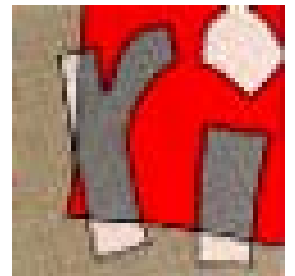
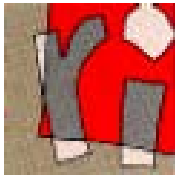


A new medical test for for atherosclerosis detection GeNo

Jérôme Azé, Noël Lucas, Michèle Sebag
(LRI, Orsay)





Approach

- Goal :

PREVENTION OF CV DISEASES



IDENTIFICATION OF RISK FACTORS

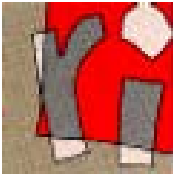


BUILD A SENSITIVE INEXPENSIVE MEDICAL TEST

- Questions :

- WHICH FACTORS ?

- HOW DO THEY INTERACT ?



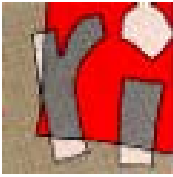
OUR GOAL

- **Usual ML & KDD**
 - Provides expert with hypotheses :
 - EXPERT says YES or NO.

- **Our approach (Visual Data Mining)**
 - Provides the expert with curves :
 - EXPERT says « let me think... »

USE VISION TO THINK

(Card, Mackinlay, Schneiderman, 1999)



OUR GOAL, cont'd

1/ We need a fine-grained perception of risk

Classification \Rightarrow Regression

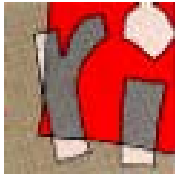
2/ Only sufficient (not necessary) conditions can be extracted from data bases.

Ex. : disease if

 In DB : your parents are diabetic

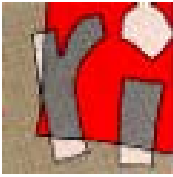
 In DB : You are smoking for long

 You don't laugh often enough



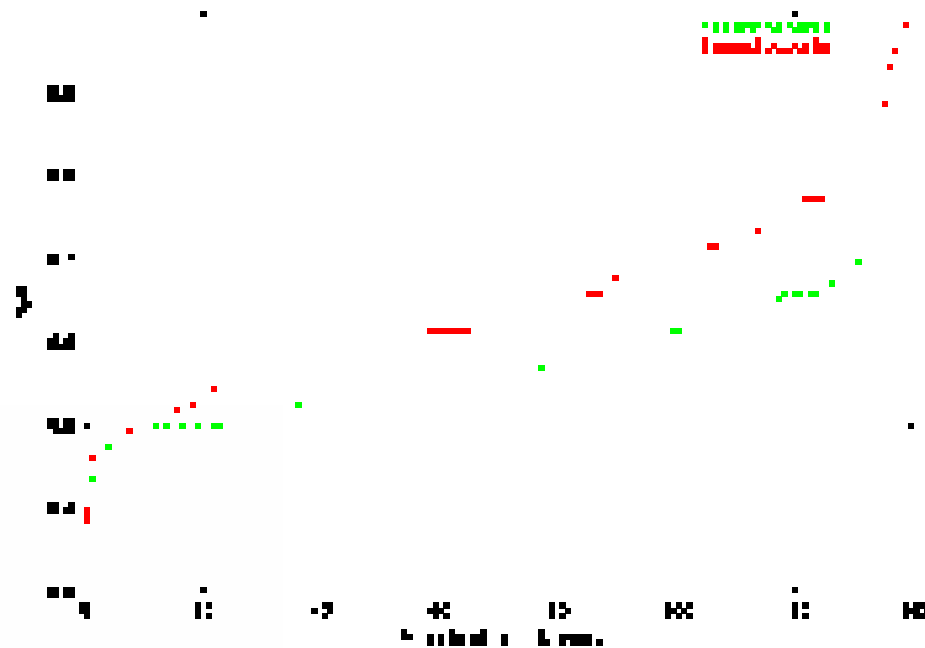
Contents

- Motivations
- Univariate and Multivariate Analysis
- Presentation of ROGER : *ROc based GENetic learner*
- Validation and construction of a Medical test
- Perspectives

