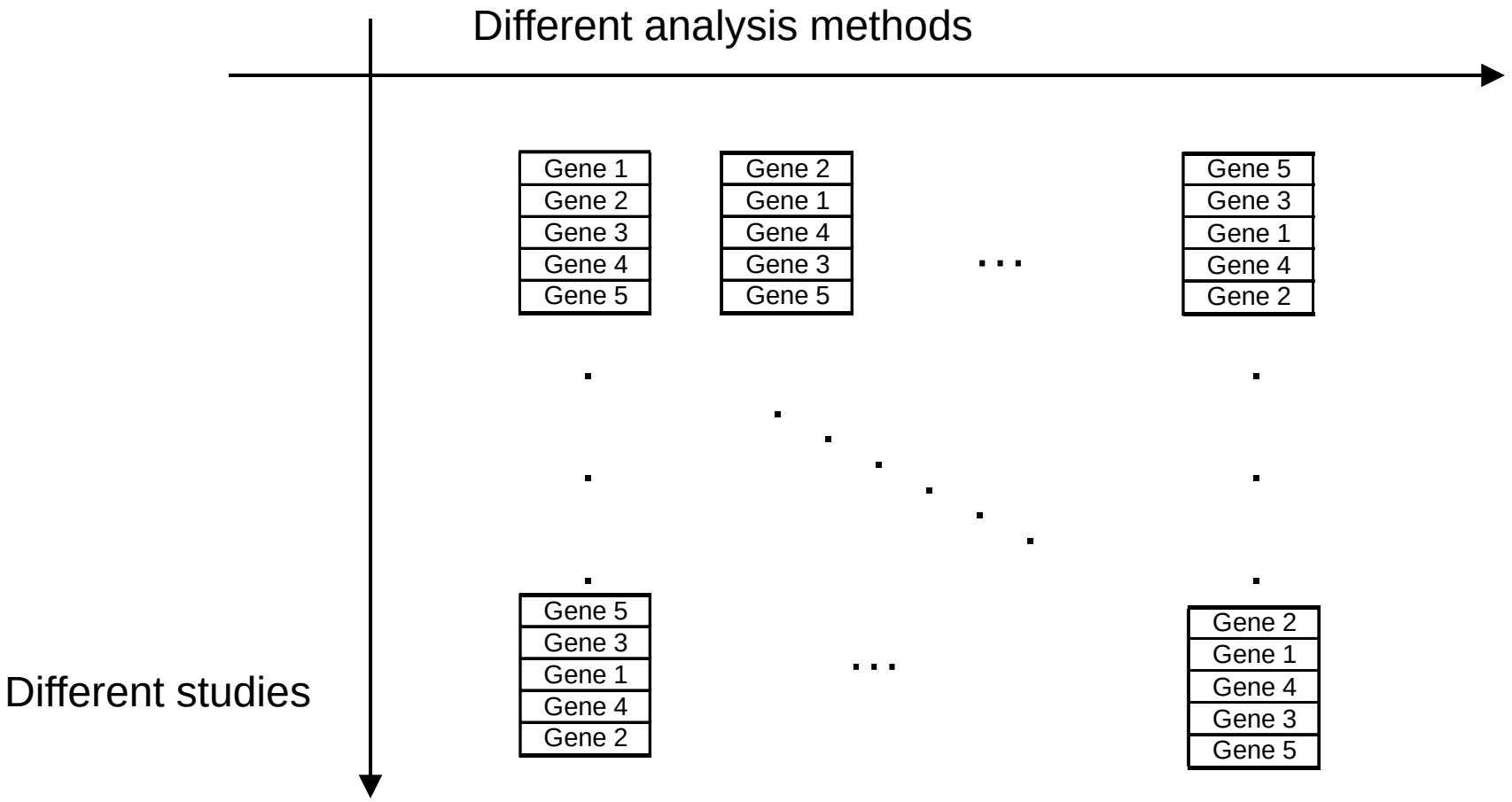
Feature **Ranking**

- **Ordered list** of features by "importance**"**
- Drug targets

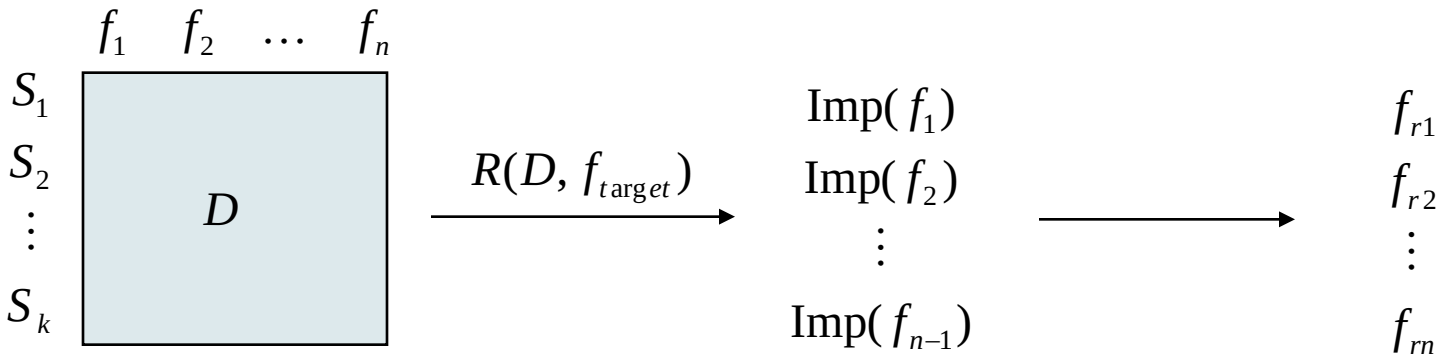
Feature **Selection**

- **Subset** of "important" features
- Diagnostic set of markers

The problem with biomarker discovery



Problem Definition



$$\mathrm{Imp}(f_{r(j-1)}) \geq \mathrm{Imp}(f_{rj}) \geq \mathrm{Imp}(f_{r(j+1)})$$

$$R_1 \rightarrow \mathrm{Imp}_1(f_j)$$
$$R_2 \rightarrow \mathrm{Imp}_2(f_j)$$
$$\vdots$$
$$R_l \rightarrow \mathrm{Imp}_l(f_j)$$

$$\mathrm{Imp}_1(f_j) \neq \mathrm{Imp}_2(f_j) \neq \ldots \neq \mathrm{Imp}_l(f_j)$$

Feature Ranking vs. Feature Selection

Feature **Ranking**

- **Ordered list** of features by "importance**"**
- Drug targets
- No explicit measure

