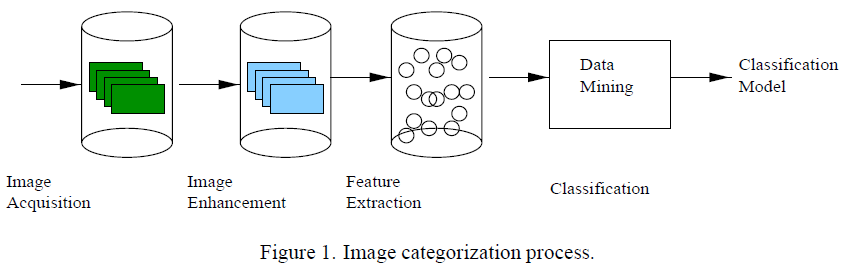
Data Mining Survey

Image Categorization Steps: Pre-classification steps and classification



from [4]

Pre classification:

1. Image Acquisition

Access to actual medical records even for educational purposes is difficult because of patients’ privacy reasons and bureaucratic hurdles. [4] relies on data collected from Mammographic Image Analysis Society for the study on breast cancer imaging. The training set consisted of images belonging to normal, benign and malign categories. They followed a hierarchical classification in the training images, where the abnormal images (with benign and malign tissues) were further classified into microcalcification, circumscribed masses, spiculated masses, ill-defined masses, architectural distortion and asymmetry. Factors such as the location, radius and type of abnormality were also recorded. In general, the quality of a dataset can be measured by the number and variety of images available. The images present must represent all possible categories or labels.

PreProcessing:

Pre processing of images before classification is necessary for improving the image quality. It is also needed for ‘data cleansing’ because the images available might be noisy, incomplete or inconsistent. For digital images, digitization process also introduces some noise into the image. Different illumination condition while clicking the image might also lead to differences between two images of the same object, taken from the same angle. Authors in [1] have reviewed how to solve problems such as different lighting conditions and ethnicity to model a person’s skin.

* **Cropping** removes non-relevant area, as well as the noise in background. Cropping may be done before image enhancement as in [4], so that the background noise does not get enhanced.
* **Feature subset Selection** may be used as a way to remove irrelevant information and reduce data dimensionality. The most discriminating set of features must be selected, weighted by the importance of features.
* **Segmentation** may be used to divide the image into smaller areas for processing.
* **Image enhancement** accentuates the areas of interest. Reduces the effect of varying illumination conditions by improving image contrast. This is done using a technique called histogram equalization. Histogram equalization makes dark regions darker and light regions lighter. This helps to standardize the images to an extent, for further processing.
* **Feature extraction** extracts relevant features and stores them, to be used as an input to classification algorithms. In [4], four parameters were computed for the extracted features:

1. mean
2. variance
3. skewness
4. kurtosis

Classification

1. **Neural networks** with back propagation was used in [4]. describe
2. **Association Rule mining** was used in [4]. describe
3. **Decision Tree was used in [10]**
4. **[11] uses Bayesian Classifiers**

Comparison of Classification methods

According to [1], image classification work is currently being done mainly by manual annotation for cataloging and indexing.

In [4] for breast cancer detection by automatic clustering, use of neural networks with back propagation was found to be less sensitive to the database imbalance at a cost of high training times. Associative mining with a much more rapid training phase, obtained better results on a well balance dataset.

Image retrieval by matching depends directly on classification, and the classification technique used must aid retrieval as well. The same distance metric that was used for classification can also be used for efficient retrieval.

Retrieval

[11] uses nearest k neighbours

**References:**

[1] W. Moudani, A.R. Sayed, “Efficient Image Classification using Data Mining,” IJCOPI Vol. 2, No. 1, Jan-April 2011

[2] J.G. Dy, C.E. Brodley, Lynn S. Broderick, A.M. Aisen, “Unsupervised Feature Selection Applied to Content-Based Retrieval of Lung Images,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 3

[3] D. Keysers, J. Dahmen, H. Ney, B.B. Wein, T.M. Lehmann, “Statisical Framework for Model-Based Image Retrieval in Medical Applications,” Journal of Electronic Imaging 12(1), 59–68 (January 2003).

[4] M.L. Antonie, O.R. Zaiane and A. Coman, “Application of Data Mining Techniques for Medical Image Classification,” In. Proc Second International Workshop on Multimedia Data Mining

[5] Z.S. Zubi, R. A. Saad, “Using Some Data Mining Techniques for Early Diagnosis of Lung Cancer”

[6] P. Rajendran, M. Madheswaran, “Hybrid Medical Image Classification Using Association Rule Mining with Decision Tree Algorithm,” In Journal of Computing, Volume 2, Issue 1, January 2010

[7] H. Muller, N. Michoux, David Bandon and Antoine Geissbuhler, “A Review of Content Based Image Retrieval Systems in Medical Applications- Clinical Benefits and Future Directions,” IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2004.

[8] N. M. Sirakovand P. A. Mlsna, “Search Space Partitioning using Convex Hull and Concavity Features for Fast Medical Image Retrieval,”

[9] T.M. Lehmann, M.O. Guld, C. Thies, B. Fischer, D. Keysers, M. Kohnen, H. Schubert, B.B. Wein , “Content Based Image Retrieval in Medical Applications for Picture Archiving and Communication Systems,” [Proceedings Vol. 5033](http://spie.org/x648.xml?product_id=459269&origin_id=x648), Medical Imaging 2003: PACS and Integrated Medical Information Systems: Design and Evaluation

[10] S.M. Khan, M.R. Islam, M.U. Chowdhury, “Medical Image Classification Using an Efficient Data Mining Technique,” in Proceedings of International Conference on Machine Learning and Applications (ICMLA'04), Louisville, KY, USA, December 2004.

[11] Y. Liu and F. Dellaert “Classification Driven Medical Image Retrieval”