Soft Computing based Medical Image Mining: A Survey

Amjad Khan[1] and Zahid Ansari[2]

[1]Department of Electronics and Communication Engineering, P.A College of Engineering, Mangalore, India

[2]Department of Computer Science and Engineering, P.A College of Engineering, Mangalore, India

Abstract *—Medical image mining is one of the most rewarding and challenging field of application in data mining and knowledge discovery. Soft computing is a consortium of methodologies that provides flexible information processing capability. Its aim is to exploit the tolerance for imprecision, uncertainty, approximate reasoning, and limited truth in order to achieve tractability, robustness, and low-cost solutions. Soft computing techniques such as fuzzy sets, neural networks, genetic algorithms, and rough sets are most widely applied for image mining. This paper presents a review on various papers on medical image mining using soft computing techniques and related issues were discussed and listed which can be resolved suitably using soft computing techniques.*

Keywords *— Soft Computing, Fuzzy Sets, Neural Networks, Image Mining.*

I. INTRODUCTION

Image mining is a vital technique which is used to mine knowledge straightforwardly from image as shown in the Figure 1. It is an interdisciplinary field that integrates techniques like computer vision, image processing, data mining, machine learning, data base and artificial intelligence that goes beyond the problem of retrieving relevant images [1], [2]. Image mining is an extension of data mining in the field of image processing which handles with the hidden knowledge extraction, image data association and additional patterns which are not clearly gathered in the images. Computer power and medical scanner data alone are not enough. It is necessary to extract the required boundaries, surfaces, and segmented volumes these organs in the spatial and temporal domains. Image segmentation is essentially a process of pixel classification, wherein the image pixels are segmented into subsets by assigning the individual pixels to classes.

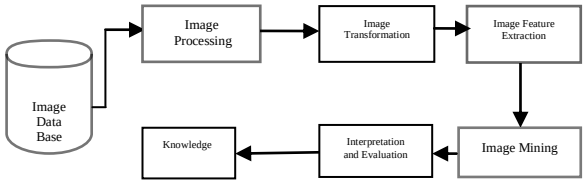

Figure 1: Image mining process

Image mining process involves the fundamental steps such as retrieving suitable image from the database whose features to be extracted, image processing, image transformation, image feature extraction, image mining, image interpretation and evaluation and knowledge discovery. Due to the rapid growth of medical data, it has become necessary to use data mining techniques such as classification and clustering to help decision support and predication systems in the field of healthcare and also helps to prepare some methods for diagnosis, prognosis, decision making. Medical image mining is the most rewarding and challenging field of application in data mining and knowledge discovery. The reason for challenging is due to the data sets of huge, complex, diverse, hierarchical, time series and varying in quality. As the available healthcare datasets are fragmented and diffused in nature, thereby making the process of data integration is a highly challenging task [16],[17]. Medical imaging techniques such as X-ray, computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), ultrasound (USG), etc. are indispensable for the precise analysis of various medical pathologies.

Clustering technique is the unsupervised classification of images into groups. The different methods of clustering are used to extract the useful patterns. The selection of data and methods for clustering is an important task in medical diagnosis and needs the knowledge of the domain. Each clustering is suitable for some medical applications. An efficient clustering method should be selected that suits desired task [3], [26]. Classification technique is a supervised learning problem which classifies a data item into one of several predefined categorical classes. In a medical domain the supervised dataset contains past patient information with a disease based on the patient's symptoms. The fundamental objective for carrying out image classification or clustering in medical image mining is to acquire content information the users are interested in from the image group label associated with the image. [4], [29].

In this paper, we present a survey on medical image mining using soft computing techniques. Organization of the rest of this article is as follows. In section II we have performed the literature review. In section III we have highlighted the related issues of the medical image mining. In section IV we present the proposed methodologies to carry out the brief analysis of the reviews performed in section II and III. Finally in section V we conclude the paper with the remarks, that applying soft computing methods on medical images helps the analysis easy and supports to prepare some methods for diagnosis, prognosis and decision, which inturns makes the health care system more accurate and efficient.

