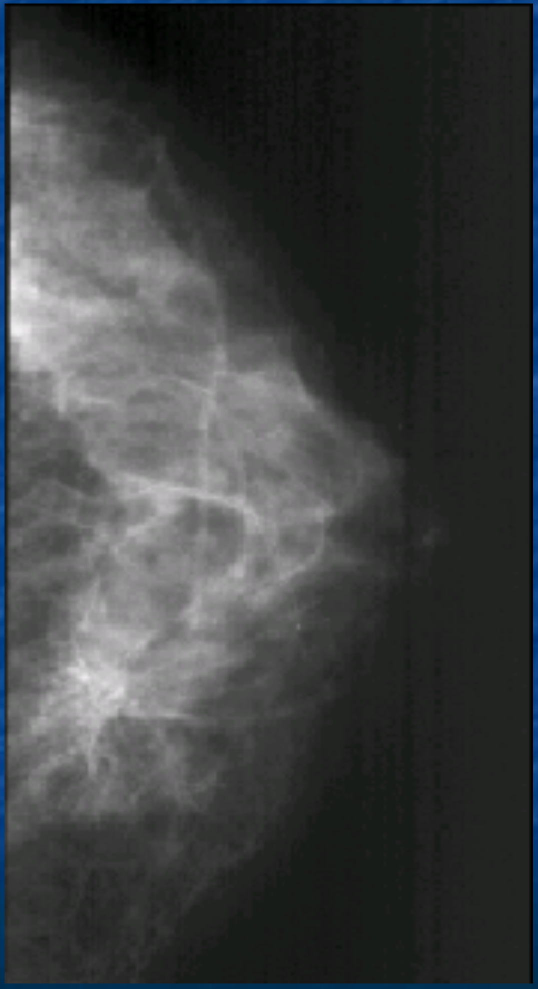
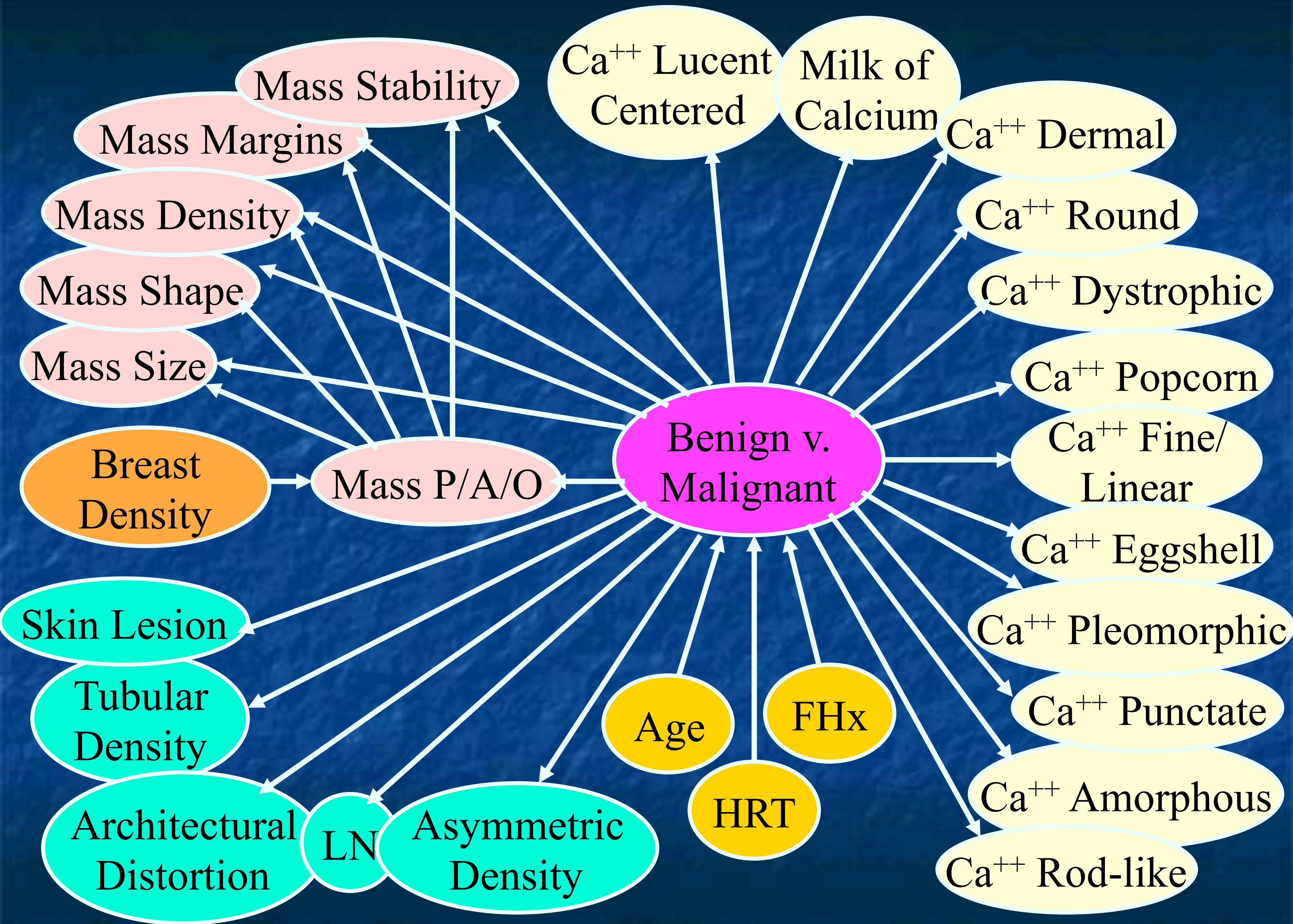
Application: Mammography

- Provide decision support for radiologists
- Variability due to differences in training and experience
- Experts have higher cancer detection and fewer benign biopsies
- **Shortage of experts**

