Logical Differential Prediction Bayes Net, Improving Breast Cancer Diagnosis for Older Women

Houssam Nassif Yirong Wu David Page
Elizabeth Burnside

University of Wisconsin, Madison, USA

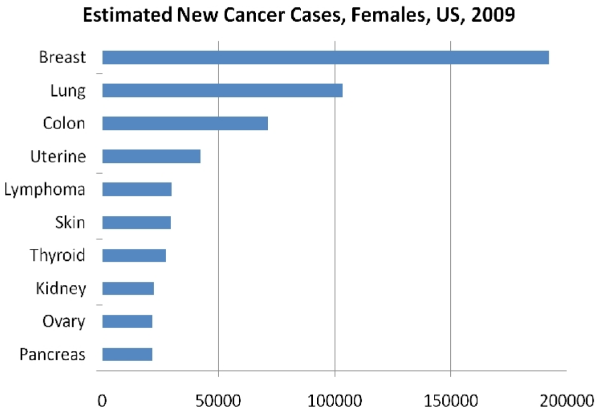
AMIA'12

Outline

1. Motivation
2. Logical Differential Prediction Bayes Net
3. Experiments and Results

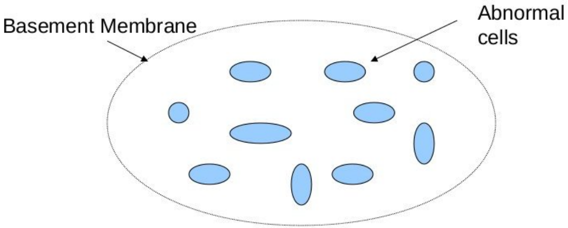
Breast Cancer Occurrences

- Lifetime risk of 14% for US women



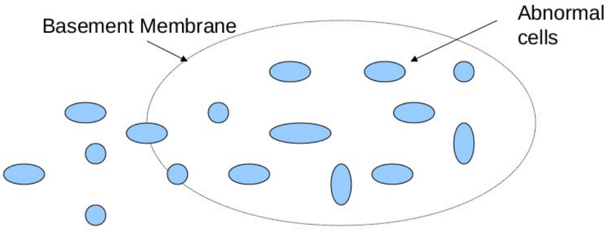
The American Cancer Society, Cancer Facts & Figures, 2009

Breast-Cancer Stages



In-Situ Cancer

Breast-Cancer Stages



Invasive Cancer

Cancer Stage Features

- In Situ can develop into Invasive
  - Current practice: Always treat In Situ
- Time to spread may be very long
  - Patient may die of other causes
  - Over-diagnosis (up to 52%), leading to over-treatment
  - Especially relevant for older patients
- What features characterize In Situ in older patients?
- How to improve diagnostic accuracy for older patients?

Cancer Stage Features

- In Situ can develop into Invasive
  - Current practice: Always treat In Situ
- Time to spread may be very long
  - Patient may die of other causes
  - Over-diagnosis (up to 52%), leading to over-treatment
  - Especially relevant for older patients

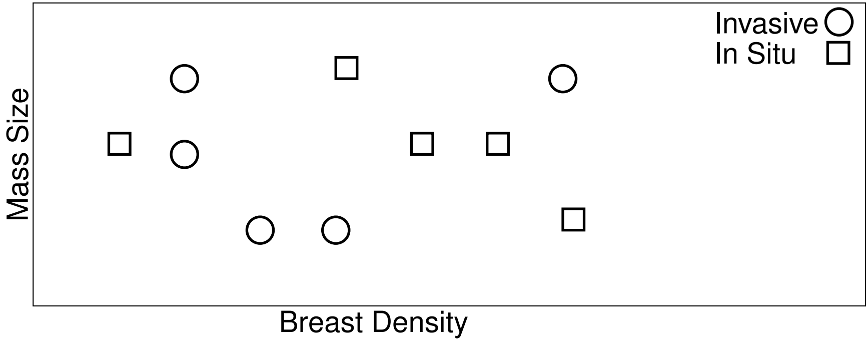
- What features characterize In Situ in older patients?
- How to improve diagnostic accuracy for older patients?

