ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

A Survey on Medical Image Classification Techniques

Manasi Kalekar

M.E 2nd Year, Dept. of Computer Engineering, Shah & Anchor Kutchhi Engineering College, Mumbai, India

ABSTRACT: With the rapid popularization of medical image, medical images have been widely applied in clinical diagnosis. They can also be useful in the education domain for healthcare students by explaining with these images will help them in their studies. It is important to classify these data efficiently and accurately. Medical Image Classification can play an important role in diagnostic and teaching purposes in Medicine. The design and selection of classification method needs a careful attention to the following issues: definitions of Class/ Group, Pattern/ Features representation (data format), Feature Extraction (parameterization), Feature Selection, selection of training and test samples (data distribution), time complexity and performance evolution (criteria). For these purposes different Classification techniques are used. In Data Mining Classification is a data mining function that allocated similar data to categories or classes. In this survey report different Medical Image Classification Techniques are explained and the Medical Image Classification using Data Mining technique is studied.

KEYWORDS: Image Classification, Texture Classification, Data Mining, Neural Network Classification, Medical Image Mining, One Class Classifier.

I. INTRODUCTION

Medical imaging is one of the most important tools in modern medicine; different types of imaging technologies such as X-ray imaging, ultrasonography, biopsy imaging, computed tomography, and optical coherence tomography have been widely used in clinical diagnosis for various kinds of diseases.Effective medical images can play an important role in aiding in diagnosis and treatment; they can also be useful in the education domain for healthcare students by explaining with these images will help them in their studies [1][2].

However, in clinical applications, it is usually time-consuming to examine an image manually. Moreover, as there is always a subjective element related to the pathological examination of an image by human physician, an automated technique will provide valuable assistance for physicians. A large focus with respect to medical image analysis has been on automated image classification [3].Although the classifiers which can provide multi-class classification such as support vector machines (SVM) and neural networks are usually selected for medical image classification [4], one-class classifiers (OCC) [5] that can work on the samples seen are, so far, more appropriate for medical image classification task.

Data Mining is the process of retrieving information / knowledge from a given data set so as to help us make patterns and predict future values. Data mining finds its applications in many fields like Business Intelligence, Decision Support Systems, Web mining and especially in the field of Medical Image Computing. Medical Image Computing is a multidisciplinary field that covers concepts from Computer Science, Electronics, Instrumentation and the medical field. It is basically a process of taking a snapshot of a particular internal part of the human body. This snapshot helps the doctors to analyze the disease a patient suffers from based on anomalies found in the Medical Image.

II. RELATED WORK

Medical imaging is the technique and process of creating visual representations of the interior of a body for clinical analysis and medical intervention, as well as visual representation of the function of some organs or tissues

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

(physiology). Medical imaging also establishes a database of normal anatomy and physiology to make it possible to identify abnormalities. Although imaging of removed organs and tissues can be performed for medical reasons, such procedures are usually considered part of pathology instead of medical imaging.

Image classification is the labelling of a pixel or a group of pixels based on its grey value [6]. Classification is one of the most often used methods of information extraction. In Classification, usually multiple features are used for a set of pixels i.e., many images of a particular object are needed. The image classification process that will be used is the following [7]:

- **Image Acquisition:** It is the action of retrieving an Image from the Source
- **Image Enhancement:** Image enhancement helps in qualitative improvement of the image with respect to a specific application
- **Feature Extraction:** It is the data cleaning phase in which the features relevant to the classification are extracted from the cleaned images
- **Classification:** Image classification is the labeling of a pixel or a group of pixels based on its grey value

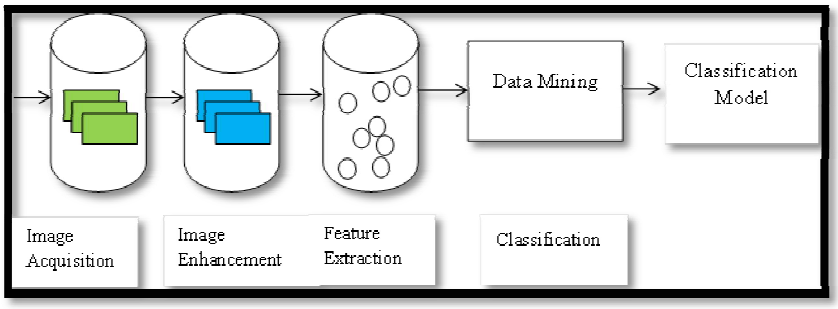

Figure 1.Image Classification Process

III. PROPOSED SYSTEM

Classification refers to as assigning a physical object or incident into one of a set of predefined categories. Medical image databases used for image classification or for teaching purposes often contain images of many different modalities, taken under varied conditions with variable accuracy of annotation. Medical Image Classification can play an important role in diagnostic and teaching purposes in medicine. For these purposes different imaging modalities are used. There are many classifications created for medical images using both grey-scale and colour medical images.There are various methods of classifying medical images and are mainly categorized in to three parts namely Texture-based classification, Neural Network classification and the Data Mining.