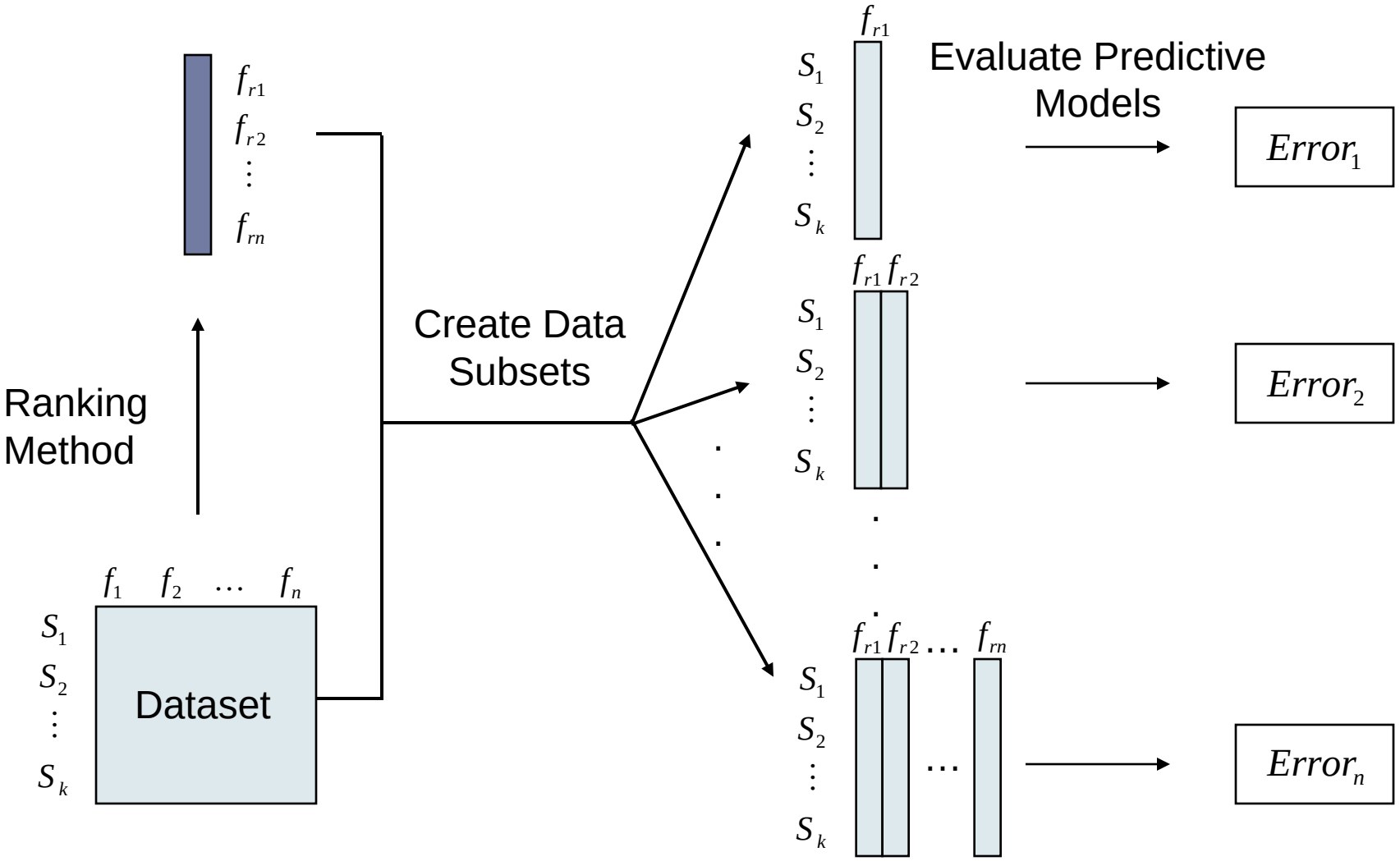
Feature **Selection**

- **Subset** of "important" features
- Diagnostic set of markers
- Accuracy of predictive model

