Using Secondary Knowledge to Support Decision Tree Classification of Retrospective Clinical Data

Dympna O'Sullivan [1], William Elazmeh [1,3], Szymon Wilk [1], Ken Farion[4,5], Stan Matwin [1,2], Wojtek Michalowski [1]and Morvarid Sehatkar [1]

[1] University of Ottawa, Canada
[2] Institute of Computer Science, Polish Academy of Sciences, Poland
[3] University of Bristol, United Kingdom
[4] Faculty of Medicine, University of Ottawa, Canada
[5] Children's Hospital of Eastern Ontario, Canada

The prediction task

- Assess pediatric asthma exacerbation with available information about patient's condition
- Patients are classified as *mild* or *other* based on length of their stay in the emergency department:
  - *mild* stays < 4 hours
  - *other (moderate/severe)* stays up to 16 hours before admission to a hospital
- *other* is the critical class (unequal misclassification cost)
- Build a prediction model to produce:
  - High Sensitivity (low false negatives)
  - High Specificity (low false positives)

Measuring performance

	Predictions	
	other	*mild*
other	True *other*	False *mild*
mild	False *other*	True *mild*

- sensitivity(*other*) $= \frac{\text{correctly classified } other}{\text{total } other}$
- specificity(*other*) $= \frac{\text{correctly classified } mild}{\text{total } mild}$
- accuracy
- area under the ROC curve (AUC)

Challenges

Medical data, often:

- contains domain-specific characteristics of complex properties (*Muller et al. 2006*).
- is described by (*Magoulas & Prentza 2001*):
  - incompleteness (missing data)
  - incorrectness (noise)
  - sparseness (inappropriate values)
  - inexactness (inappropriate parameters)
- is heterogeneous in source and structure (*Cios & Moore 2002*)

which can present real problems for a learning method.

Furthermore...

- Interpretability of predictions for domain users is important (*Lavrac 1999*)
- This limits our choices of learning methods (must provide systematic explanation of predictions)
- These include models that are:
  - Probabilistic
  - Case-based
  - Rule-based, or
  - Tree-based
- We choose the latter by using Decision Trees

In this work

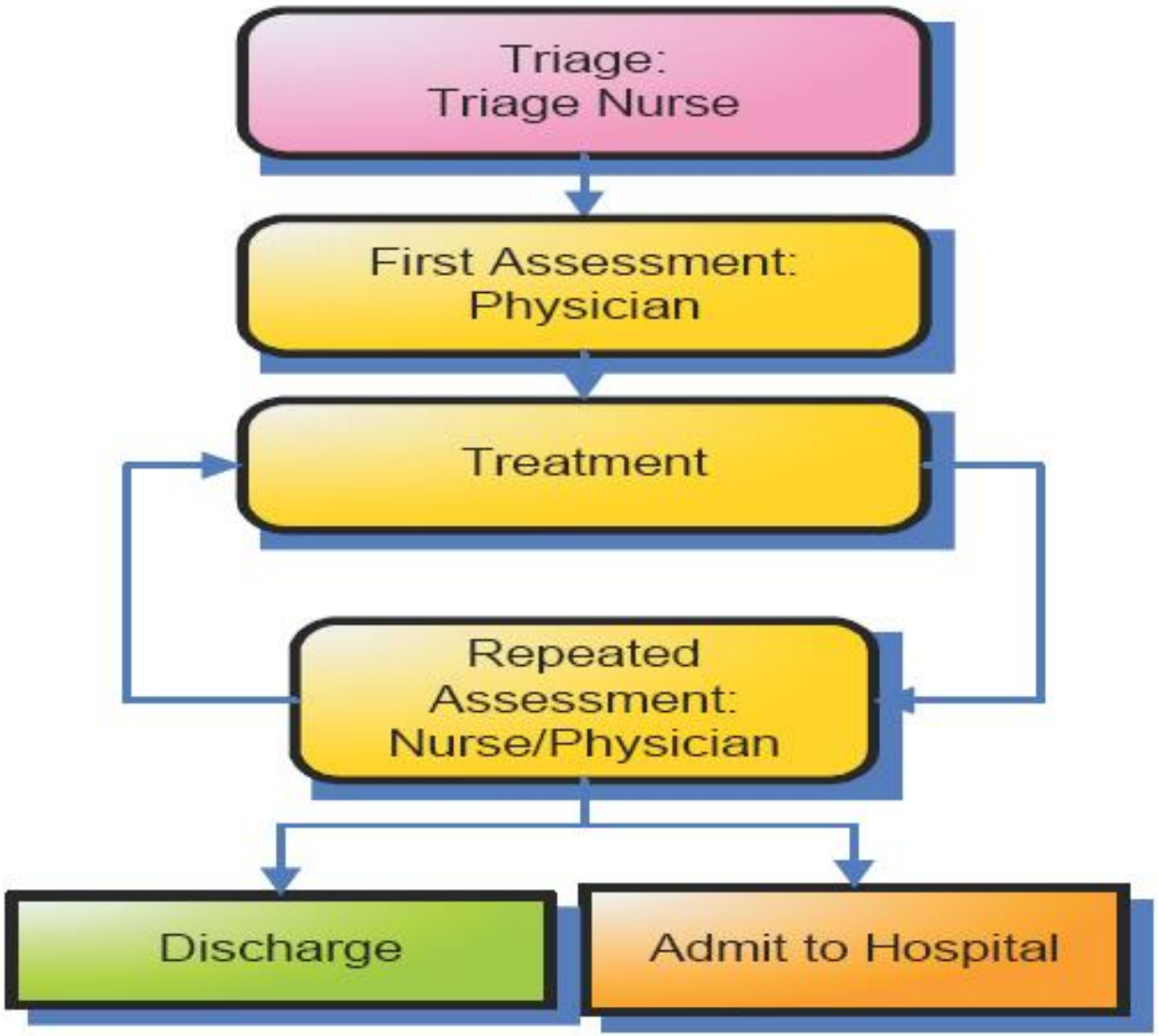
- We use a decision tree-based prediction model trained and tested on the whole data (baseline model) which yields unacceptable results
- We use additional clinical knowledge (Preschool Respiratory Assessment Measure PRAM) to partition the data into "typical (PRAM)" and "atypical (Non-PRAM)" patients
- we build a separate decision tree model on each set
- These two models perform significantly better than the baseline model

