

Detecting Cardiac Pathologies from Annotated Auscultations

CBMS 2012 – June 20th 2012 – Rome, Italy



Pedro Ferreira
Inês Dutra
Daniel Pereira
Felipe Mourato
Sandra Mattos
Miguel Coimbra
Ricardo Cruz-Correia

Outline

- Cardiovascular Diseases
- DigiScope Project
 - **Machine Learning / Data Mining**
 - **Objectives**
 - Data
 - Feature Selection
 - Association Rules
 - Classification
 - Non – Relational
 - Relational (ILP)
- Conclusions and Future Work

Outline

- Cardiovascular Diseases
- DigiScope Project
 - Machine Learning / Data Mining
 - Objectives
 - Data
 - Feature Selection
 - Association Rules
 - Classification
 - Non – Relational
 - Relational (ILP)
- Conclusions and Future Work

Cardiovascular Diseases



- **17.3 million**
 - died from CVDs in 2008
- **80%**
 - of CVD deaths take place in low and middle-income countries
- **23.6 million**
 - will die from CVDs by 2030

Outline

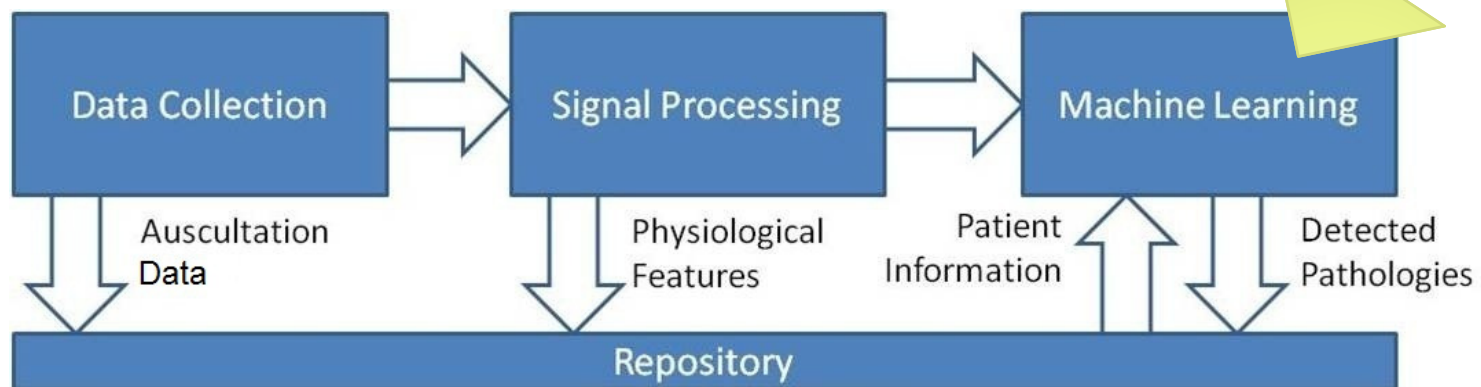
- Cardiovascular Diseases
- DigiScope Project
 - Machine Learning / Data Mining
 - Objectives
 - Data
 - Feature Selection
 - Association Rules
 - Classification
 - Non – Relational
 - Relational (ILP)
- Conclusions and Future Work

DigiScope Project



- **Help General Practitioners (GPs)** in their daily medical routine
- Capable of automatically **extract clinical features** from collected data
- May provide **clinical second opinion** on specific heart pathologies

DigiScope Project



Outline

- Cardiovascular Diseases
- DigiScope Project
 - **Machine Learning / Data Mining**
 - **Objectives**
 - Data
 - Feature Selection
 - Association Rules
 - Classification
 - Non – Relational
 - Relational (ILP)
- Conclusions and Future Work

Objectives

- Automatically **learn classifiers** that distinguish **normal patients** from **patients with a cardiac pathology**
 - Our **classifiers rely only on** the cardiologist provided **annotation** and not on the raw sound data itself
- Automatically **extract** new and relevant **knowledge** from the dataset