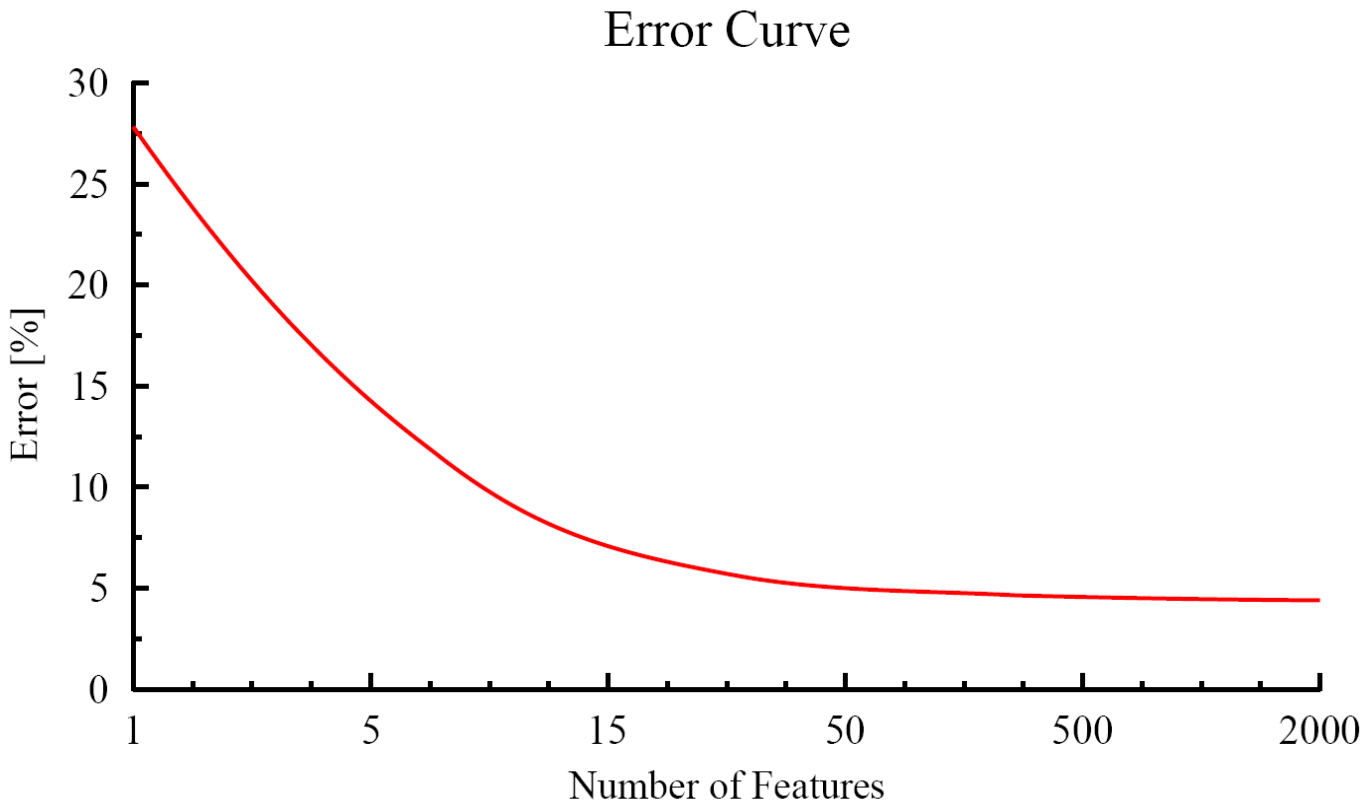
Outline

- Introduction
- Problem Definition
- Evaluation Method for Feature Rankings
- Experiments
- Results
- Conclusions

