ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

A Novel Segmentation and Features Extraction with Features Selection Methods for Classification of Medical Images.

Suhashini M*, and Veeramuthu A.

Department of IT, Sathyabama University, Chennai – 600 119, Tamil Nadu, India.

ABSTRACT

The segmentation is an important research task has been done on the images, which split the images into more segments and this segmented process has been used to identify the affected areas in bronchi which gives an efficient and understandable results. The drawback of previous method, it that is difficult to distinguish critical tissues from normal tissues and it is used to normalize the visual pattern by using computed tomography images. In the previous work, the exactness of finding the affected area is bit low and the proposed method will outperform significantly better which leads to improve the performance of identifying the affected areas in bronchi. In this paper morphological technique is proposed to avoid the noise in image and canonical computes the medical image. This preprocessing techniques are used to analyses those images and normalize the intensity of the individual particles images using the new super pixel segmentation. After segmentation process, feature will be extract. In feature extraction the images are extracted in various colors and select a clear image using feature selection process. PSO (Particle Swarm Optimization) is proposed in this paper to identify affected area in lungs accurately. The Support Vectors Machines (SVM) classifier and Fuzzy c-means are used to identify the exactness of the diseases in bronchi. The affected areas have been characterized using different intensity of color and examine the size of the affected areas. Depend on the intensity the ratio of the affected areas can be initiated. The performance of the current work is considerably better.

Keywords: Morphological, Canonical, Super pixel segmentation, SVM, Fuzzy C-Means, PSO.

Corresponding author

ISSN: 0975-8585

INTRODUCTION

The cancer caused in lung is the main problem in the human body. Compare to the different types of cancer, lung cancer shows the highest death rate. It is the serious cancer, in which survival rate of the human is small after the analysis. Lung cancer survival is straightly related to growth at its time of detection. Earlier detection is the best way to provide the successful treatment. To detect the cancer in lung at starting stage, CAD (computer aided diagnosis) system will helps to achieve the task and this technique will be implemented by a technique of image processing. In order to examine the parts of human body the modalities of medical imaging are used. The image characteristic can be extracted by applying the resultant images with the support of difficult image processing method. For the detection of semi automated or automated lung cancer, the method of image analysis is necessary. The fields of mathematical morphology which are all based on an easy mathematical concept from a set theory, contributes to a broad range of operators to the image processing. The operators are mostly helpful for the binary image analysis and generally used in noise removal, image segmentation, edge detection and image enhancement. The only way in finding the difference in morphological operations is by the way how they are carrying the comparison. Once the effect of these methods is measured then it should be introduced into the process of multiple transformation method which is extracting the characteristic of an exact image. For the pixel classification process, the segmentation of super pixel image is very important in which the individual pixels are assigned to classes for segmenting the pixels of image in to subsets. The boundaries and their segmented organs are very serious in the process of quantification for medical surgeons and physicians, in every division of medicine, which deal with imaging [1].

In several visions of computer, one of the application that is broadly used is the application of the feature extraction in which the desired images is retrieved from a huge collection on the basis of the features that is routinely extracted from the images. The extraction of the shape and edge information starts the process of feature extraction from unique medical x- ray images. A severe awareness in the literature of the image information recovery is received by the systems known as Content- Based Image Retrieval (CBIR). The algorithm in this system is classified in to two types called Selection and Extraction. The rich image content is transformed in to several content features by the extraction task. The process of producing the features for classification and selection [2] tasks is known as feature extraction. The particular part of an image can be extracted by the stage of image feature extraction. Once the segmentation occurred on lung area, the features will be extracted and the diagnosis rule will be designed so that the cancerous nodules inside the lungs will be detected. The number of features that are offered to the classification task [3] will be reduced by the feature selection. The presentation of image recognition and image classification can be affected by the Image feature selection. The technique of image reduction which is been broadly used for knowledge detection and image mining and allowing the elimination of redundant and irrelevant features while preserving the relevant features is known as the feature selection. The effective image mining and a smaller amount of image transmission are implied by the feature selection and it selects the image recognition system quality and removes the redundant and irrelevant features.