Outline

Differential Prediction

Definition

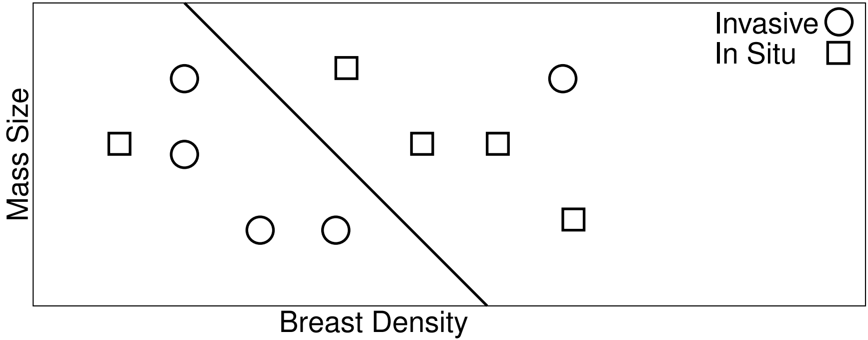
Differential Prediction (DP): Classifier exhibits significant performance differences over particular instance subgroups



Differential Prediction

Definition

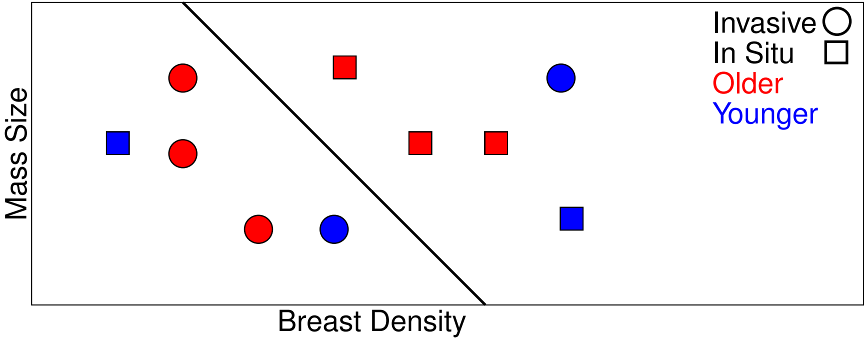
Differential Prediction (DP): Classifier exhibits significant performance differences over particular instance subgroups



Differential Prediction

Definition

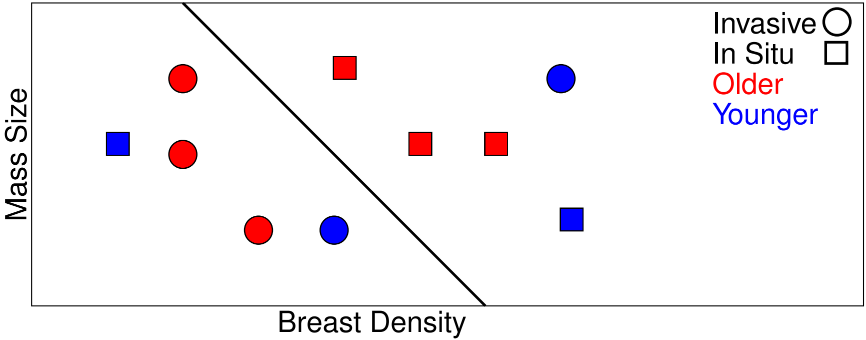
Differential Prediction (DP): Classifier exhibits significant performance differences over particular instance subgroups