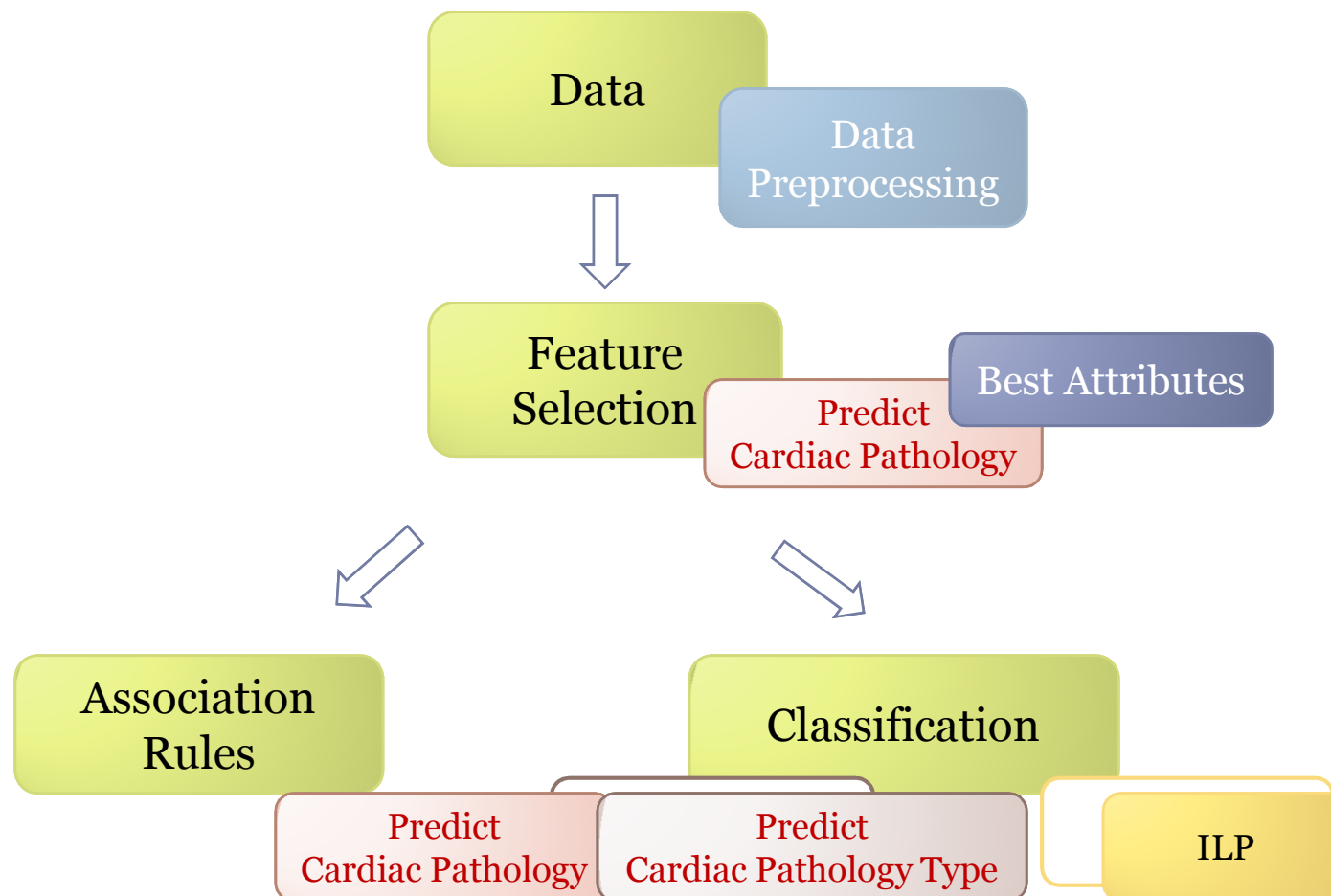
Machine Learning/Data Mining

Data

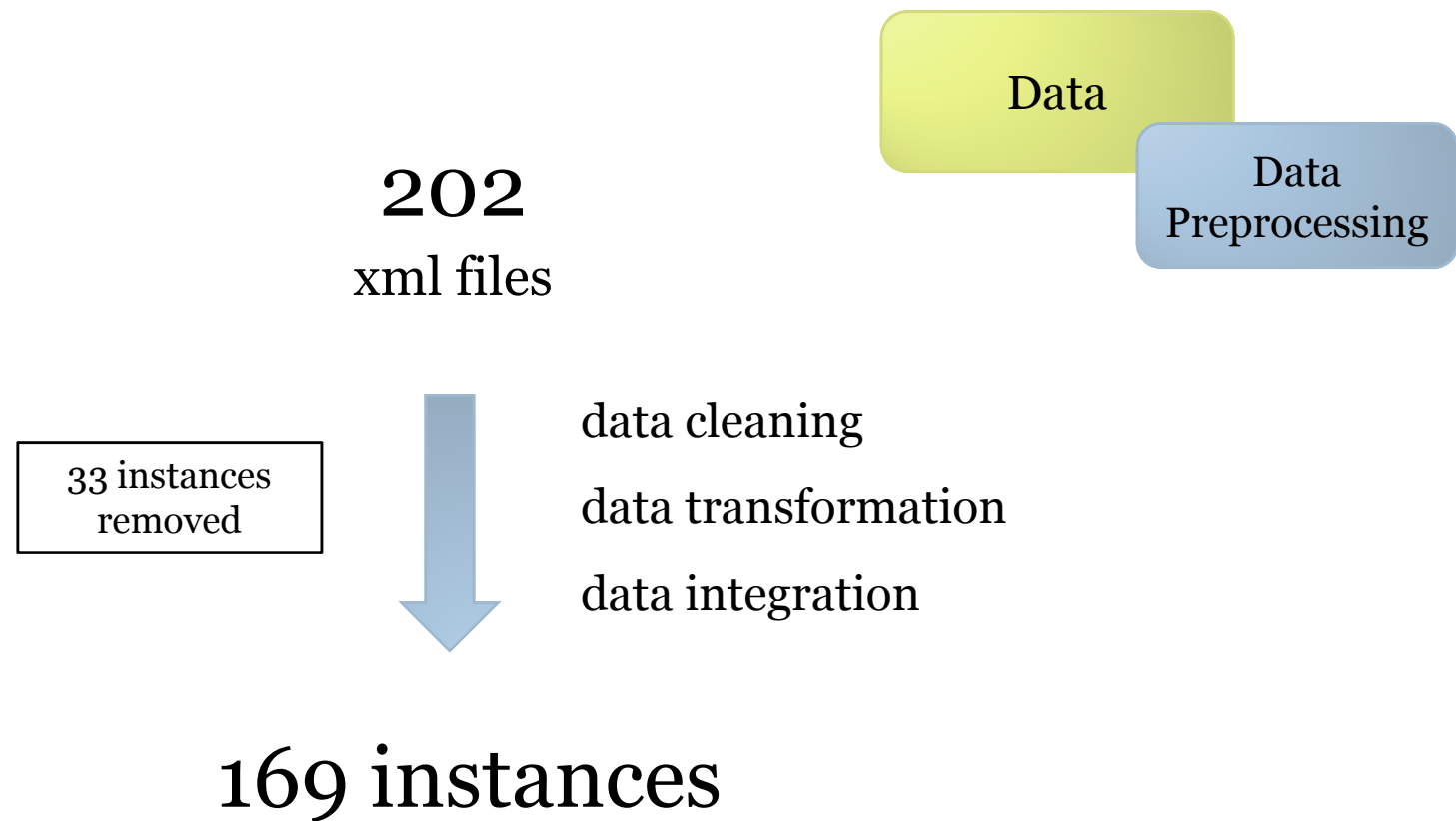


- 202 cases from children
 - Pernambuco, Recife – Brazil
 - Collected between **June** to **September 2011**
 - [0-19] years old
 - Average age: 7.31

Machine Learning/Data Mining



Machine Learning/Data Mining



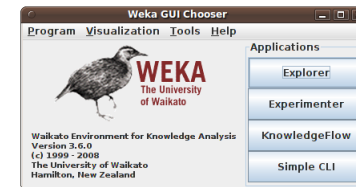