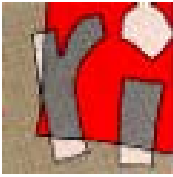


Univariate Analysis

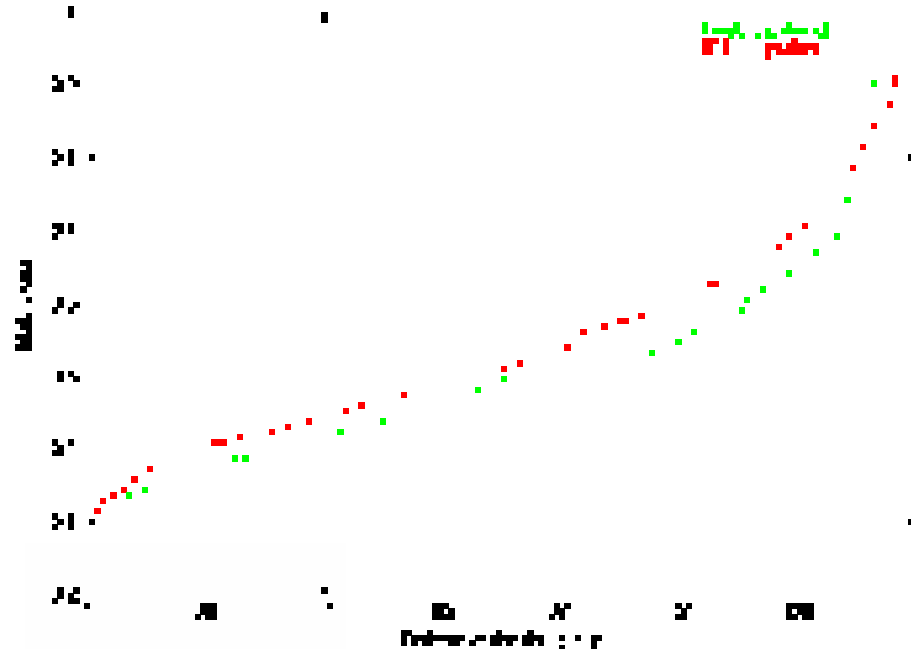
- graphical analysis of one risk factor for the healthy and ill patient (of the risk group)



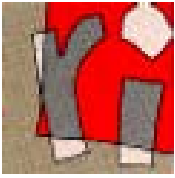
- limitations : factors interact



Multivariate Analysis



- limitations : manual selection of factors



Presentation of ROGER

ROc based GENetic learner

Search space / hypothesis space

linear function of attributes

$$h(\text{Ex}) = \sum \alpha_i \times \text{att}_i(\text{Ex}) \text{ avec } (\text{Ex}, +/-)$$

Evaluation of an hypothesis : area under the ROC curve

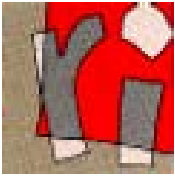
$h \rightarrow (\text{risk}(\text{Ex}), \text{Etiq}(\text{Ex}))$

Sort patient by increasing risk

++++-+----++-----++++-----

+ : ill

- : healthy

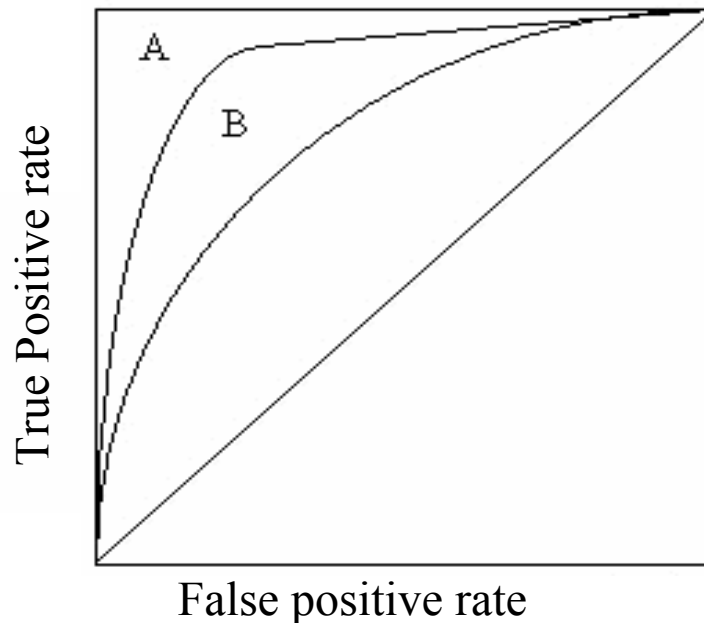


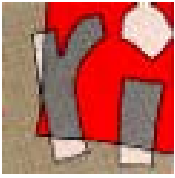
The ROC curve

(Receiver Operating Characteristics)

Evaluation of a medical test :

