Feature Extraction and Classification of Mammographic Masses

Presented by,

Jignesh Panchal
Anuradha Agatheeswaran

Introduction

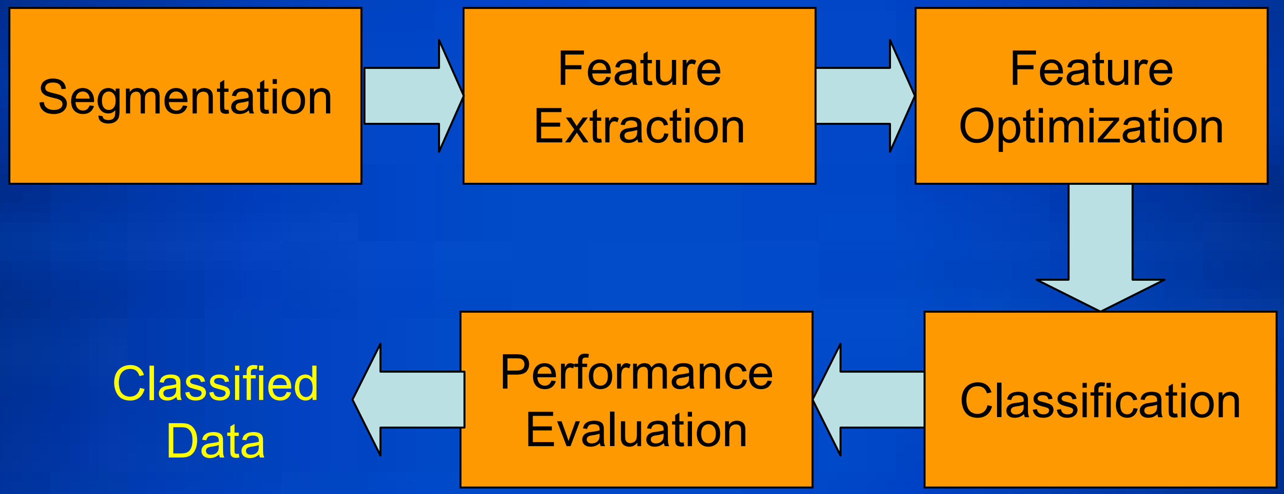
- Breast cancer is a leading cause in women deaths.

- Computer-Aided Systems are efficient tools in early detection of cancer.

- Generally the tumors are of two types:
  - Benign : Round
  - Malignant : Spiculated.

- A computer-aided classification system has been developed which classifies the mammographic tumors in two classes: benign or malignant.

System Overview



System Overview (Contd.)

- Segmentation: Images are manually segmented by the expert radiologists and the boundaries marked by them are assumed to be correct.

- Feature Extraction: In this study, total 9 features are extracted.
  - 5 Texture features
  - 3 Shape features
  - 1 Age feature

- Features are further optimized by using Stepwise Linear Discriminant Analysis.

- Maximum Likelihood Classifier is used for the classification and the performance is evaluated using leave-one-out testing method.

Mammographic Dataset

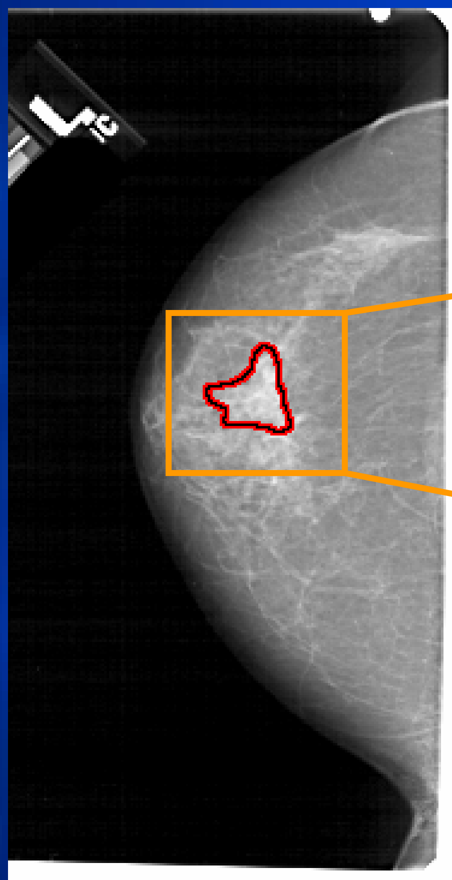
- Mammographic database for this system is obtained from the ‘Digital Database for Screening Mammography’, University of South Florida, Tampa.