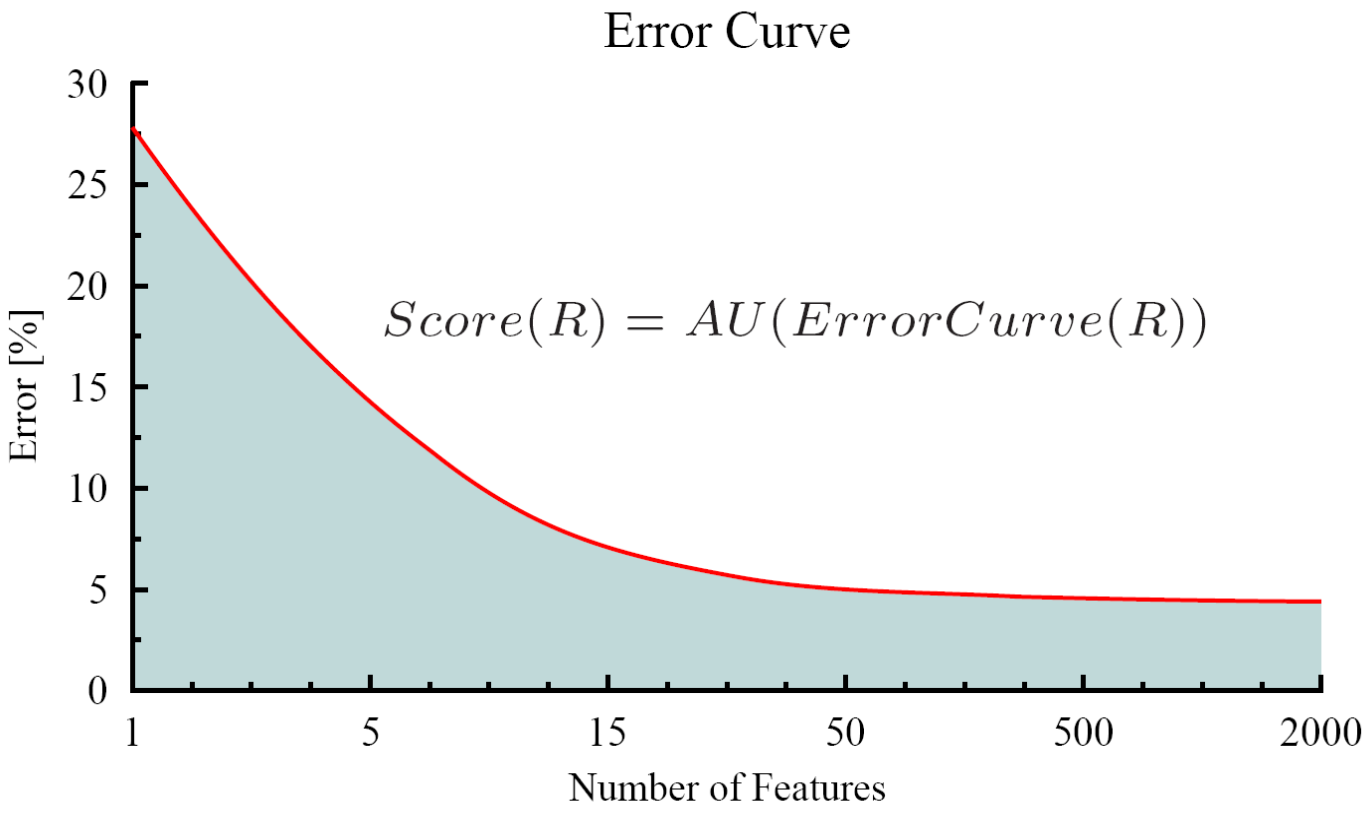
Evaluation method



Error Curve


Error Curve
30
25
20
15
10
5
0
Error [%]
1
5
15
50
500
2000
Number of Features

Error Curve


Error Curve
30
25
20
15
10
5
0
Error [%]
1
5
15
50
500
2000
Number of Features
$Score(R) = AU(ErrorCurve(R))$

Outline

- Introduction
- Problem Definition
- Evaluation Method for Feature Rankings
- Experiments
- Results
- Conclusions