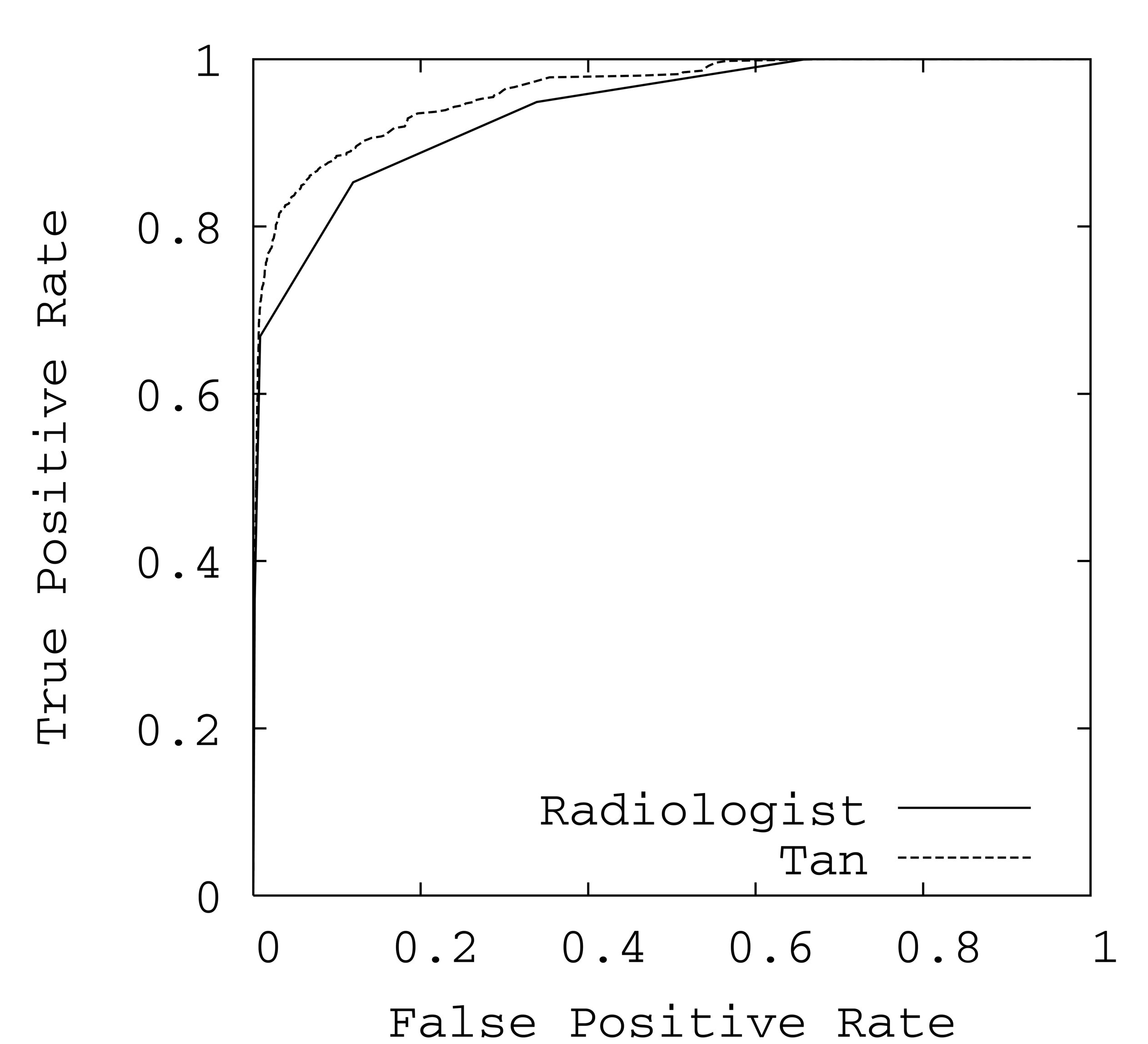


Bayes Net for Mammography

- Kahn, Roberts, Wang, Jenks, Haddawy (1995)
- Kahn, Roberts, Shaffer, Haddawy (1997)
- Burnside, Rubin, Shachter (2000)
- Bayes Net can now outperform general radiologists and perform at level of expert mammographers: area under ROC curve of 0.94


Mass Stability
Mass Margins
Mass Density
Mass Shape
Mass Size
Breast Density
Mass P/A/O
Ca++ Lucent Centered
Milk of Calcium
Ca++ Dermal
Ca++ Round
Ca++ Dystrophic
Ca++ Popcorn
Ca++ Fine/ Linear
Ca++ Eggshell
Ca++ Pleomorphic
Ca++ Punctate
Ca++ Amorphous
Ca++ Rod-like
Benign v. Malignant
Skin Lesion
Tubular Density
Architectural Distortion
LN
Asymmetric Density
Age
HRT
FHx

ROC: Radiologist vs. BN (TAN)



Technical Issue for Rest of Talk

- *Q: Can learning improve the expert constructed Bayes Net?*

- Learning Hierarchy
  - Level 1: Parameter
  - Level 2: Structure
  - Level 3: Aggregate
  - Level 4: **View**

Standard ML

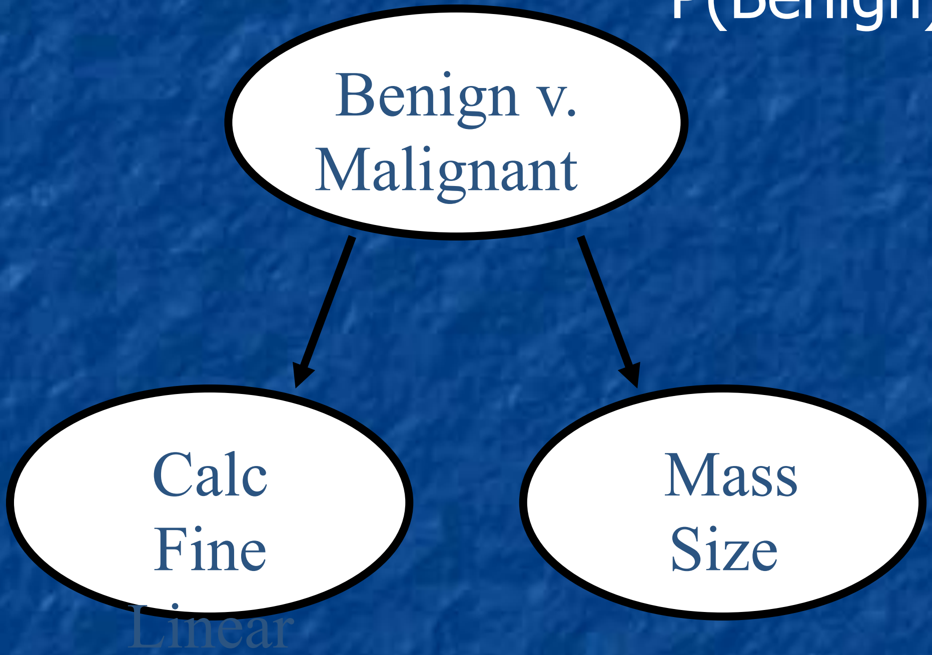
New Capabilities

Mammography Database

Patient	Abnormality	Date	Calcification Fine/Linear	…	Mass Size	Loc	Benign/ Malignant
P1	1	5/02	No		0.03	RU4	B
P1	2	5/04	Yes		0.05	RU4	M
P1	3	5/04	No		0.04	LL3	B
P2	4	6/00	No		0.02	RL2	B
…	…	…	…		…	…	…

Level 1: Parameters

P(Benign) = .99



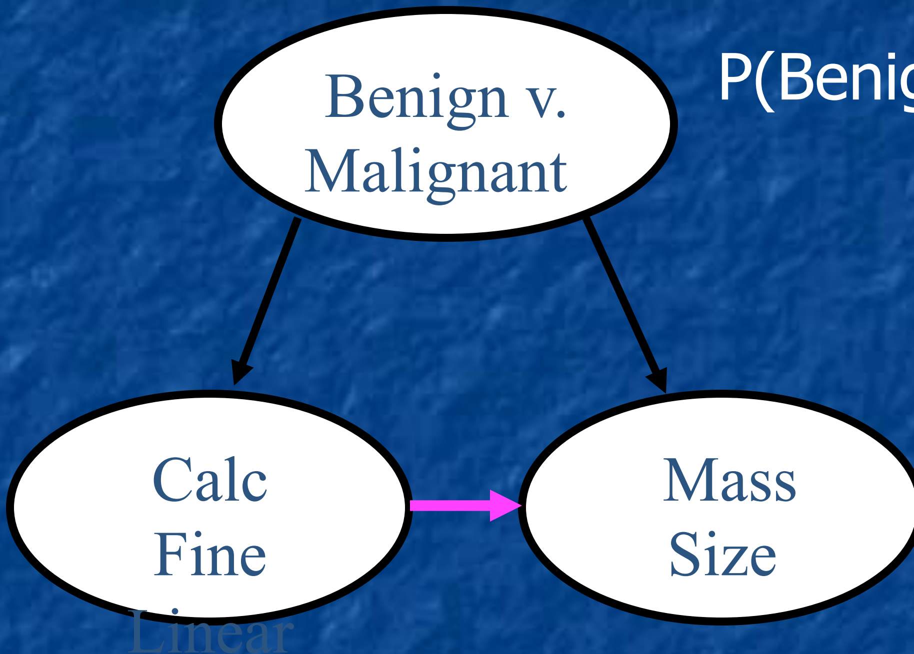
P(Yes| Benign) = .01

P(Yes| Malignant) = .55

P(size > 5| Benign) = .33

P(size > 5| Malignant) = .42

Level 2: Structure + Parameters



P(Benign) = .99

P(Yes| Benign) = .01

P(Yes) = .02

P(Yes| Malignant) = .55

P(size > 5| Benign ^ Yes) = .4

P(size > 5| Benign) = .33

P(size > 5) = .1

P(size > 5| Malignant ^ Yes) = .6

P(size > 5| Malignant) = .42

P(size > 5| Benign ^ No) = .05

P(size > 5| Malignant ^ No) = .2

Data

- Structured data from actual practice
- National Mammography Database
  - Standard for reporting all abnormalities
- Our dataset contains
  - 435 malignancies
  - 65,365 benign abnormalities
- Link to biopsy results
  - Obtain disease diagnosis – our ground truth