More specifically

Our asthma exacerbation data contains 362 records which:

- are collected retrospectively
- contains:
  - information about patient history
  - initial triage assessment data, and
  - the first round of physician's reassessment data (after aprox. 2 hours)
- have attributes with plenty (98%) missing values,
- can possible have a temporal representation but of inconsistent intervals, and
- have the problem of "values as attributes" problem

Pediatric asthma workflow in the ED



Preschool Respiratory Assessment Measure Asthma Index

Signs	0	1	2	3
Suprasternal indrawing	absent		present	
Scalene retractions	absent		present	
Wheezing	absent	expiratory	insiratory and expiratory	Audible without stethoscope /absent with no air entry
Air entry	normal	decreased bases	widespread decrease	absent/ minimal
Oxygen saturation	$\geq$95%	92-95%	$<$92%	

Table: PRAM Scores by Chalut D, Ducharme F, Davis GM Journal of Pediatrics 2000;137:762-768

Preschool Respiratory Assessment Measure Asthma Index

Signs	0	1	2	3
Suprasternal indrawing	absent		present	
Scalene retractions	absent		present	
Wheezing	absent	expiratory	insiratory and expiratory	Audible without stethoscope /absent with no air entry
Air entry	normal	decreased bases	widespread decrease	absent/ minimal
Oxygen saturation	$\geq$95%	92-95%	$<$92%	

Table: PRAM Scores by Chalut D, Ducharme F, Davis GM Journal of Pediatrics 2000;137:762-768

Preschool Respiratory Assessment Measure Asthma Index

Signs	0	1	2	3
Suprasternal indrawing	absent		present	
Scalene retractions	absent		present	
Wheezing	absent	expiratory	insiratory and expiratory	Audible without stethoscope /absent with no air entry
Air entry	normal	decreased bases	widespread decrease	absent/ minimal
Oxygen saturation	$\geq$95%	92-95%	$<$92%	

Retraction

Table: PRAM Scores by Chalut D, Ducharme F, Davis GM Journal of Pediatrics 2000;137:762-768