Hence for this purpose Particle Swarm Optimization method was effectively useful for this activity. The image recognition task quality can be enhanced by the means of accuracy classification for the purpose of improving the recovery performance. SVM (Support Vectors Machines) has exposed their capacity in pattern classification and recognition. The most popular method known to be the Fuzzy c-means (FCM) algorithm [4] is used in image segmentation due to its robust feature for uncertainty and will preserve more information than methods of hard segmentation. Though the conventional FCM algorithm performs well on noise-free images, serious limitation will occur. It does not include any information regarding spatial context, making it to be sensitive to imaging artifacts and noise.

RELATED WORKS

By using mutual information, the paper [1] introduced a greedy feature selection method for combining the feature-feature mutual information and feature-class information to find an optimal subset of features. These features which are help to minimize redundancy and to maximize relevance among features. The effectiveness of the selected features is evaluated using multiple classifiers on multiple datasets.

The author proposed an image classification for leveraging the non-negative sparse coding, correlation constrained low rank and sparse matrix decomposition techniques. Extraction of features

information for representing the images by using max pooling and spatial pyramid matching method. Collected the columns of both matrices as the bases through the locality -constrained linear coding. For final classification, a linear SVM classifier is trained [2].

In Image processing, feature subset selection plays a vital role for data reduction with the help of removing unrelated and redundant dimensions. From a given set of data features are extracted to find the subsets of features that are relevant to data mining. Based on these criteria, subsets of features are used spatial gray level difference method, feature extraction algorithm and Correlation based Feature Selection. Projected Classification algorithm [3] is used for classifying brain image data in an efficient manner.

The procedure for segmenting and classifying scanned legume leaves based on the analysis of the veins. Most important legume species are studied namely soybean, red and white beans. By using Un-constrained hit -or-miss transform and adaptive thresholding, segmentation process is carried out the performance. Segmented venation is computed by several morphological features and it will be classified by four alternative classifier namely support vector machine, penalized discriminant analysis and random forest [4].

The author proposed a new feature selection method which is depend on fisher criterion and genetic optimization for recognition problem [5]. It is mainly used for evaluating feature subsets. By using various classifiers such as support vector machine, Bagging, Naive Bayes, K-nearest neighbor and Ada-Boost for classifying the region of interests in lungs. The FIG method gives the better recognition performance and accurate rate.

The mining is highly efficient to extract knowledge from the data. The main aim of this paper is to describe the key papers and gives some information to the medical practitioners. Data mining approaches such as classification, regression, clustering, association and hybrid are used in each task. According to CRISP-DM is adapted to manage all activities [6]. A hybrid intelligent system that includes Fuzzy Min-Max neural network, the classification and Regression tree, and the random forest model was proposed. It is mainly support the decision tool for medical data classification [7]. Three Benchmark medical data sets such as Breast cancer, Diabetes and liver disorders is used for evaluating the effectiveness of the hybrid intelligent systems.

Mainly focusing on image retrieval techniques which include the features like texture and shape. The use of CBIR is insisting the lesser computational time. There is some issue resolved by CBIR techniques, basically some traditional drawback like heavy workload. The CBIR techniques can be used for analyzed and extracted automatically by the system. Recall and precision values are used for calculating the overall performance. Depends upon the analyzed values the comparison graph was drawn [8]. The author proposed that positron emission tomography imaging gives the more information about many diseases which is used for evaluating inflammation, infection and cancer with the help of emitted photon from a radiotracer localized to abnormal cells [9].

The author is mainly focusing on analysis the structure of the brain images by using Computer Aided Diagnosis. The input images are decomposed and extracted by using Multi Level Discrete Wavelet Transform method. The PNN-RBF training and classification is applied to classify the brain images which is help to detect the diseases in early stage. The fisher discriminant ratio method is used to reduce the computational cost. Segmenting the images based on the region of interest by using morphological filter techniques [10].

The author proposed an automatic system for diagnosis of the brain tumor [11]. By using preprocessing step, remove the noise from MR images.SVM classification is used for getting more accuracy result in tumor. Separating the tumor region from normal brain cells with the help of FCM clustering. Finally it gives an accurate tumor region in brain.

