

# Logical Differential Prediction Bayes Net, improving breast cancer diagnosis for older women

Houssam Nassif, MS<sup>1</sup>, Yirong Wu, PhD<sup>1</sup>, David Page, PhD<sup>1</sup>, and Elizabeth Burnside, MD,  
MPH, MS<sup>1</sup>

<sup>1</sup> University of Wisconsin, Madison, USA

## Abstract

Overdiagnosis is a phenomenon in which screening identifies cancer which may not go on to cause symptoms or death. Women over 65 who develop breast cancer bear the heaviest burden of overdiagnosis. This work introduces novel machine learning algorithms to improve diagnostic accuracy of breast cancer in aging populations. At the same time, we aim at minimizing unnecessary invasive procedures (thus decreasing false positives) and concomitantly addressing overdiagnosis. We develop a novel algorithm, Logical Differential Prediction Bayes Net (LDP-BN), that calculates the risk of breast disease based on mammography findings. LDP-BN uses Inductive Logic Programming (ILP) to learn relational rules, selects older-specific differentially predictive rules, and incorporates them into a Bayes Net, significantly improving its performance. In addition, LDP-BN offers valuable insight into the classification process, revealing novel older-specific rules that link mass presence to invasive, and calcification presence and lack of detectable mass to DCIS.

## 1 Introduction

Breast cancer is the most common type of cancer among women, with a 12% probability of incidence in a lifetime<sup>4</sup>. Breast cancer has two basic stages: an earlier *in situ* stage where cancer cells are still confined where they developed, and a subsequent *invasive* stage where cancer cells infiltrate surrounding tissue. Since nearly all ductal carcinoma in situ (DCIS) cases can be cured<sup>3</sup>, current practice is to treat DCIS occurrences in order to avoid progression into invasive tumors<sup>4</sup>. However, the time required for a DCIS tumor to reach invasive stage may be sufficiently long for a woman to die of other causes<sup>10,46</sup>; raising the possibility that the diagnosis may not have been necessary, a phenomenon called *overdiagnosis*.

Estimates of cancer overdiagnosis range from 25% – 52% and likely occur more frequently in women with limited life expectancy and indolent (non-invasive and low grade) disease<sup>56,27</sup>. This is mostly the case of older women, since their breast cancers tend to be less aggressive<sup>18,26</sup> and may accompany co-morbidities<sup>25,57</sup>. Furthermore, less aggressive disease as seen in DCIS may not cause morbidity or mortality in older women because of limited life expectancy<sup>17</sup>. While DCIS incidence and detection rate significantly increased over the years in all age groups<sup>6</sup>, its increase rate was most notable in women > 50<sup>54</sup>.

On one hand, it is clear that screening mammography in aging populations diagnoses disease at an earlier stage and reduces recognized age disparities in breast cancer mortality<sup>34,44</sup>. On the other, inherent problems with screening mammography, specifically false positive rates<sup>16</sup> and overdiagnosis<sup>56</sup>, have a substantial influence on the efficacy of screening in older age groups. False positives lead to high rates of breast biopsy, a costly, invasive and potentially painful procedure<sup>42</sup>. In the US, women > 65 are estimated to undergo 140,000 biopsies per year, most of which reveal a benign finding<sup>21</sup>.

The above shortcomings are amplified by the fact that, by 2050, the number of women > 65 is projected to be more than double, and the number of women > 85 to be more than triple, their numbers in 2010<sup>53</sup>. For these reasons, the 2009 US National Institutes of Health consensus conference on DCIS highlighted the need for methods that can accurately identify patient subgroups that would benefit most from treatment, as well as those who do not need treatment<sup>2</sup>. For the latter, the risk of progression would be low enough to employ watchful waiting (mammographic evaluation at short term intervals) rather than biopsy<sup>48</sup>.

In order to advocate for watchful waiting rather than biopsy in women > 65, risk prediction of benign, DCIS, and invasive disease based on mammographic features must be accurate. The literature confirms that the mammographic appearance as described by the radiologist can predict the histology of breast cancer<sup>52,51</sup>. Fortunately, mammography performs superiorly in older women<sup>45</sup>, and mammography features are based on a standardized Breast Imaging

Reporting and Data System (BI-RADS) lexicon<sup>5</sup>. In fact, Bayes Net models built using BI-RADS mammography features can accurately determine breast disease in a general population<sup>8,7</sup>.

Nevertheless, to personalize and optimize breast cancer diagnosis in aging women, we need multirelational algorithms that can address the reality of disease heterogeneity (in our case, based on age), while learning predictive variables for risk prediction in the target population. In this work, we introduce our Logical Differential Prediction Bayes Net (LDP-BN) algorithm, which integrates three machine learning techniques to optimize breast cancer risk prediction in women > 65. These techniques include: 1) leveraging multi-relational data to discover predictive rules via Inductive Logic Programming (ILP), 2) addressing breast cancer heterogeneity by performing differential prediction over age, and 3) incorporating these predictive logical rules, tailored to women > 65, into a Bayes Net for classification/risk prediction.