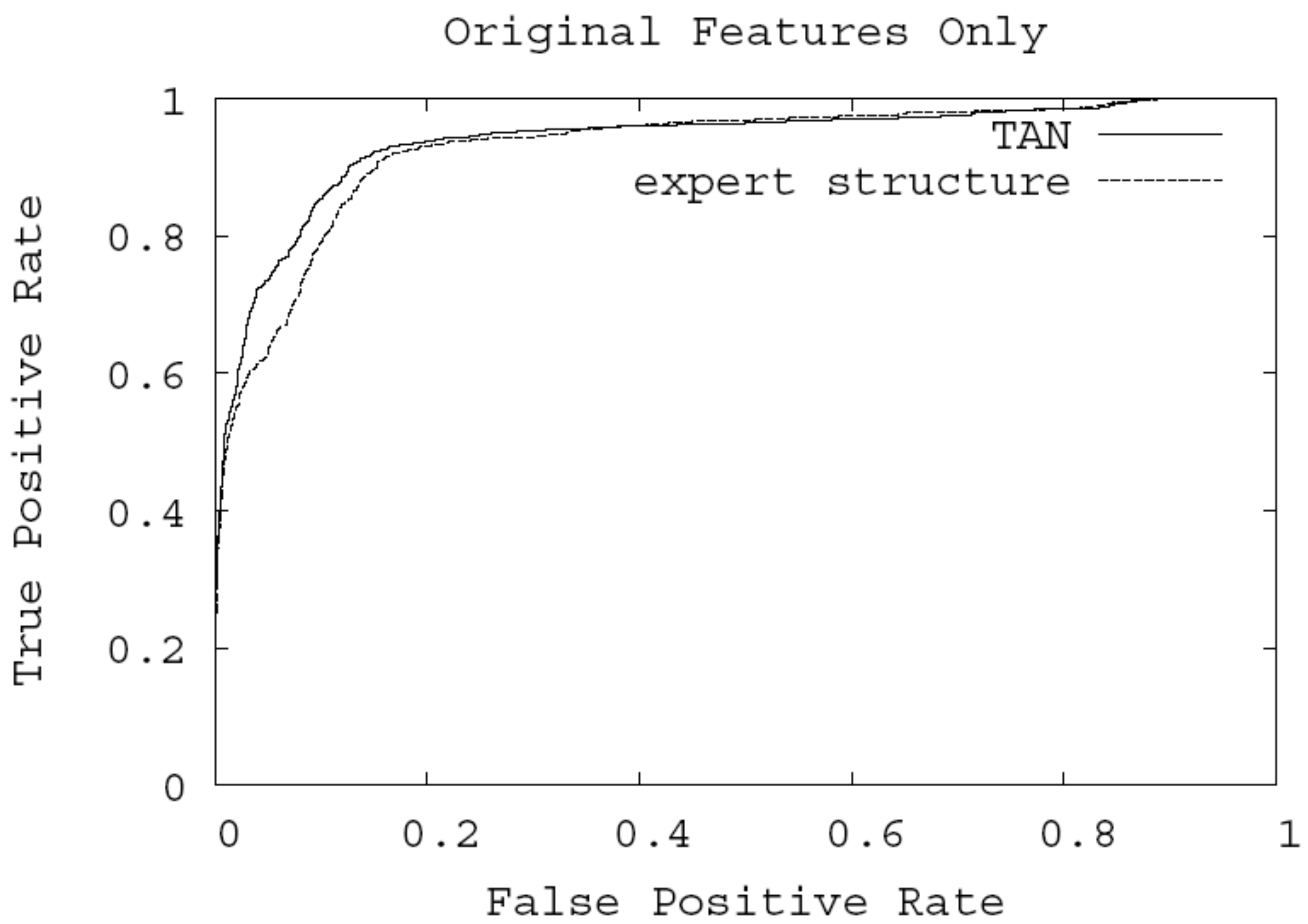
Hypotheses

- Learn relationships that are useful to radiologist
- Improve by moving up learning hierarchy

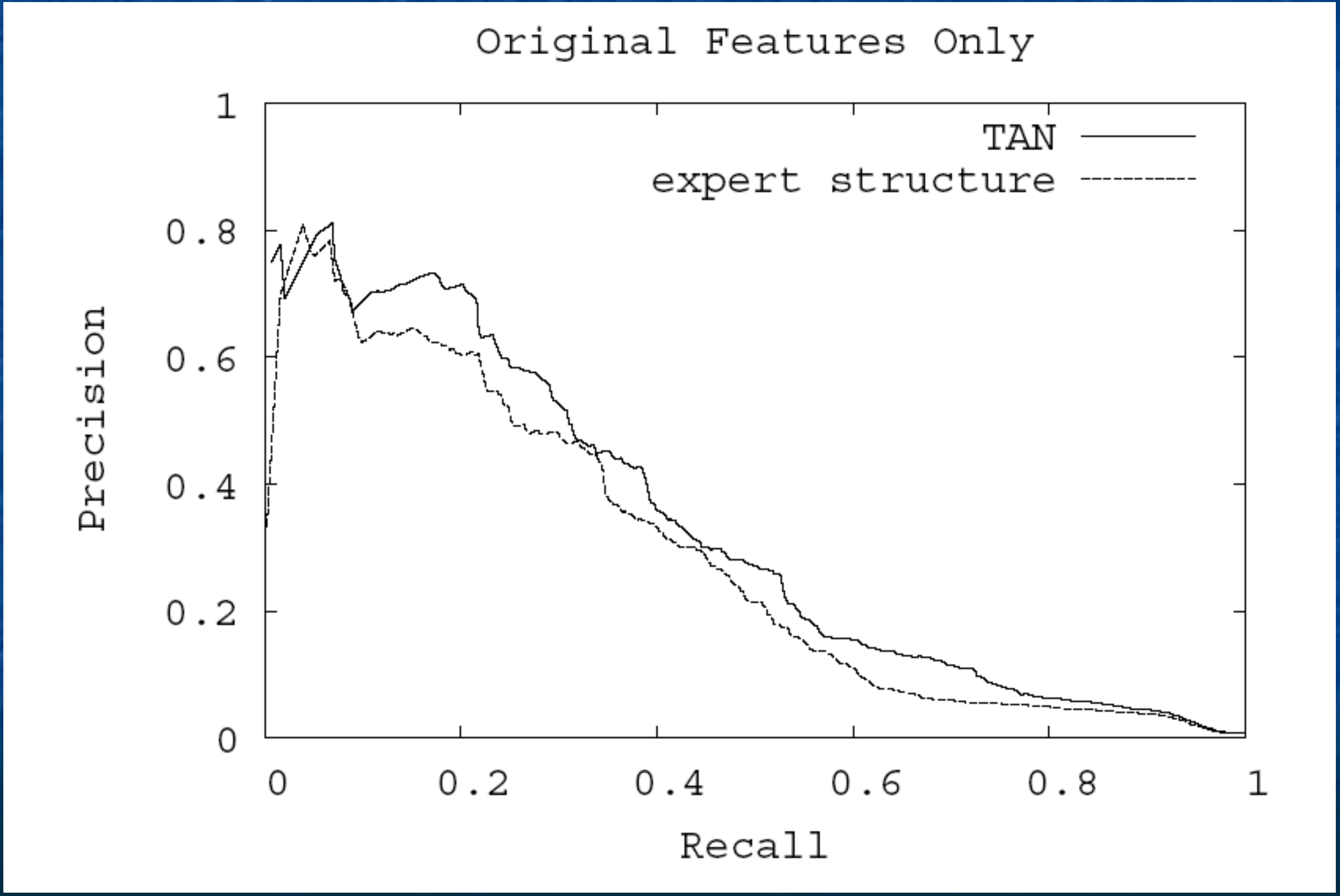
Results

- Trained (Level 2, TAN) Bayesian network model achieved an AUC of 0.966 which was significantly better than the radiologists' AUC of 0.940 (P = 0.005)
- Trained BN demonstrated significantly better sensitivity than the radiologist (89.5% vs. 82.3% —P = 0.009) at a specificity of 90%
- Trained BN demonstrated significantly better specificity than the radiologist (93.4% versus 86.5%—P = 0.007) at a sensitivity of 85%

ROC: Level 2 (TAN) vs. Level 1



Precision-Recall Curves



Mammography Database

Patient	Abnormality	Date	Calcification Fine/Linear	…	Mass Size	Loc	Benign/ Malignant
P1	1	5/02	No		0.03	RU4	B
P1	2	5/04	Yes		0.05	RU4	M
P1	3	5/04	No		0.04	LL3	B
P2	4	6/00	No		0.02	RL2	B
…	…	…	…		…	…	…