Preschool Respiratory Assessment Measure Asthma Index

Signs	0	1	2	3
Suprasternal indrawing	absent		present	
Scalene retractions	absent		present	
Wheezing	absent	expiratory	insiratory and expiratory	Audible without stethoscope /absent with no air entry
Air entry	normal	decreased bases	widespread decrease	absent/ minimal
Oxygen saturation	$\geq$95%	92-95%	$<$92%	

Exp_Wheeze
Insp_Wheeze
Absent/Present

Table: PRAM Scores by Chalut D, Ducharme F, Davis GM Journal of Pediatrics 2000;137:762-768

Preschool Respiratory Assessment Measure Asthma Index

Signs	0	1	2	3
Suprasternal indrawing	absent		present	
Scalene retractions	absent		present	
Wheezing	absent	expiratory	insiratory and expiratory	Audible without stethoscope /absent with no air entry
Air entry	normal	decreased bases	widespread decrease	absent/ minimal
Oxygen saturation	$\geq$95%	92-95%	$<$92%	

Good/Reduced

Table: PRAM Scores by Chalut D, Ducharme F, Davis GM Journal of Pediatrics 2000;137:762-768

Preschool Respiratory Assessment Measure Asthma Index

Signs	0	1	2	3
Suprasternal indrawing	absent		present	
Scalene retractions	absent		present	
Wheezing	absent	expiratory	insiratory and expiratory	Audible without stethoscope /absent with no air entry
Air entry	normal	decreased bases	widespread decrease	absent/ minimal
Oxygen saturation	$\geq$95% SAO2	92-95%	$<$92%	

Table: PRAM Scores by Chalut D, Ducharme F, Davis GM Journal of Pediatrics 2000;137:762-768

Mapping PRAM to our data

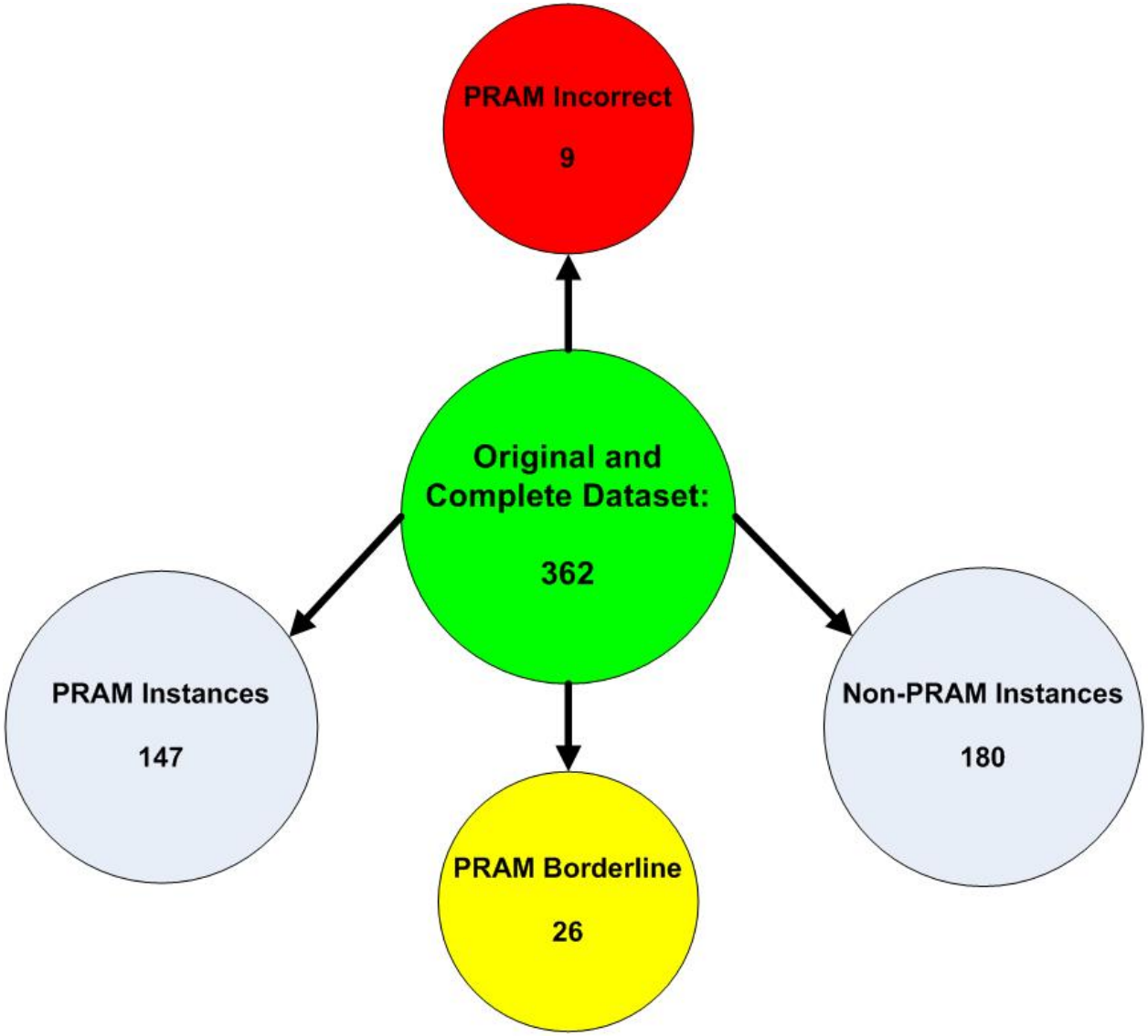
- Need to infer some of the values because:
  - No Mapping for PRAM *Suprasternal Indrawing*
  - `Retractions` in our dataset mapped to *Scalene Retractions* in PRAM
  - `Exp_Wheeze` and `Insp_Wheeze` (two attributes in our dataset) mapped to one PRAM attribute, *Wheezing*
  - `Air_Entry` in our dataset has 2 values but PRAM *Air Entry* has 4 possible values
  - `SAO2` in our dataset maps to PRAM *Oxygen Saturation*
- All mappings approved by ED physician

