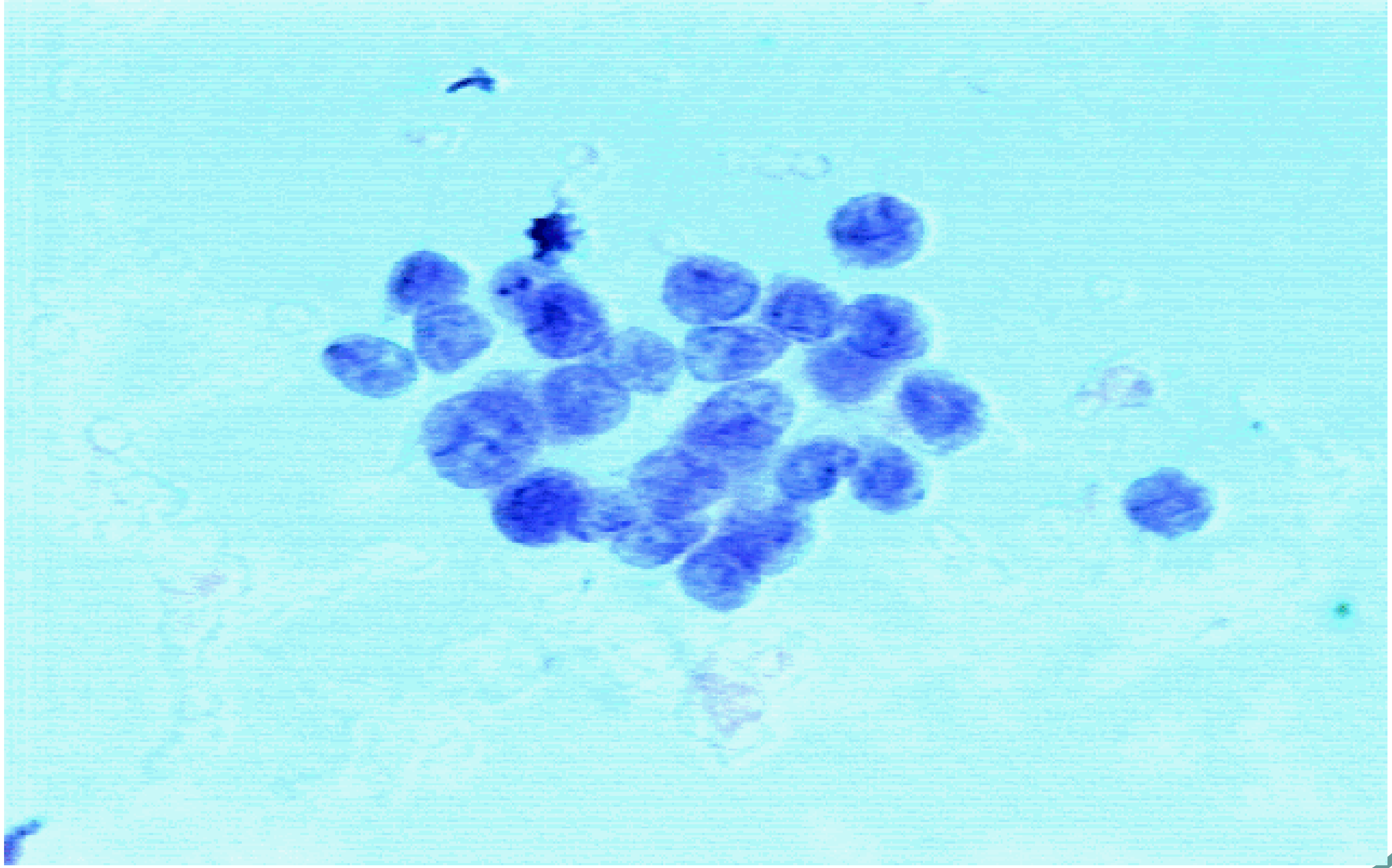