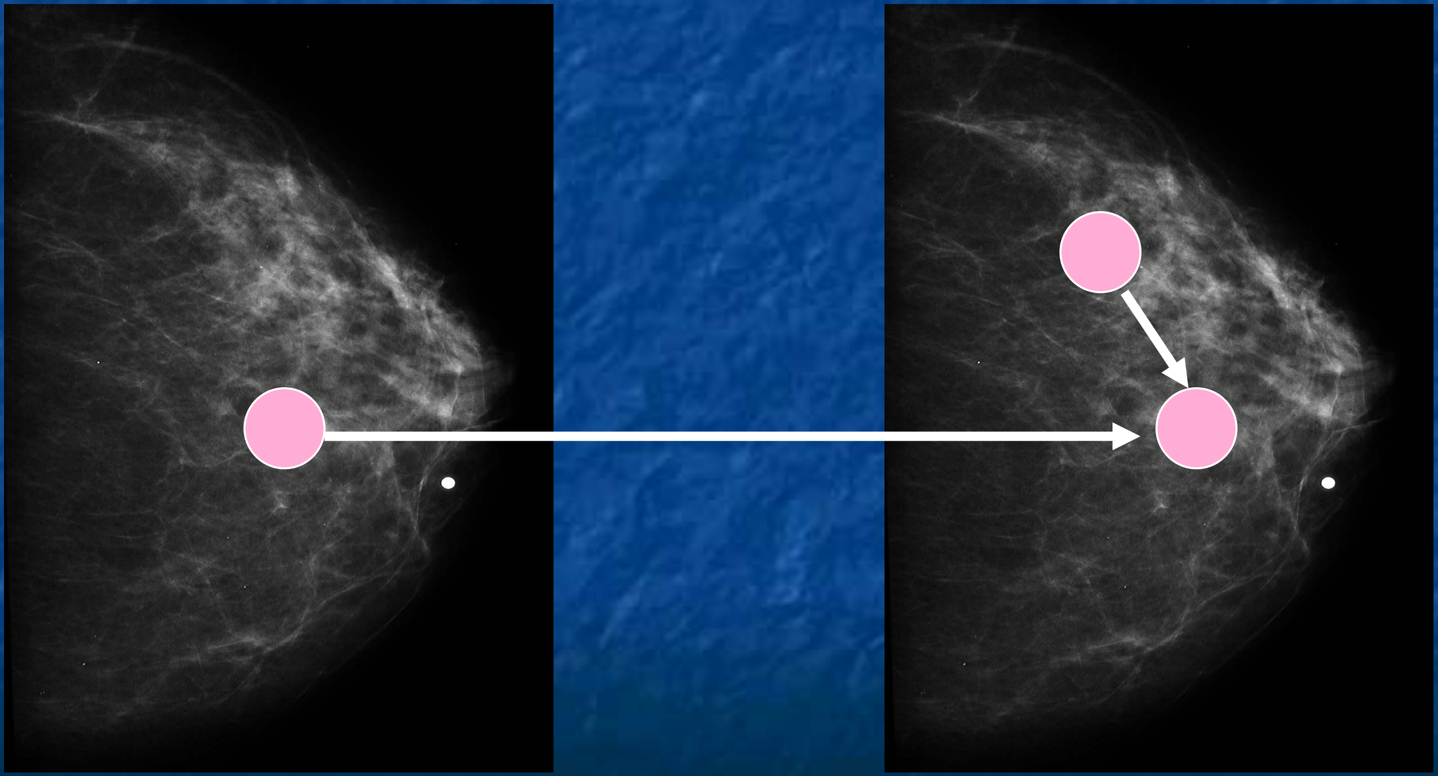
Statistical Relational Learning

- Learn probabilistic model, but don't assume iid data: there may be relevant data in other rows or even other tables

- Database schema: defines set of features

Connecting Abnormalities

May 2002 | Patient 1 | May 2004



SRL Aggregates Information from Related Rows or Tables

- Extend probabilistic models to relational databases
- Probabilistic Relational Models (Friedman et al. 1999, Getoor et al. 2001)
  - Tricky issue: one to many relationships
  - Approach: use aggregation
- PRMs cannot capture all relevant concepts

Aggregate Illustration

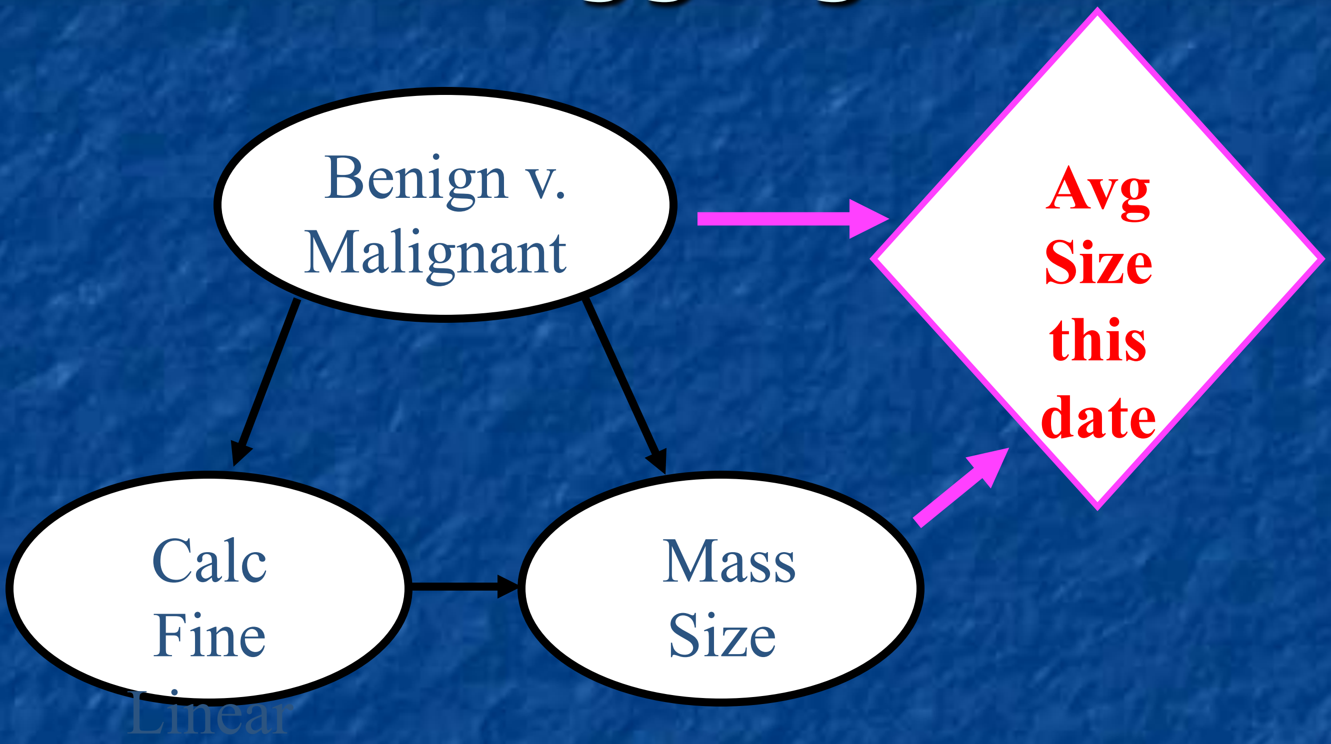
Aggregation Function:
Min, Max, Average, etc.

Patient	Abnormality	Date	Calcification Fine/Linear	…	Mass Size	Loc	Benign/ Malignant
P1	1	5/02	No		0.03	RU4	B
P1	2	5/04	Yes		0.05	RU4	M
P1	3	5/04	No		0.04	LL3	B
P2	4	6/00	No		0.02	RL2	B
…	…	…	…		…	…	…

New Schema

Patient	Abnormality	Date	Calcification Fine/Linear	…	Mass Size	**Avg Size this Date**	Loc	Benign/ Malignant
P1	1	5/02	No		0.03	**0.03**	RU4	B
P1	2	5/04	Yes		0.05	**0.045**	RU4	M
P1	3	5/04	No		0.04	**0.045**	LL3	B
P2	4	6/00	No		0.02	**0.02**	RL2	B
…	…	…	…		…	**…**	…	…

Level 3: Aggregates



Note: Learn parameters for each node

Database Notion of View

- New tables or fields defined in terms of existing tables and fields known as ***views***
- A view corresponds to alteration in database schema
- Goal: automate the learning of views