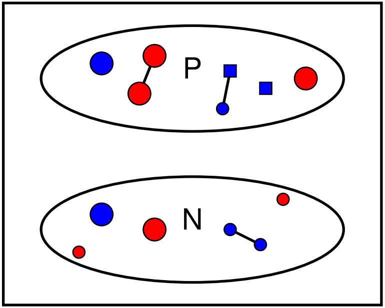
Differential Prediction

Definition

Differential Prediction (DP): Classifier exhibits significant performance differences over particular instance subgroups



Inductive Logic Programming (ILP) Example

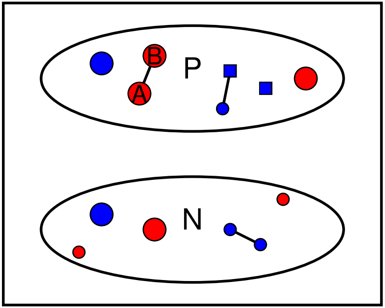


Example

$P(A), red(A), big(A), round(A)$
$sibling(A, B)$

- Pick a positive instance
- $P(X)$ *if* $square(X)$
- $P(X)$ *if* $red(X) \wedge big(X)$
  - 1 false positive
- $P(X)$ *if* $sibling(X, Y) \wedge square(Y)$
- 1 false negative
- Form theory

Inductive Logic Programming (ILP) Example

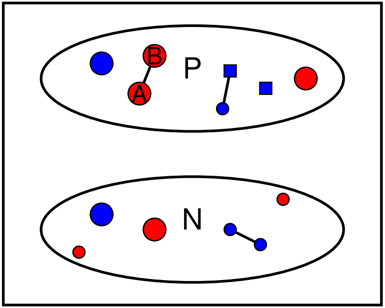


Example

$P(A), red(A), big(A), round(A)$
$sibling(A, B)$

- Pick a positive instance
- $P(X)$ *if* $square(X)$
- $P(X)$ *if* $red(X) \wedge big(X)$
  - 1 false positive
- $P(X)$ *if* $sibling(X, Y) \wedge square(Y)$
- 1 false negative
- Form theory

Inductive Logic Programming (ILP) Example

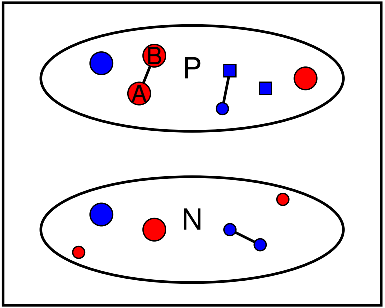


Example

P(*A*), *red*(*A*), *big*(*A*), *round*(*A*)
sibling(*A*, *B*)

- Pick a positive instance
- $P(X)$ if $square(X)$
- $P(X)$ if $red(X) \wedge big(X)$
  - 1 false positive
- $P(X)$ if $sibling(X, Y) \wedge square(Y)$
- 1 false negative
- Form theory

Inductive Logic Programming (ILP) Example

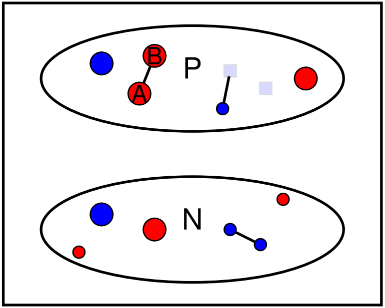


Example

$P(A), red(A), big(A), round(A)$
$sibling(A, B)$

- Pick a positive instance
- $P(X)$ *if* $square(X)$
- $P(X)$ *if* $red(X) \wedge big(X)$
  - 1 false positive
- $P(X)$ *if* $sibling(X, Y) \wedge square(Y)$
- 1 false negative
- Form theory