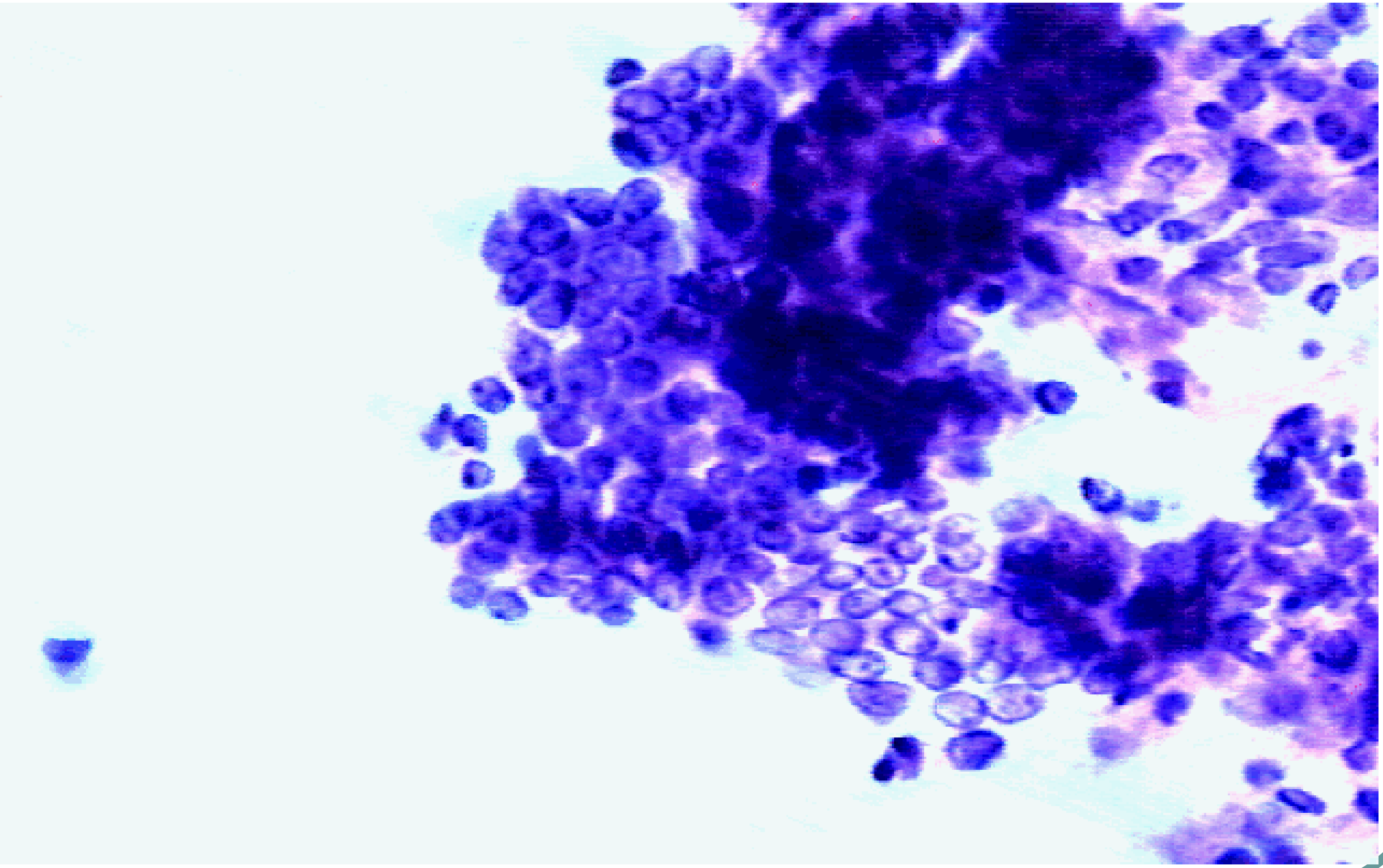