Our aim is to improve the diagnosis of invasive breast cancer while minimizing unnecessary invasive procedures (decrease false positives) and concomitantly address overdiagnosis. On a dataset composed of biopsy-confirmed invasive and DCIS mammography records, LDP-BN shows a statistically significant improved predictive power in women > 65 as compared to the base case Bayes Net model in this age group.

## 2 Background

Classical classification problems focus on segregating between two or more target classes, while maximizing a given statistic (e.g., accuracy, precision, recall, area under the curve)<sup>35</sup>. Nevertheless, the predictive power of a classifier can vary across the input space and the classifier may exhibit significant differences over particular instance subgroups. A classifier is *differentially predictive* when it behaves differently over the input space, making consistent nonzero errors of prediction for members of a given subgroup<sup>9</sup>. Capturing and modeling this differential prediction allows for a deeper understanding of the underlying problem, context-specific decision making, and identification of diverging data subsets.

Differential prediction was originally used in psychology to assess the fairness of cognitive and educational tests. It is detected by fitting a common regression equation and checking for systematic prediction discrepancies for given subgroups, or by building regression models for each subgroup and testing for differences between the resulting models<sup>30,58</sup>. An example is assessing how college admission test scores predict first year cumulative grades for males and females. For each gender group, we fit a regression model. We then compare the slope, intercept and/or standard errors for both models. If they differ, then the test exhibits differential prediction and may be considered unfair.

An important application of differential prediction is in marketing studies, where it can be used to understand the best targets for an advertising campaign and is often known as uplift modeling. Seminal work includes true response modeling<sup>43</sup>, true lift model<sup>32</sup>, and incremental value modeling<sup>24</sup>. As an example, one group combined a regression and a decision tree model to identify customers for whom direct marketing has sufficiently large impact<sup>24</sup>. The splitting criterion is obtained by computing the difference between the estimated probability increase for the attribute on the treatment set and the estimated probability increase on the control set.

The classification literature, especially in the medical domain, has extended the differential prediction concept to differences in predicted performance when an instance is classified into one condition rather than into another<sup>49</sup>. Hence differential prediction is detected by comparing the performance of different classifiers on the same subgroup (e.g.<sup>15</sup>), or the same classifier on different subgroups (e.g.<sup>41,55</sup>).

To the best of our knowledge, our group was the first to explicitly identify relational rules that achieve a differential prediction across given subsets<sup>37,38</sup>. In a prior work, given a 2-strata (older/younger) 2-class (invasive/DCIS) data, we divided each stratum into a training and a testing set<sup>37</sup>. Aiming at uncovering age-specific invasive and DCIS breast cancer rules, we trained an ILP model on one stratum training subset, and tested its resulting rules on the same-stratum and different-stratum testing subsets. We reported rules which exceed precision and recall thresholds over their same-stratum testing set, and whose precision is significantly better on the same-stratum as compared to the different-stratum testing sets.

### 3 Logical Differential Prediction Bayes Net

In this work, we present the Logical Differential Prediction Bayes Net (LDP-BN) algorithm, which extends our previous ILP-based differential rule learner and incorporates the differential rules in a Bayes Net.

Inductive Logic Programming (ILP) is a commonly used machine learning approach for relational rule learning<sup>13</sup>. ILP generates a hypothesis composed of a set of logical if-then rules that cover most of the positive examples, and as few negative examples as possible.

We use the ILP system Aleph<sup>50</sup>, which is based on the Progol algorithm<sup>36</sup>. Progol’s main advantage is the use of a bottom clause to guide the search. Aleph randomly selects a positive example  $pos(x_i)$  and searches for the most specific hypothesis  $\perp_i$  that, together with the background knowledge  $B$ , entails  $x_i$ :  $(B \wedge \perp_i \wedge x_i) \vdash pos(x_i)$ . This is the “saturation” step, and  $\perp_i$  is the bottom clause for example  $i$ . The use of a bottom clause ensures that, by construction, all clauses in a refinement graph search are guaranteed to cover at least the example associated with the bottom clause.

Aleph then performs a general-to-specific top-down hypothesis space search, bounded by the most general possible hypothesis and by the bottom clause. To do so, Aleph guides the search using the bottom clause. Starting with the most general hypothesis  $pos(\mathbf{X})$ , Aleph refines the clause by repeatedly adding literals from the bottom-clause. This process is the “reduction” step. Algorithm 1 highlights the major steps of Aleph.

---

#### Algorithm 1 Aleph

---

**Require:** Examples  $E$ , mode declarations  $M$ , background knowledge  $B$ , Scoring function  $S$

```

Learned_rules  $\leftarrow \{\}$ 
Pos  $\leftarrow$  all positive examples in  $E$ 
while Pos do
    Select example  $e \in Pos$ 
    Construct bottom clause  $\perp_e$  from  $e$ ,  $M$  and  $B$  ▷ Saturation step
    Candidate_literals  $\leftarrow$  Literals( $\perp_e$ )
    New_rule  $\leftarrow pos(\mathbf{X})$  ▷ Most general rule
    repeat ▷ Top-down reduction step
         $Best\_literal \leftarrow \underset{L \in Candidate\_literals}{\operatorname{argmax}} S(New\_rule \text{ with precondition } L)$ 
        add Best_literal to preconditions of New_rule
    until No more  $S(New\_rule)$  score improvement
    Learned_rules  $\leftarrow Learned\_rules + New\_rule$ 
    Pos  $\leftarrow Pos - \{\text{members of } Pos \text{ covered by } New\_rule\}$ 
end while
return Learned_rules

```

---

To learn older-specific rules, we start by constructing an ILP model that learns rules discriminating between DCIS and invasive over the older cohort. By construction, the resulting rules perform well over the older stratum. We then test each one of these rules on the younger stratum, and keep rules that perform poorly. The greater a rule’s performance difference between the older and younger strata, the more differential predictive this rule is. Figure 1 is a flowchart of this ILP-based differentially predictive algorithm.

