A Survey on Feature Selection Techniques in Medical Image Processing

K.Baskar
Dept of Computer Science & Engg,
S.K.P. Engineering College
Tiruvannamalai, India
hodcse@skpec.in

D.Seshathiri
Dept of Computer Science & Engg,
S.K.P. Engineering College
Tiruvannamalai, India
seshathiri@skpec.in

Abstract— This paper analyses features selection method used in medical image processing. How image is selected by using diverse sort of method similarly: screening, scanning and selecting. We discussed on feature selection procedure which is extensively used for data mining and knowledge discovery and it carryout elimination of redundant features, concomitantly retaining the fundamental bigoted information, feature selection implies less data transmission and efficient data mining. It accentuates the need for further research in the field of pattern recognition that can effectively determine the situation with captured portion of human body.

Keywords—feature selection, CBIR, medical image, screening, scanning, selecting.

I. INTRODUCTION

Medical imaging modalities are used to probe the human body. Interpretation of the resulting images requires sophisticated image processing methods that enhance visual interpretation, and image analysis methods that provide automated or semi-automated tissue detection, measurement and characterization, multiple transformations will be needed in order to extract the data of interest from an image, and a hierarchy in the processing steps will be evident, e.g. enhancement will precede restoration, which will precede analysis [1].

II. MEDICAL IMAGE PROCESSING

Image processing in medical diagnosis involve stages such as image capture, image enhancement, image segmentation and feature extraction [2, 3] Figure 1 shows a general description of lung cancer detection system that contains four basic stages. As depicted in fig 1, medical image processing contains different stages. The first stage starts with taking a collection of image (normal and abnormal) from the available patient. The second stage applies several techniques of image enhancement, to get best level of quality and clearness. The third stage implies image segmentation algorithms which play an effective rule in image processing stages, and the fourth stage obtain the general features from enhanced segmented image which gives indicators of normality or abnormality of images.

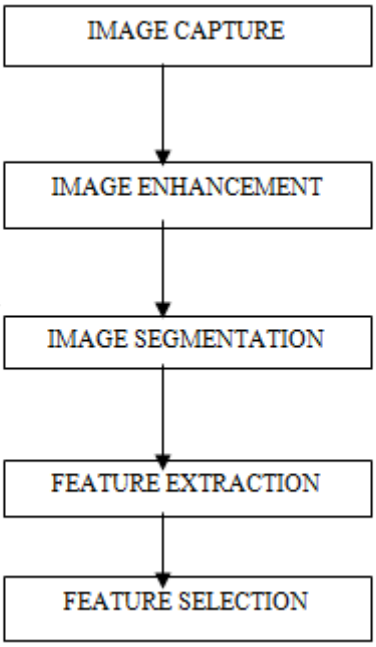

Fig 1. Process Flow in Medical Image processing

III. FEATURE SELECTION IN MEDICAL IMAGE PROCESSING

Feature selection is a dimensionality reduction technique widely used for data mining and knowledge discovery and it allows exclusion of redundant features, concomitantly retaining the underlying hidden information, feature selection entails less data transmission and efficient data mining. It also brings potential communication advantages in terms of packet collisions, data rate, and storage [4]. Feature selection is one of the key topics in machine learning and other related fields. It can eliminate the irrelevant noisy features and thus improve the quality of the data set and the performance of learning systems [5]. Expeditious growth of digital image databases motivated Content Based Image Retrieval (CBIR) which in turn requires efficient search schemes. Low level visual features including color, texture and shape, are automatically selected to represent images [6].

A. *Fundamental Feature Selection Techniques in Medical Image Processing*

The feature selection method discussed on three steps when selecting image which are: screening, ranking and selecting. In screening, it removes insignificant and problematic predictors and records or cases, such as predictors with too many missing values or predictors with too much or too little variation to be useful. Ranking, Sorts remaining predictors and assigns ranks based on importance. Selecting: It identifies the subset of features by preserving only the most significant predictors and filtering or excluding all others [7]. The Feature Selection screens, ranks, and selects are the predictors that are most significant.