II. LITERATURE REVIEW

Here, the reviews on various image mining techniques that are carried out earlier are described briefly. In paper [20], the work done combines some ideas of image segmentation into content based image classification. The concept of image segmentation for medical images using techniques of clustering is being proposed. It is concluded that for the clustering of image data, fuzzy k-means is better than k-means by increasing the fuzzy factor and takes lesser time to cluster the images. In paper [21], various data mining techniques employed for medical data mining are discussed in paper [22]. Comparison of various classification techniques that are implemented for the x-ray images and various related methods that are used to classify and indexed x-ray images are discussed. It is concluded that Quality and speed of retrieving medical images from large collections of databases are more important and essential for the better functioning of health organizations. In paper [23], the Medical Image Segmentation and classification of segmentation methods are discussed. It is noted that Region classification is an essential process in the visualization of brain tissues of MRI. Segmentation and characterization of Brain MRI image regions using self organizing map and neuro fuzzy techniques to automatically extract tumour region of brain. In paper [24], to identify high active neuron analysis for Lysosomal storage disorder (LSD), k-means clustering is applied and it is noted that the MRI cluster analysis of the communication of neuron process will increase the speed. The density based as well as hierarchal clustering algorithms are compared for high active neurons for the further mapping process. In paper [25], an overview of different methods used to extract the Image features and different mining techniques applied on it are discussed. Mining techniques aim at reducing the time of searching and can be effectively used by combining different algorithm and concepts. In paper [14], Decision tree based image processing and image mining technique were projected. Pixel-wise image characteristics were extracted and changed into a database which permits a variety of data mining algorithms to make explorations on it. In paper [26], different clustering techniques and their advantages and drawbacks are discussed. In paper [27], bacterial image classification for all types of bacteria and color segmentation has been done with histogram thresholding. The morphological features like erosion, reconstruction and dilation with binarization and thresholding has been done for the edge detection using freeman contour. SOM and k-means have been used for clustering of bacteria. This method will assist the doctors in deciding the intensity of the diseases or infection that can save critically ill patients. In paper [28], various clustering techniques and parameter are discussed. The shortcoming of k-means clustering algorithm to find optimal k value and initial centroid for each cluster is discussed and validity of N-cut algorithm, mean shift algorithm and M-step algorithm is discussed for overcoming the shortcoming of k-means algorithm [34]. In paper [29], data mining and statistical techniques used in medicine are discussed. This paper deals on the application of data mining for skin diseases. Diagnosis of erythemato-squamous diseases is worked using dermatology datasets. In paper [30], various clustering techniques applied in data mining are presented. It is identified two major drawbacks that influence the feasibility of cluster analysis in real world applications in data mining includes the inefficiency of clustering algorithms on handling arbitrarily shaped clusters in extremely large datasets directly impacts the effect of cluster validation and very high computational cost of statistics-based methods for assessing the consistency of cluster structure between the sampling subsets [35].

III. RELATED ISSUES

In the literature review of the various medical image mining, some issues were observed which is depicted as below.

- The real world images usually contain missing, inaccurate, noisy or inconsistent data.
- The imperfect data from real world images.
  - Negatively impact on pattern discovery and results obtained are not trustworthy
  - Increase the chances of discovering spurious patterns.

To overcome the above mentioned issues soft computing techniques can be employed.

IV. PROPOSED METHODOLOGIES

Soft computing is a consortium of methodologies that provides flexible information processing capability [4]. Its aim is to exploit the tolerance for imprecision, uncertainty, approximate reasoning, and partial truth in order to achieve tractability, robustness, and low-cost solutions [1]. The role model for soft computing is the human mind. The guiding principle of soft computing is: Exploit the tolerance for imprecision, uncertainty, partial truth, and approximation to achieve tractability, robustness and low solution cost. Soft computing may be viewed as a foundation component for the emerging field of conceptual intelligence. Few soft computing tools are: Fuzzy Systems, Neural Networks, Evolutionary Computation, Machine Learning and Probabilistic Reasoning.