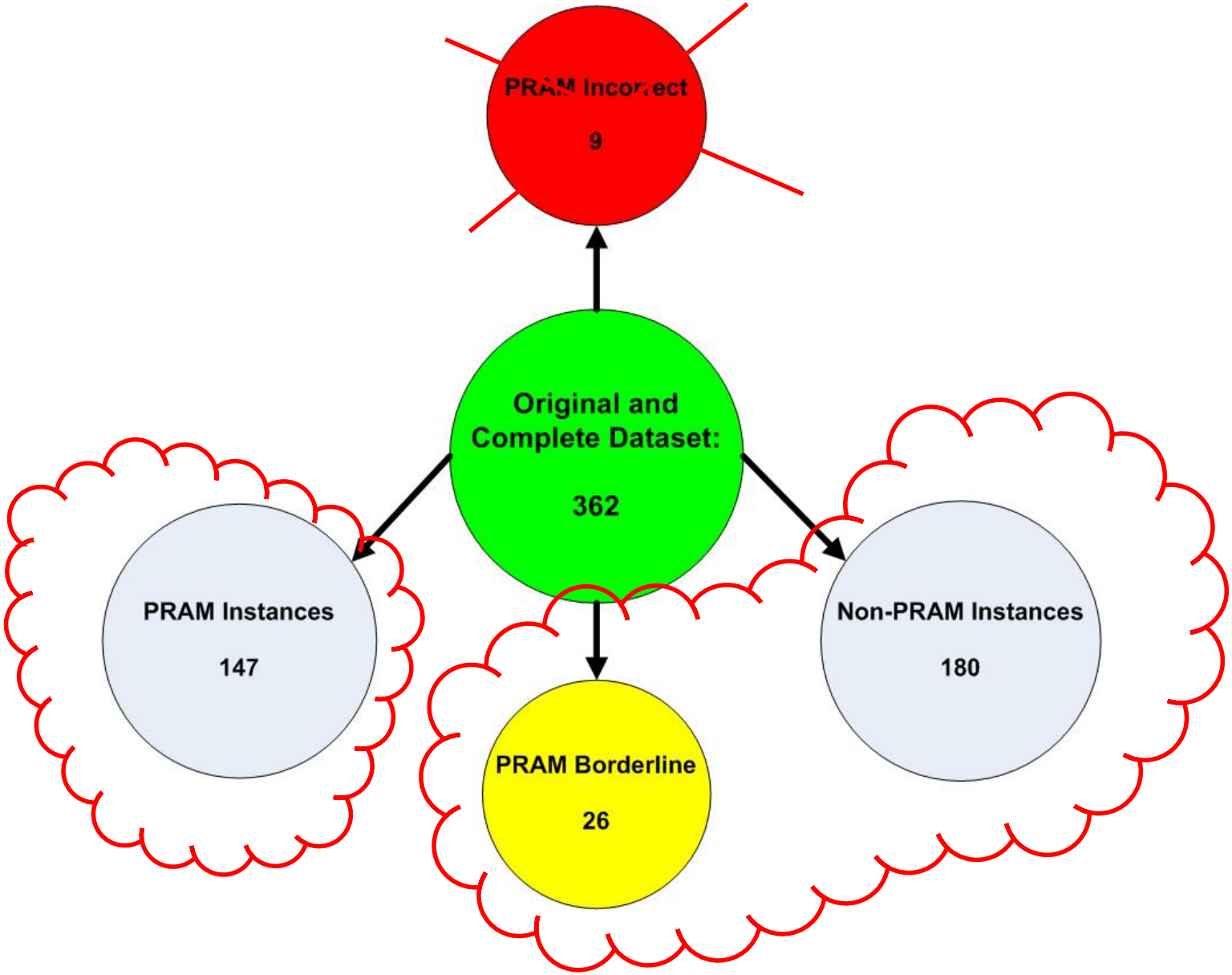
Rules to compute PRAM Scores

```
RETRACTIONS=absent, AIR_ENTRY=good          --> 0
RETRACTIONS=absent, AIR_ENTRY=reduced       --> 1
RETRACTIONS=absent, AIR_ENTRY=?             --> 1
RETRACTIONS=present                         --> 2

EXP_WHEEZE=absent,  INSP_WHEEZE=absent      --> 0
EXP_WHEEZE=present, INSP_WHEEZE=absent      --> 1
EXP_WHEEZE=present, INSP_WHEEZE=present     --> 2
EXP_WHEEZE=absent,  INSP_WHEEZE=present **** Undefined

AIR_ENTRY=good                              --> 0
Class=mild, AIR_ENTRY=reduced              --> 1
Class=other, AIR_ENTRY=reduced              --> 3

SAO2=ge_95                                  --> 0
SAO2=ge_93_lt_95                            --> 1
SAO2=ge_88_lt_93                            --> 2
SAO2=lt_88                                  --> 2
```

