Using Machine Learning to Identify Benign Cases with Non-Definitive Biopsy

USA:
- 1 woman dies of breast cancer every 13 minutes
- In 2011:
  - 230,480 with cancer
  - 39,520 (~17%) died
Source: U.S. Breast Cancer Statistics - October 2011

Portugal:
- Per year:
  - 4500 new cases
  - 1500 deaths (33%)
Source: Liga Portuguesa Contra o Cancro - November 2011

Mammography:
- The cheapest and most efficient method to detect cancer in a preclinical stage

Methodology
- Learn rules that characterize benign biopsy cases
- Machine learning technique: Inductive Logic Programming (ILP)
- Aleph
- Data represented as Prolog facts
- Data collected between Dec 31 2005 and Dec 31 2009 From UW-Hospital

Data
- Objective
  Reduce the number of patients that go to excision

Results
- ILP can derive rules that accurately predict when a woman may not require excision after a non-definitive core breast biopsy.
- All five rules predict a substantial number of cases that are benign, and only two miss a single malignancy each.