We use  $m$ -estimate to represent the probability of an example given a rule. We set both  $m$  and the minimum number of positive examples to be covered by an acceptable rule to 10% of the number of positive examples per stratum and class. Given a rule  $R$  covering  $P(R)$  positives and  $N(R)$  negatives over data  $D$ , with  $Prior$  being the fraction of positive examples in the data  $D$ ,  $m$ -estimate is computed as:

$$mestimate(R|D) = (P(R) + m \times Prior) \div (P(R) + N(R) + m). \quad (1)$$

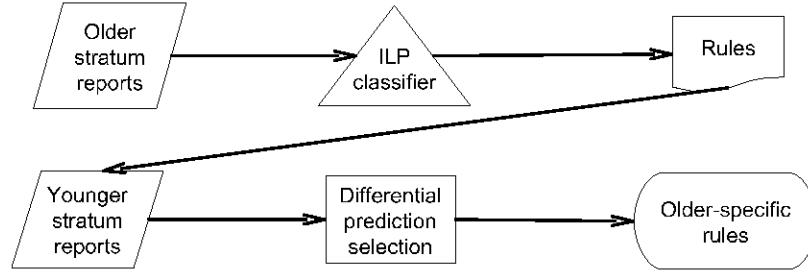


Figure 1: Differential Prediction approach to identify older-specific logical rules

When constructing rules over the older dataset, we score each rule  $R$  by considering its positive cover and  $m$ -estimate:

$$S(R|older) = poscover(R|older) \times mestimate(R|older). \quad (2)$$

The benefit of using ILP in differential prediction is twofold. First, we can use a first-order logic formulation to represent complex relational patterns. Second, we shall take advantage of ILP’s ability to learn easy-to-understand logical if-then rules, whereas each individual rule can be viewed as a concept. We can thus identify rules that only apply to particular data subsets, isolate subgroups covered by a particular rule, and incorporate the differential rules into a Bayes Net as additional features.

Bayesian Belief Networks (Bayes Nets) are informatics tools used for predicting the probability (risk) of an outcome based on observed variables. Bayes Nets predict the probability of an outcome using a graphical structure encoding variables (nodes), conditional dependence relationships (arcs) and probabilities quantified in conditional probability tables associated with each node<sup>35</sup>.

Given feature vectors composed of discrete variables, Bayes Nets can be learned directly from data. Using various heuristic search techniques, the objective is to infer a network that best represents the training data probability distribution. After the Bayes Net structure is determined, conditional probability tables are computed using standard occurrence counting techniques<sup>35</sup>.

We use the Tree Augmented Naive Bayes (TAN) algorithm<sup>19</sup>. TAN starts with a Naive Bayes structure: the class variable has no parents, and is itself the sole parent of each attribute. TAN then adds arcs between variables to approximate the interactions between attributes. It uses a tree structure to ensure that each attribute has at most one other attribute augmenting edge pointing to it.

Once we generate ILP differential rules, we incorporate them as additional variables in the original data feature vector. Each rule can be seen as a binary variable: a given example is either covered or not covered by that rule. We then learn a Bayes Net over the rule-augmented data. This resulting Bayes Net is a Logical Differential Prediction Bayes Net.

#### 4 Materials and Methods

We apply LDP-BN to the same 2-strata (older/younger) 2-class (invasive/DCIS) dataset containing consecutive mammography studies spanning 1997-2004 from the University of California, San Francisco<sup>37</sup>. The dataset age-based separation is correlated with menopausal status. Whereas the younger cohort ( $< 50$ ) is mostly premenopausal, the middle cohort ( $50 - 65$ ) contains most perimenopausal, and the older cohort ( $> 65$ ) is mostly postmenopausal.

To accentuate age-based differences, we limit our age-based analysis to the younger and older cohorts. The older cohort is composed of patients aged  $> 65$ , while the younger cohort groups patients  $< 50$ . If age based differences exist, they are most likely explained by steady and gradual changes rather than an abrupt shift at any single age. In fact, early work showed that the assignment of mammography exams into specific age cohorts with a certain cut point (usually at age 50) may not be desirable unless outcomes abruptly change at this cut point<sup>29</sup>. Changes due to menopause do not appear as sharp changes at any specific age when averaged over a population of women. Removing the middle-age group helps impose a more marked distinction between older and younger age groups making potential

| Structured                     | Extracted using NLP <sup>39</sup> |
|--------------------------------|-----------------------------------|
| 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                |
| Left BI-RADS category          | Asymmetric breast tissue          |
| Right BI-RADS category         | Focal asymmetric density          |
| Combined BI-RADS category      | Tubular density                   |
| Principal finding              | Mass size                         |

Table 1: Dataset structured and extracted features

observed differences clearer.