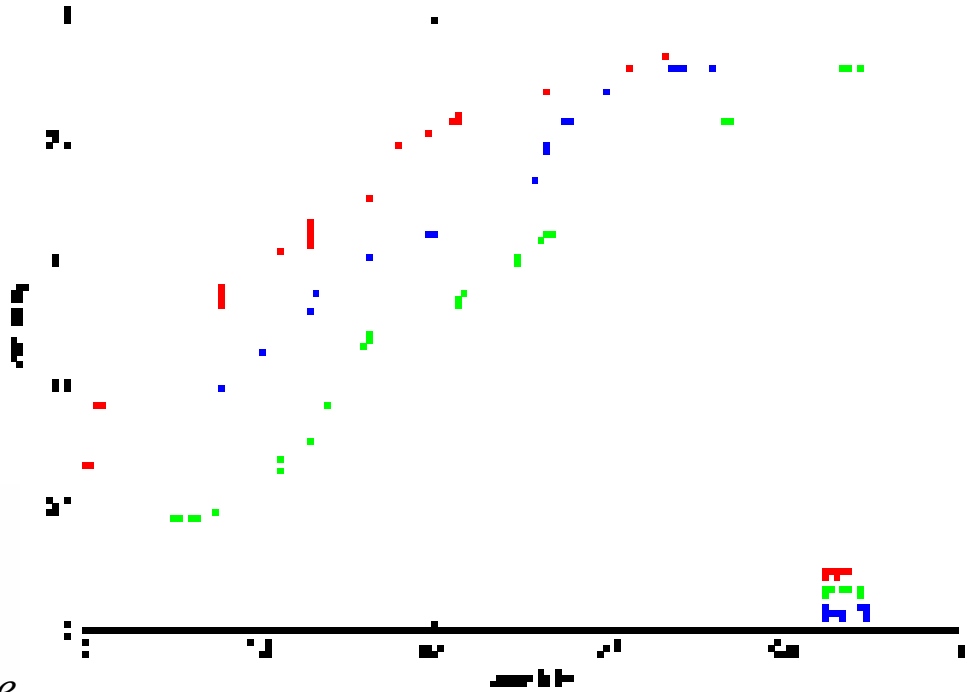
- True positive rate (medical sensitivity)
- True negative rate (medical specificity)





Validation and construction of a Medical test

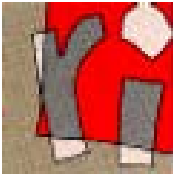
- goal: an inexpensive test (exploiting available information)



Test 1: the whole set of attributes

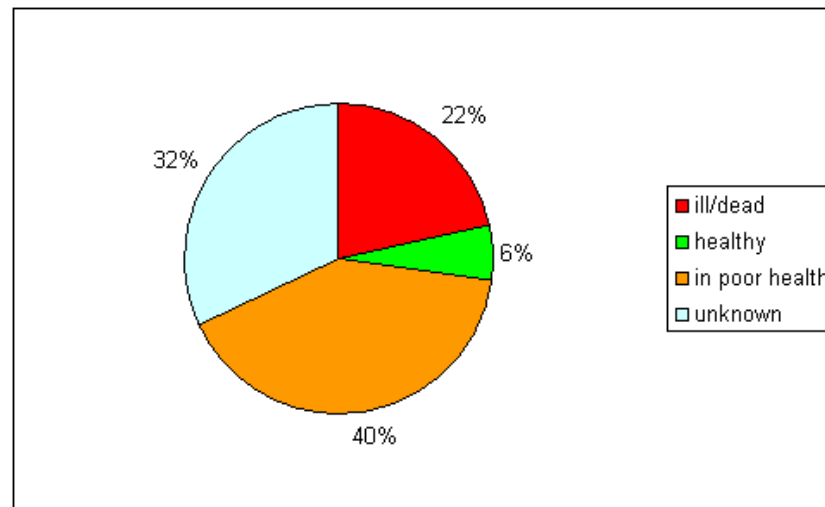
Test 2: no medical attributes

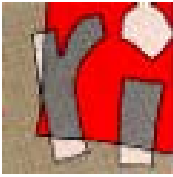
Test 3: medical attributes + bmi, age



validation of the first test

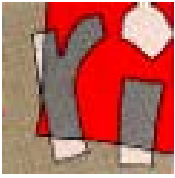
- target : patient of the normal group, with maximal risk estimate
- study of the patients with risk $>$ risk of the target





Conclusions

- from the learning and mining view point
 - relevant approach for feature selection and sensitivity analysis
- from a medical perspective
 - satisfactory performance
 - Demonstrate the contribution of habitus to CV disease



Perspectives

- from the learning and mining view point
 - the variability of evolutionary solution can be use in a visual data mining framework
- from a medical view point
 - hypotheses suggested by learning and mining must be validated within medical procedure