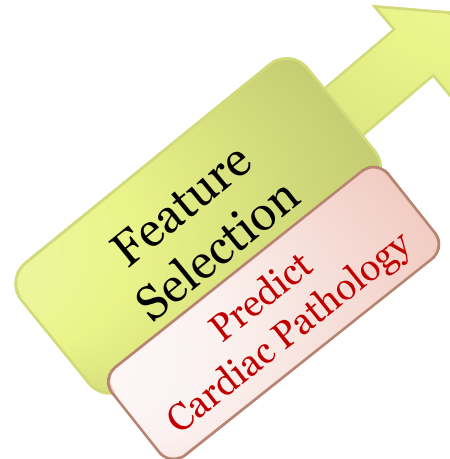
Machine Learning/Data Mining

169 instances

13 attributes



53 attributes
(41 original)



Machine Learning/Data Mining

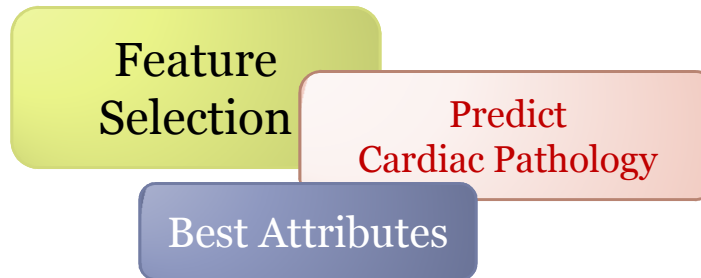
Attribute
BMI_def
Age_def
Sex
SystolicSystemicPressure_def
DiastolicSystemicPressure_def
Hypertension
Murmur
Grading
S2Status
IfAbnormal
PulmonaryComponent
CardiacPathology
CardiacPathologyType