Using PRAM to partition the data

- PRAM complete and correct cases are 'typical' and correspond to the first set (PRAM set)
- All other cases (PRAM incomplete and PRAM incorrect) are 'atypical' and correspond to the second set (Non-PRAM set)





Experiments design

- Train & test a decision tree on the whole data (baseline)
- Apply PRAM and partition data into PRAM and Non-PRAM data
- Train & test a decision tree for each subset of data
- Use leave-one-out test strategy (relatively small data)
- Later, apply feature selection techniques and compare

Performance comparison

Set	Size	Sens	Spec	Acc	AUC
Entire	362	73	63	69	69
PRAM Set	**147**	**93**	**96**	**95**	**98**
Non-PRAM Set	**206**	**89**	**53**	**74**	**77**

- Performance is better on both sets individually
- High sensitivity (lower false mild) is achieved
- High specificity (lower false other) is not necessarily attainable on the non-PRAM set

Decision tree produced on entire data

```
REASSESSED_EXP_WHEEZE = ABSENT
|   TRIAGE_SAO2 = GE_95: MILD (60.84/11.86)
|   TRIAGE_SAO2 = GE_88_LT_93: OTHER (9.04/3.5)
|   TRIAGE_SAO2 = LT_88: OTHER (1.03/0.02)
|   TRIAGE_SAO2 = GE_93_LT_95: MILD (16.11/7.34)
REASSESSED_EXP_WHEEZE = PRESENT
|   REASSESSED_TEMP = LT_38
|   |   REASSESSED_HEART_RATE = ABNORMAL: OTHER (66.31/19.34)
|   |   REASSESSED_HEART_RATE = MILD_ABNORMAL: MILD (12.13/4.54)
|   |   REASSESSED_HEART_RATE = NORMAL: OTHER (0.0)
|   REASSESSED_TEMP = GE_38_LT_39
|   |   REASSESSED_RETRACTIONS = PRESENT: OTHER (46.45/9.64)
|   |   REASSESSED_RETRACTIONS = ABSENT: MILD (9.74/4.84)
|   REASSESSED_TEMP = GE_39: OTHER (20.35/6.27)
```

