Leveraging Expert Knowledge to Improve Machine-Learned Decision Support Systems

Finn Kuusisto, MS[1]; Inês Dutra, PhD[2]; Mai Elezaby, MD[1];
Eneida Mendonça, MD, PhD[1]; Jude Shavlik, PhD[1];
Elizabeth Burnside, MD, MPH, MS[1]

[1]University of Wisconsin, Madison, USA
[2]University of Porto, Portugal

Disclosure

Finn Kuusisto discloses that he has no relationships with commercial interests.

Learning Objective

After participating in this activity, the learner should be better able to:

Collaborate with clinical and/or machine learning experts in decision support system development

Opportunity & Problem

Great opportunities for machine-learned decision support systems

But…

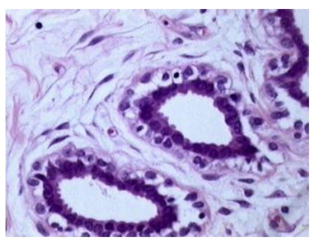
Standardized, complete, and sufficient training data is rarely available

Upgrade Prediction

1
Mammogram

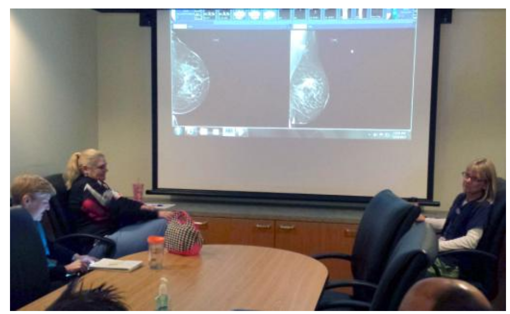
Abnormality

2
Needle Biopsy



Benign Tissue

3
Radiologic-Histologic Correlation



Non-definitive Diagnosis

4
Excision

Final Diagnosis

Malignant
=
“Upgrade”

Image Sources:
1. NIH - wikimedia.org/wiki/File:Woman_receives_mammogram.jpg
2. Itayba - wikimedia.org/wiki/File:Normal.jpg
3. UW Hospital and Clinics
4. NIH - wikimedia.org/wiki/File:Surgical_breast_biopsy.jpg

Upgrade Prediction

- 5-15% of core needle biopsies non-definitive
- Approximately 35,000-105,000* per year
- 80-90% of non-definitive biopsies are **benign**

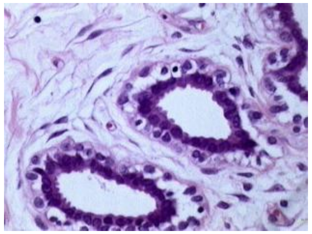
* Based on 2010 annual breast biopsy utilization rate

Upgrade Prediction

1
Mammogram

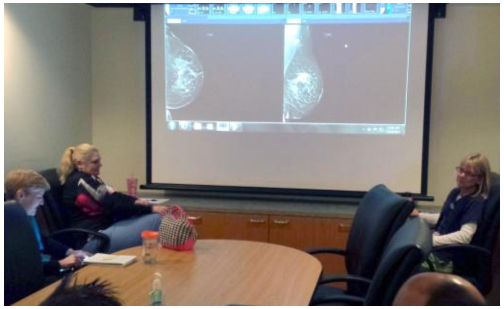
Abnormality

2
Needle Biopsy



Benign Tissue

3
Radiologic-Histologic Correlation



Non-definitive Diagnosis

4
Excision

Final Diagnosis

Malignant
=
"Upgrade"

Image Sources:
1. NIH - wikimedia.org/wiki/File:Woman_receives_mammogram.jpg
2. Itayba - wikimedia.org/wiki/File:Normal.jpg
3. UW Hospital and Clinics
4. NIH - wikimedia.org/wiki/File:Surgical_breast_biopsy.jpg

ABLe

Comprises two parts

1) Definitions of advice sources

2) Iterative process for model refinement

ABLe - Advice Definitions

Task

- What is the problem and scope?
- What predictor variables are important?
- How should the problem be modeled?

ABLe - Advice Definitions

Task

- What is the problem and scope?
- What predictor variables are important?
- How should the problem be modeled?