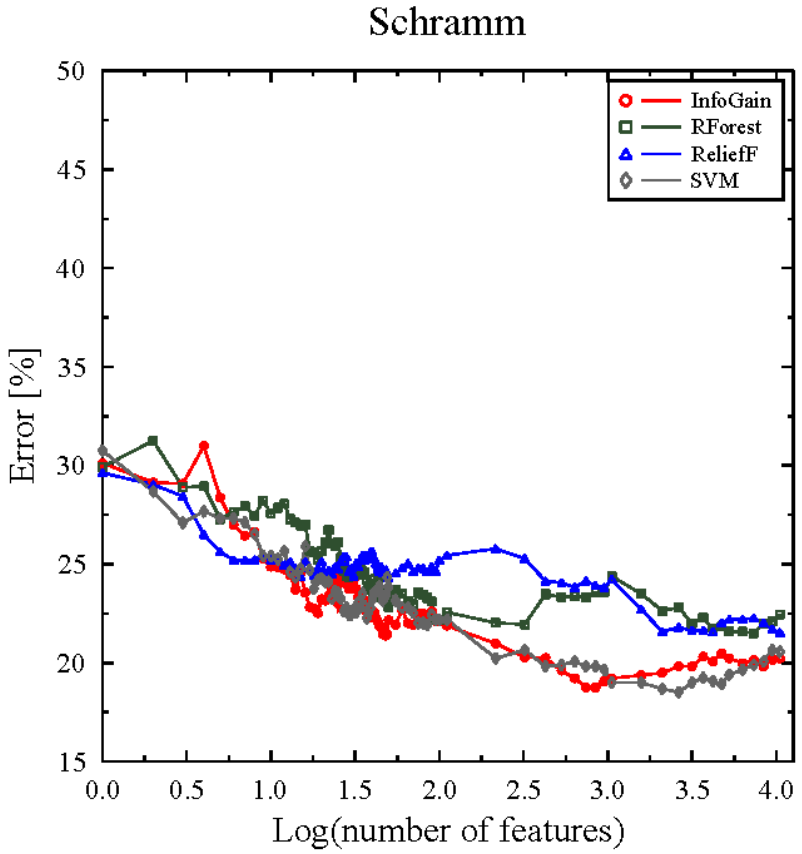
Experimental Datasets

- Neuroblastoma expression data:
  - **Neuroblastoma** is the most common extracranial solid tumor of childhood
  - Three public studies: De Preter et al. (17), Schramm et al. (63) and Wang et al. (100)
  - Target of interest: "**Relapse/No Relapse"** status of a patient