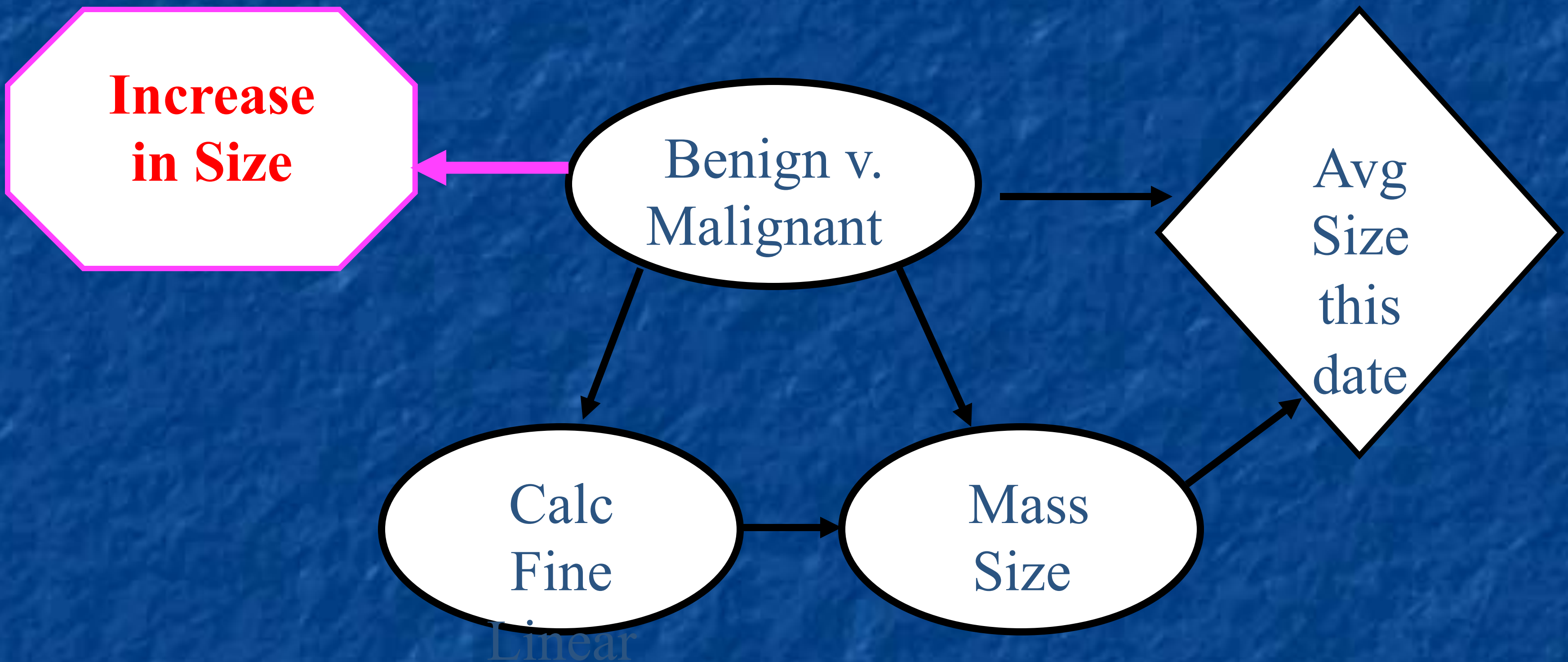
Possible View

Patient	Abnormality	Date	Calcification Fine/Linear	…	Mass Size	Loc	Benign/ Malignant
P1	1	5/02	No		0.03	RU4	B
P1	2	5/04	Yes		0.05	RU4	M
P1	3	5/04	No		0.04	LL3	B
P2	4	6/00	No		0.02	RL2	B
…	…	…	…		…	…	…

New Schema

Patient	Abnormality	Date	Calcification Fine/Linear	…	Mass Size	**Increase In Size**	Loc	Benign/ Malignant
P1	1	5/02	No		0.03	**No**	RU4	B
P1	2	5/04	Yes		0.05	**Yes**	RU4	M
P1	3	5/04	No		0.04	**No**	LL3	B
P2	4	6/00	No		0.02	**No**	RL2	B
…	…	…	…		…	**…**	…	…

Level 4: View Learning



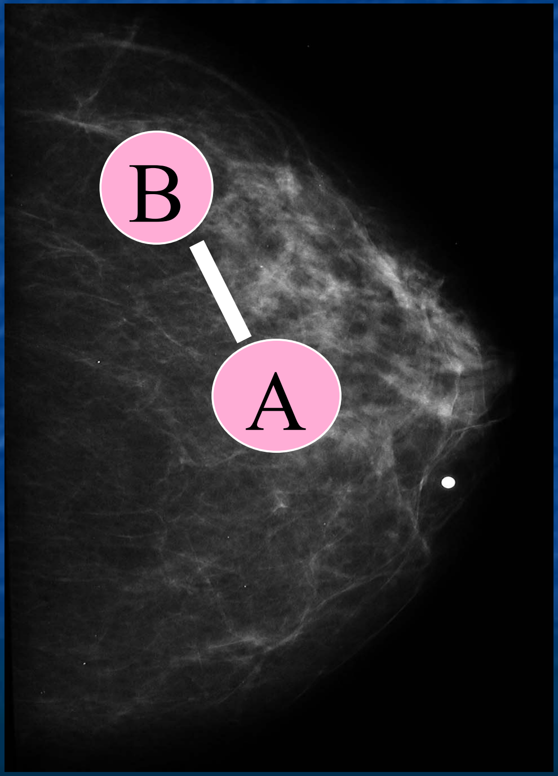
Note: Include aggregate features
Learn parameters for each node

Level 4: View Learning

- Learn rules predictive of “malignant”
  - We used Aleph (Srinivasan)
- Treat each rule as a new field
  - 1 if abnormality matches rule
  - 0 otherwise
- New view consists of original table extended with new fields

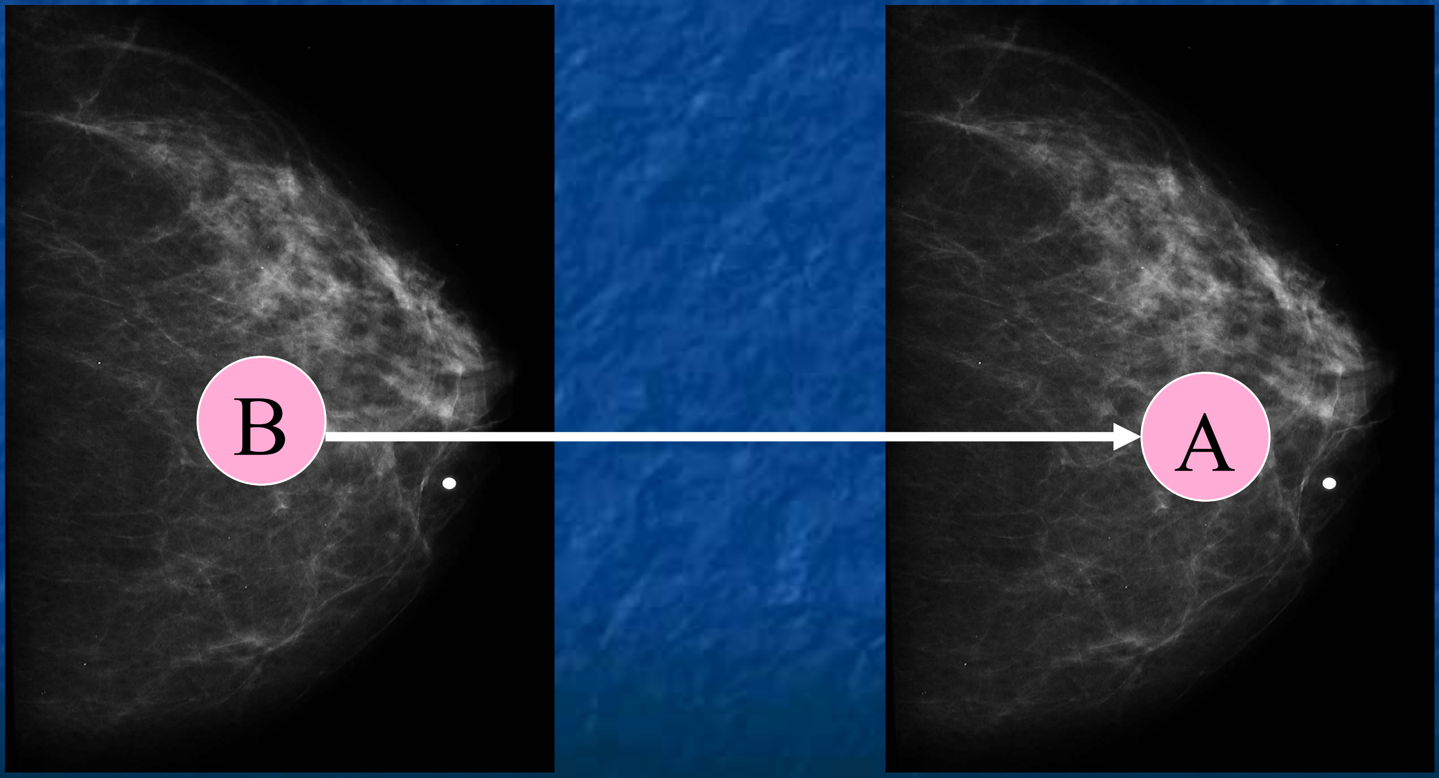
Key New Predicate I

in_same_mammogram(A,B)



Key New Predicate II

prior_mammogram(A,B)



