Sugandha (sugandh2@illinois.edu)

Title: A survey on Automatic Classification and Retrieval Techniques for Medical Image Mining

Abstract:

Introduction:

The use of data mining techniques for medical images in health care is on a rise. Doctors are increasingly relying on mining existing images for computer aided detection (CAD) of new cases. These include X-ray, MRI, CT, mammograms and skin region images of existing cases. Data mining of such existing images can provide relevant knowledge for diagnosis. Some of the challenges being faced in using such techniques are a large data set of available images, low contrasts, inherent noise in medical images, and the problem of organization and retrieval from the image database. These problems can be solved by introduction of efficient medical image classification and retrieval methods.

Importance of the study:

In cases such as cancer detection, any human error can prove very costly. Moreover, current methods of detection do not provide a 100% certain diagnosis. A false positive in can lead to an unnecessary biopsy, whereas a false negative can lead to a patient's death. It has been shown that the use of image mining techniques in conjunction with a doctor's diagnosis can significantly improve the accuracy of the diagnosis, thus reducing costs in terms of patients' lives as well as health care spending.

Methodology and Topics Covered:

This survey aims to study and compare the various methods being used for automatic classification and retrieval of medical images for the purpose of data mining. The general process for all the image classification methods broadly includes the following steps:

Image Preprocessing: Involves image segmentation, feature subset selection to isolate the region of interest and image enhancement (normalization using histogram equalization).

Classification algorithm: This includes methods such as associative rule mining, decision trees, neural networks, and statistical and probabilistic frameworks.

Most hospitals use textual description for retrieving images from a database. This method is not efficient because it depends on individuals' perception and interpretation, and alphanumeric queries are often insufficient to describe an image. Text-based description tends to be incomplete, imprecise, and inconsistent in specifying visual information [1]. The first step in most of the automatic retrieval methods is also the classification of the query input image to recognize the relevant category and therefore, methods similar to the classification algorithms discussed above can be used. A study and comparison of Knowledge Discovery Question Language (KDQL), content based image retrieval (CBIR), query by example (QBE), customized query approach[2] and image retrieval in medical applications (IRMA)[3] methods for medical image retrieval is being done as a part of this survey. Most of these methods evaluate color, texture and shape of images and are built upon the classification methods discussed.

Conclusion:

The survey concludes with a summary of the current research and future scope in this area.

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. Sirakov and 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, 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.