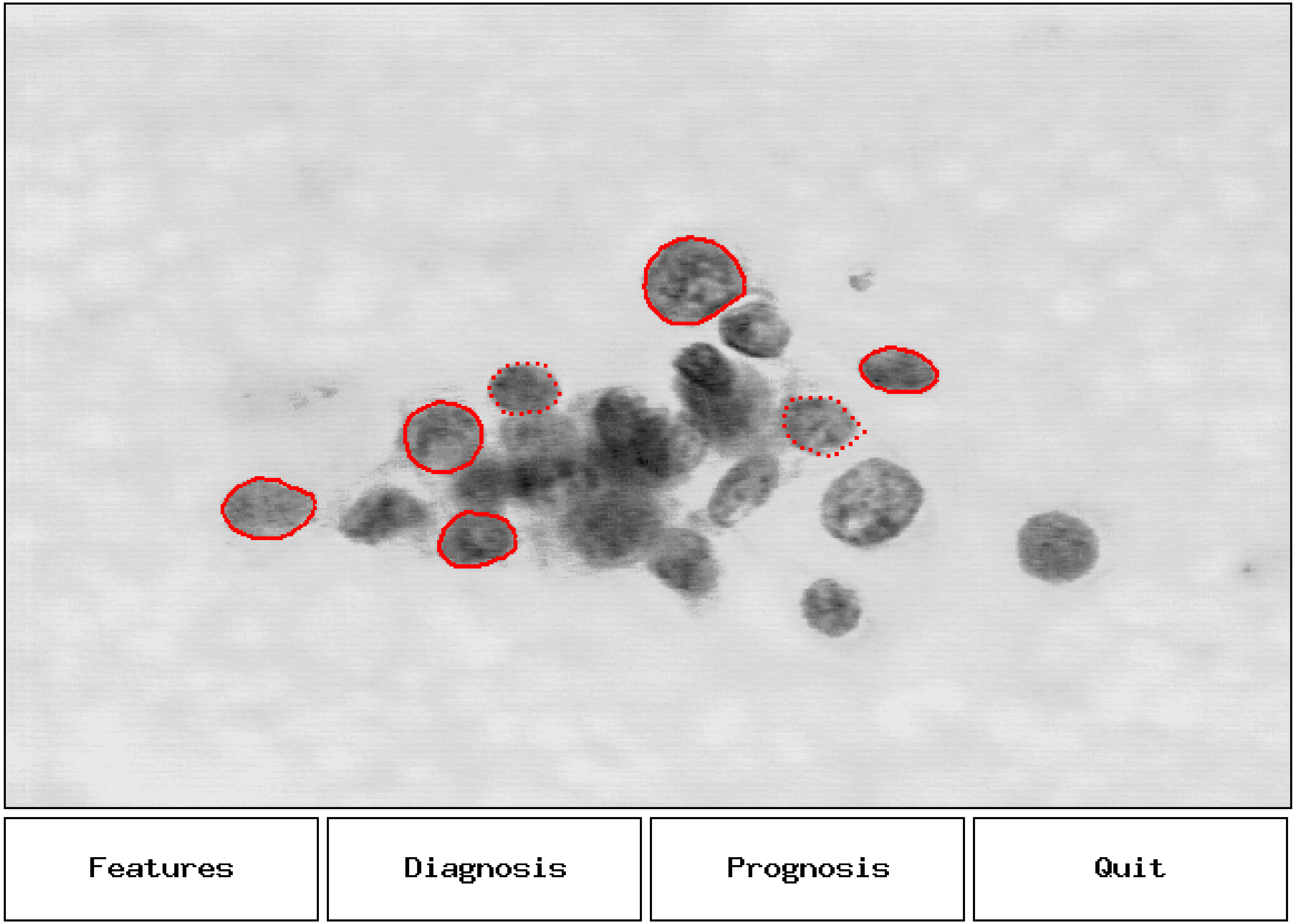
Breast Cancer Diagnosis