Our dataset contains 401 mammograms in older women diagnosed invasive breast cancer, 132 mammograms in older women diagnosed DCIS, 264 mammograms in younger women diagnosed invasive, and 110 mammograms in younger women diagnosed DCIS. All invasive and DCIS cancers were biopsy-proven. The mammography reports use a structured format that records patient characteristics and examination findings (Table 1). BI-RADS descriptors were extracted from the dictated text<sup>39</sup>.

Breast cancer data is multi-relational, combining clinical data on multiple levels including the patient, mammogram, abnormality, biopsy, pathology, etc. In addition, temporal relationships influence prediction accuracy. For example, a patient’s prior mammogram can profoundly influence the likelihood of breast cancer on a subsequent examination based on changes seen<sup>11</sup>. ILP can naturally represent such relational and temporal relationships without information loss, changes in data frequencies, nor explosion in database size. For that, we need to augment the original data with relational and temporal predicates that link multiple records together (Table 2). Each such predicate is analogous to a table in an entityrelationship database scheme. For example, adding the *old study (id, old id)* predicate allows a rule to link a given patient’s mammogram to her prior observations and records.

|  |
|--|
| 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)   |

Table 2: List of ILP temporal and relational extensional predicates

We use ILP to generate older-specific invasive and DCIS differentially predictive rules. Since we are only interested in older-specific differential prediction in this project, we do not revert the stratum order of Figure 1 and hence we do not learn younger-specific rules. We define older-specific differentially predictive rules as those having a good performance on the older-stratum (recall > 10%, precision > 60%), and a worse performance on the younger-stratum

(both older-stratum precision and recall results are no worse than younger's, one of them being statistically significantly better at the 95% confidence level<sup>22</sup>). We then incorporate all learned differential rules into a Bayes Net for older-specific invasive/DCIS prediction. Given a patient mammography record, our LDP-BN, like the standard Bayes Net, outputs the probability of this record being invasive versus in situ.

To train and test the LDP-BN, we use conventional stratified 10-fold cross validation, which ensures that cases used to train the model are never used for testing that model. Since the rule-learning component of LDP-BN can represent temporal relationships by relating distinct records, we require that all records of the same patient be in the same fold. We construct the ROC curves with the final curve being the result of vertically averaging the 10 curves from the 10 folds. We perform ILP experiments with the YAP Prolog compiler<sup>47</sup>, and construct the Bayes Net using Weka<sup>23</sup>.

## 5 Results

We compare the Logical Differential Prediction Bayes Net with the Bayes Net built without the added differential prediction rules. Table 3 shows the Area Under the ROC Curve for each of the 10 folds. Figure 2 depicts the final ROC curves, with LDP-BN constantly outperforming Bayes Net. The area under the curve is 0.8304 for Bayes Net, and 0.8911 for LDP-BN. The difference is statistically significant, with a paired two-tailed t-test giving  $p$ -value  $< 0.0001$ .

| Fold      | 1      | 2      | 3      | 4      | 5      | 6      | 7      | 8      | 9      | 10     |
|-----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Bayes Net | 0.8067 | 0.8208 | 0.8714 | 0.8438 | 0.7990 | 0.8769 | 0.7183 | 0.9154 | 0.8490 | 0.8154 |
| LDP-BN    | 0.8846 | 0.8996 | 0.8973 | 0.8714 | 0.8615 | 0.9346 | 0.8163 | 0.9654 | 0.9317 | 0.8779 |

Table 3: Area under the ROC curve results for the Bayes Net and LDP-BN algorithms over the 10 folds

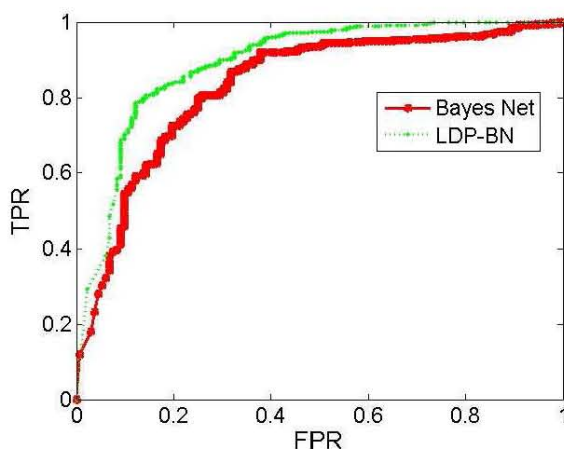


Figure 2: Final ROC curves for the Bayes Net and LDP-BN algorithms

LDP-BN returned several rules, some of which characterize DCIS and others characterize invasive cancers. All these rules were incorporated in the final LDP-BN model. We hereby present their English translation. A mammogram is older-specific DCIS differentially predictive if:

1. The principal finding is calcification or single dilated duct, and the patient had no prior surgery.
2. The principal finding is calcification or single dilated duct, and the currently examined breast had a BI-RADS category of 1 in a previous study.
3. The principal finding is calcification or single dilated duct, and the other side breast had a BI-RADS category of 1 in a previous study.