The morphology technique is treating the image such as set of images and detects it by structural elements. The algorithmic essence for the mathematical morphology is producing the expression by interacting the objects and element structure. The extraction of features could be viewed within some steps like preprocessing that removing the variance distraction from the dataset, so the downstream regression or classifiers estimators are performing very well. The area of starting the regression and completing features

extraction are really dark, one proper technique of feature extraction is being used for mapping the data into label of classes for classification process [12].

The distribution of features across the overlap region affects the accuracy, when the matching images are applied. The author proposed the spatial statistics [13] and it will be measured by Ripley's k-function. Experiments on stitching images into mosaics confirm that better coverage values yield better quality outputs. To evaluate the spatial coverage in an image and to determine whether points are aggregated at multiple scales. Two measures have been established, the ability of the operator to produce the same descriptor for the same feature in different images, and coverage, the distribution of detected features around the image.

Bag of features model is the most successful algorithm that is used in image classification tasks. BOF has many advantages such as simplicity, generality and scalability and it still affected from various drawbacks. Various techniques have been proposed here is for overcome these shortcomings. BOF model [14] is used to improve some extent and integrate each individual module. Combination of texture and edge based local features a new scheme will be introduced. By using spatial weighing is performed to capture the image saliency.

The author is describing the reduction dimensionality as the preprocessing step of widespread in the analysis of high dimensional data, modeling and visualization and the easiest way for reducing the dimensionality through extracting the features. The enhanced images are being processed by the edge detectors which being followed through segmentation technique being used in the radon transform. These all features are transformed into a multi-dimensional space through the transformation of Gabor for obtaining the required vector features. The preservation and uniqueness feature of the images have been found for the maximum representation form which acquired the transmission of vector through the technique of matrix-based encryption. The learning form of information is being required over the accurate classification and recognition of image. The technique has achieved more and better enhancement within decision tree and Bayesian classifiers. But, if the images are being covered by the noise and the patterns not clear or incomplete, then the modeling of the image is too tough for the issues of feature and noise extraction, so it's needed a novel algorithm for feature learning [15].

The author has described about the concept of developing a system of CAD (computer-aided diagnosis) that is more useful for the diagnosing the appropriate radiologist medical patterns and the efficient feature selection over the system of CAD. This system has been implemented within four stages such as ROI (Region of Interest) selection for provided image, the second is extracting features on the basis of ROI, third is the stages of the feature selection are selecting the similar feature and last is the stage of classification has been used for the image recognition [16].

PROPOSED SYSTEM

System Architecture

The lung image will be given as an input in the computer-aided diagnosis to extract. Then the image will be preprocessed. Preprocessing consists of morphological process, canonical and super pixel segmentation process. These preprocessing is used to extract a feature of input medical image. In morphological process, the noise will be suppressed and the medical image information will be studied and retrieved. The morphological process is widely used in an image processing. The canonical process, compute the information of a medical image. The computed information is segmented in the super pixel segmentation. In super pixel segmentation, a medical image will be segmented pixel by pixel. Through morphological, canonical and super pixel segmentation process a feature of a medical image will be extracted. In feature extraction the various colors will be used to extract an image. After extraction the feature selection process will be done. In feature selection, the clear image will be selected, which are shown in Figure 1.

In this paper the modified Particle Swarm Optimization (PSO) technique is proposed. This PSO technique is used in feature extraction and feature selection in order to get accurate image features. Finally, used Support Vectors Machines (SVM) and Fuzzy c-means (FCM) algorithms are used for classification. SVM is used to recognize an image and FCM is used to simulate a problem in medical image. Using the proposed algorithms tumor size in the lungs was found efficiently at early stage itself.