Variable Relationships

- What combinations of variables are important to the task?

ABLe - Advice Definitions

Task

- What is the problem and scope?
- What predictor variables are important?
- How should the problem be modeled?

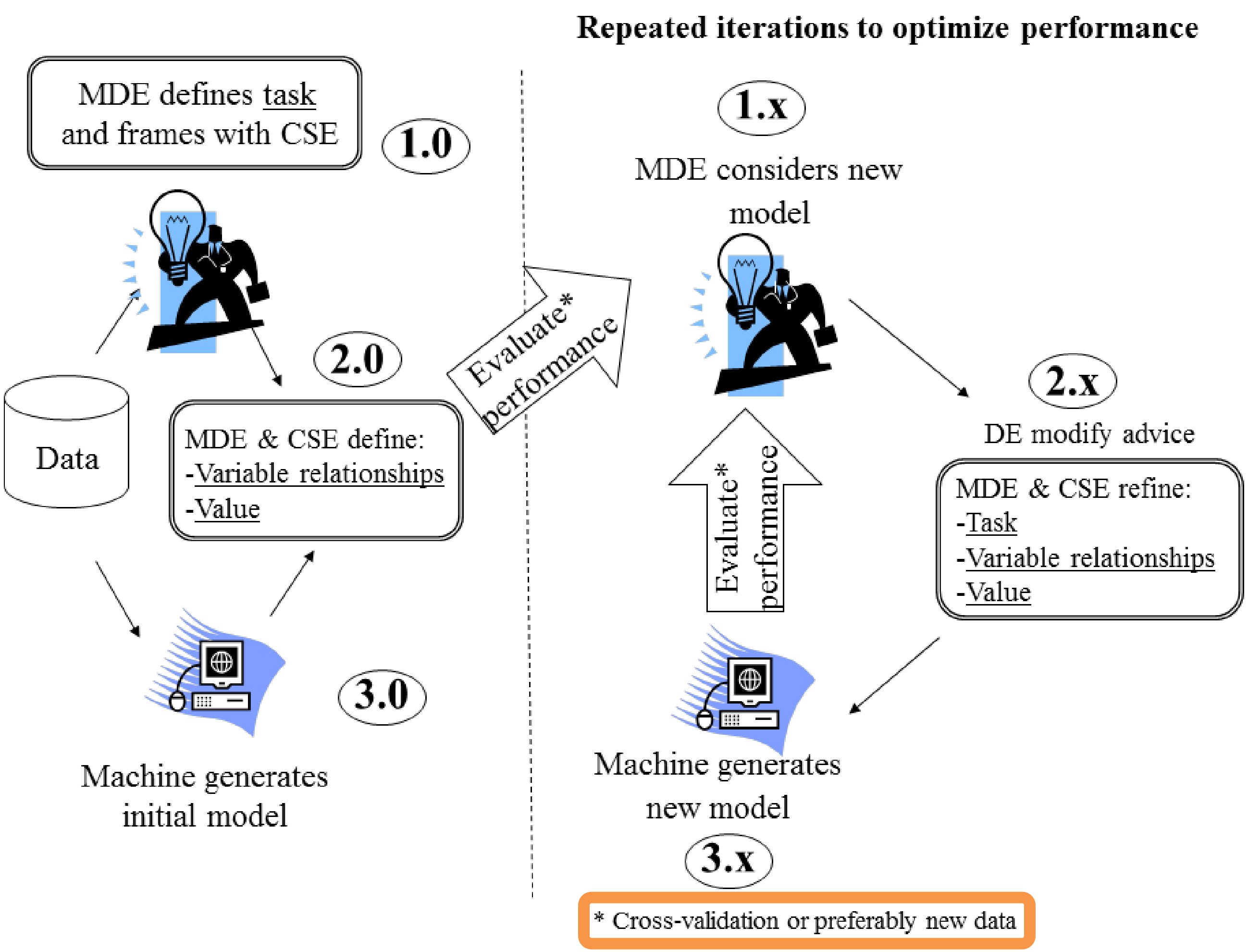
Variable Relationships

- What combinations of variables are important to the task?