- Soft computing is a collection of methodologies that
  - Exploit the tolerance for imperfection and uncertainty.
  - Provide capability to handle real life ambiguous situations.
  - Try to achieve robustness against imprecision.

- Principal Soft computing techniques includes fuzzy sets, neural networks, genetic algorithms, and rough sets are most widely applied in the medical image mining. Generally fuzzy sets are suitable for handling the issues related to understandability of patterns, incomplete/noisy data, and mixed media information which provide approximate solutions faster. Neural networks are nonparametric, robust, and exhibit good learning and generalization capabilities in data-rich environments. Genetic algorithms provide efficient search algorithms to select a model from mixed media data, based on some objective function. Rough sets are suitable for handling different types of uncertainty in medical image data [5],[8],[13].
  - Soft computing based classification technique consists of fuzzy classification, neural classification etc. Classification is a supervised learning problem which classifies a data item into one of several predefined categorical classes in which the output is a discrete classification and the possible mutually exclusive classes of the problem[4], [7].
  - Soft computing based clustering technique consists of fuzzy clustering, artificial neural networks for clustering etc. Clustering is the unsupervised classification of images into groups. Classification of similar objects into different groups, or more precisely, the partitioning of a data set into subsets called clusters. Clustering is a common technique for statistical data analysis, which is used in many fields, including machine learning, data mining, pattern recognition, image analysis and bioinformatics [3]. Among the fuzzy clustering methods, fuzzy *c*-means (FCM) algorithm is the method used in image segmentation because it has robust characteristics for ambiguity and it has a good performance in a large class of images. The *k*-Means clustering algorithm is one of the most commonly used methods for partitioning the data. [5], [6], [33].
    - A fuzzy clustering: This method assigns degrees of membership in several clusters to each input pattern. A fuzzy clustering can be converted to a hard clustering by assigning each pattern to the cluster with the largest measure of membership [31], [32].
    - Artificial neural networks (ANNs) for clustering: the features of ANNs are inherently parallel and distributed processing architectures. They can act as pattern normalizers and feature selectors by appropriate selection of weights [31], [3].

V. CONCLUSION

From the survey it is depicted that applying soft computing techniques for medical image mining may improve the performance and fetch better results. Related issues were listed and discussed from the review which can be resolved suitably using soft computing techniques. Soft computing features can help us to make the medical image mining more reliable, more effective and efficient. Its aim is to exploit the tolerance for imprecision, uncertainty, approximate reasoning, and partial truth in order to achieve tractability, robustness, and low cost solutions. Different soft computing tools can be used in different phases of medical image mining using soft computing. The phases include denoising, classification, clustering, filtering, searching, matching, customized searching and filtering. Hence help to prepare some methods for diagnosis, prognosis and decision making most accurate and efficient.

REFERENCES

[1] Hsu, Wynne, Mong Li Lee, and Ji Zhang. "Image mining: Trends and developments." Journal of Intelligent Information Systems 19, no. 1 (2002): 7-23.

[2] Hema, a., and e. annasaro. "A survey in need of image mining techniques." international journal of advanced research in computer and communication engineering (ijarcce) issn (print):2319-5940.

[3] S.Zulaikha Beevi , M.Mohammed Sathik , K.Senthamaraikannan, "A Robust Fuzzy Clustering Technique with Spatial Neighborhood Information for Effective Medical Image Segmentation" (IJCSIS) International Journal of Computer Science and Information Security, Vol. 7, No. 3, March 2010.

[4] P.K.Vaishali , A.Vinayababu , "Application of Data mining and Soft Computing in Bioinformatics" International Journal of Engineering Research and Applications (IJERA), ISSN: 2248-9622, Vol. 1, Issue 3, pp.758-771.