Experimental Methodology

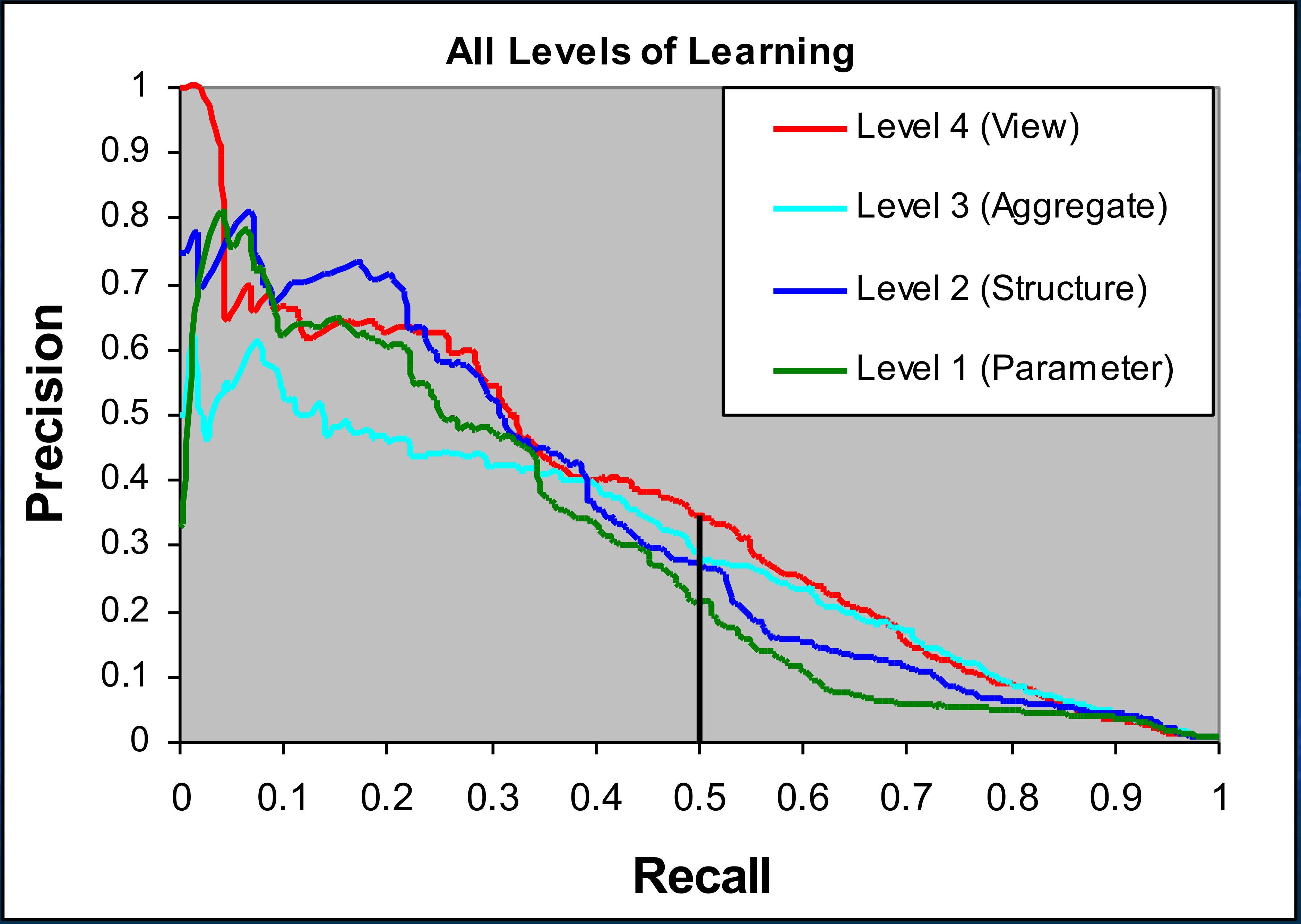
- 10-fold cross validation
- Split at the patient level
- Roughly 40 malignant cases and 6000 benign cases in each fold
- Tree Augmented Naïve Bayes (TAN) as structure learner (Friedman,Geiger & Goldszmidt ’97)

Approach

- Level 3: Aggregates
  - 27 features make sense to aggregate
  - Aggregated over patient and mammogram
- Level 4: View
  - 4 folds to learn rules
  - 5 folds for training set

Sample View
[Burnside et al. AMIA05]

```
malignant(A) :-
    birads_category(A,b5),
    massPAO(A,present),
    massesDensity(A,high),
    ho_breastCA(A,hxDCorLC),
    in_same_mammogram(A,B),
    calc_pleomorphic(B,notPresent),
    calc_punctate(B,notPresent).
```

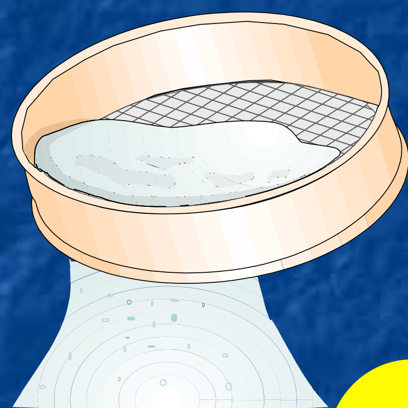

All Levels of Learning
Precision
Recall
Level 4 (View)
Level 3 (Aggregate)
Level 2 (Structure)
Level 1 (Parameter)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

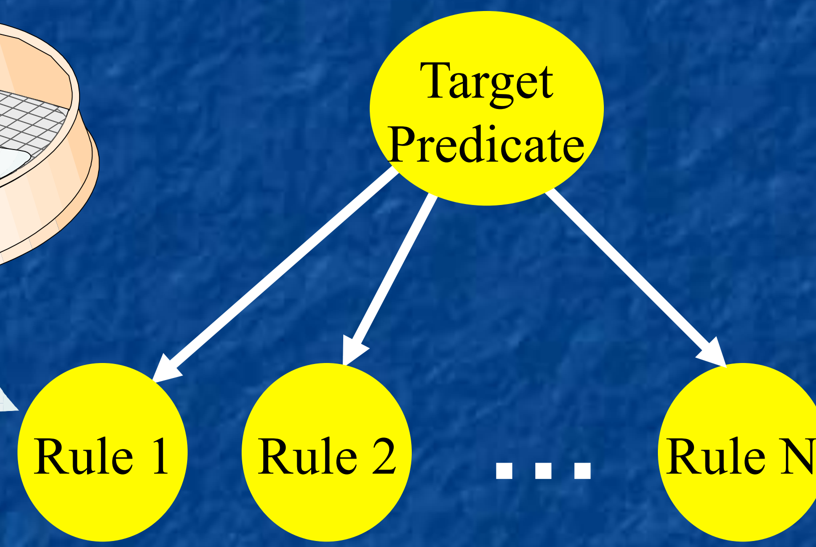
View Learning: First Approach

[Davis et al. IA05, Davis et al. IJCAI05]

Step 1 | Step 2 | Step 3

Rule Learner





Learn | Select | Build Model

Drawback to First Approach

- Mismatch between
  - Rule building
  - Model’s use of rules
- Should Score As You Use (SAYU)

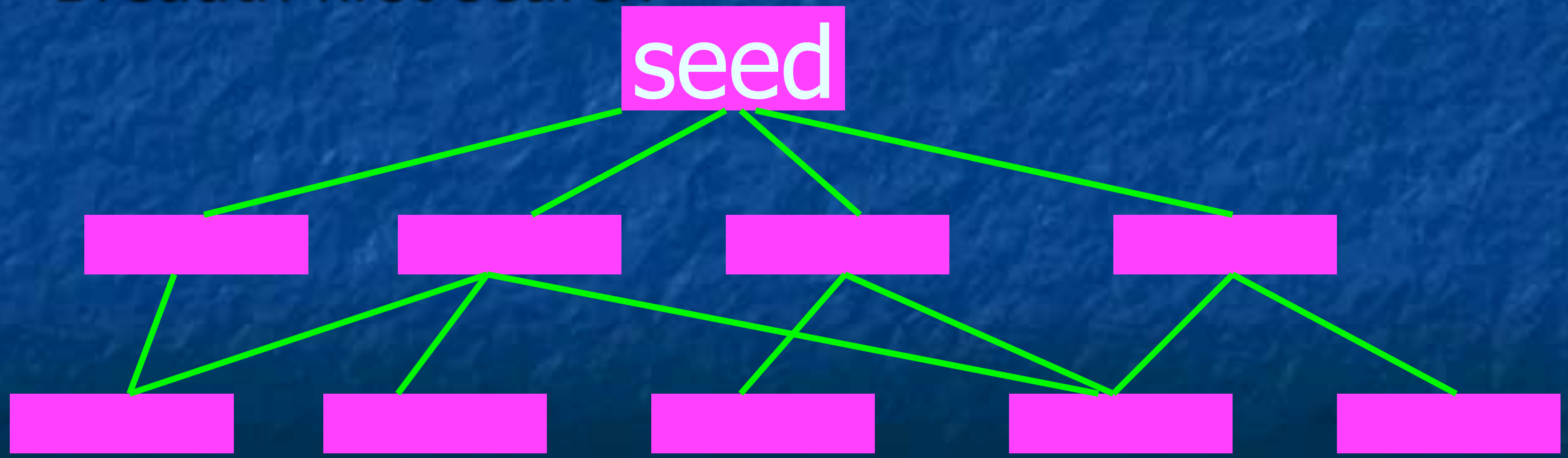
SAYU

[Davis et al. ECML05]