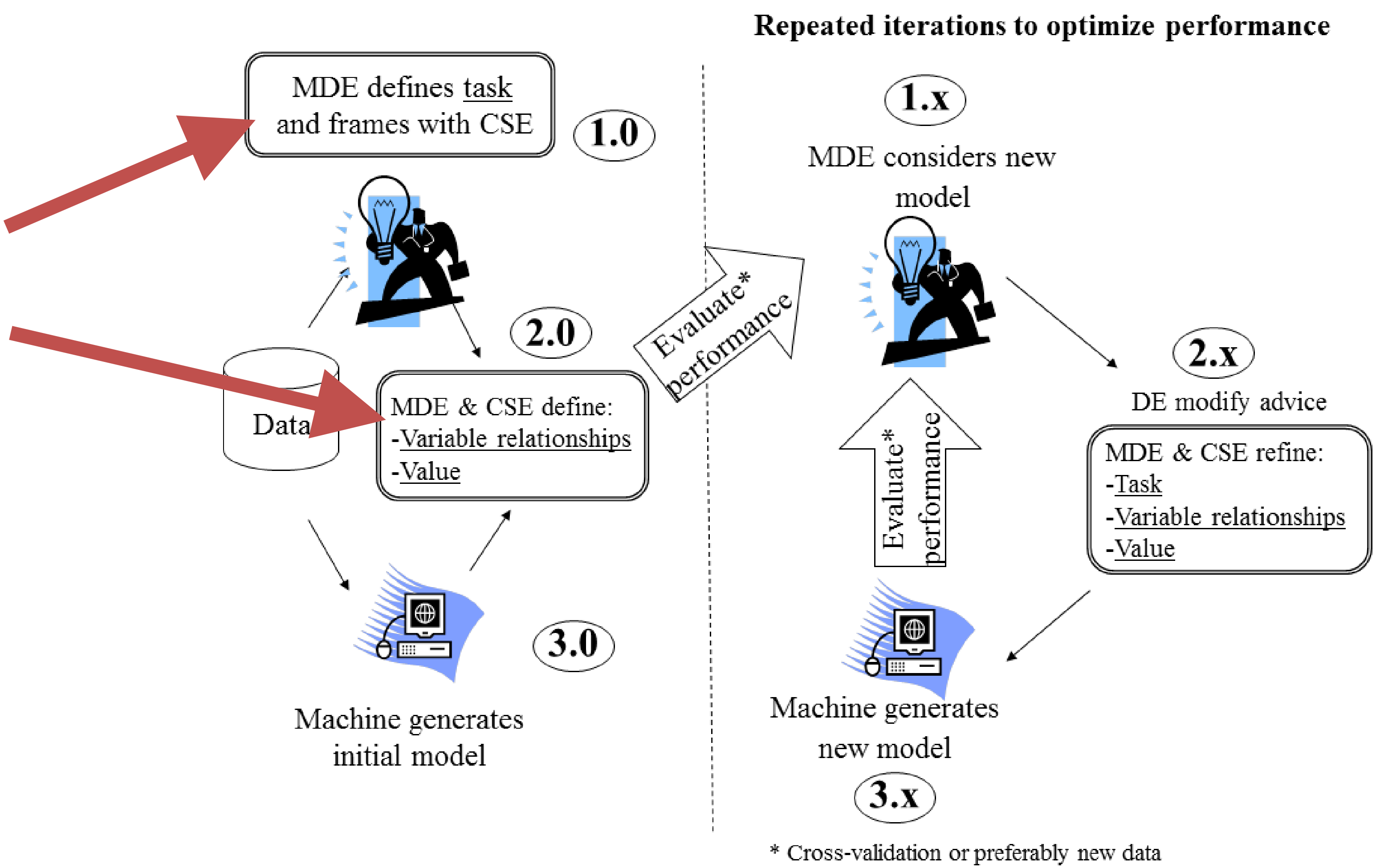
Parameter Values

- What is the clinical objective?
- What model parameters best represent that objective?