Inductive Logic Programming (ILP) Example

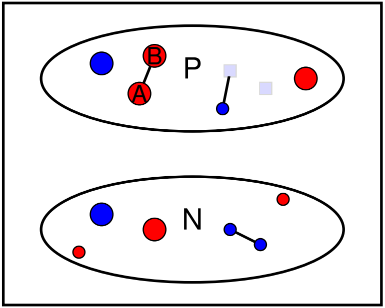


Example

$P(A), red(A), big(A), round(A)$
$sibling(A, B)$

- Pick a positive instance
- $P(X)$ *if* $square(X)$
- $P(X)$ *if* $red(X) \wedge big(X)$
  - 1 false positive
- $P(X)$ *if* $sibling(X, Y) \wedge square(Y)$
- 1 false negative
- Form theory

Inductive Logic Programming (ILP) Example

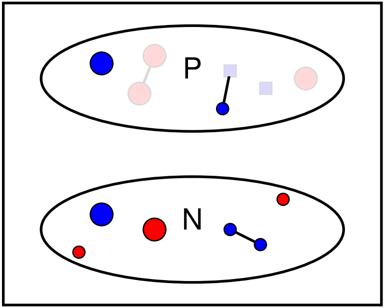


Example

$P(A), red(A), big(A), round(A)$
$sibling(A, B)$

- Pick a positive instance
- $P(X)$ *if* $square(X)$
- $P(X)$ *if* $red(X) \wedge big(X)$
  - 1 false positive
- $P(X)$ *if* $sibling(X, Y) \wedge square(Y)$
- 1 false negative
- Form theory

Inductive Logic Programming (ILP) Example

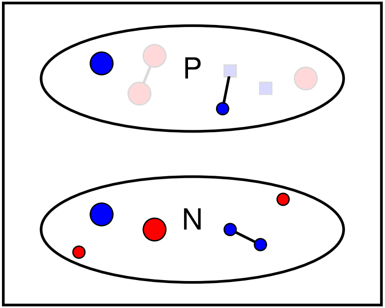


Example

$P(A), red(A), big(A), round(A)$
$sibling(A, B)$

- Pick a positive instance
- $P(X)$ *if* $square(X)$
- $P(X)$ *if* $red(X) \wedge big(X)$
  - 1 false positive
- $P(X)$ *if* $sibling(X, Y) \wedge square(Y)$
- 1 false negative
- Form theory

Inductive Logic Programming (ILP) Example

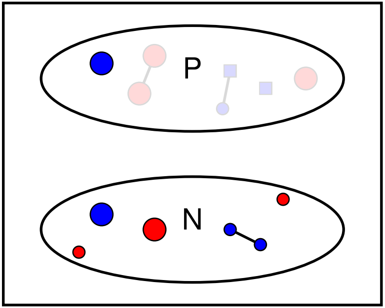


Example

$P(A), red(A), big(A), round(A)$
$sibling(A, B)$

- Pick a positive instance
- $P(X)$ *if* $square(X)$
- $P(X)$ *if* $red(X) \wedge big(X)$
  - 1 false positive
- $P(X)$ *if* $sibling(X, Y) \wedge square(Y)$
- 1 false negative
- Form theory

Inductive Logic Programming (ILP) Example

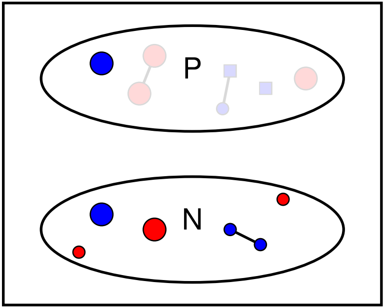


- Pick a positive instance
- $P(X)$ *if* $square(X)$
- $P(X)$ *if* $red(X) \wedge big(X)$
  - 1 false positive
- $P(X)$ *if* $sibling(X, Y) \wedge square(Y)$
- 1 false negative
- Form theory