B. *Survey on Feature Selection Techniques*

Haleh and Kenneth describes part of a larger attempt to apply machine learning techniques to such problems in an effort to automatically generate and progress the classification rules needed for various recognition tasks, image recognition presents a diversity of difficult classification problems involving the identification of significant scene components in the presence of noise, adopting lighting conditions, and shifting viewpoints [8]. Since each feature used as part of a classification procedure can increase the cost and running time of a recognition system, there is strong motivation within the image processing community to design and implement systems with small feature sets. At the same time there is a potentially opposing need to include a sufficient set of features to achieve high recognition rates under difficult conditions. This has led to the development of a variety of techniques within the image processing community for finding an "optimal" subset of features from a larger set of possible features. Sérgio *et al.*, described the advantage of a single-valued functions that evaluate rankings to develop a family of feature selection methods based on the genetic algorithm, it improve the accuracy of content-based image retrieval systems and it also evaluate the ranking quality allows improving retrieval performance [9]. Medical images play a central role in patient diagnosis, therapy, surgical planning, medical reference, and training. With the recent boom in the availability of filmless radiology equipment, the management of digital medical mages is receiving more and more attention. Picture Archiving and Communication Systems (PACS) have been successfully introduced in many hospitals and specialized clinics, providing quick access to screening exams and integrating the actors involved in the enterprise's workflow. The radiological databases originally built for storing digital images have evolved from simple storage servers of past exams, kept for legal reasons, to active and easily accessible repositories for research and decision support. Jaba and Shanthi reviewed previously on continuous feature discretization and identified defining characteristics of the methods. Then suggest a new supervised approach which merges discretization and feature selection to select the most relevant features which can be used for classification purpose. The classification method to be used is Associative Classifiers [10]. Medical images are a primary part of medical diagnosis and treatment. These images are unlike from typical photographic images primarily as they disclose internal anatomy as contrasting to an image of surfaces. Sasi and Kumaraswamy, said with various techniques proposed in literature for feature extraction, classification and retrieval, Content-based image retrieval (CBIR) is a widely researched area. Also discussed that Information Gain is used to achieve the structure of a feature sets to find a subset of the original feature vector for efficient computation and features are optimized using Particle Swarm Optimization (PSO) [11]. Yong Fan, *et al.*, presented a framework for brain classification based on multi-parametric medical images, and described the method advantage of multi-parametric imaging to provide a set of discriminative features for classifier construction by using a regional feature extraction method which takes into account joint correlations among different image parameters [12]. Ling-Chen *et al.*, discussed a feature selection algorithm rooted on ant colony optimization (ACO), and said Image feature selection (FS) is a significant task which can affect the presentation of image classification and recognition [13]. Ant colony optimization (ACO) is an evolution simulation algorithm proposed by Dorigo *et al.*, It has been successfully used for system fault detecting, job-shop scheduling, network load balancing, graph coloring, robotics and other combinational optimization problems. Pushpalata and Jyoti, described feature selection technique and an ensemble model proposed to improve classification accuracy. Feature selection technique is used for selecting subset of relevant features from the data set to build robust learning models and discussed further more that Classification accuracy is improved by removing most irrelevant and redundant features from the dataset and stated that Ensemble model is proposed for improving classification accuracy by combining the prediction of multiple classifiers, Three decision tree data mining classifiers were considered for classification which are CART, CHAID and QUEST [7]. Jin Yu *et al.*, presented an approach that involves the analysis of Co focal Scanning Laser Tomography (CSLT) images using moment techniques to obtain abstract image defining features, and then the use of these features to train classifiers for automatically differentiating CSLT images of healthy and diseased optic nerves, and exploration in feature subset selection methods for reducing the comparatively large input space produced by the moment methods [14].

Vasantha *et al.*, discussed that Breast cancer is the most common type of cancer found in women, and they proposes a image classifier to classify the mammogram images, mammogram image is classified into normal image, benign image and malignant image. A hybrid approach of feature selection was proposed in reduction of about 75% of the features [15]. Saravana *et al.*, discussed about feature selection and an efficient method for feature extraction was proposed for image retrieval process and described Content-Based Image Retrieval as a technique that utilizes the visual content of an image to search for similar images in large scale image databases. Feature selection and feature extraction method were the significant tasks that were considered in image retrieval process [16].

Huanzhang *et al.* discussed about Feature subset selection as a significant subject when training classifiers in Machine Learning (ML) problems and illustrated the information that the complexity of the classifier parameters adjustment during training swells exponentially with the number of features. So they introduced a novel embedded feature selection method, called ESFS, which was simulated from the wrapper method SFS as it relies on the simple standard to add incrementally most relevant features [17]. Georgia *et al.,* discussed the study of investigated information theoretic approach to feature selection for computer-aided diagnosis, the approach was based on the mutual information (MI) concept. MI measures the general dependence of random variables without making any assumptions about the nature of their underlying relationships. They described MI that it can potentially offer some advantages over feature selection techniques that focus only on the linear relationships of variables [18]. Mohamed *et al.,* discussed an approach which was proposed to develop a computer-aided diagnosis (CAD) system that can be very helpful for radiologist in diagnosing microcalcifications' patterns in digitized mammograms earlier and faster than typical screening programs and showed the efficiency of feature selection on the CAD system, and implemented the proposed method in four stages which are [19]:

a) The region of interest (ROI) selection of 32x32 pixels size which identifies clusters of microcalcifications,

b) The feature extraction stage based on the wavelet decomposition of locally processed image (region of interest) to compute the significant features of each cluster,

c) The feature selection stage, which select the most significant features to be used in next stage, and

d) The classification stage, which classify between normal and microcalcifications' patterns and then classify between benign and malignant microcalcifications.