- Build network as we learn rules
  [Landwehr et al. AAAI 2005]
- Score rule on whether it improves network
- Results in tight coupling between
  rule generation, selection and usage

SAYU Details

- Based on Aleph algorithm
- Randomly pick positive example as seed
- Build ‘bottom’ clause
- Breadth first search

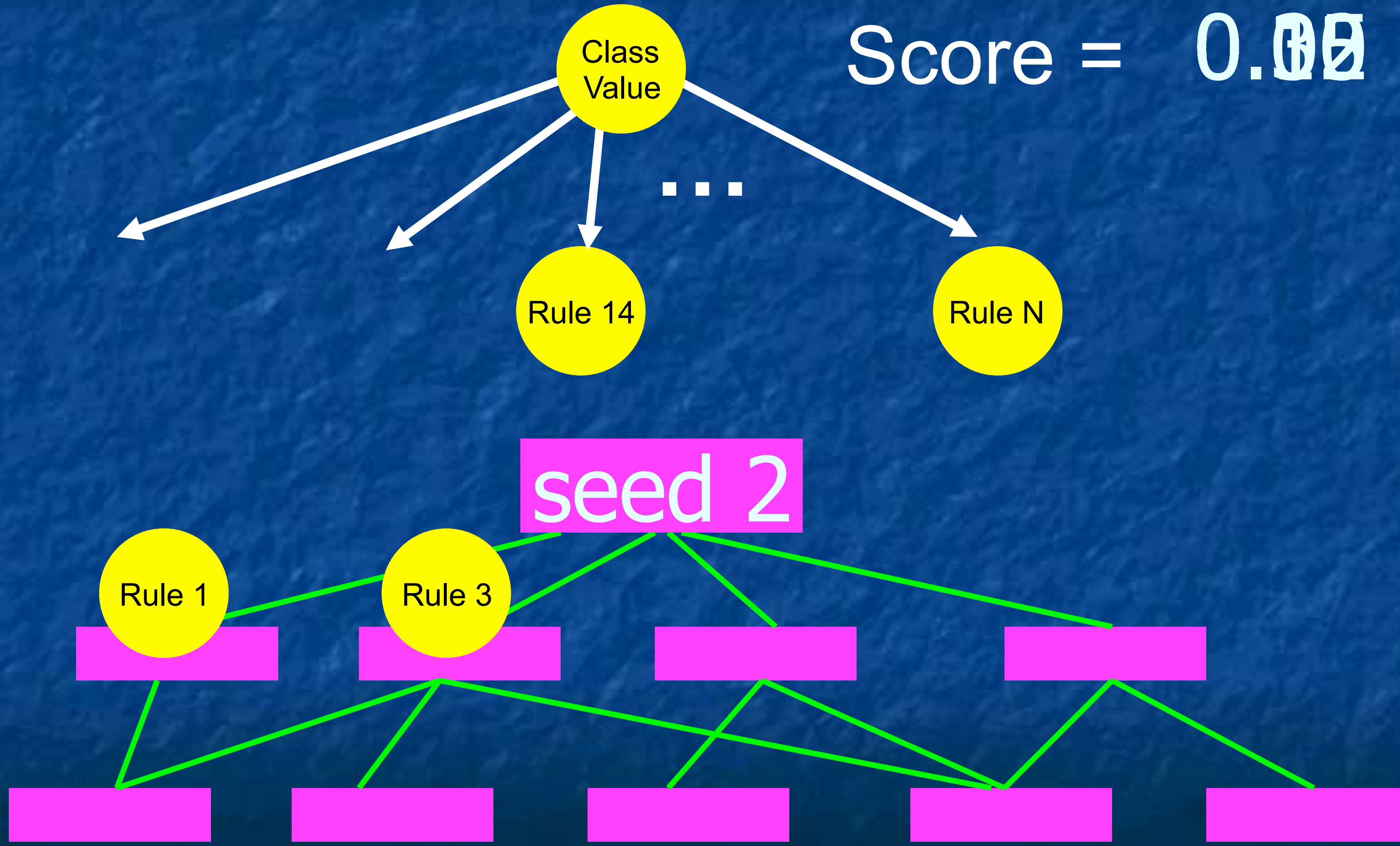


Differences from Standard Rule Learner (Aleph)

- Score rule by adding it to network
- Switch seeds after incorporating a rule into the network

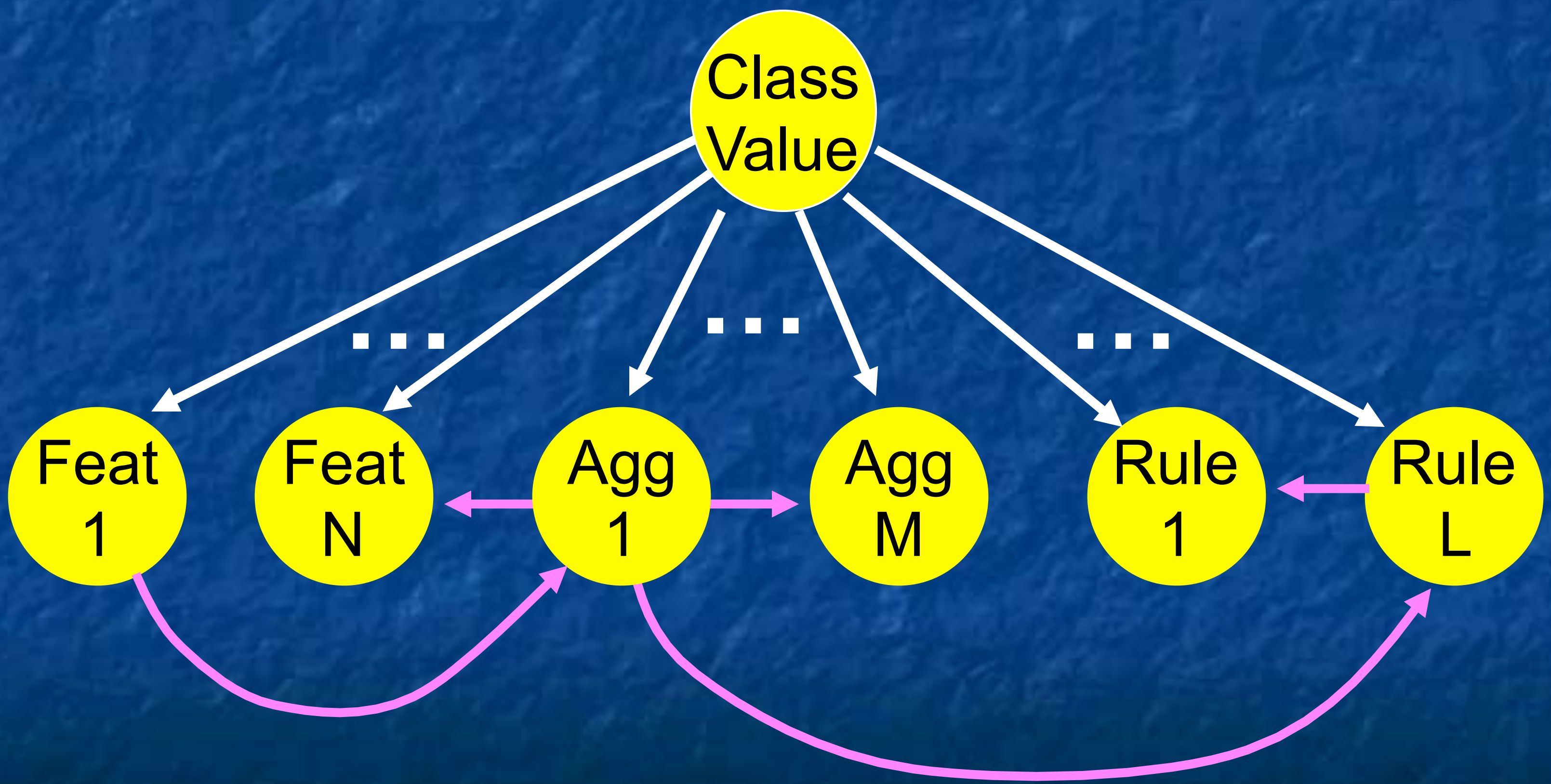
SAYU-NB

Score = 0.