ISSN: 0975-8585

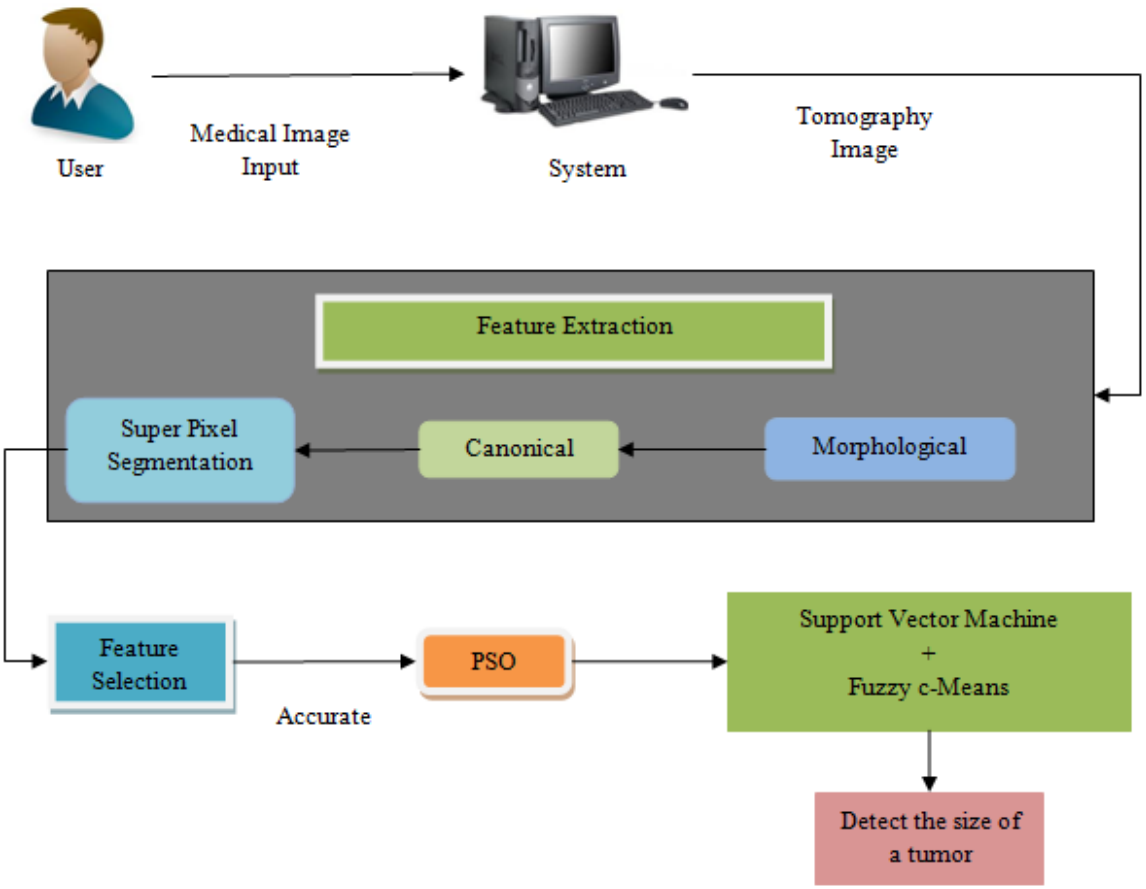

Fig 1: System architecture

Feature Extraction

The analysis of the images and objects to extract the most relevant features that are the representative of several classes of objects is done by the methodologies of feature extraction. The main aim of this feature extraction is to reduce the exact data by the means of measuring the certain features or properties. This will differentiate the pattern of one input from other pattern. The characteristic of the pattern will be provided to the classifier by the means of extracted feature. The stage of image feature extraction will extract the parts of the image that are preferred.

Morphology

For image processing the nonlinear filter method called as morphology is used, this also includes feature extraction, image segmentation, image compression, shape recognition, noise suppression, image reconstruction and restoration, texture analysis etc. This method is broadly used in the image processing area.

Canonical

These canonical functions are translated to the equivalent functionality of data source for the provider. Hence the function invocations are allowed, that are expressed in a normal form across the sources of data.

Super-pixel segmentation

An over-segmentation of an image is formed by the super pixel algorithm. In order to form the super pixel cluster and to produce the final segmentation, the segmentation algorithm processes the super pixels.

Feature Selection

The technique of image reduction which is been broadly used for knowledge detection and image mining and allowing the elimination of redundant and irrelevant features while preserving the relevant features is known as the feature selection. To improve the performance of the learning system the feature selection removes the image recognition system, quality, redundant and irrelevant features. The support vector machine helps in doing this process. Thus it is one of the pre-processing steps provided to the machine learning for image recognition is the feature extraction.

Particle Swarm Optimization (PSO)

The Particle Swarm Optimization (PSO) is a computational method that optimizes a problem and produce an appropriate solution to an image features. It is used to find an optimal threshold to segment and extract the images. The PSO algorithm could be applied in both feature extraction and feature selection method for getting an exact image features.