A. Texture Classification

In texture classification the goal is to assign an unknown sample image to one of a set of known texture classes. Texture classification is one of the main domains in the field of texture analysis. Texture analysis is important in many applications of computer image analysis for classification or segmentation of images based on local spatial variations of intensity or colour [8].

B. Neural Network Classification

Neural networks have emerged as an important tool for the classification. The recent research activities in neural classification have established that neural networks are a promising alternative to different conventional classification methods. The advantage of neural networks resides in the following theoretical aspects [9].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

C. Data Mining Technique

Classification in data mining technique is used to predict group membership for data instances. Data mining involves the use of sophisticated data analysis tools to discover the relationships in large data set. Data mining never means a collection and managing data, it also includes analysis and prediction of data. Classification techniques are capable of processing a wider variety of data than regression and also growing in popularity.

The overall proposed model of classifying the medical images is mentioned in Figure 2. The first step of classification is based on Image pre-processing, Feature extraction and selection. Once the feature selection and extraction process is carried out the algorithm is then implemented on the resultantreduced data set.

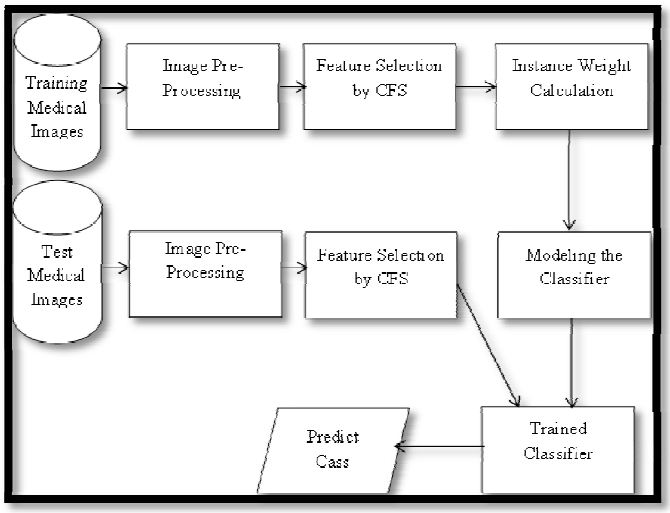

Figure2. Architecture of Medical Image Classification

The steps of classifying the Medical Images are mentioned as:

I. Image Pre-processing and Feature Extraction, Selection:

For Image Processing author has used fundus images where there is high variation in the pigment colors of the eye and illumination of the image. As a result the image is converted to gray scale version and an adaptive histogram equalization process is carried to adjust the variation in contrast.

The features from the images are extracted using statistical features. For building the classification model author has used images of size 576x720 pixels each. These images were divided into 128 smaller sub images each of size 36x90 pixels. After this the four statistical features such as Mean, Variance, Skewness and Kurtosis are extracted. Overall the feature set extracted from each image contained a feature vector of 512 attributes.

Thirdly, the sizes of these feature vectors were further reduced using a feature selection technique called the CfssubsetEval. This step was carried out to improve the quality of data to be used for classification. This method determines the predictive power of a given feature and the amount of redundancy that exists between the various features. It then uses a correlation-based method to reduce the number of features by selecting those having the most correlation with the given class labels and least amount of correlation with each other.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

II. Medical Image Classification Using Improved KNN

The weight of each training instance was calculated based on the similarity of the training instance with each of the test instances. Using a neighbourhood size, which was usually 5%- 10% of the training set, the number of nearest neighbours to each training instances was chosen. The procedure of calculating the instance weights is mentioned below. It takes as input the training set of medical images and gives as output the weight of each image in the training set.