ABLe - Iterative Process



Phase 1


Repeated iterations to optimize performance
MDE defines task and frames with CSE
1.0
Data
2.0
MDE & CSE define:
-Variable relationships
-Value
3.0
Machine generates initial model
Evaluate* performance
1.x
MDE considers new model
2.x
DE modify advice
MDE & CSE refine:
-Task
-Variable relationships
-Value
Evaluate* performance
Machine generates new model
3.x
* Cross-validation or preferably new data

Phase 1

Task

- Simple probabilistic model (Naïve Bayes)
- Standardized BI-RADS descriptor features
- Some non-standard pathology features and demographics
- Predict probability of **malignancy**
- Assume excision at 2% model score

Variable Relationships

- Rules predicting **increase**/**decrease** risk of **malignancy**

Parameter Values

- None

Variable Relationships

If-Then rules that suggest **increase**/**decrease** risk of **upgrade**.

High-risk mass rule:

```
IF
Irregular mass shape is present   OR
Spiculated mass margin is present OR
High density mass is present      OR
Abnormality is increasing
THEN
Risk of upgrade increases
```

Biopsies in Practice (2006-11)

Core Needle Biopsies
2,808

Core Needle Biopsies + Dx Mammogram
1,910

Malignant Biopsy
601

Benign Biopsy
1,309

Non-definitive
157

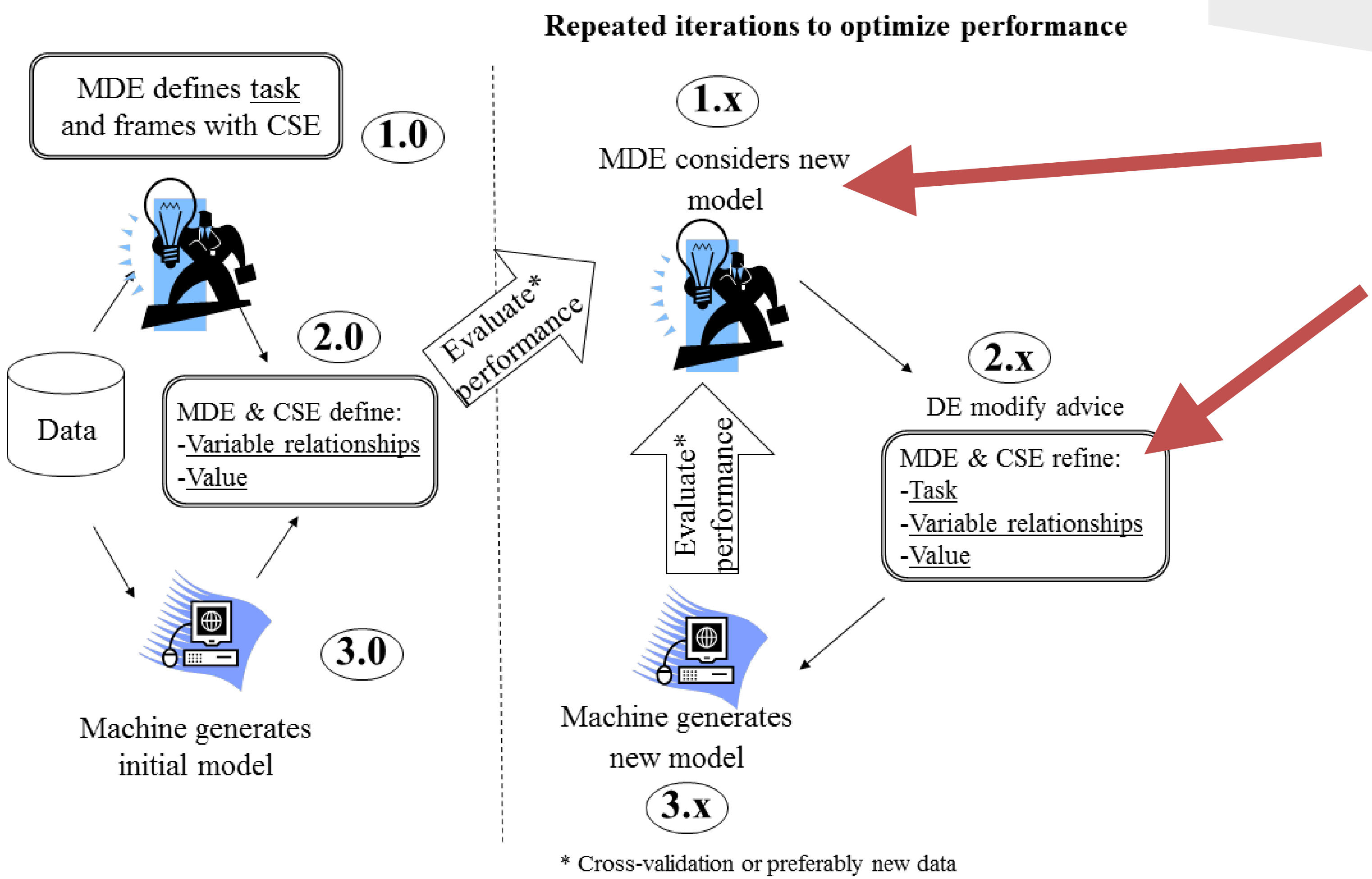
Malignant (upgrade)
29

Benign (non-upgrade)
128

Phase 1 Results

	Data	Rules	Data + Rules
Malignant Excisions Missed (%)	8 (27.6%)	1 (3.4%)	9 (31.0%)
Benign Excisions Avoided (%)	46 (35.9%)	5 (3.9%)	63 (49.2%)

Phase 2