Wisconsin Breast Cancer Database

- Walberg and Mangasarian 1990
- 699 instances, 2 classes
  - 458 instances in the benign category
  - 241 instances in the malignant category






Features
Diagnosis
Prognosis
Quit

Predictor

- 9 features to predict one of benign and malignant categories.
  - clump thickness
  - uniformity of cell size
  - uniformity of cell shape
  - marginal adhesion
  - single epithelial cell size
  - bare nuclei
  - bland chromatin
  - normal nucleoli and mitoses

Features

(a) Radius was computed by averaging the length of radial line segments from the center of the nuclear mass to each of the points of the nuclear border.

(b) Perimeter was measured as the distance around the nuclear border.

Features

(c) Area was measured by counting the number of pixels in the interior of the nuclear border and adding one-half of the pixels on the perimeter.

(d) Perimeter and area were combined to give a measure of the compactness of the cell nuclei using the following formula: $\text{perimeter}^2/\text{area}$.

Features

(e) Smoothness was quantified by measuring the difference between the length of each radius and the mean length of adjacent radii.

(f) Concavity was determined by measuring the size of any indentations in the nuclear border.

Features

(g) Concave points counted the number of points on the nuclear border that lie on an indentation.

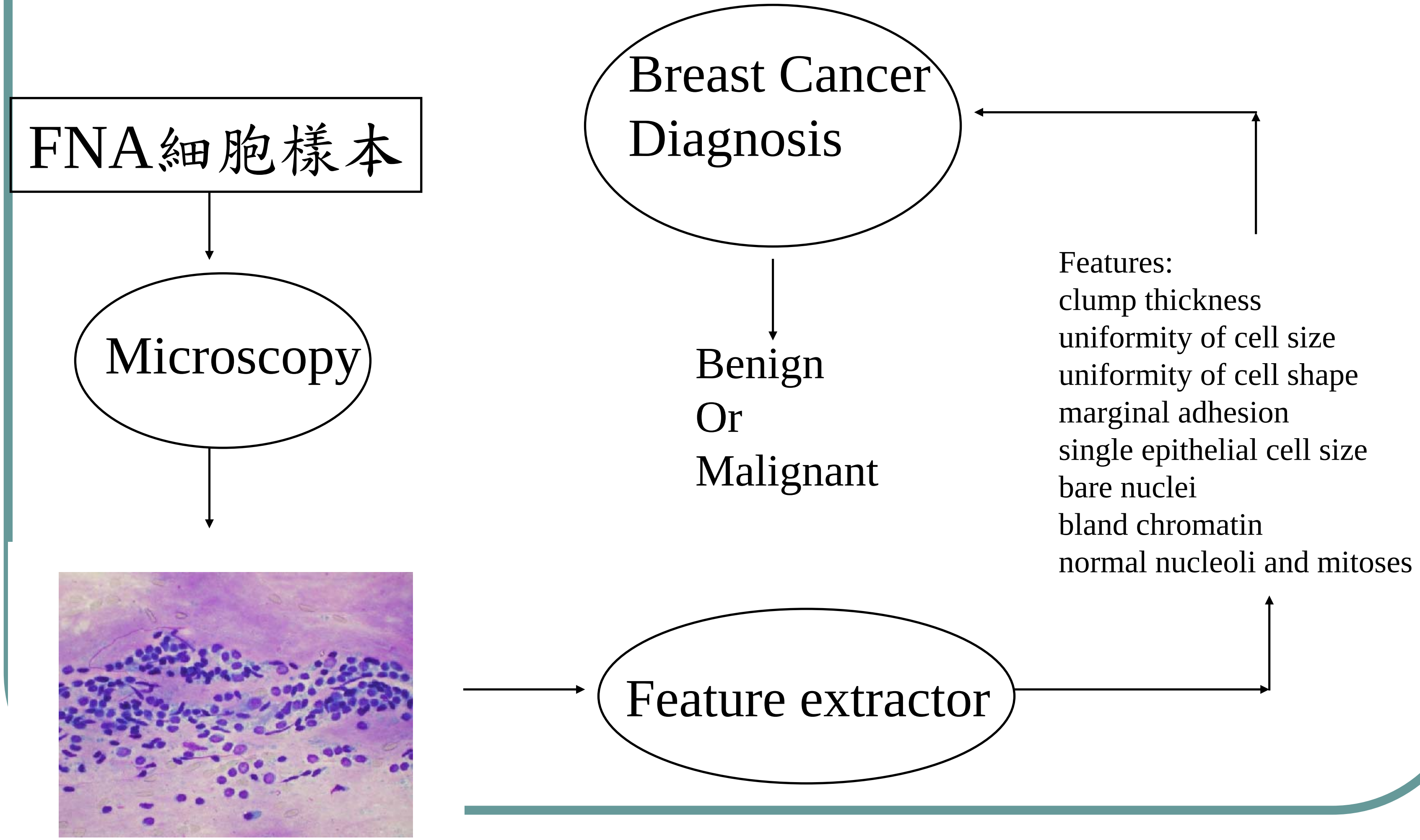
(*h*) Symmetry was measured by finding the relative difference in length between line segments perpendicular to and on either side of the major axis.