Feature
Selection

Predict
Cardiac Pathology

13 attributes

Machine Learning/Data Mining



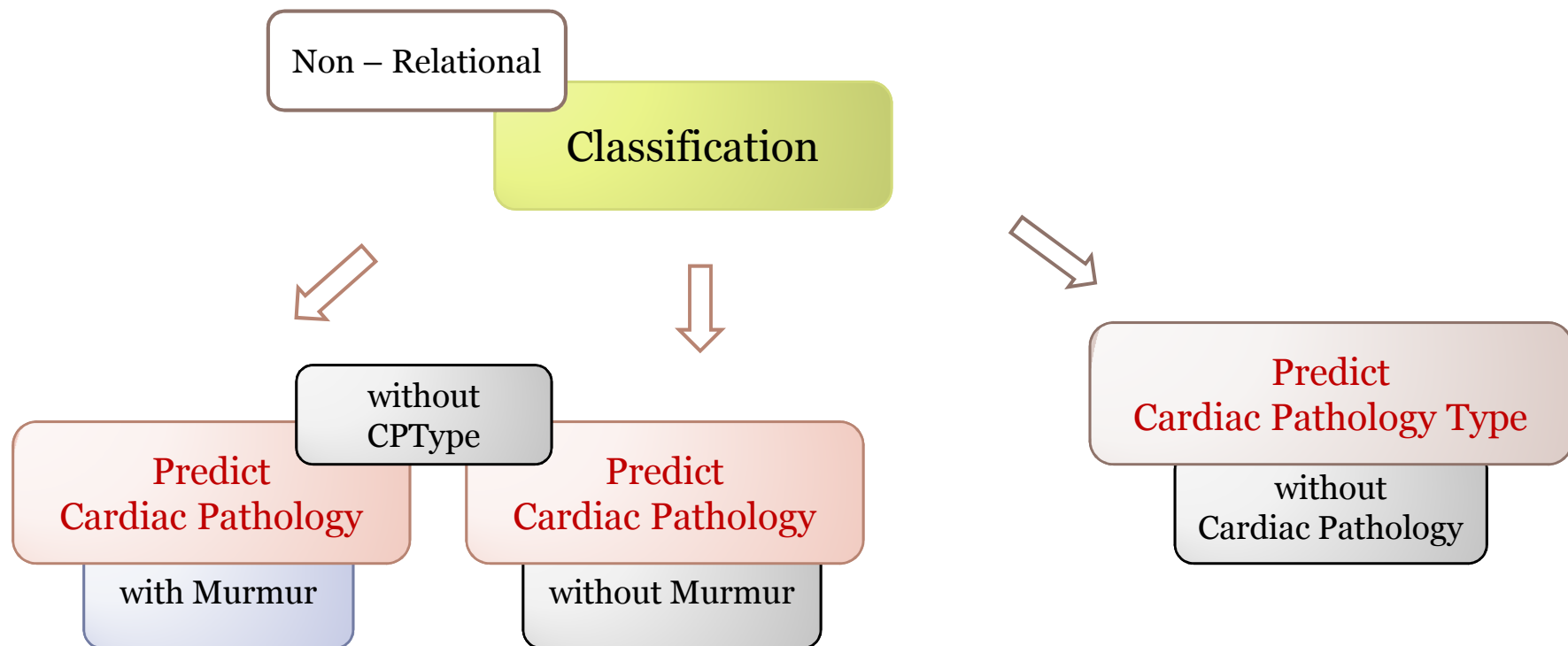
- **With all 13 attributes:**

- **Murmur**

- **In the absence of Murmur:**

- **S2Status**
 - **IfAbnormal**
 - **SystolicSystemicPressure_def**

Machine Learning/Data Mining



for the 169 cases

- 40 (+)
- 129 (-)