The tree uses:

- a mix of Triage and Reassessment attributes
- PRAM and Non-PRAM attributes

Decision tree produced on PRAM data

```
REASSESSED_INSP_WHEEZE = ABSENT
|   REASSESSED_AIR_ENTRY = GOOD: MILD (61.0)
|   REASSESSED_AIR_ENTRY = REDUCED
|   |   REASSESSED_SAO2 = GE_95
|   |   |   ALLG_FOOD = NO: MILD (6.0/1.0)
|   |   |   ALLG_FOOD = YES: OTHER (2.0)
|   |   REASSESSED_SAO2 = GE_93_LT_95: OTHER (2.0)
|   |   REASSESSED_SAO2 = GE_88_LT_93: OTHER (5.0)
|   |   REASSESSED_SAO2 = LT_88: OTHER (0.0)
REASSESSED_INSP_WHEEZE = PRESENT
|   REASSESSED_SAO2 = GE_95
|   |   REASSESSED_AIR_ENTRY = GOOD: MILD (13.0)
|   |   REASSESSED_AIR_ENTRY = REDUCED: OTHER (4.0)
|   REASSESSED_SAO2 = GE_93_LT_95: OTHER (31.0/1.0)
|   REASSESSED_SAO2 = GE_88_LT_93: OTHER (20.0)
|   REASSESSED_SAO2 = LT_88: OTHER (3.0)
```

The tree uses:

- only Reassessment attributes
- PRAM attributes and ALLG_FOOD

Decision tree produced on Non-PRAM data

```
REASSESSED_EXP_WHEEZE = ABSENT
|   REASSESSED_RETRACTIONS = PRESENT
|   |   TRIAGE_HEART_RATE = MILD_ABNORMAL: MILD (8.4/3.56)
|   |   TRIAGE_HEART_RATE = ABNORMAL: OTHER (22.98/6.82)
|   |   TRIAGE_HEART_RATE = NORMAL: OTHER (0.0)
|   REASSESSED_RETRACTIONS = ABSENT
|   |   TRIAGE_SAO2 = GE_95: MILD (39.24/8.45)
|   |   TRIAGE_SAO2 = GE_88_LT_93: OTHER (3.19/0.43)
|   |   TRIAGE_SAO2 = LT_88: OTHER (0.25/0.0)
|   |   TRIAGE_SAO2 = GE_93_LT_95: MILD (9.42/3.95)
REASSESSED_EXP_WHEEZE = PRESENT
|   REASSESSED_HEART_RATE = ABNORMAL: OTHER (101.05/22.52)
|   REASSESSED_HEART_RATE = MILD_ABNORMAL
|   |   PREV_ED_LAST_YEAR = 2_VISITS: OTHER (4.09/0.7)
|   |   PREV_ED_LAST_YEAR = 1_VISIT: OTHER (5.31/1.6)
|   |   PREV_ED_LAST_YEAR = NONE: MILD (8.54/1.01)
|   |   PREV_ED_LAST_YEAR = 3_VISITS: OTHER (0.44)
|   |   PREV_ED_LAST_YEAR = GE_4_VISITS: MILD (2.3/0.1)
|   REASSESSED_HEART_RATE = NORMAL: OTHER (0.81/0.1)
```

The tree uses:

- a mix of Triage and Reassessment attributes
- PRAM and Non-PRAM attributes

Performance with feature selection

Feature Selection	Mode	Size	Sens	Spec	Acc	AUC
Information Gain	Auto.	362	72	63	68	69
Chi-squared	Auto.	363	72	63	68	69
Combinatorial	Auto.	362	72	65	69	71
Wrap Naive Bayes	Auto.	362	71	60	70	77
On All Att.	Expert	362	72	66	70	73
On Only PRAM Att.	Expert	362	77	78	70	71

Conclusions

- Retrospectively collected clinical data provides many difficulties for data mining and machine learning
- Performance is medically unacceptable
- Our experiments show that PRAM is reliable
- Using PRAM assists the learning model to focus on better groups of examples
- Machine learning methods can potentially extend PRAM
- Feature selection (automatic or by expert) fails to improve the performance
- We will use prospective data being collected