1. The Neighbourhood size of the training set, Size is calculated, to find the weight of each training instance.
2. The instance weights for each of the training sets are calculated as follows.

$$I\_Weight(a) = \frac{1}{Ssize} f(\text{class (a)}, \text{classSi(a)})$$

Where class (a) is the class label of the medical image, Si(a) is the ith medical image in the neighbourhood of a class(Si(a)) is the class label of Si(a).The function used in the above weighting formula is calculated as follows

$$f(x,y) = \begin{cases} 1, & x = y \\ 0, & x \neq y \end{cases}$$

Improved KNN Algorithm

Every new medical image is classified as follows:

1) Firstly, the distance between the test medical image and the training medical image is found by using the Euclidean distance measure as mentioned below,

$$D(a,b) = \sum_{i=1}^{A} \sqrt{(ai - bi)}$$

Where A is the total number of features ai and bi are two separate training instances

2) Using the distance measure formula mentioned above, the k-nearest neighbors for the medical test image are found.

Now the weights that were calculated before are used to assign weights to the nearest neighbors of the given test instance.

$$W(i) = I\_Weight(i) * \frac{1}{D+0.5}$$

3) Using the above equation the weights are then summed up for every nearest neighbor that belongs to the same class type.
4) The given test medical image is assigned the class label having the highest weight.

The main reason for using the 0.5 value in above equation was to avoid cases where the distance between the test case and the training instance is zero. Without the use of the 0.5 value, W(i) is such cases could be infinite leading to erroneous results.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

E. One-Class KernelClassifiers

In machine learning, one-class classification, also known as unary classification, tries to identify objects of a specific class amongst all objects, by learning from a training set containing only the objects of that class. This is different from and more difficult than the traditional classification problem, which tries to distinguish between two or more classes with the training set containing objects from all the classes.

F. One-Class Kernel Subspace Ensemble

Classification scheme based on One-Class KPCA model ensemble which is used for classification of Medical Images.

- The ensemble is built with the feature subset strategy.
- Each One-class classifier is trained with one type of feature extracted from the training images.
- Kernel Principal Component Analysis(KPCA) model was chosen as the classifier of the ensemble
- If m-class classification tasks and n different kinds of image features, the ensemble will consists of mxn KPCA models
- For an unlabeled image its n-types of features will first be mapped into the kernel space by the corresponding n-trained KPCA models from each class.

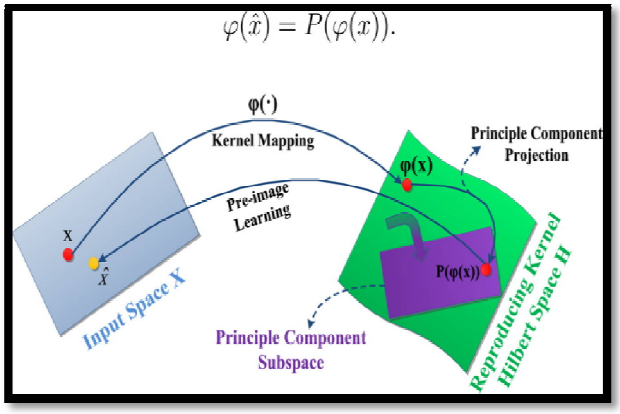

Figure3. Illustration of KPCA pre-image learning

The mapped features will then be reconstructed from the high dimensional kernel space into the original space by pre-image learning, the distances between the original features and reconstructed features will be measured.

The distances given by the KPCA models will be combined to output confidence score describing the probability of the sample belonging to a class. For a m-class classification task, the m confidence scores will be obtained, one for each class.

The image will be classified into the class with the maximum confidence Score.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

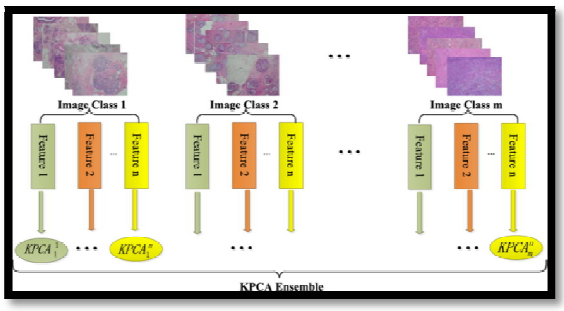

Figure 4. Construction of one-class KPCA ensemble from different feature sets

G. Cascading One-Class Kernel Subspace Ensembles

Classification system built as a serial fusion of two different classifier ensembles with reject options to enhance the classification reliability. The first ensemble consists of a set of Kernel Principal Component Analysis one-class classifiers trained for each image class with different image features. The second ensemble consists of a Random Subspace Support Vector Machine(SVM) ensemble. During Classifier generation, randomly sampled subsets of features, following the Random Subspace procedure, were used. For both of the ensembles the reject option was implemented using a confidence threshold.