Machine Learning/Data Mining

- ZeroR (baseline classifier)
- OneR
- DTNB
- PART

rules

- NaiveBayes
- BayesNet (TAN)

bayes

- SMO

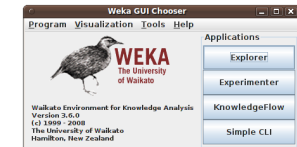
functions

- J48
- DecisionStump
- RandomForest
- SimpleCart
- NBTree

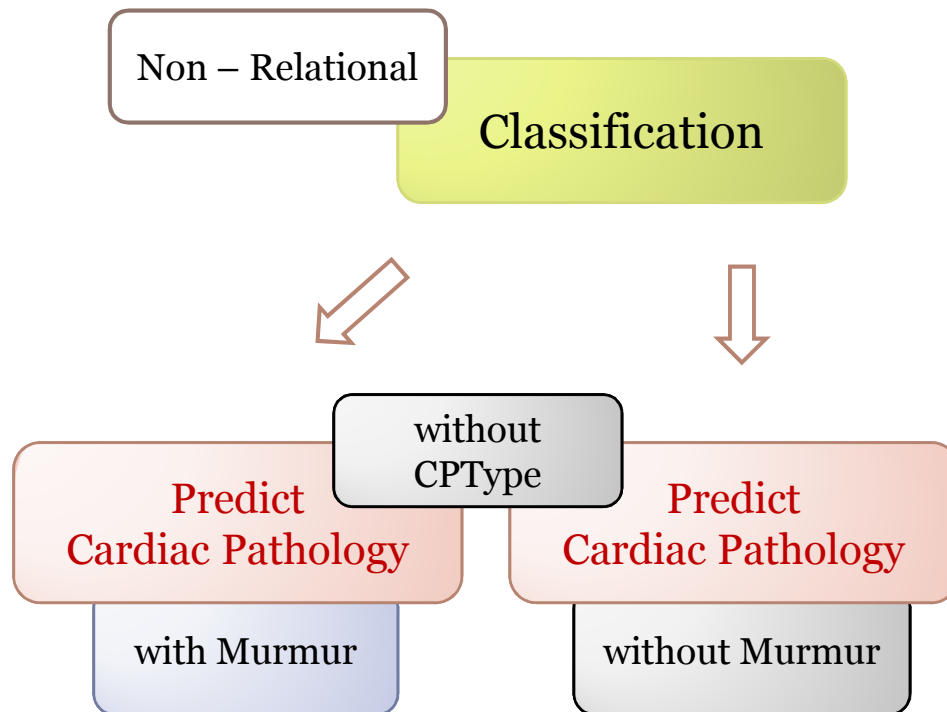
trees

- AdaBoostM1
- Bagging
- Dagging
- Grading
- Stacking
- Vote

meta-learning



Machine Learning/Data Mining



for the 169 cases

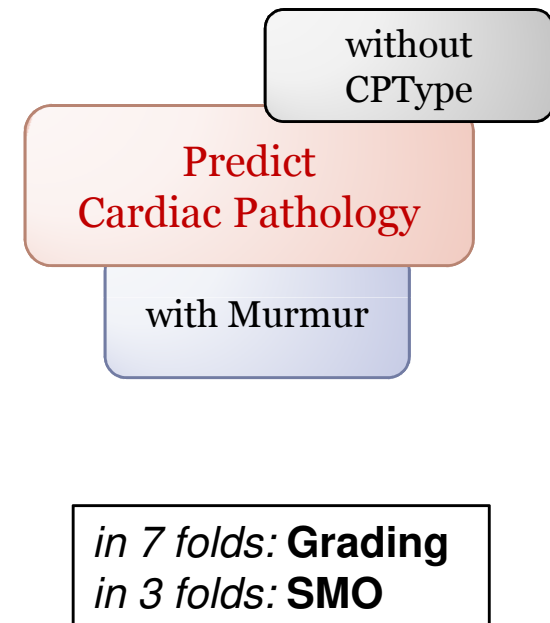
- 40 (+)
- 129 (-)



- **10 x 10 fold stratified cross validation with tuning sets**
- Paired Corrected T-Tester
 - **Significance level: 0.05**