Guo-Zheng *et al.,* discussed the feature selection methods with support vector machines which contains obtained satisfactory results, and propose a prediction risk based on feature selection method with multiple classification support vector machines. The performance of the projected method is compared with the earlier methods of optimal brain damage rooted feature selection methods with binary support vector machines [4]. Shuqin *et al.,* said feature selection techniques has been widely used in various fields and discussed a new refined feature selection module which utilizes two-step selection method in computer-aided diagnosis (CAD) system for liver disease, the method used was filter and wrapper method, Support Vector Machine (SVM) and Genetic Algorithm (GA) And stated that the advantage was to show the ability of accommodating multi feature selection search strategies and combining filter and wrapper method, especially in identifying optimal and minimal feature subsets for building the classifier [20]. Yong and Ding-gang described feature extraction and selection are of great importance in neuro image classification for identifying informative features and reducing feature dimensionality, which are generally implemented as two separate steps and presented an integrated feature extraction and selection algorithm with two iterative steps: constrained subspace learning based feature extraction and support vector machine (SVM) based feature selection [21]. Haleh and Kenneht discussed an approach being explored to develop the usefulness of machine learning techniques for generating classification rules for complex, real world data. An approach has been implemented and tested on difficult texture classification problems.

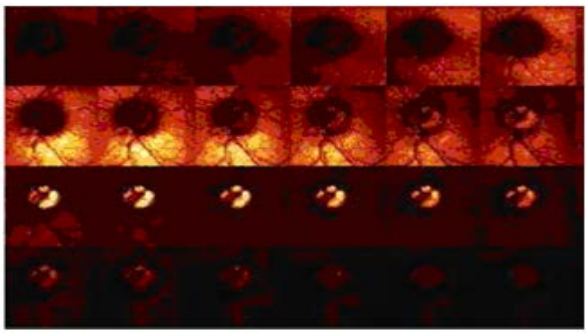

Fig2. CSLT Image of optic disc.

. The approach involves the use of genetic algorithms as a "front end" to traditional rule induction systems in order to identify and select the best subset of features to be used by the rule induction system [8]. Feature Selection (FS) algorithms aim at choosing a reduced number of features that preserves the most relevant information of the dataset. FS is usually applied as a preprocessing step in data mining tasks by removing irrelevant or redundant features (dealing with the dimensionality issue), therefore leading to more efficient (reducing the computational cost and the amount of memory required) and accurate classification, clustering and similarity searching processes. Since each feature used as part of a classification procedure can increase the cost and running time of a recognition system, there is strong motivation within the image processing community to design and implement systems with small feature sets. At the same time there is a potentially opposing need to include a sufficient set of features to achieve high recognition rates under difficult conditions. This has led to the development of a variety of techniques within the image processing community for finding an "optimal" subset of features from a larger set of possible features. Images have a large number of features. It is significant to identify and extract interesting features for a particular task in order to reduce the complexion of processing. These are attributes or portion of the image being analyzed that is most likely to give interesting rules for that problem. Not all the attributes of an image are useful for knowledge extraction. An image can be adequately represented using the attributes of its features.

The extraction of the features from an image can be done using a variety of image processing techniques. We localize the extraction process to very small regions in order to ensure that we capture all areas. Feature selection helps to reduce the feature space which improves the prediction accuracy and minimizes the computation time. This is achieved by removing irrelevant, redundant and noisy features .i.e., it selects the subset of features that can achieve the best performance in terms of accuracy and computation time. It performs the Dimensionality reduction. Features are generally selected by search procedures.

A number of search procedures have been proposed. Popularly used feature selection algorithms are Sequential Forward Selection (SFS), Sequential Backward selection (SBS), Genetic Algorithm (GA) and Particle Swarm Optimization. In this work a combined approach of Greedy stepwise method and Genetic Algorithm is proposed to select the optimal features. The selected optimal features are considered for classification.

IV. CONCLUSION

From this survey, it is discovered that selection algorithm determines the authenticity of a medical image process decisions. The selection algorithms are primarily used for the screening, ranking, and selection of the images, which are the predictors that are most significant in removing insignificant and problematic predictors and records or cases, such as predictors with too many missing values or predictors with too much or too little variation to be useful. In medical image processing, a robust and sophisticated method will be necessary such that two or three of the existing selection methods can be hybridized for better performance in real time.