4. The principal finding is calcification or single dilated duct, and in a previous study both breasts had the same BI-RADS category.
5. The principal finding is calcification or single dilated duct, the breast density is 2, and there is no focal asymmetric density.
6. The currently examined breast BI-RADS category increased by at least three since a previous unilateral exam of the same breast, no reported mass.
7. The currently examined breast had a BI-RADS category of 2 during a prior screening visit, no reported mass.
8. The indication of exam is breast problem other, patient has no prior surgery, and no reported mass.

A mammogram is older-specific invasive differentially predictive if:

9. The patient has a prior invasive biopsy.
10. The currently examined breast's prior first diagnostic exam didn't confirm a malignancy.
11. The principal finding is mass, and the patient does not have a family history of breast cancer.
12. Mass size is between 8 and 20 mm inclusive, and no reported calcification morphology.
13. Mass size is less than or equal to 18 mm, no reported mass shape, and no reported calcification morphology.

## 6 Discussion

Logical Differential Prediction Bayes Net, which incorporates differentially predictive rules into a Bayes Net, clearly improves its classification/risk prediction. The improvement is statistically significant ( $p$ -value  $< 0.0001$ ), and the LDP-BN ROC constantly outperforms Bayes Net's ROC (Figure 2). In addition to improving invasive/DCIS pre-biopsy detection, the differential rules themselves provide valuable insight into each disease's differential features. Several observations emerge from the returned rules.

First we note that the principal mammographic finding is a calcification or a single dilated duct in several older-specific DCIS rules (rules 1 – 5, some of which are redundant). A single dilated duct is a rare finding and was combined with calcification in our data for convenience. Based on these rules, calcification—the more common finding—is a differential predictor of DCIS older patients, which is a novel and interesting result. A possible explanation is that, in asymptomatic women, DCIS disease is often associated with screen-detected micro-calcifications; while in symptomatic women, DCIS is associated with a palpable mass or pathological nipple discharge<sup>40</sup>. DCIS tends to be more indolent, non-palpable, and manifest as micro-calcification in older patients; in contrast to younger women who tend to have more rapidly proliferating cancers that develop into a palpable mass<sup>20</sup>. This previously unreported finding merits further investigation.

Interestingly, rule 5 combines DCIS and a specific type of breast density, breast density class 2 out of an increasing density scale of 1 – 4. This is a relatively low breast density, more common in older women, since breast density decreases with age<sup>28</sup>. A lower breast density significantly increases mammogram sensitivity<sup>33</sup>, allowing for easier micro-calcification detection, as is captured by rule 5.

In rule 6, the increase in the examined breast BI-RADS category indicates DCIS. The BI-RADS category summarizes the examining radiologist's opinion and findings concerning the mammogram<sup>5</sup>. It takes values {1, 2, 3, 0, 4, 5}, in increasing order of malignancy probability. An increase in the BI-RADS category over multiple visits reflects increasing suspicion of malignancy. This may be a more pronounced feature in older women because they tend to have more prior mammograms.

The next observation spans both invasive and DCIS rules, and links the current mammogram with previous ones (rules 2, 3, 4, 6, 7, 10). This too may be a more pronounced feature in older women because they tend to have more prior



mammograms than younger women. In fact, regular screening mammography is usually recommended for women aged 40 and above.

Unlike DCIS rules which require a lack of prior surgery (rules 1, 8), older-specific invasive rule 9 specifies a prior invasive biopsy. A prior biopsy revealing invasive disease is thus a predictor of invasive recurrence in older women. This may reflect the higher risk of proliferation and recurrence of invasive tumors<sup>31</sup> which, combined with a longer life-span for the recurrence to manifest itself, is more common in older women.

The last observation is the presence of a mass for older-specific invasive differential prediction (rules 11 – 13), and its absence for DCIS (rules 6 – 8). Studies have shown that breast cancer in younger women is pathophysiologically more aggressive and has a poorer prognosis<sup>14,20</sup>. Younger women tend to have higher proportions of poorly differentiated, rapidly proliferating tumors (be it invasive or DCIS) that tend to be larger<sup>1</sup>. This increases the likelihood of a mass associated with a DCIS tumor in younger women; explaining why the lack of a reported mass is differentially predictive of DCIS in older women. Concomitantly, tumors in older women tend to grow at a slower rate, and once it is detectable as a mass, it may more likely be invasive. This novel finding merits further investigation.

## 7 Future Work

This work can be extended in multiple directions. First, we note that LDP-BN rules are learned for their differential predictive potential, separately from the Bayes Net. Integrating the differential rules identification and the Bayesian Network construction into a global optimization framework may result in a better performance<sup>12</sup>.

Second, our differential prediction approach is a generate-then-test method. Target stratum rules are generated by training solely on the target stratum subset, and then filtered by testing on the other stratum. A more rigorous approach is to use test-incorporation, by altering the ILP search space to be differential-sensitive<sup>38</sup>.