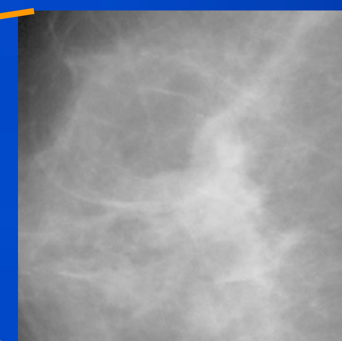
- In this study, total 73 mammograms are used
    - 41 Benign
    - 32 Malignant

- The images are compressed to 8 bits/pixel using the software “heathusf v1.1.0”, provided by USF.

- Region of interest is cropped to a size of 1024 x 1024 pixels, rather than using the entire mammograms.

Mammographic Dataset (Contd.)



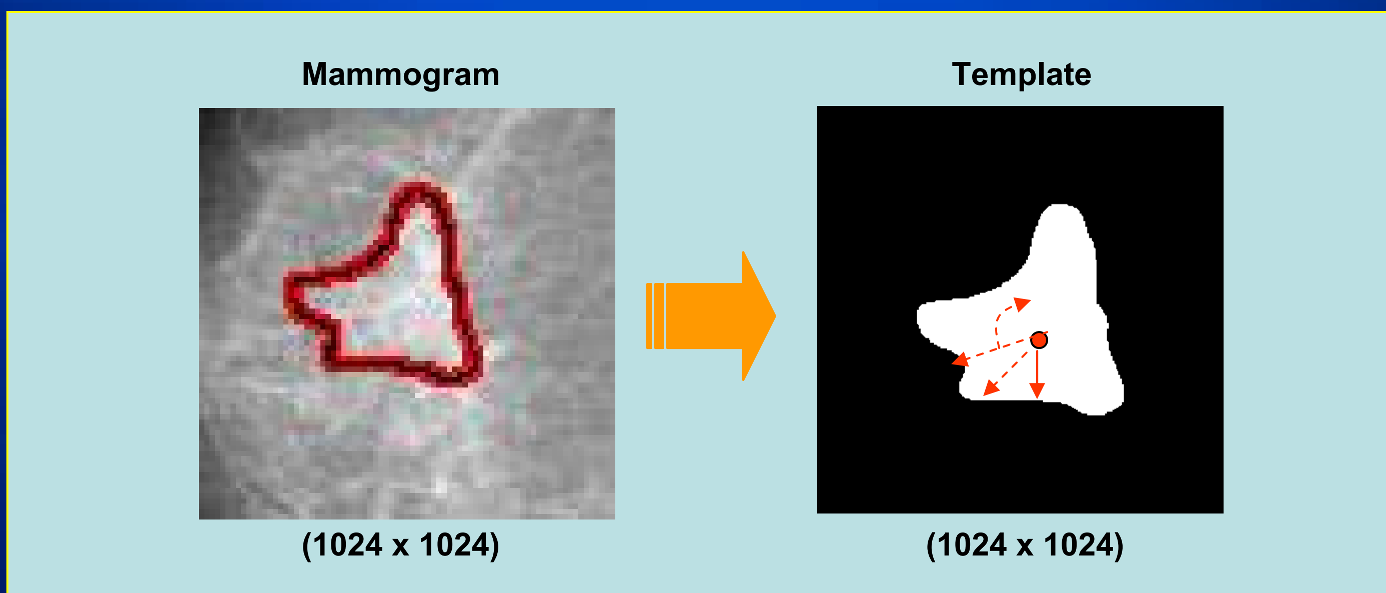


(1024 x 1024)

Feature Extraction: Shape Features

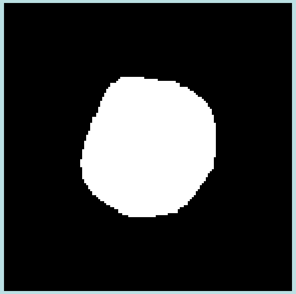
- Radial Distance Measure (RDM) is a very useful term in the shape analysis.