V. REFRENCES

[1] Dougherty G, (2010). Image analysis in medical imaging: recent advances in selected examples. Biomedical Imaging and Intervention Journal(BIIJ),pp 1-8.

[2] Loganathan R and Kumaraswamy, (2013). Performance Evaluation of Image Compression for Medical Image. Indian Journal of Computer Science and Engineering (IJCSE), 4(2), 138-143.

[3] Hana, B., Yunjo, L. Cheryl L. G and Morris M., (2013). Age-related dedifferentiation and Compensatory Changes in the functional network underlying face processing. Neurobiology of Aging, 34(12), 2759-2767.

[4] Thiemjarus .S, B. P. L. Lo, Laerhoven K.V and G. Z. Yang, (2005). Feature Selection for Wireless Sensor Networks. In Proc of 1st International Workshop on wearable and Implatable Body Sensors Networks.

[5] Guo-Zheng Li, Jie Yang, Guo-Ping Liu and Li Xue, (2004). Feature Selection for Multi-Class Problems Using Support Vector Machines. Image Processing and Pattern Recognition, 109-111.

[6] Gnanaseker, P., Nagappan, A, Sharavanan, S, Saravanan O, Vinodkumar D., Elayabharathi T and Karthik G, (2011). Investigation on feature extraction and classification of Medical images. World Academy of Science, Engineering and Technology, 60, 327-332.

[7] Pushpalata P. and Jyoti B. G., (2012). Improving Classification Accuracy by Using Feature Selection and Ensemble Model. International Journal of Soft Computing and Engineering (IJSCE), 2(2), 380 – 386.

[8] Haleh V. and Kenneth D. J, (1992). Genetic Algorithms as a Tool for Feature Selection in Machine Learning. Artificial Intelligence, *102-109.*

[9] Sérgio Francisco da Silva , Marcela Xavier Ribeiro , João do E.S. Batista Neto, Caetano Traina-Jr. ,Agma J.M. Traina, (2011). Improving the ranking quality of medical image retrieval using a genetic feature selection method. Medical Imaging, 1012-1017.

[10] Jaba S. L. and Shanthi V., (2009), International Journal of Computer Theory and Engineering, 1(2), 154-158.

[11] Sasi K. M and Y.S. Kumaraswamy (2012). Medical Image Retrieval System Using PSO for Feature Selection. International Conference on Computational Techniques and Mobile Computing (ICCTMC'2012). 182-186.

[12] Yong F., Susan M., ResnicResnick, Christos Davatzikos, (2008). Section on Biomedical Image Analysis, Department of Radiology, University of Pennsylvania, Philadelphia, PA Lab of Personality and Cognition, Intramural Research Program, National Institute on Aging, Baltimore, MD

[13] Ling C., Bolun C., and Yixin C., (2011). Image Feature Selection Based on Ant Colony Optimization.

[14] Jin Y., Syed S. R. A., and Paul H. A., (2005). A hybrid feature selection strategy for image defining features: towards interpretation of optic nerve images,

[15] Vasantha M., Dr.V.S. Bharathi, Dhamodharan R, (2010). Medical Image Feature, Extraction, Selection and Classification, International Journal of Engineering Science and Technology. 2(6), 2071-2076.

[16] Saravana K., Sumathi A., and Latha K, (2012). Feature Selection and Extraction for Content-Based Image Retrieval, *International* Journal of Mathematics Trends and Technology. 3(2), 70-73.

[17] Huanzhang F, Zhongzhe X, Emmanuel D, Weibei D, and Liming C, (2009). Image Categorization Using ESFS: A New Embedded Feature Selection Method Based on SFS. 288–299.

[18] Georgia D. T., Erik D. Frederick, Mia K. Markey, and Carey E. F., (2001). Department of Biomedical Engineering, Duke University, Durham.

[19] Mohamed A. A, Wael A. M, Abo-Bakr M. Y, Yasser M. K, and Ahmed S. M, (2009). Feature Selection in Computer Aided Diagnosis System for Microcalcification Detection in Digital Mammograms. *Proc of 26th National Radio Science Conference (NRC2009).*1-9.

[20] Shuqin W, Yan S, Qionghua W, Jun Y, Lixu G, Lijun Q and Jianrong X, (2009). Improvement of feature Selection in multi-phase CT images of hepatic lesions. *Proc of the International Symposium on Bioelectronics & Bioinformatics*, 72-75.

[21] Yong F. and Dinggang S., (2009). Integrated Feature Extraction and Selection for Neuroimage classification. *Proc of SPIE of Medical Imaging*, 7259, 1-8.

[22] Adegoke.bo,Ola.b.o and omatayo Me(2014),Review of Feature Selection Methods in Medical Image Processing, IOSR journal of Engineering,vol .04, issue 1(2014) ,pp01-05