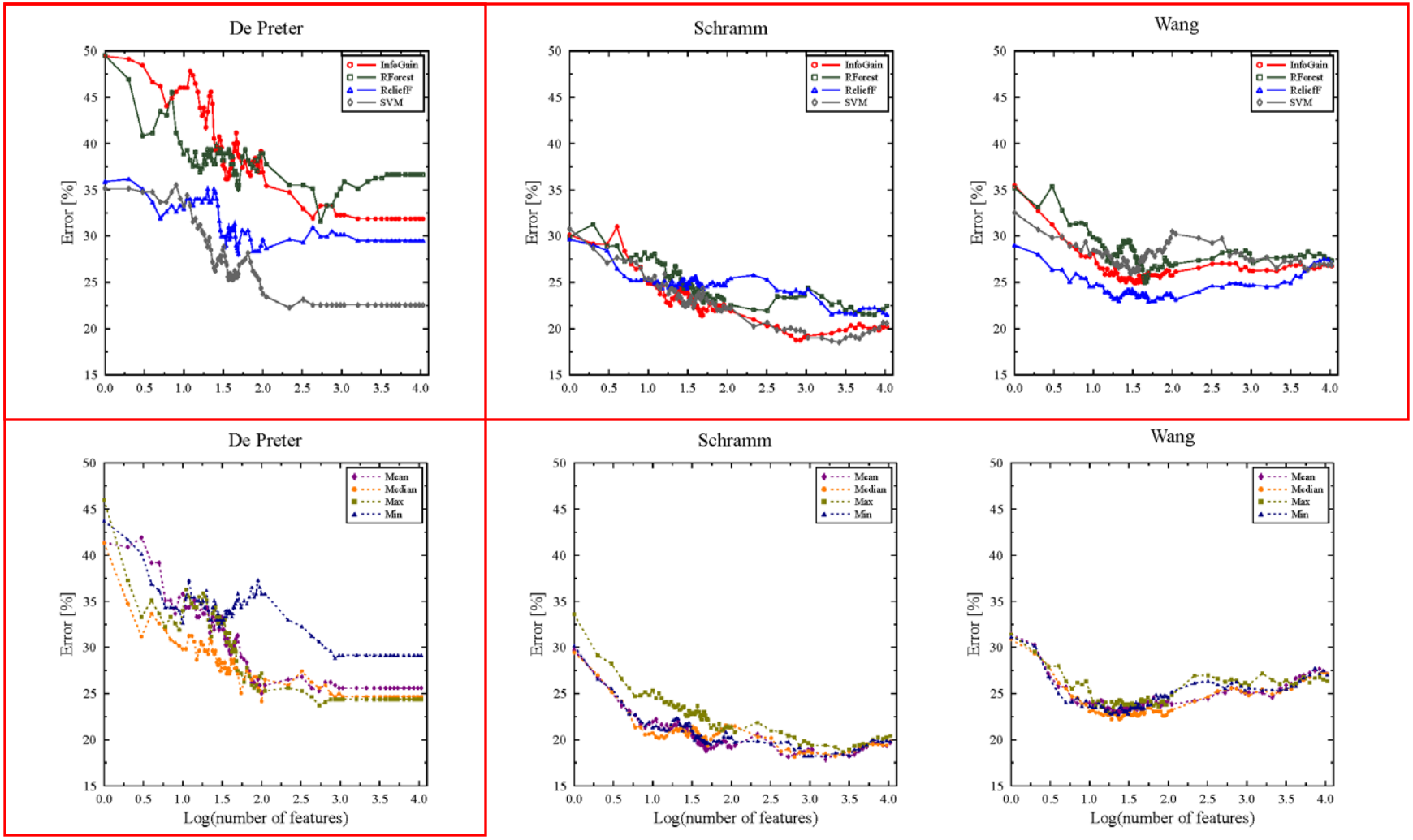
Experimental Scenarios

- Individual studies
  - Ranking algorithms: Info Gain, ReliefF, Random Forests and SVM-RFE
  - Aggregation functions: Mean, Median, Min and Max
- Multiple studies
  - Compare rankings from **single** studies to **aggregated rankings** from multiple studies
- .632+ Bootstrap for error estimation (Naïve Bayes)

Results Sample



Individual Studies: Error Curves



Individual Studies: AU Error Curves

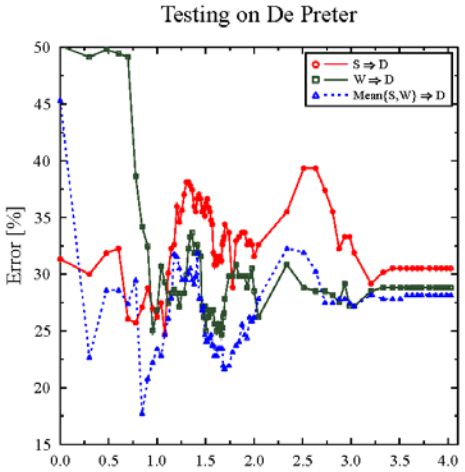
	Info Gain	R Forest	ReliefF	SVM	Mean	Median	Min	Max