[5] Ramadass Sudhir , " A Survey on Image Mining Techniques: Theory and Applications" Computer Engineering and Intelligent Systems, ISSN 2222-1719, Vol 2, No.6, 2011.

[6] J. Priya, "A Survey on Image Mining Techniques for Image Retrieval" International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), ISSN: 2278 – 1323, Volume 2, Issue 7, July 2013.

[7] Yushu Xiong, "A Clustering Algorithm Based on Rough Set and Genetic Algorithm" TELKOMNIKA, Vol. 11, No. 10, October 2013, pp. 5782 ~ 5788, e-ISSN: 2087-278X.

[8] Mignotte, M. "A de-texturing and spatially constrained K-means approach for image segmentation" Pattern Recognition Lett. (2010), doi:10.1016/j.patrec.2010.09.016. Elsevier.

[9] S. Belhassen and H. Zaidi, "A novel fuzzy C-means algorithm for Unsupervised segmenation of heterogeneous tumors in PET" Medical Physics, Vol. 37, No. 3, March 2010.pp.1309-1323.

[10] Caroline Petitjean, Jean-Nicolas Dacher. A review of segmentation methods in short axis cardiac MR images. Medical Image Analysis, Else vier, 2010.

[11] Krishna Kant Singh, Akansha Singh, "A Study of Image Segmentation Algorithms for Different Types of Images" IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 5, September 2010.