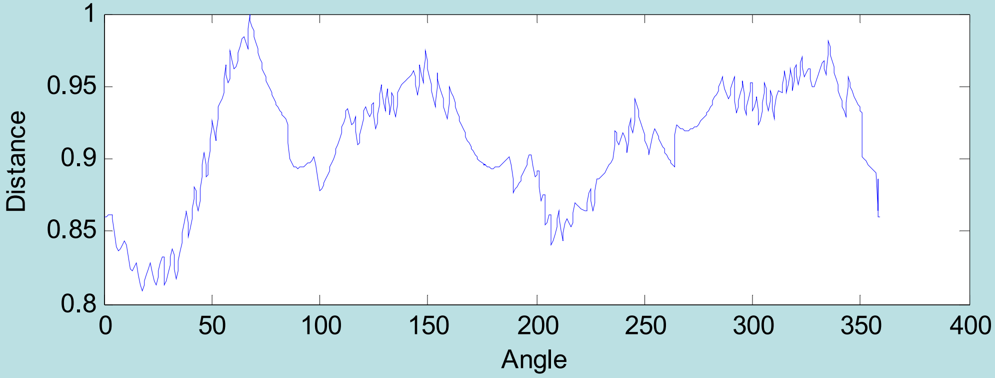
- RDM: It is basically the Euclidean distance calculated from the center of the tumor to the boundary pixels and normalized by dividing with the maximum length.



Shape Features (Contd.)

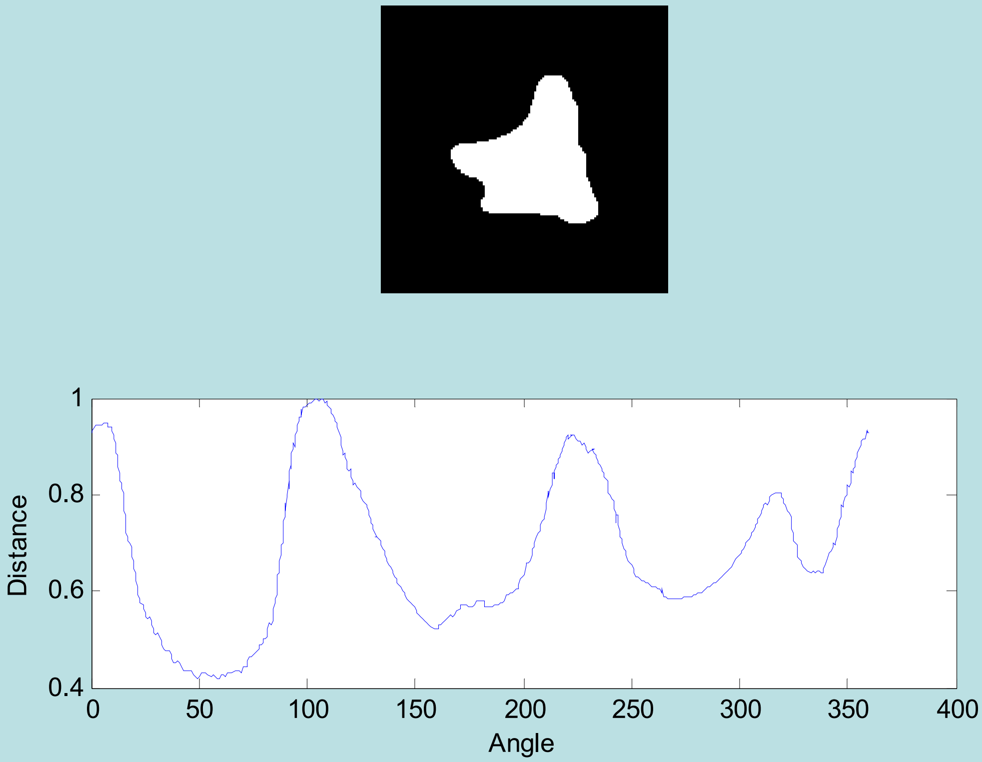
Benign





Shape Features (Contd.)

Malignant



Shape Features (Contd.)