Features

(*i*) Fractal dimension was approximated using the "coastline approximation" described by Mandelbrot (12) that measured nuclear border irregularity.

(*j*) Texture was measured by finding the variance of the gray scale intensities in the component pixels.

Breast Cancer Classification



FNA細胞樣本

Microscopy

Feature extractor

Features:
clump thickness
uniformity of cell size
uniformity of cell shape
marginal adhesion
single epithelial cell size
bare nuclei
bland chromatin
normal nucleoli and mitoses

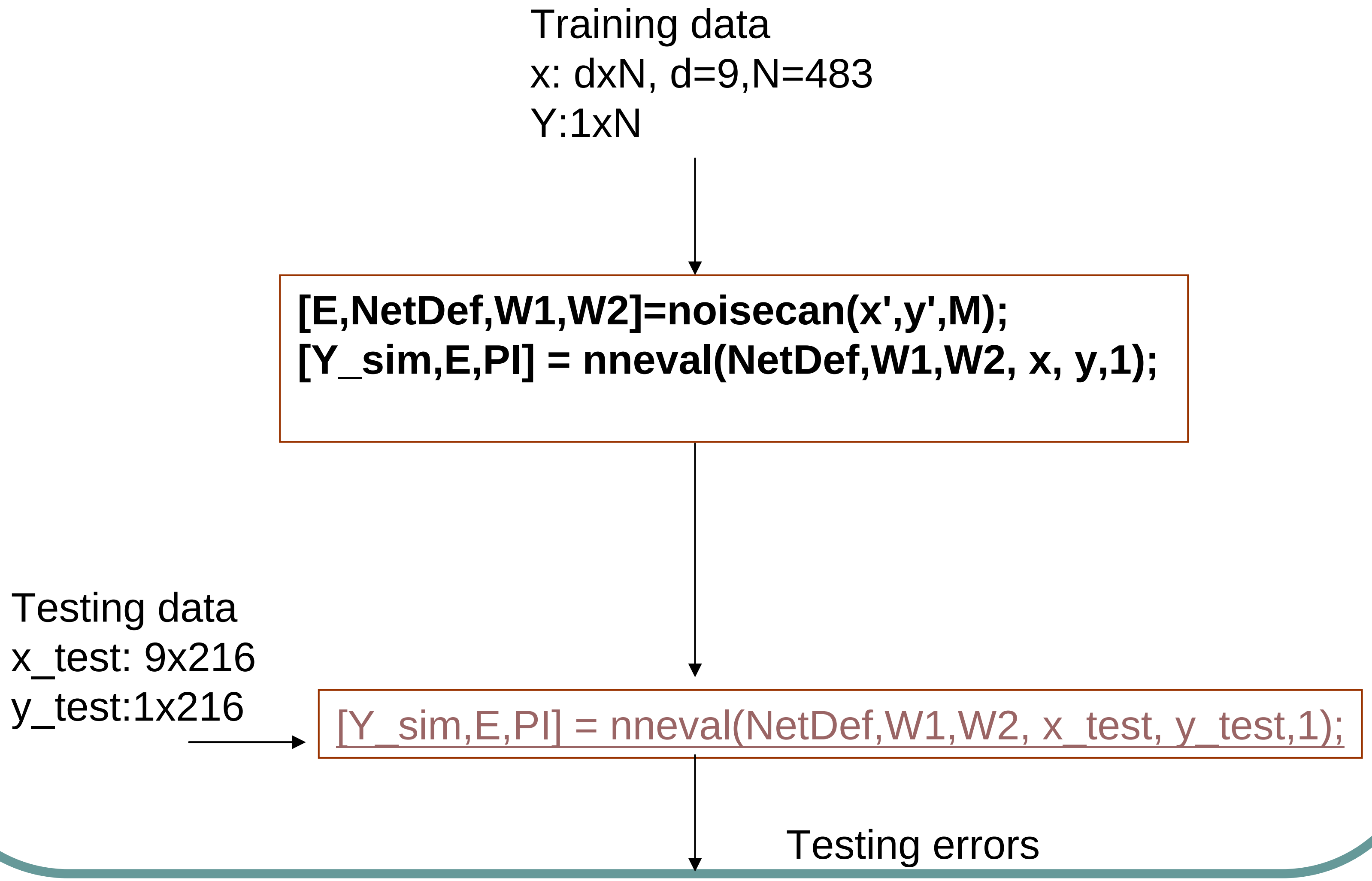
Breast Cancer Diagnosis

Benign
Or
Malignant

Input Data

```
TrainFile = 'data\\breast_train.dat';
TestFile = 'data\\breast_test.dat';
TrainSet = load(TrainFile, '-ascii');
TestSet= load(TestFile,  '-ascii');
```

Data flow



Training data
x: dxN, d=9,N=483