Machine Learning/Data Mining

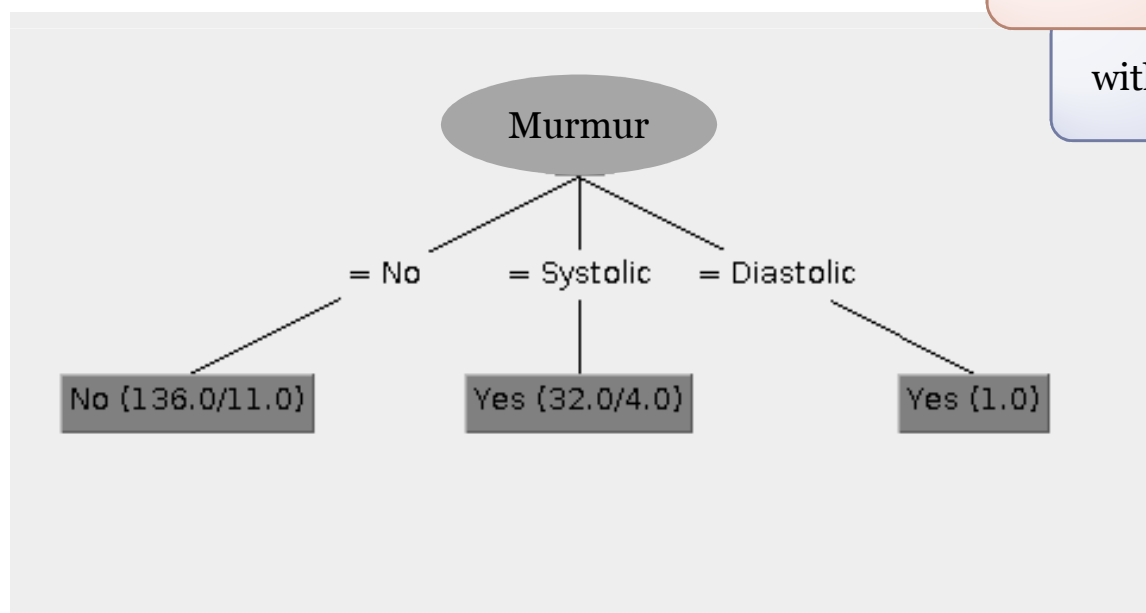
Metrics	Tuning	Test
CCI (%)	91.56 v	90.53
Sensitivity	0.72 v	0.70
Specificity	0.98	0.97
AUC	0.85	0.83



v - Results **statistically better** than ZeroR algorithm (baseline), **with $p=0.05$**