Repeated iterations to optimize performance
MDE defines task and frames with CSE
1.0
2.0
Data
MDE & CSE define:
-Variable relationships
-Value
3.0
Machine generates initial model
Evaluate* performance
1.x
MDE considers new model
2.x
DE modify advice
MDE & CSE refine:
-Task
-Variable relationships
-Value
Evaluate* performance
Machine generates new model
3.x
* Cross-validation or preferably new data

Observations & Refinements

Observations

- No output threshold with acceptable performance

Observations & Refinements

Observations

- No output threshold with acceptable performance
- Non-definitive biopsies broken into 3 categories at diagnosis
    - Atypical / Radial Scar (ARS)
    - Insufficient (I)
    - Discordant (D)

Observations & Refinements

Observations

- No output threshold with acceptable performance
- Non-definitive biopsies broken into 3 categories at diagnosis
    - Atypical / Radial Scar (ARS)
    - Insufficient (I)
    - Discordant (D)
- ARS and I cases consistently mislabeled
    - ARS and I more dependent on pathology
    - D more dependent on imaging descriptors

Observations & Refinements

Observations

- No output threshold with acceptable performance
- Non-definitive biopsies broken into 3 categories at diagnosis
    - Atypical / Radial Scar (ARS)
    - Insufficient (I)
    - Discordant (D)
- ARS and I cases consistently mislabeled
    - ARS and I more dependent on pathology
    - D more dependent on imaging descriptors

Refinements

- Focus exclusively on discordant cases

Discordant Biopsies (2006-11)