Support Vector Machine (SVM)

The SVM (Support Vector Machine) is implementing the complicated decision rules through non-linear function Φ for mapping the training points to high-dimensional space of feature, where every point of the label contents are separate. A hyper separate plane is finding the maximum distance a mid of the training point and itself.

Fuzzy c-Means Algorithm

The system of fuzzy logic was used for problem simulation. The fuzzy-c means is too useful when the cluster number is pre-defined; so it always tries to combine every data into one cluster. This behavior of the fuzzy-c means is making it too different, which is not able to define the exact membership of the given point cluster; instead of, it makes an exact calculation over the likelihood from the data point is belongs to a cluster.

Algorithm Description

Particle Swarm Optimization (PSO) Pseudo Code

```
Input: segmented image
Output: most relevant image features
Begin
        [p*] = PSO()
        Q = Particle_Initialization();
        For i=1 to it_max
        For each particle q in Q do
        fq = f(q);
        If fq is better than f(qBest)
        qBest = q;
        end
        end
        kBest = best q in Q;
        For each particle q in Q do
         x = x + a1*rand*(qBest – q) + a2*rand*(kBest – q);
        Q=Q+X;
        end
        end
        End
```

Fuzzy C-Means Algorithm

Step 1: Set the cluster centroids vi according to the segment of the image,

ISSN: 0975-8585

fuzzification parameter z ($1 \le z < \infty$), the values of c and $\varepsilon > 0$.

Step 2: Calculate the segment using Seg(g)= $\sum_{s=0}^{S-1}\sum_{t=0}^{T-1}\delta(f(s,t)-g)$

Step 3: Calculate the membership function using $U_{ig}^{(b)} = \frac{1}{\sum_{j=1}^{c}[\frac{d(g,vi)}{d(g,vj)}]^{2/(q-1)}}, \forall\, i,g$

Step 4: Calculate the cluster centroids using $V_i^{(b+i)} = \frac{\sum_{g=Lmin}^{Lmax}(U_{ig}^{(b)})^q seg(g)g}{\sum_{g=Lmin}^{Lmax}(U_{ig}^{(b)})^q seg(g)}$.

Step 5: Go to step 3 and repeat until convergence.

Step 6: Calculate the a priori probability using $P_{ik} = \frac{\#N_k^i}{\#N_k}$ with the obtained results of membership function and centroids.

Step 7: Recalculate the membership function and cluster centroids using

$U_{ik}^{*(b)} = \frac{P_{ik}}{\sum_{j=1}^{c}(\frac{d_{ik}}{d_{jk}})^{2/(q-1)}}$ and $V_i^{*(b+1)} = \frac{\sum_{k=1}^{n}(u_{ik}^{*(b)})^q x_k}{\sum_{k=1}^{n}(u_{ik}^{*(b)})^q}$ with the probabilities.

Step 8: If the algorithm is convergent, go to step 9; otherwise, go to step 6.

Step 9: Image segmentation after defuzzification using $C_k = arg_i\{max(u_{ik})\}$, i=1,2….,C and then a region labeling procedure is performed.

SVM Algorithm

1. Repossess the dataset values from user database
2. Train the normal data in to processed data
3. Exceed the values into svm libsvm code
4. Discover the hyper plane that enlarges the distance by using SVM formulae:

$$min\,\frac{1}{2}||w||^2 + C\sum_i \xi_i$$

subject to $y_i(w^T\emptyset(x_i)+b) \ge 1-\xi_i$

$\xi_i \ge 0, i = 1, \ldots\ldots, l$

If data cannot be normalized find the attack margin value

Where, Φ (xi) maps xi into a higher-dimensional space and C > 0 is the regularization parameter. Due to the possible high dimensionality of the vector variable w, usually we solve the following dual problem. $\min_\alpha Q = \frac{1}{2}\alpha^T Q\alpha - e^T\alpha$ Subject to $Y^T\alpha = 0$ find the average attack margin value of each dataset