Machine Learning/Data Mining

J48 Decision Tree



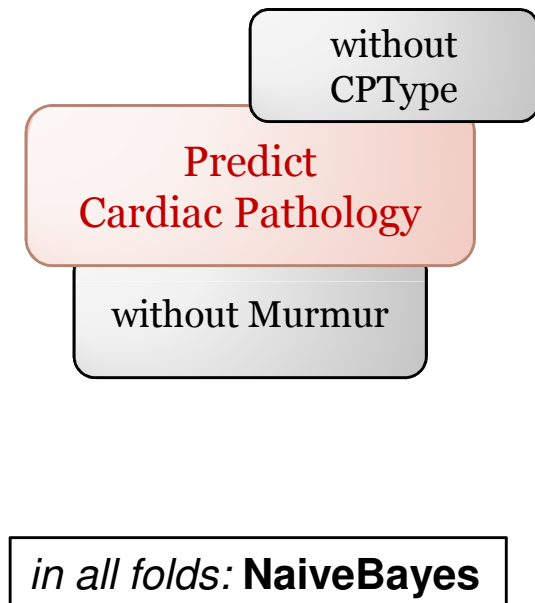
without
CPTYPE

Predict
Cardiac Pathology

with Murmur

Machine Learning/Data Mining

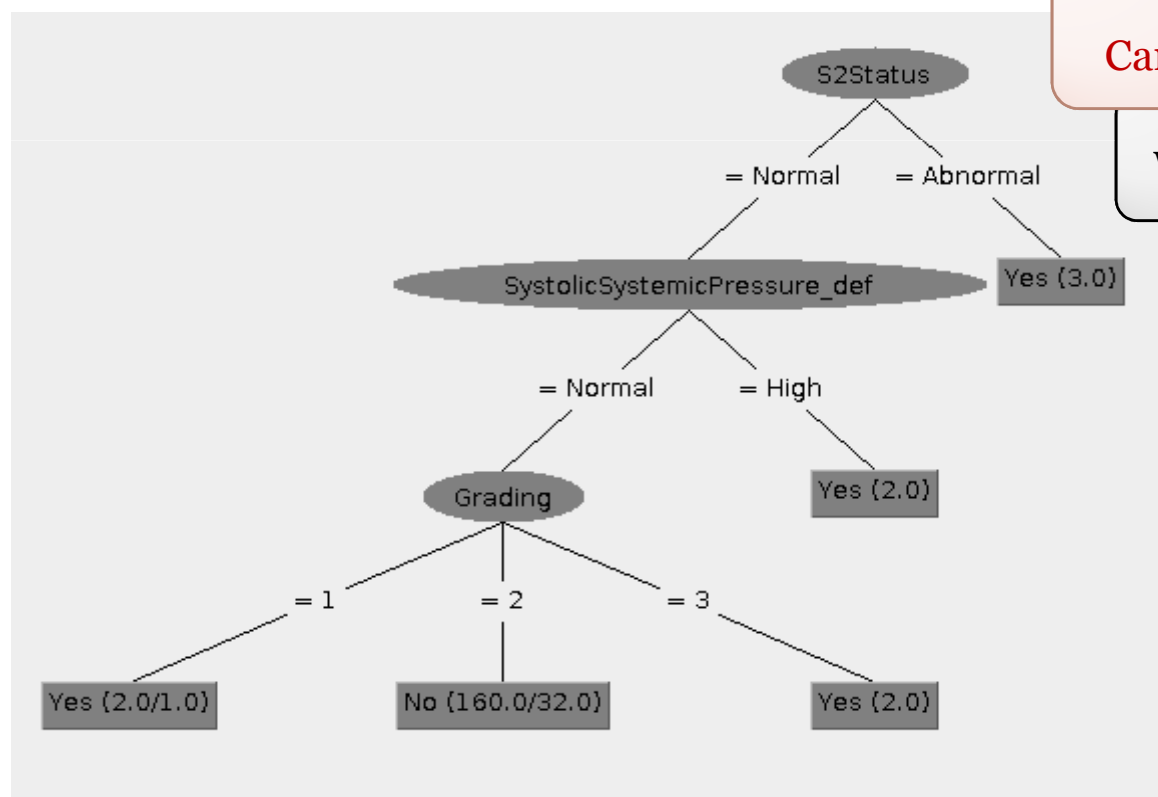
Metrics	Tuning	Test	
CCI (%)	91.56 v	90.53	with Murmur
Sensitivity	0.72 v	0.70	
Specificity	0.98	0.97	
AUC	0.85	0.83	
			Test
	CCI (%)	79.37	79.29
	Sensitivity	0.28 v	0.28
	Specificity	0.95	0.95
	AUC	0.65	0.60



v - Results **statistically better** than ZeroR algorithm (baseline), **with $p=0.05$**

Machine Learning/Data Mining

J48 Decision Tree

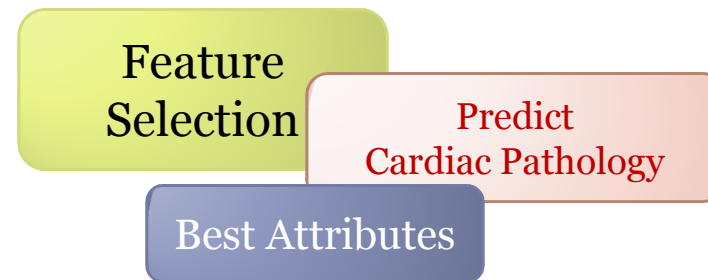


Predict
Cardiac Pathology

without
CPTYPE

without Murmur

Machine Learning/Data Mining



- With all 13 attributes:

- Murmur

- In the absence of Murmur:

- S2Status
- IfAbnormal
- SystolicSystemicPressure_def

Machine Learning/Data Mining



Non – Relational

Classification



Predict
Cardiac Pathology Type

without
Cardiac Pathology

- ▣ **10 x 10 fold stratified cross validation**
- ▣ Paired Corrected T-Tester
 - **Significance level: 0.05**

Machine Learning/Data Mining

Predict
Cardiac Pathology Type

without
Cardiac Pathology

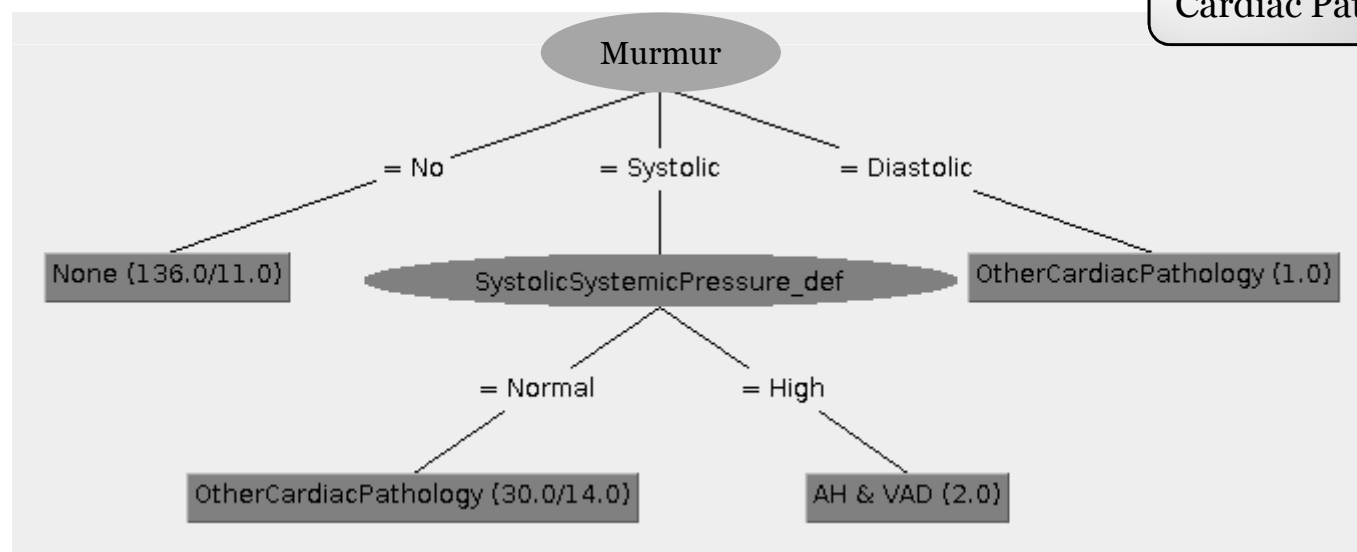
Metrics	AdaBoostM1 algorithm
CCI (%)	84.04 (+/- 5.30) v
Sensitivity	0.75 (+/- 0.27) v
Specificity	0.89 (+/- 0.06) *