Third, we only proposed solutions for the 2-strata 2-class LDP-BN problem. We plan on exploring multi-strata and multi-class LDP-BN. For instance, a natural extension of our work is to build an LDP-BN that predicts invasive, in situ and benign findings.

Finally, a practical extension of this work would be a nomogram or an online calculator. Such a tool can be used by patients and physicians alike for a more personalized assessment.

## 8 Conclusion

In this work, we present our novel Logical Differential Prediction Bayes Net (LDP-BN) algorithm. LDP-BN uses ILP to discover relational logical rules that predict cancer diagnosis. Furthermore, since breast cancer pathophysiology and prognosis differs based on age, LDP-BN uses differential prediction to extract tailored rules optimizing accuracy for women > 65. Finally, LDP-BN incorporates the differentially predictive logical rules as variables in a Bayes Net. The need is clear: our breast cancer invasive/DCIS data is multi-relational (combining predictive variables on different levels including the patient, mammogram, and pathology report) and contains important temporal relationships (a patient's prior mammogram can profoundly influence the likelihood of breast cancer on a subsequent examination). Our results show that LDP-BN significantly outperforms Bayes Net.

In addition, our differentially predictive rule-discovery approach offers interesting insight into the invasive/DCIS differential space. It matches many of the prior invasive versus DCIS knowledge, and also generates new rules that merit further investigation. Namely, we find novel older-women specific rules that link mass presence to invasive, and lack of detectable mass as well as calcification presence to DCIS.

## References

- [1] S. Acbi and M. Castiglione. The enigma of young age. *Ann. Oncol.*, 17(10):1475–1477, 2006.
- [2] C. J. Allegra, D. R. Aberle, P. Ganschow, S. M. Hahn, C. N. Lee, S. Millon-Underwood, M. C. Pike, S. Reed, A. F. Saftlas, S. A. Searvalone, A. M. Schwartz, C. Slomski, G. Yothers, and R. Zon. National Institutes of Health State-of-the-Science Conference Statement: Diagnosis and Management of Ductal Carcinoma In Situ, September 22–24, 2009. *J. Natl. Cancer Inst.*, 102(3):161–169, 2010.