Y:1xN

```
[E,NetDef,W1,W2]=noisecan(x',y',M);
[Y_sim,E,PI] = nneval(NetDef,W1,W2, x, y,1);
```

Testing data
x_test: 9x216
y_test:1x216

```
[Y_sim,E,PI] = nneval(NetDef,W1,W2, x_test, y_test,1);
```

Testing errors

Exercise

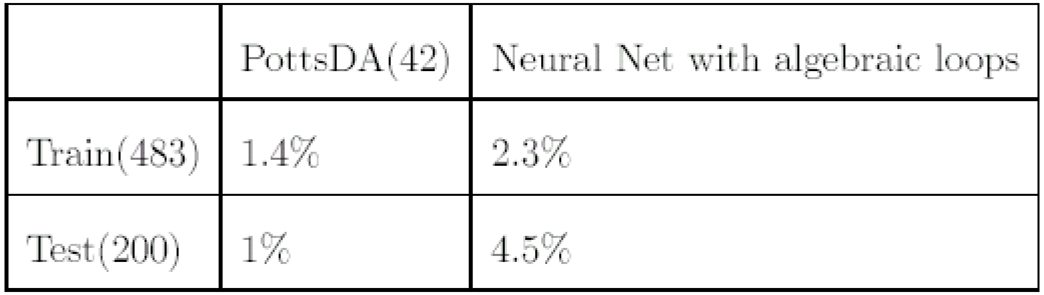
Breast cancer classification by learning MLP

- Input data
- Form training data: x and y
- Form testing data: x_test, y_test
- Train an MLP network
- Test
- Output testing errors

Simulation Results

- Walberg and Mangasarian 1990

  error rate for testing > 6%
- 683 instances of the database by Malini Lamego(2001)



	PottsDA(42)	Neural Net with algebraic loops
Train(483)	1.4%	2.3%
Test(200)	1%	4.5%

- For the 219-case test set, the RBF method with 80 kernels and the SVM method result in error rates 4.17% and 4.63% for testing.