SAYU-View

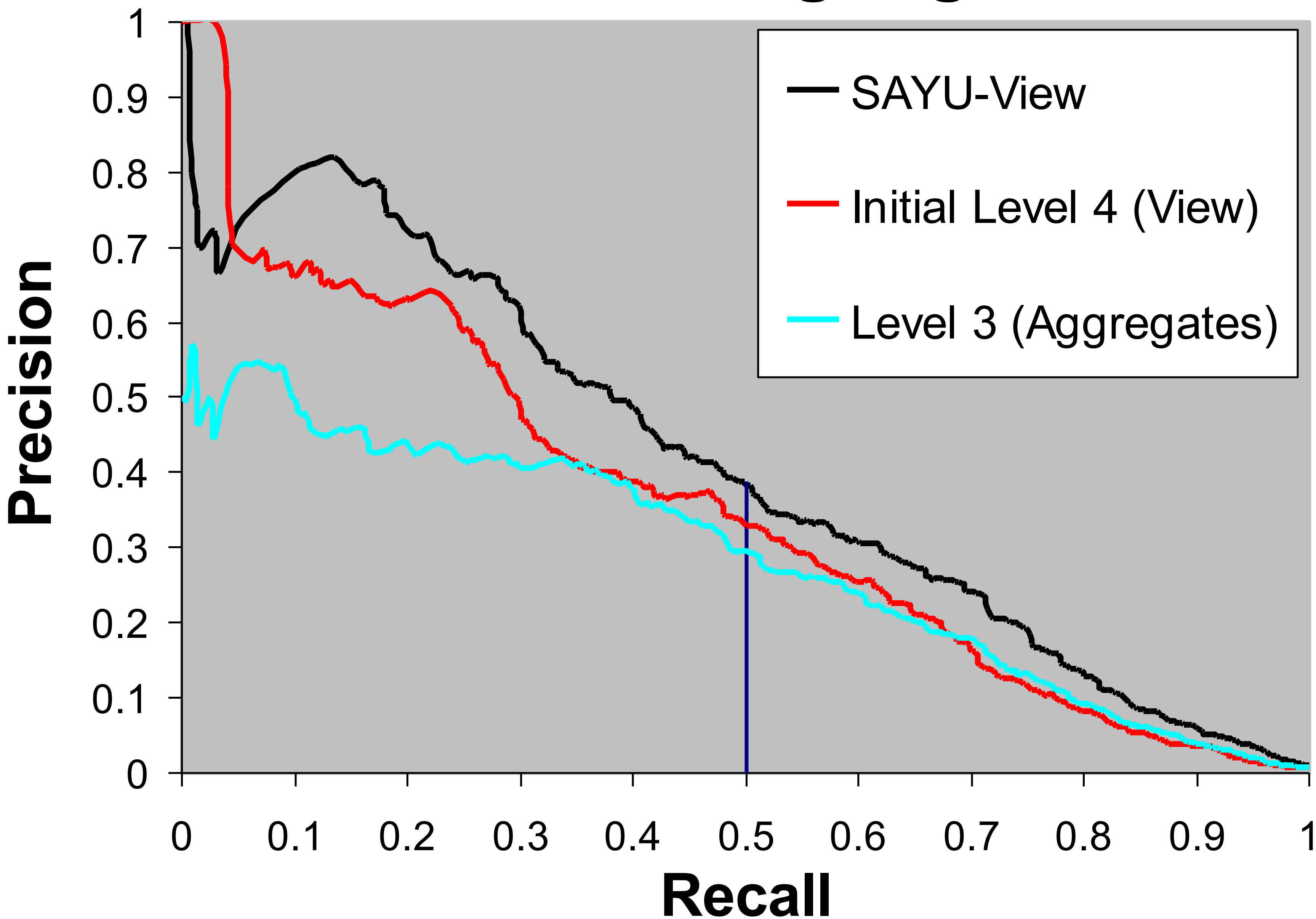
[Davis et al. Intro to SRL 06]



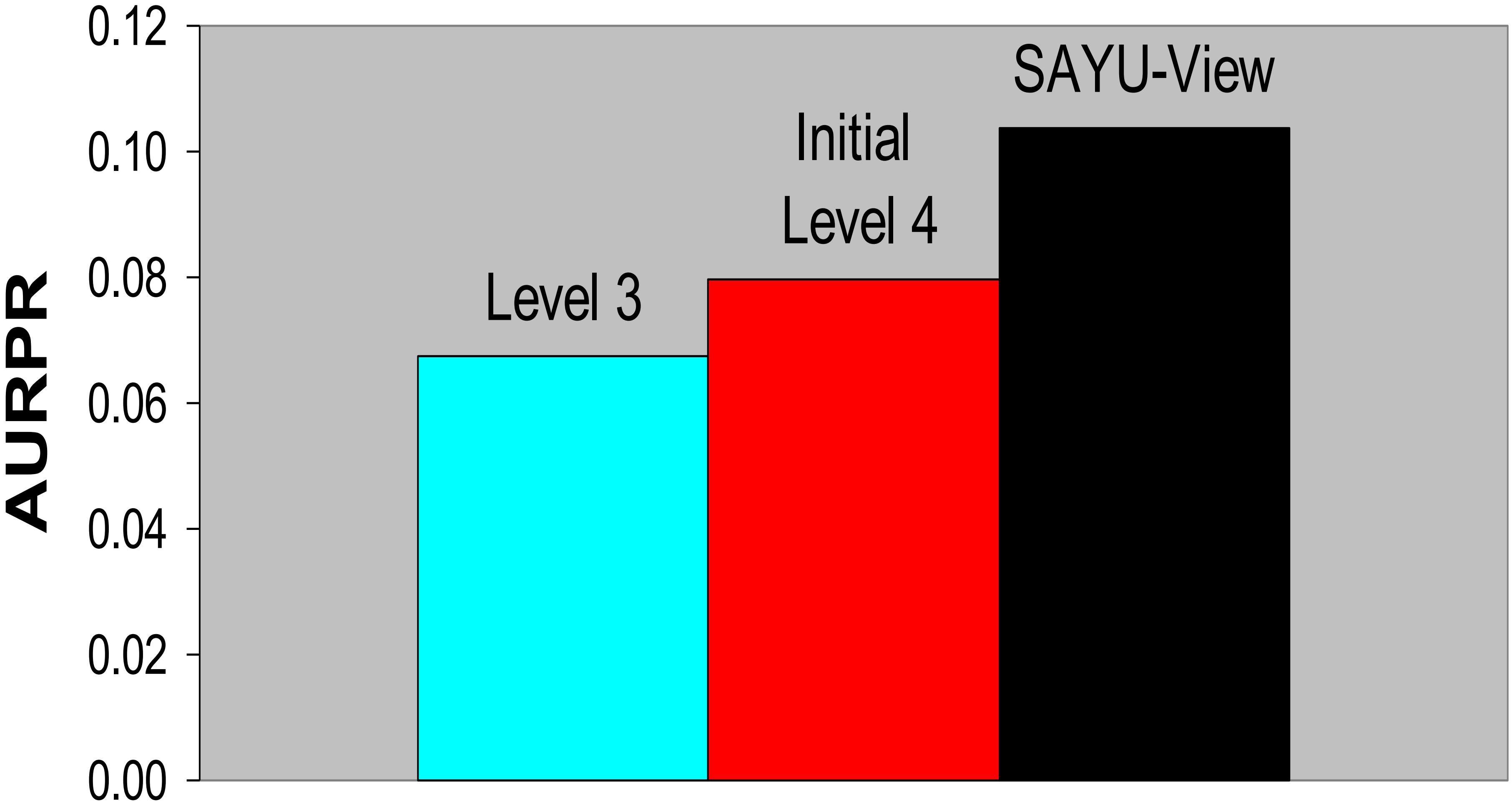
Parameter Settings

- Score using AUC-PR (recall >= .5)
- Keep a rule: 2% increase in AUC
- Switch seeds after adding a rule
- Train set to learn network structure and parameters
- Tune set to score structures

Relational Learning Algorithms



Average Area Under PR Curve For Recall >= 0.5



Acknowledgements

- Jesse Davis (his thesis work)
- Beth Burnside, MD, MPH, Chuck Kahn, MD
- Vitor Santos Costa, Jude Shavlik, Raghu Ramakrishnan
- Funding
  - NCI (R01, UWCCC core grant)
  - NLM (training grant in biomedical informatics)
  - NSF (relational learning)
  - DOD (Air Force – relational learning)

Using Views

```
malignant(A) :-
    archDistortion(A,notPresent),
    prior_mammogram(A,B),
    ho_BreastCA(B,hxDCorLC),
    reasonForMammo(B,s).
```