5. Pass them to Radial basis function to derive the predicted value of normal and attacker dataset.

RESULTS AND DISSCUSIONS

Comparison of Classification Accuracy

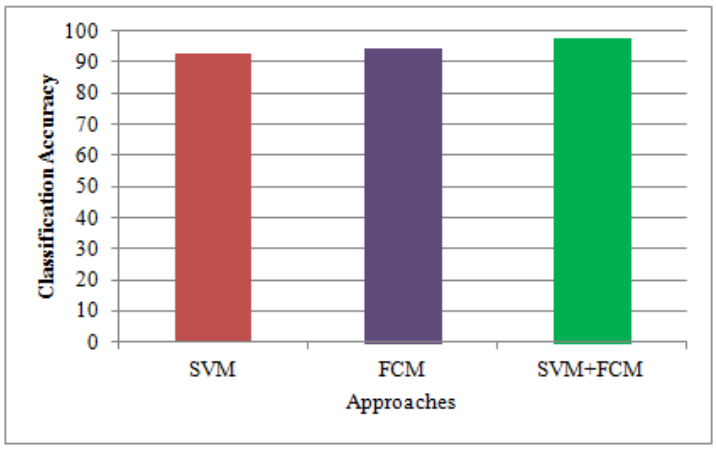

Fig. 2 Classification accuracy with different approaches

ISSN: 0975-8585

The accuracy of the proposed technique with the combination of SVM (Support Vectors Machines) and FCM (Fuzzy c-means) which is better than other techniques like SVM, FCM and etc. The proposed technique is providing more accuracy over the existing issues, which is shown in Figure 2.

S.No	Approaches	Classification accuracy (%)
1	SVM	92
2	FCM	94
3	SVM+FCM	97

Table 1: Classification accuracy

The SVM, FCM and SVM+FCM approaches and its classification accuracy information are shown in Table 1. When compare to existing techniques SVM+FCM classification accuracy was significantly better.

Comparison of Noise Level

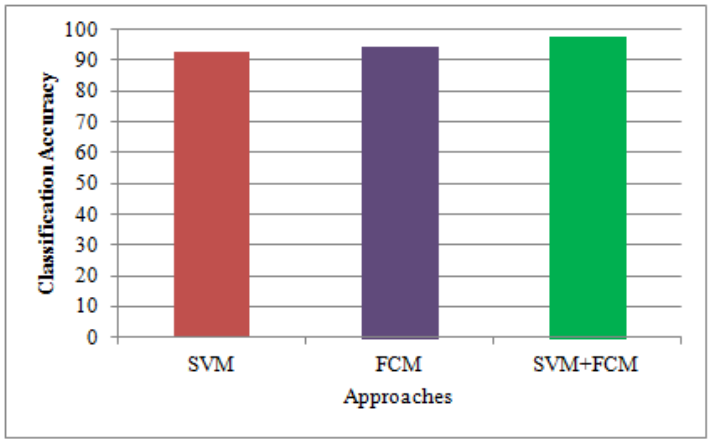

Fig. 3 Comparison of noise level

The noise level of the proposed technique with a combination of SVM (Support Vectors Machines) and FCM (Fuzzy c-means) which is better than other techniques like SVM, FCM and etc. The proposed technique is reduces the noise level over the existing issues, which is shown in Figure 3.

S. No.	Approaches	Noise level (%)
1	SVM	20
2	FCM	15
3	SVM+FCM	10

Table 2: Noise level

The SVM, FCM and SVM+FCM approaches and its noise level are tabulated in Table 2. When compare to existing techniques SVM+FCM noise level is considerably less.

CONCLUSION

The proposed technique of this Paper is Particle Swarm Optimization and the Support Vector Machine (SVM) with Fuzzy-C Means (FCM), which providing a better solution over the existing issues. The identification of the lung cancer is having a better enhancement within proposed technique; here the proposed technique is extracting the features of the lung cancer based on the proposed parameters like morphological, canonical and super pixel segmentation. The extracted features are being selected in the process of feature selection which provides more accuracy within the proposed Particle Swarm Optimization (PSO); The PSO is defining the fitness of each iteration for the features extraction based on the proposed parameters with SVM and FCM. The

Support Vector Machine (SVM) and the combined FCM (Fuzzy-C Means) algorithm is collecting the feature after extraction and combining the parameters of symptoms of the lung cancer. The parameters are being considered like color, size and etc. The overall proposed system is producing a better output for the detection of lung cancer.