De Preter	0.386	0.385	0.313	0.270	0.306	0.282	0.336	0.293
Schramm	0.228	0.246	0.246	0.226	0.206	0.208	0.209	0.225
Wang	0.273	0.288	0.249	0.283	0.253	0.249	0.255	0.259

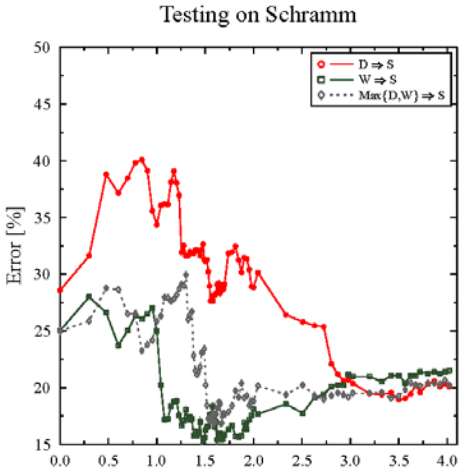
Individual Studies: AU Error Curves

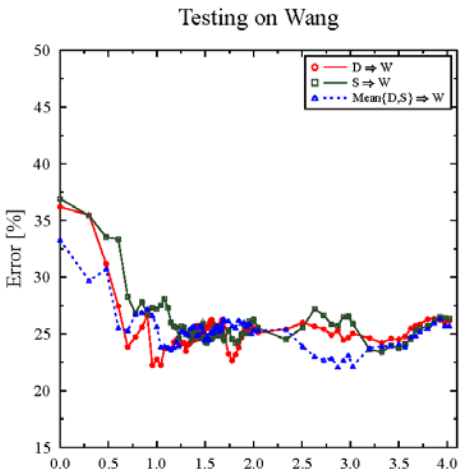
	Info Gain	R Forest	ReliefF	SVM	Mean	Median	Min	Max
De Preter	0.386	0.385	0.313	**0.270**	0.306	**0.282**	0.336	0.293
Schramm	0.228	0.246	0.246	**0.226**	**0.206**	0.208	0.209	0.225
Wang	0.273	0.288	**0.249**	0.283	0.253	**0.249**	0.255	0.259