v - Results **statistically better** than ZeroR algorithm (baseline), **with p=0.05**;

* - Results **statistically worst** than ZeroR algorithm (baseline), **with p=0.05**.

Machine Learning/Data Mining

J48 Decision Tree



Predict
Cardiac Pathology Type

without
Cardiac Pathology

Machine Learning/Data Mining

Relational

Classification

ILP

```
'CardiacPathology' (A)  if  
    bmi (A, obese)  and  
    'Murmur' (A, 'Systolic').
```

Rule holds for **6 (15%) out of the 40 patients** with a **Cardiac Pathology**

Does not apply to any healthy patient (129)

Outline

- Cardiovascular Diseases
- DigiScope Project
 - **Machine Learning / Data Mining**
 - **Objectives**
 - Data
 - Feature Selection
 - Association Rules
 - Classification
 - Non – Relational
 - Relational (ILP)
- **Conclusions and Future Work**

Conclusions and Future Work

- a) Train a classifier with performance of **90.5%**, sensitivity of **0.70** and specificity of **0.97** to predict pathologies on unseen cases
- b) In order to correctly classify Cardiac Pathologies:
 - i. **Murmur annotated** by physicians
 - ii. **Murmur** extracted by **signal processing**
- c) Intriguing **rule found** that relates **BMI** with **Murmur** and **Cardiac Pathology**
 - BMI usually not considered relevant to predict Cardiac Pathologies in children

Conclusions and Future Work

- a) Try to extract relevant **knowledge** from data regarding **adults** and **pregnant women**
- b) **Final Goal: Integrated tool**, capable of online **predicting cardiac pathologies** and **recommending** additional screening

Thank you!

pedroferreira@dcc.fc.up.pt
ines@dcc.fc.up.pt



Appendices

Machine Learning/Data Mining

Attribute	Value
BMI_def	{Normal Weight, Underweight, Overweight, Obese}
Age_def	{Baby, Preschool, Scholar Age, Teenager}
Sex	{Female, Male}
SystolicSystemicPressure_def	{Normal, High}
DiastolicSystemicPressure_def	{Normal, High}
Hypertension	{No, Yes}
Murmur	{No, Systolic, Diastolic}
Grading	{1, 2, 3}
S2Status	{Normal, Abnormal}
IfAbnormal	{NA, Single, Fixed Split}
PulmonaryComponent	{Normal, Hyperfonetic}
CardiacPathology	{Yes, No}
CardiacPathologyType	{None, IntraventricularCommunication (IC), ArterialHypertension (AH), ValvularAorticDisease (VAD), PulmonaryHypertension (PH), OtherCardiacPathology (OCP), AH & VAD, IC & OCP, VAD & OCP}

Machine Learning/Data Mining

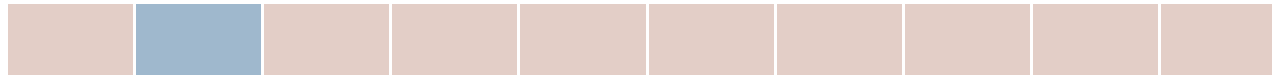
10 x 10 fold stratified cross-validation

Iteration

1



2



3



4



5



(...)

(...)

Training

Test



Machine Learning/Data Mining

Metrics
CCI
K
MAE
Sensitivity
Specificity
Precision
F-Measure
AUC