

# Image Mining

## Implementation of Classification Algorithms on Mammograms

Bassem Fadlia  
Veena Sridhar

Dec 6 2000

CMPUT 695

1

# Contents

1. Introduction
2. Phase I ( Global feature extraction )
3. Phase II ( Local feature extraction)
4. The Algorithms used
5. Results
6. Conclusions and Future work

Dec 6 2000

CMPUT 695

2

# Introduction

Aim: To Classify Mammograms into

- Malignant
- Benign
- Normal

Data Set:

- DDSM database

Basic Problem

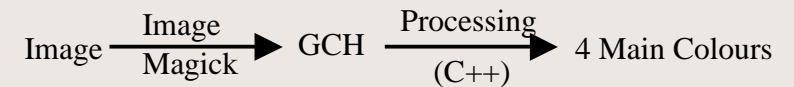
- Lack of domain knowledge

Dec 6 2000

CMPUT 695

3

# Pre-Processing ( Phase I )



Visual features

- Global colour histogram
- Further Processing to get the percentage of colours

Non-Visual features

- Age

Dec 6 2000

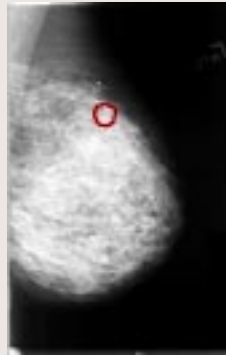
CMPUT 695

4

## Phase I contd.

### Problems faced

- The red markings in the image. Solved this problem by converting into gray-scale
- Unavailability of other important attributes like marital status and number of children



Dec 6 2000

CMPUT 695

5

## Pre-Processing ( Phase II )

Gray-Scale Image



Segmentation Algorithm



### Spatial Feature Extraction

- Number of segments
- Distance of the most significant segment
- Position of the most significant segment relative to the mammogram edge

Dec 6 2000

CMPUT 695

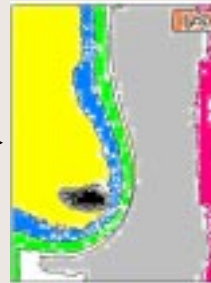
6

## Pre-Processing ( Phase II )

Gray-Scale Image



Segmentation Algorithm



### Spatial Feature Extraction

- Number of segments
- Distance of the most significant segment
- Position of the most significant segment relative to the mammogram edge

Dec 6 2000

CMPUT 695

7

## Phase II contd.

### Problems Faced

1. Noise being considered a segment. Solved by using boundary detection
2. The most significant segment did not always match the one that was marked.
3. Determination of threshold value.

Dec 6 2000

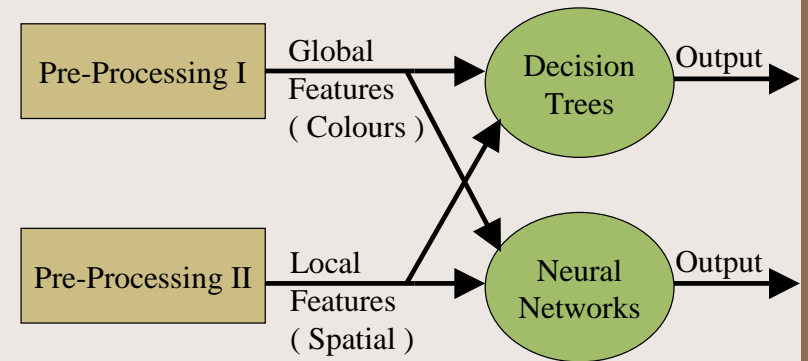
CMPUT 695

8

# The Classification Algorithms Used

1. Decision Tree Algorithm (ID3)
  - The learning and classification steps are generally fast.
2. Neural Network with Back propagation
  - High tolerance to noisy data.

# Block Diagram of the Project



# Results

	Global Features Visual only	Global Features Visual + age	Local features
Neural Networks	40%	60%	50%
Decision Trees	55%	44%	66%

The percentages of correct Classification

# Conclusions and Future Work

## Conclusions

- Pre-Processing is an important and time consuming step in Image Mining
- Both global and local features are important in the classification of Mammograms.
- More domain knowledge is needed to obtain realistic attributes.

## Future Work

- To work with Spiculated lesions
- To formulate the threshold value.