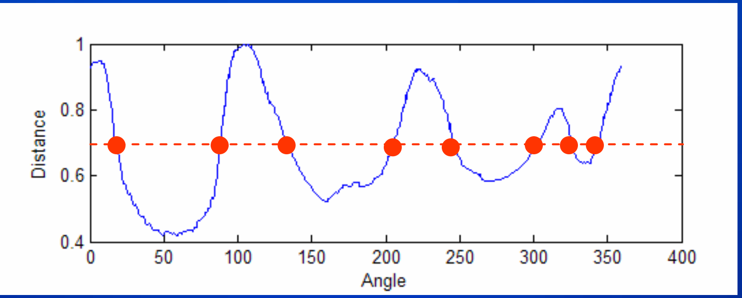
- Features Extracted:
  - Mean:

$$d_{avg} = \frac{1}{N}\sum_{I=1}^{N} d\,(i)$$

  - Variance:

$$\sigma^2 = \frac{1}{N}\sum_{I=1}^{N} (d\,(i) - d_{avg})^2$$

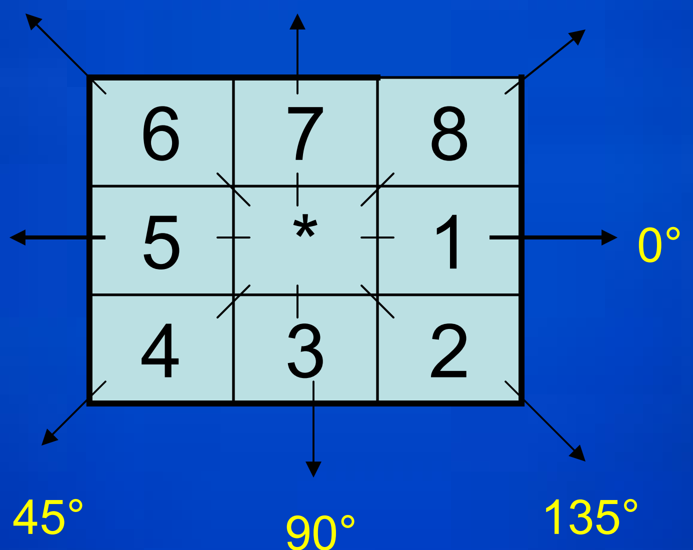
  - Zero crossings



Texture Analysis

- Texture features contains the information about the tonal variations in the spatial domain.

- Gray-tone spatial-dependence matrices



Direction considered

Texture Analysis (Cont.)

- Calculation of all four distance 1 gray-tone spatial-dependence (GTSD) matrices

0	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

4 X 4 image with 4 gray tone values

	0	1	2	3
0	#(0,0)	#(0,1)	#(0,2)	#(0,3)
1	#(1,0)	#(1,1)	#(1,2)	#(1,3)
2	#(2,0)	#(2,1)	#(2,2)	#(2,3)
3	#(3,0)	#(31)	#(3,2)	#(3,3)

General form of GTSD matrix

4	2	1	0
2	4	0	0
1	0	6	1
0	0	1	2

0°

6	0	2	0
0	4	2	0
2	2	2	2
0	0	2	0

90°

4	1	0	0
1	2	2	0
0	2	4	1
0	0	1	0

45°

2	1	3	0
1	2	1	0
3	1	0	2
0	0	2	0

135°

Texture Analysis (Cont.)

- Texture features extracted from different directions are

Energy Uniformity of the region	$f_1 = \sum_i \sum_j \{p(i,j)\}^2$
Contrast Amount of local variations	$f_2 = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \right\}$
Correlation Gray tone linear dependence	$f_3 = \dfrac{\sum_i \sum_j (ij)\, p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$
Inertia Degree of fluctuations of image intensity	$f_4 = \sum_i \sum_j (i-j)^2\, p(i,j)$
Homogeneity	$f_5 = \sum_i \sum_j \dfrac{1}{1+(i-j)^2}\, p(i,j)$

- For better accuracy, each texture feature in all direction are summed. Therefore there are 5 texture features instead of 20.

Feature optimization and Classification

- To optimize the feature , stepwise LDA is used.

Forward Selection

Features $f_1, f_2, \ldots\ldots\ldots f_N$

Performance measure (PM) of N features

Sort according to PM values

Loop N times to get the optimum set of feature so that the performance measure improves.