Example

$P(A), red(A), big(A), round(A)$
$sibling(A, B)$

Inductive Logic Programming (ILP) Example

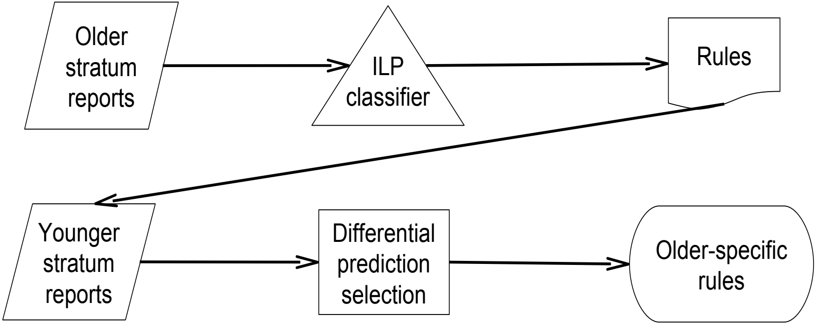


Example

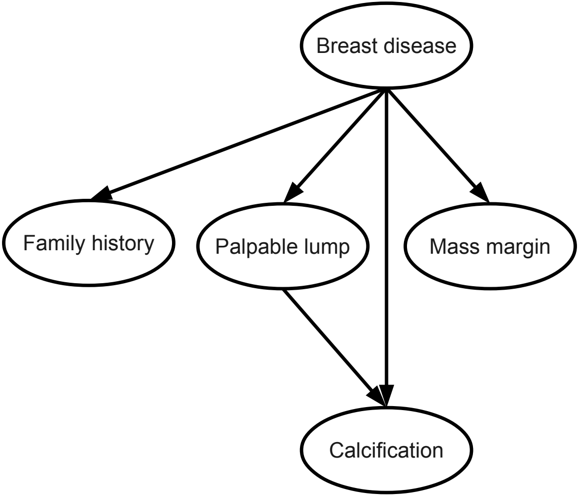
$P(A), red(A), big(A), round(A)$
$sibling(A, B)$

- Pick a positive instance
- $P(X)$ *if* $square(X)$
- $P(X)$ *if* $red(X) \wedge big(X)$
  - 1 false positive
- $P(X)$ *if* $sibling(X, Y) \wedge square(Y)$
- 1 false negative
- Form theory