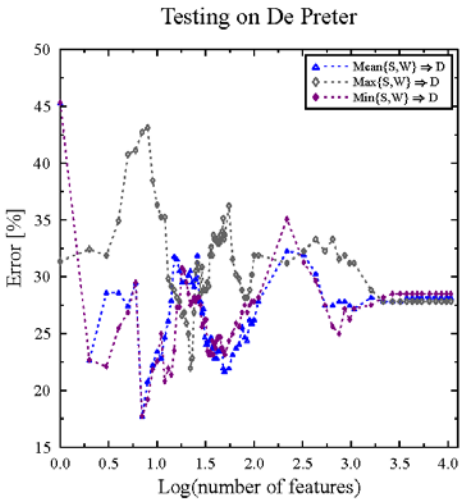
- Feature rankings from different algorithms:
  - Have similar error curves, unless small data sizes
- Aggregating feature rankings from different algorithms:
  - Using the median provides the best results
  - Area under the curve is comparable to individual algorithms, but much less variable

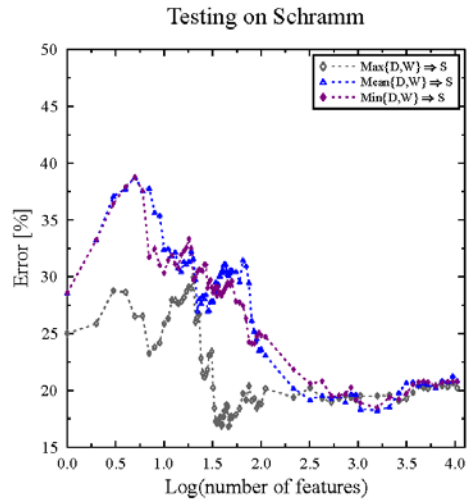
Multiple Studies: Error Curves

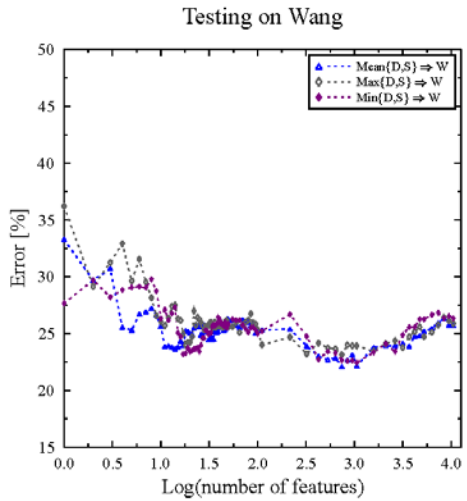












Multiple Studies: AU Error Curves

	De Preter	Schramm	Wang
De Preter	\	0.283	0.263
Schramm	0.326	\	0.269
Wang	0.337	0.207	\
Mean	0.305	0.260	0.254
Min	0.301	0.260	0.256
Max	0.321	0.221	0.262
	WaggS	**DaggW**	**SaggD**

Test Set

Ranking Method

Aggregated Datasets

Multiple Studies: AU Error Curves

	De Preter	Schramm	Wang
De Preter	\	0.283	**0.263**
Schramm	**0.326**	\	0.269
Wang	0.337	**0.207**	\
Mean	0.305	0.260	**0.254**
Min	**0.301**	0.260	0.256
Max	0.321	**0.221**	0.262
	WaggS	**DaggW**	**SaggD**

- Aggregating feature rankings from different studies:
  - Has a generally beneficial effect on reducing the error size of the curves, but
  - More sophisticated aggregation methods are needed (ex. that take into account different study sizes)

Summary

- Biomarker discovery as problem of evaluating feature rankings
- Evaluation methodology for feature rankings:
  - Relates rankings to predictive performance via the so-called error curve (AU error curve as numerical indicator of quality)
- Experimental results demonstrate that:
  - The evaluation method is useful for comparing feature rankings
  - Aggregation of rankings (different algorithms, different studies) is beneficial for deriving more robust biomarker signatures