[12] Foschi, Patricia G., Deepak Kolippakkam, Huan Liu, and Amit Mandvikar. "Feature Extraction for Image Mining." In Multimedia Information Systems, pp. 103-109. 2002.
[13] Piero P. Bonissone, "Fuzzy Logic and Soft Computing: Technology Development and Applications" Schenectady NY 12309, USA. July 10, 1997.
[14] Lu, Kun-Che, and Don-Lin Yang. "Image Processing and Image Mining using Decision Trees." J. Inf. Sci. Eng. 25, no. 4 (2009): 989-1003.
[15] de Oliveira, J. Valente, and Witold Pedrycz, eds. Advances in fuzzy clustering and its applications. New York: Wiley, 2007.
[16] K.J. and G.W. Moore Cios, "Uniqueness of medical data mining ," Artificial Intelligence in Medicine, pp. 1-24, 2002.
[17] Zhang, Y., Katta, "Medical Data Mining," *Data Mining and Knowledge Discovery*, pp. 305-308, 2002.
[18] Irene M. Mullins et al., "Data mining and clinical data repositories: Insights from a 667,000 patient data set," *Computers in Biology and Medicine*, vol. 36, pp. 1351-1377, 2006.
[19] J. C., Lobach, D. F., Goodwin, L. K., Hales, J. W., Hage, M. L., EdwardHammond, W. Parther, "Medical Data Mining: Knowledge Discovery in a Clinical Data Warehouse," 1997.
[20] Md. Khalid Imam Rahmani, Naina Pal, Kamiya Arora, " Clustering of Image Data Using K-Means and Fuzzy K-Means" International Journal of Advanced Computer Science and Applications ,Vol. 5, No. 7, 2014.
[21] V.Gayathri, M.Chanda Mona, S.Banu Chitra, "a survey of data mining techniques on medical diagnosis and research" Singaporean Journal of Scientific Research (SJSR), Vol.6. No.6, Pp. 301-310,2014.
[22] Priti Sahu, Ashish Kumar Jain, "A review on new data mining techniques for x-ray image classification" International Journal of Advanced Technology & Engineering Research (IJATER),Vol.1, page 55-60,2014.
[23] Anamika Ahirwar, "Study of Techniques used for Medical Image Segmentation and Computation of Statistical Test for Region Classification of Brain MRI" *I.J. Information Technology and Computer Science*, 2013, 05, 44-53.
[24] Dr.D. Gladis, S.Rani, "k-means clustering to identify high active neuron analysis for LSD" International Journal of Innovative Research in Science Engineering and Technology, Vol. 2, Issue 9, September 2013
[25] Hemavathy R, Dr. Shobha G," Image Mining Techniques: A Review", International Journal of Computer Science and Information Technology & Security. Vol. 3, No.1, February 2013.
[26] [26] N.S.Nithya, Dr.K.Duraiswamy, P.Gomathy, "A Survey on Clustering Techniques in Medical Diagnosis" International Journal of Computer Science Trends and Technology (IJCST) Vol. 1 Issue2, Nov-Dec 2013
[27] Chayadevi ML, Raju GT, "Data mining, Classification and Clustering with Morphological features of Microbes" International Journal of Computer Applications, Volume 52– No.4, August 2012.
[28] Ashwini Gulhane, Prashant L. Paikrao, D. S. Chaudhari, "A Review of Image Data Clustering Techniques" International Journal of Soft Computing and Engineering (IJSCE), Volume-2, Issue-1, March 2012
[29] E. Barati, M. Saraee, A. Mohammadi, N. Adibi and M. R. Ahamadzadeh, " A Survey on Utilization of Data Mining Approaches for Dermatological (Skin) Diseases Prediction", Journal of Selected Areas in Health Informatics (JSHI): March Edition, page.1-11,2011.
[30] S.R.Pande, Ms. S.S.Sambare, V.M.Thakre, "Data Clustering Using Data Mining Techniques" International Journal of Advanced Research in Computer and Communication Engineering, Vol. 1, Issue 8, October 2012.
[31] A.K. jain, M.N. murty, P.J. flynn, "Image Segmentation Using Clustering" 1996 IEEE Computer Society Press.
[32] Zahid Ansari, Syed Abdul Sattar, A. Vinaya Babu, M. Fazle Azeem, Mountain density-based fuzzy approach for discovering web usage clusters from web log data, *Fuzzy Sets and Systems*, Volume 279, 15 November 2014, Pages 40-63.
[33] Zahid Ansari and Amjad Khan, "Fast Global k-Means Method To Discover User Session Clusters from Web Log Data", *International Journal of Computer Engineering and Applications* (IJCEA), pp. 26-35, Vol. VIII No. III, Dec.2014.
[34] Zahid Ansari, A. Vinaya Babu, Waseem Ahmed and Mohammed Fazle Azeem. "A Comparative Study of Mining Web Usage Patterns Using Variants of k-Means Clustering Algorithm". International Journal of Computer Science and Information Technologies, pp. 1407-1413 Vol. 2, No. 4, July 2011.
[35] Zahid Ansari, Waseem Ahmed, M.F. Azeem and A.Vinaya Babu. "Discovery of Web Usage Profiles Using Various Clustering Techniques". International Journal of Computer Information Systems, pp. 18-27 Vol. 1, No. 3, July 2011. (ISSN 2229-5208.
[36] Anamika Ahirwar and R.S. Jadon," FCSOFM Technique for Medical Image Segmentation", International Journal of Bio-Science and Bio-Technology, Vol.7, No.4 (2015), pp.277-282
[37] Rupinder Pal Singh, Varsha Varma, Pooja Chaudhary, " A Hybrid Technique for Medical Image Denoising using NN, Bilateral filter and LDA", International Journal of Research Development & Innovation (IJRDI) Volume 1, Issue 1, January 2015, Pg. 40-44.
[38] vinothkumar d, subashini s, "a hybrid approach for medical image retrival system using FOCS framework", International Journal of Computer Science and Mobile Computing, Vol. 4, Issue. 4, April 2015, pg.142 – 152.
[39] Stefano, Loris, Sheryl, and Emanuele, "Texture Descriptors Based on Dijkstra's Algorithm for Medical Image Analysis", Innovation in Medicine and Healthcare 2014, pg 74-82.
[40] Sahil J Prajapati , Kalpesh R Jadhav, "Brain Tumor Detection By Various Image Segmentation Techniques With Introducation To Non Negative Matrix Factorization", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, Issue 3, March 2015.