Optimum features $\tilde{f}_1, \tilde{f}_2, \ldots\ldots\ldots \tilde{f}_M$

Backward Rejection

Optimum features $\tilde{f}_1, \tilde{f}_2, \ldots\ldots\ldots \tilde{f}_M$

Loop M times to get the "most" optimum set of features so as to improve the PM compared to the forward selection

"Most" optimum features $\tilde{f}_1, \tilde{f}_2, \ldots\ldots\ldots \tilde{f}_K$

Feature optimization and Classification (Cont.)

- Maximum likelihood is used as a performance measure used to evaluate the features

- The classifier used is a maximum likelihood with LDA and method of testing was leave-one out

Results and Discussions

Feature	Accuracy
Energy	0.5625
Inertia	0.40625
Entropy	0.65625
Homogeinety	0.46875
Correlation	0.46875
RDM mean	0.53125
RDM variance	0.46875
Zero-crossings	0.375
Age	0.65625

Table 1: Accuracies of individual features

Table 2 (a): Confusion Matrix for Texture Features

	Benign	Malignant	
Benign	24	17	0.5854
Malignant	19	13	0.4063
	0.5581	0.4333	0.5068

Table 2 (b): Confusion Matrix for Shape Features

	Benign	Malignant	
Benign	38	3	0.9268
Malignant	27	5	0.1563
	0.5846	0.625	0.589

Table 3: Confusion Matrix for the optimum set of features after performing stepwise LDA

	Benign	Malignant	
Benign	34	7	0.8293
Malignant	9	23	0.7188
	0.7907	0.7667	0.7808

Conclusion and Future Work

- Accuracy of 78% is achieved with the combination of texture, shape and age feature

- Future work:
    - Better segmentation method
    - Implementations of rubber band straightening algorithm
    - Different algorithms for texture feature like gray-level run length method, gray level difference method can be implemented

References

- "Normal mammogram classification based on regional analysis" -Yajie Sun; Babbs, C.F.; Delp, E.J.; Circuits and Systems, 2002. MWSCAS- 2002. The 2002 45th Midwest Symposium on, Volume: 2 , 4-7 Aug 2002
- http://marathon.csee.usf.edu/Mammography/Database.html
- "Classification of Linear Structures in Mammographic Images - Reyer Zwiggelaar and Caroline R.M. Boggis, Division of Computer Science, University of Portsmouth, Greater Manchester Breast Screening Service, Withington Hospital, Manchester
- "Gradient and texture analysis for the classification of Mammographic masses" *Mudigonda, N.R.; Rangayyan, R.; Desautels, J.E.L.; Medical Imaging, IEEE Transactions on, Volume: 19, Issue: 10, Oct. 2000 Pages: 1032 – 1043*
- http://marathon.csee.usf.edu/Mammography/software/heat heathusf_v1.1.0.html
- "Texture Features for image Classification" *Haralick , R.M; Shanugam k; Dinstein, I;* Systems,Man and Cybernetics,IEEE transactions on Vol.SMC- 3,No. 6 Nov. 1973 Pages 610 – 621
- "Classifying Mammograhic Lesions Using Computerized Image Analysis" *Kilday, J; Palmieri, F; Fox, M.D;* Medical Imaging, IEEE Transactions on, Volume: 12, No.4, 1993, Pages: 664 – 669
- "Classifying Mammographic Mass Shapes Using the wavelet transform Modulus-Maxima Method" *Bruce, L.M; Adhami, R.R;* Medical Imaging, IEEE Transactions on, Volume: 18,No.12,Dec 1999, Pages: 1170 – 1177
- "Discrimination of subtly different vegetative species via hyperspectral data" *Mathur, A.; Bruce, L.M.; Byrd, J;* Geoscience and Remote Sensing Symposium, 2002. IGARSS '02. 2002 IEEE International Volume: 2 , 2002 Page(s): 805 –808
- "A Theoretical Comparison of Texture Algorithms " *Conners, R.W Harlow, C.A;* Pattern Analysis and Machine Intelligence, IEEE Transactions on, Vol: PAMI-2, No. 3, May 1980, Pages 204 - 222


Questions

Table 4: Confusion Matrix for all the features without age

	Benign	Malignant	
Benign	24	17	0.5854
Malignant	19	13	0.4063
	0.5581	0.4333	0.5068