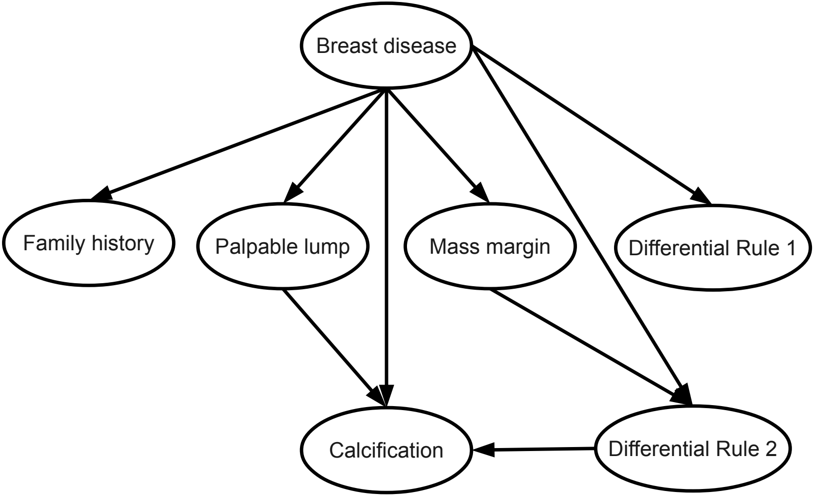
Differential Prediction ILP Approach



Bayes Net


Breast disease
Family history
Palpable lump
Mass margin
Calcification

Logical Differential Prediction Bayes Net



Outline

1. Motivation
2. Logical Differential Prediction Bayes Net
3. Experiments and Results

Dataset

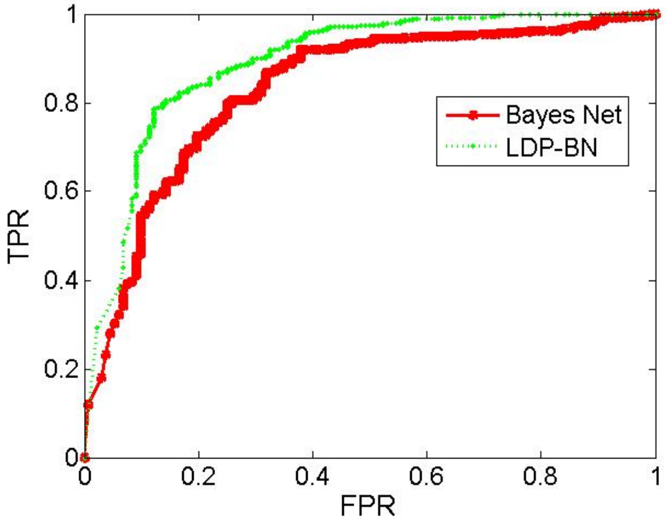
Stratify our data (*Nichols'06*):

Younger: $<$ 50 years, mostly pre-menopausal

Older: $>=$ 65 years, post-menopausal

Stratum	Invasive	In-Situ	Total
Younger	264	110	374
Older	401	132	533
Total	665	242	907

LDP-BN Outperforms BN



Older-Specific In Situ Rules

1. Calcification
   - Tumor indolent in older women
   - Asymptomatic in situ detected due to micro-calcifications
   - Novel finding
2. Class 2 breast density
   - Lower breast density increases mammogram sensitivity, easier micro-calcification detection
3. BI-RADS score increase
4. Prior mammograms
   - Regular screening age > 40

Older-Specific Invasive Rules

1. Mass presence
    - Tumor indolent in older women
    - Once detectable as a mass, is likely invasive
    - Novel finding
2. Prior invasive biopsy
    - Invasive higher risk of proliferation
3. Prior mammograms
    - Regular screening age > 40

Contribution Summary

- Extended Differential Prediction to Bayes Nets
- Proposed a Logical Differential Prediction Bayes Net
- LDP-BN outperforms Bayes Net
- Mined novel older-specific differential predictive mammography rules

- This work is supported by US National Institute of Health (NIH) grant R01-CA165229.

Outline

Rule Selection

ILP rule generation

- $Score(Rule|Older) = positive\_cover(Rule|Older) \times m\_estimate(Rule|Older)$
- $m = 10\%$ in m-estimate
- Recall(Rule|Older) $>= 10\%$
- Precision(Rule|Older) $>= 60\%$

Differential prediction filter

- Recall(Rule|Older) $>=$ Recall(Rule|Younger)
- Precision(Rule|Older) $>=$ Precision(Rule|Younger)
- Precison or recall significantly better

List of Mammography Features

Structured	NLP Extracted (*Nassif'09*)
Family breast cancer history	Mass margin
Personal breast cancer history	Mass shape
Prior surgery	Calcification distribution
Palpable lump	Calcification morphology
Screening v/s diagnostic	Architectural distortion
Indication for exam	Associated findings
Breast Density	Mammary lymph node
BI-RADS code left	Asymmetric breast tissue
BI-RADS code right	Focal asymmetric density
BI-RADS code combined	Tubular density
Principal finding	Mass size

List of ILP extended predicates

first diagnostic mammogram (id)
old study (id, old id)
old biopsy (id, old id, result)
old biopsy same location (id, old id, result)
mass size decrease (id, old id)
mass size increase (id, old id)
this side BI-RADS old study (id, old id, old BI-RADS)
other side BI-RADS old study (id, old id, old BI-RADS)
combined BI-RADS old study (id, old id, old BI-RADS)
this side BI-RADS decrease (id, old id)
other side BI-RADS decrease (id, old id)
combined BI-RADS decrease (id, old id)
this side BI-RADS increase by at least X (id, old id)
other side BI-RADS increase by at least X (id, old id)
combined BI-RADS increase by at least X (id, old id)