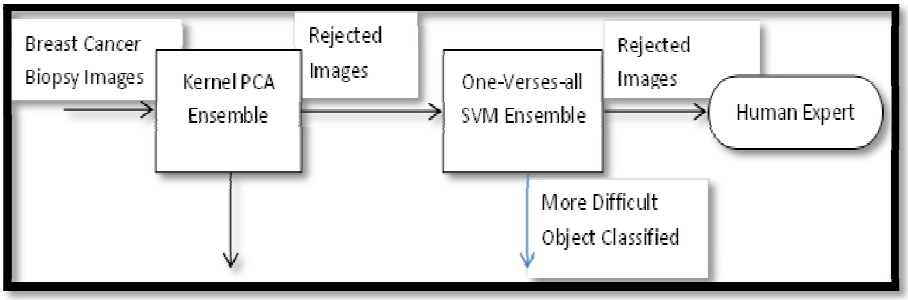

Figure 5. Operation of the hybrid classification scheme comprised of a cascade of two classifier ensemble

Comparative Analysis

Medical Image Classification Method	Related work	Pros	Cons
Medical Image Classification using improved KNN	Weighted average method for distance between test and training instance	Simple to implement and use Easy to explain prediction	Need a lot of space to store all examples Takes more time to classify new example
Medical Image Classification using one-class kernel subspace ensemble	Ensemble of one-class classifier built with feature subset strategy	High Classification accuracy	The no of principal component and width of Gaussian kernel is fixed
Medical Image Classification using Cascading one-class kernel subspace ensemble	Cascade of two classifier ensembles	High classification reliability	Ensemble size and rejection threshold is decided empirically

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

IV. CONCLUSION

Classification is a significant and major machine learning area and it has been renewed due to promising application such as data mining.
In this paper three types of Medical Image Classification using different data mining techniques are studied

- Modified KNN algorithm uses an instance-weighing scheme based on the distance measure. This model gives better classification accuracy compared to other existing Classifiers.
- One-class KPCA model ensemble is used for classification of medical images. This model obtains high classification accuracy.
- Cascading of One-class KPCA model ensemble uses a reject option in order to minimize the cost of misclassification so as to ensure high classification reliability

REFERENCES

1. LE Boucheron, Object- and spatial-level quantitative analysis of multispectral histopathology images for detection and characterization of cancer. Thesis, University of California Santa Barbara, 2008
2. C Loukas, A survey on histological image analysis-based assessment of three major biological factors influencing radiotherapy: proliferation, hypoxia and vasculature. Comput. Methods Programs Biomed. **74**(3), 183–199 (2004)
3. N Orlov, L Shamir, T Macura, J Johnston, DM Eckley, IG Goldberg, WND-CHARM: multi-purpose image classification using compound image transforms. Pattern Recognit. Lett. **29**(11), 1684–1693 (2008)
4. L Kuncheva, J Rodriguez, C Plumpton, D Linden, S Johnston, Random subspace ensembles for FMRI classification. IEEE Trans. Med. Imaging.**29**(2), 531–542 (2010)
5. D Tax, One-class classification. Thesis, Delft University of Technology, 2001
6. Digital Image Processing - Chellappa, 2nd Edition, IEEE Computer Society Press, 1992.
7. Maria-LuizaAntonie, Osmar R. Za¨ıane, AlexandruComan "Application of Data Mining Techniques for Medical Image Classification"
8. A. Kassner, R.E. Thornhill, "*Texture Analysis: A Review of Neurologic MR Imaging Applications*", A Journal of NeuroRadiology 31:809, May 2010.
9. K.V.Ramana, Raghu.B.Korrapati, "*Neural Network Based Classification and Diagnosis of Brain Hemorrhages"*, International Journal of Artificial Intelligence and Expert Systems (IJAE), Volume (1): Issue (2) 2009

BIOGRAPHY

Ms.Manasi Kalekar is a M.E 2nd year student in Shah & Anchor Kutchhi Engineering College Chembur, University of Mumbai. Her research interests are Image Processing, Computer Networks etc.