- [3] American Cancer Society. *Breast Cancer Facts & Figures 2009-2010*. American Cancer Society, Atlanta, USA, 2009.
- [4] American Cancer Society. *Cancer Facts & Figures 2009*. American Cancer Society, Atlanta, USA, 2009.
- [5] American College of Radiology, Reston, VA, USA. *Breast Imaging Reporting and Data System (BI-RADS<sup>TM</sup>)*, 4th edition, 2003.
- [6] W. F. Anderson, K. C. Chu, and S. S. Devesa. Distinct incidence patterns among in situ and invasive breast carcinomas, with possible etiologic implications. *Breast Cancer Res Treat*, 88(2):149–159, 2004.
- [7] E. S. Burnside, J. Davis, J. Chhatwal, O. Alagoz, M. J. Lindstrom, B. M. Geller, B. Littenberg, K. A. Shaffer, C. E. Kahn, and D. Page. Probabilistic computer model developed from clinical data in national mammography database format to classify mammographic findings. *Radiology*, 251:663–672, 2009.
- [8] E. S. Burnside, D. L. Rubin, and R. D. Shachter. Using a Bayesian network to predict the probability and type of breast cancer represented by microcalcifications on mammography. *Stud Health Technol Inform*, 107(Pt 1):13–17, 2004.
- [9] T. A. Cleary. Test bias: Prediction of grades of negro and white students in integrated colleges. *Journal of Educational Measurement*, 5(2):115–124, 1968.
- [10] L. C. Collins, R. M. Tamimi, H. J. Bacr, J. L. Connolly, G. A. Colditz, and S. J. Schnitt. Outcome of patients with ductal carcinoma in situ untreated after diagnostic biopsy: results from the Nurses' Health Study. *Cancer*, 103(9):1778–1784, 2005.
- [11] J. Davis, E. S. Burnside, I. de Castro Dutra, D. Page, R. Ramakrishnan, V. Santos Costa, and J. Shavlik. View Learning for Statistical Relational Learning: With an application to mammography. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, pages 677–683, Edinburgh, Scotland, 2005.
- [12] J. Davis, E. S. Burnside, I. de Castro Dutra, D. Page, and V. Santos Costa. An integrated approach to learning Bayesian Networks of rules. In *Proceedings of the 16th European Conference on Machine Learning*, pages 84–95, Porto, Portugal, 2005.
- [13] L. De Raedt. *Logical and Relational Learning*. Springer, 2008.
- [14] P. C. Dubsky, M. F. X. Gnant, S. Taucher, S. Roka, D. Kandioler, B. Pichler-Gebhard, I. Agstner, M. Seifert, P. Sevelada, and R. Jakesz. Young age as an independent adverse prognostic factor in premenopausal patients with breast cancer. *Clin. Breast Cancer*, 3:65–72, 2002.
- [15] S. L. Dugdaleby, A. A. Jackson, K. M. Godfrey, S. M. Robinson, H. M. Inskip, and the Southampton Womens Survey Study Group. Cut-off points for anthropometric indices of adiposity: differential classification in a large population of young women. *British Journal of Nutrition*, 101:424–430, 2009.
- [16] J. G. Elmore, M. B. Barton, V. M. Mocerri, S. Polk, P. J. Arena, and S. W. Fletcher. Ten-year risk of false positive screening mammograms and clinical breast examinations. *N Engl J Med*, 338(16):1089–1096, 1998.
- [17] V. L. Ernster, R. Ballard-Barbash, W. E. Barlow, Y. Zheng, D. L. Weaver, G. Cutter, B. C. Yankaskas, R. Rosenberg, P. A. Carney, K. Kerlikowske, S. H. Taplin, N. Urban, and B. M. Geller. Detection of ductal carcinoma in situ in women undergoing screening mammography. *J Natl Cancer Inst*, 94(20):1546–1554, 2002.
- [18] B. L. Fowble, D. J. Schultz, B. Overmoyer, L. J. Solin, K. Fox, L. Jardines, S. Orel, and J. H. Glick. The influence of young age on outcome in early stage breast cancer. *Int J Radiat Oncol Biol Phys*, 30(1):23–33, 1994.
- [19] N. Friedman, D. Geiger, and M. Goldszmidt. Bayesian network classifiers. *Machine Learning*, 29:131–163, 1997.
- [20] C. Gajdos, P. I. Tartert, I. J. Bleiweiss, C. Bodian, and S. T. Brower. Stage 0 to stage III breast cancer in young women. *J. Am. Coll. Surg.*, 190(5):523–529, 2000.
- [21] K. Ghosh, L. J. r. Melton, V. J. Suman, C. S. Grant, S. Sterioff, K. R. Brandt, C. Branch, T. A. Sellers, and L. C. Hartmann. Breast biopsy utilization: a population-based study. *Arch Intern Med*, 165(14):1593–1598, 2005.
- [22] C. Goutte and E. Gaussier. A probabilistic interpretation of precision, recall and  $F$ -score, with implication for evaluation. In *Proc. of the 27th European Conference on IR Research*, pages 345–359, Santiago de Compostela, Spain, 2005.
- [23] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The WEKA Data Mining Software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18, 2009.
- [24] B. Hansotia and B. Rukstales. Incremental value modeling. *Journal of Interactive Marketing*, 16(3):35–46, 2002.
- [25] E. F. Harris, W. T. Hwang, S. L. Urtishak, J. Plastaras, B. Kinonian, and S. L. J. The impact of comorbidities on outcomes for elderly women treated with breast-conservation treatment for early-stage breast cancer. *Int J Radiat Oncol Biol Phys*, 70(5):1453–1459, 2008.
- [26] U. W. Jayasinghe, R. Taylor, and J. Boyages. Is age at diagnosis an independent prognostic factor for survival following breast cancer? *ANZ J Surg*, 75(9):762–767, 2005.
- [27] K. J. Jorgensen and P. C. Gotzsche. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. *BMJ*, 339:b2587, 2009.
- [28] L. F. Kelemen, V. S. Pankratz, T. A. Sellers, K. R. Brandt, A. Wang, C. Janney, Z. S. Fredericksen, J. R. Cerhan, and C. M. Vachon. Age-specific trends in mammographic density. *American Journal of Epidemiology*, 167(9):1027–1036, 2008.
- [29] D. B. Kopans, R. H. Moore, K. A. McCarthy, D. A. Hall, C. A. Hulka, G. J. Whitman, P. J. Slanetz, and E. F. Halpern. Positive predictive value of breast biopsy performed as a result of mammography: there is no abrupt change at age 50 years. *Radiology*, 200(2):357–360, August 1996.