Discordant Biopsy
60

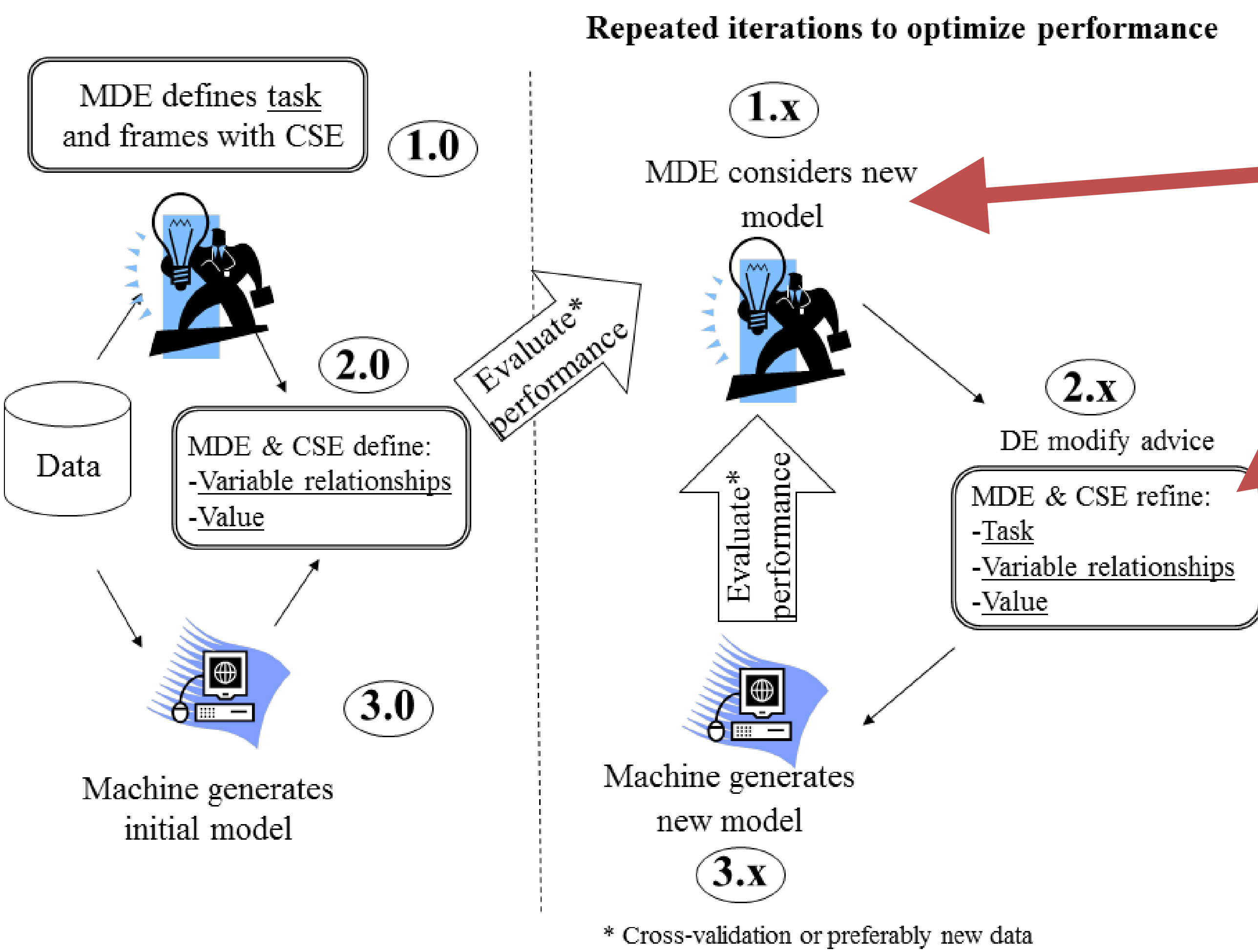
Malignant (upgrade)
10

Benign (non-upgrade)
50

Phase 2 Results

	Data	Rules	Data + Rules
Malignant Excisions Missed (%)	3 (30.0%)	1 (10.0%)	3 (30.0%)
Benign Excisions Avoided (%)	29 (58.0%)	17 (34.0%)	27 (54.0%)

Phase 3


Repeated iterations to optimize performance
MDE defines task and frames with CSE
1.0
Data
2.0
MDE & CSE define:
-Variable relationships
-Value
3.0
Machine generates initial model
Evaluate* performance
1.x
MDE considers new model
2.x
DE modify advice
MDE & CSE refine:
-Task
-Variable relationships
-Value
Evaluate* performance
Machine generates new model
3.x
* Cross-validation or preferably new data

Observations & Refinements

Observations

- Good ranking of cases by output probabilities
- Most cases assigned less than 2% risk

Observations & Refinements

Observations

- Good ranking of cases by output probabilities
- Most cases assigned less than 2% risk

Refinements

- Make model more conservative
    - Specify different costs for false negatives (FN) versus false positives (FP)
    - Take from utility analysis literature in mammography

Phase 3 Results

	Data	Rules	Data + Rules
Malignant Excisions Missed (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Benign Excisions Avoided (%)	5 (10.0%)	5 (10.0%)	12 (24.0%)

Conclusions

- Presented a framework for collaboration and leveraging domain expert advice
- Demonstrated ABLe on important task
- Achieved best results using ABLe

Future Work

- Use inductive logic programming (ILP) to automatically infer if-then rules from data
  - Allows automated feature construction/selection
  - Easily control constraints on features
- Evaluate model on unseen data
  - From our own institution
  - At collaborating institutions
- Grow model development data using natural language processing methods

Thanks

Questions?