REFERENCES

[1] N.Hoque, D.K.Bhattacharyya, J.K.Kalita, "MIFS-ND: A mutual information-based feature selection method", Expert Systems with Applications, Elsevier, pp.6371-6385, 2014.

[2] Chunjie Zhag, Jing Liu, Chao Liang, Zhe Xue, Junbiao Pang, Qingming Huang, "Image classification by non-negative sparse coding, correlation constrained low-rank and sparse decomposition" Computer Vision and Imaging Understanding, Elsevier, pp.14-22, 2014.

[3] Veeramuthu A,Meenakshi S, and Kameshwaran A, "A Plug-in Feature Extraction and Feature Subset Selection Algorithm for Classification of Medical Brain Image Data", IEEE International Conference on Comunication and Signal Processing, pp. 1545-1551, 2014.

[4] Monica G.Larese, Rafael Namias, Rogue M.Craviotto, Miriam R.Arango, Carina Gallo, Pablo M.Granitto, "Automatic Classification of legumes using leaf vein image features", Pattern Recognition, pp.158-168, 2014.

[5] Xiabi Liu, Ling Song, Yanfeng Zhao, Xinming Zhoa, and Chunwu zhou, "Recognizing Common CT Imaging Signs Of Lung Diseases Through a New Feature Selection Method based on Fisher Criterion and Genetic Optimization", IEEE Journal of Biomedical and Health Informatics, Vol.19, No.2, 2015.

[6] Nura Reza Babavalian, Amir-Masoud Eftekhari Moghadam, and Vahid Kashani Tabar, "Knowledge discovery in medicine: Current issue and future trend", Expert Systems with applications, Elsevier, pp.4434-4463, 2014.

[7] Manjeevan Seera, and Chee Peng Lim, "A hybrid intelligent system for medical data classification", Expert Systems with Applications, Elsevier, pp.2239-2249, 2014.

[8] A. Veeramuthu, Dr. S. Meenakshi, Dr. S. Justin Samuel, and S. Reena Esther, “Content Based Image Classification based on Texture and Shape Analysis using Wavelet Feature and Clustering Model”, International Journal of Applied Engineering Research, Research India Publications, Vol. 9, No.12, PP. 1787-1799, 2014.

[9] Brent Foster, Ulas Bagci, Awais Mansoor, Ziyue Xu, and Daniel J.Mollura, "A Review on segmentation of positron emission tomography images", Computer in Biology and Medicine, pp.76-96, 2014.

[10] A. Veeramuthu, S. Meenakshi, V. Priya Darsini, “Brain Image Classification Using Learning Machine Approach and Brain Structure Analysis”, Procedia Computer Science, Elsevier ScienceDirect, pp. 388-394, 2015.

[11] Quratul Ain, M.Arfan Jaffar, Tae-Sun Choi, "Fuzzy anisotropic diffusion based segmentation and texture based ensemble classification of brain tumor", Applied Soft Computing, pp.330-340, 2014.

[12] Ryszard S. Chora’s,“Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems”, International Journal of Biology and Biomedical Engineering, Issue 1, Vol. 1, pp. 7-16, 2007.

[13] Erkan Bostanci, Nadia Kanwal, and Adrian F. Clark,”Spatial Statistics of Image Features for Performance Comparison” , IEEE Transactions on Image Processing, Vol. 23, No. 1,pp.153-162, 2014.

[14] Lingxi Xie, Qi Tian, Meng Wang, and Bo Zhang, ”Spatial Pooling of Heterogeneous Features for Image Classification”, IEEE Transactions on Image Processing, Vol. 23, No.5, pp.1994-2008, 2014.

[15] Nikolay C., Marc M. Van Hulle, “Comparison of Two Feature Extraction Methods Based On Maximization of Mutual Information”, Laboratorium for Neuro- en physiologies, 2000.

[16] Mohamed A. A, Wael A. M, Abo-Bakr M. Y, Yasser M. K, and Ahmed S. M,”Feature Selection in Computer Aided Diagnosis System for Micro calcification Detection in Digital Mammograms”, Proc of 26th National Radio Science Conference, pp.1-9, 2009.