- [30] R. L. Linn. Single-group validity, differential validity, and differential prediction. *Journal of Applied Psychology*, 63:507–512, 1978.
- [31] Y. Liu, M. Pérez, M. Schootman, R. L. Aft, W. E. Gillanders, M. J. Ellis, and D. B. Jeffe. A longitudinal study of factors associated with perceived risk of recurrence in women with ductal carcinoma in situ and early-stage invasive breast cancer. *Breast Cancer Res. Treat.*, Epub ahead of print, 2010.
- [32] V. S. Lo. The true lift model - a novel data mining approach to response modeling in database marketing. *SIGKDD Explorations*, 4(2):78–86, 2002.
- [33] M. T. Mandelson, N. Oestreicher, P. L. Porter, D. White, C. A. Finder, S. H. Taplin, and E. White. Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. *J. Natl. Cancer Inst.*, 92(13):1081–1087, 2000.
- [34] E. P. McCarthy, R. B. Burns, K. M. Freund, A. S. Ash, M. Shwartz, S. L. Marwill, and M. A. Moskowitz. Mammography use, breast cancer stage at diagnosis, and survival among older women. *J Am Geriatr Soc*, 48(10):1226–1233, 2000.
- [35] T. M. Mitchell. *Machine Learning*. McGraw-Hill International Editions, Singapore, 1997.
- [36] S. H. Muggleton. Inverse entailment and Progol. *New Generation Computing*, 13:245–286, 1995.
- [37] H. Nassif, D. Page, M. Ayvaci, J. Shavlik, and E. S. Burnside. Uncovering age-specific invasive and DCIS breast cancer rules using Inductive Logic Programming. In *1st ACM International Health Informatics Symposium*, pages 76–82, Arlington, VA, 2010.
- [38] H. Nassif, V. Santos Costa, E. S. Burnside, and D. Page. Relational differential prediction. In *ECML-PKDD 2012*, Bristol, UK, 2012. Accepted.
- [39] H. Nassif, R. Wood, E. S. Burnside, M. Ayvaci, J. Shavlik, and D. Page. Information extraction for clinical data mining: A mammography case study. In *ICDM Workshops*, pages 37–42, Miami, Florida, 2009.
- [40] N. Patani, B. Cutuli, and K. Mokbel. Current management of DCIS: a review. *Breast Cancer Res Treat*, 111(1):1–10, 2008.
- [41] B. Phibbs and W. Nelson. Differential classification of acute myocardial infarction into ST- and Non-ST segment elevation is not valid or rational. *Annals of Noninvasive Electrocardiology*, 15(3):191–199, 2010.
- [42] S. P. Poplack, P. A. Carney, J. E. Weiss, L. Titus-Ernstoff, M. E. Goodrich, and A. N. Tosteson. Screening mammography: costs and use of screening-related services. *Radiology*, 234(1):79–85, 2005.
- [43] N. J. Radcliffe and P. D. Surry. Differential response analysis: Modeling true response by isolating the effect of a single action. In *Credit Scoring and Credit Control VI*, Edinburgh, Scotland, 1999.
- [44] W. M. Randolph, J. S. Goodwin, J. D. Mahnen, and J. L. Freeman. Regular mammography use is associated with elimination of age-related disparities in size and stage of breast cancer at diagnosis. *Ann Intern Med*, 137(10):783–790, 2002.
- [45] R. D. Rosenberg, W. C. Hunt, M. R. Williamson, F. D. Gilliland, P. W. Wiest, C. A. Kelsey, C. R. Key, and M. N. Linver. Effects of age, breast density, ethnicity, and estrogen replacement therapy on screening mammographic sensitivity and cancer stage at diagnosis: review of 183,134 screening mammograms in Albuquerque, New Mexico. *Radiology*, 209(2):511–518, 1998.
- [46] M. E. Sanders, P. A. Schuyler, W. D. Dupont, and D. L. Page. The natural history of low-grade ductal carcinoma in situ of the breast in women treated by biopsy only revealed over 30 years of long-term follow-up. *Cancer*, 103(12):2481–2484, 2005.
- [47] V. Santos Costa. The life of a logic programming system. In M. G. de la Banda and F. Pontelli, editors, *Proceedings of the 24th International Conference on Logic Programming*, pages 1–6, Udine, Italy, 2008.
- [48] S. J. Schnitt. Local outcomes in ductal carcinoma in situ based on patient and tumor characteristics. *J Natl Cancer Inst Monogr*, 2010(41):158–161, 2010.
- [49] Society for Industrial and Organizational Psychology. *Principles for the Validation and Use of Personnel Selection Procedures*, 4th edition, 2003.
- [50] A. Srinivasan. *The Aleph Manual*, 4th edition, 2007.
- [51] L. Tabar, H. H. Tony Chen, M. F. Amy Yen, T. Tot, T. H. Tung, L. S. Chen, Y. H. Chiu, S. W. Duffy, and R. A. Smith. Mammographic tumor features can predict long-term outcomes reliably in women with 1-14-mm invasive breast carcinoma. *Cancer*, 101(8):1745–1759, 2004.
- [52] M. G. Thurfjell, A. Lindgren, and E. Thurfjell. Nonpalpable breast cancer: Mammographic appearance as predictor of histologic type. *Radiology*, 222(1):165–170, 2002.
- [53] G. K. Vincent and V. A. Velkoff. *THE NEXT FOUR DECADES, The Older Population in the United States: 2010 to 2050*. Number P25-1138 in Current Population Reports. U.S. Census Bureau, Washington, DC, 2010.
- [54] B. A. Virnig, S. Y. Wang, T. Shamliyan, R. L. Kane, and T. M. Tuttle. Ductal carcinoma in situ: Risk factors and impact of screening. *J Natl Cancer Inst Monogr*, 2010(41):113–116, 2010.
- [55] A. M. M. Vlaar, A. Bouwmans, W. H. Mess, S. C. Tromp, and W. E. J. Weber. Transcranial duplex in the differential diagnosis of parkinsonian syndromes: a systematic review. *Journal of Neurology*, 256(4):530–538, 2009.
- [56] H. G. Welch and W. C. Black. Overdiagnosis in cancer. *J Natl Cancer Inst*, 102(9):605–613, 2010.
- [57] R. Yancik, M. N. Wesley, L. A. Ries, R. J. Havlik, B. K. Edwards, and J. W. Yates. Effect of age and comorbidity in postmenopausal breast cancer patients aged 55 years and older. *JAMA*, 285(7):885–892, 2001.
- [58] J. W. Young. Differential validity, differential prediction, and college admissions testing: A comprehensive review and analysis. Research Report 2001-6, The College Board, New York, 2001.