A Hybrid Approach for Classification of DICOM Image

J. Umamaheswari  
Research Scholar, Department of Computer Science  
Dr. G.R.D College of Science,  
Coimbatore, Tamilnadu, India

Dr. G. Radhamani  
Director, Department of Computer Science  
Dr. G.R.D College of Science,  
Coimbatore, Tamilnadu, India

Abstract— Image classification is a most important step for image analysis. As the same in medical area especially for diagnosing the disease of the patient, classification plays a great role for the doctors to treat the patient according to the severeness of the diseases. In case of DICOM images it is very tough for optimal identification and early detection of diseases. Classification is a computational procedure that separates the images into groups according to their features that extracted. DICOM is latest medical imaging technology. DICOM is used for brain scans and it is very useful and effective technique to detect the dissimilarity in brain images. In this paper a hybrid approach is proposed for DICOM image classification. The approach consists of feature extraction and classification. The classification consists of Multi Linear Discriminent Analysis (MLDA) and Support Vector Machine (SVM). Classification is done on the base of parameter extracted by Gray Level Co-occurrence Matrix (GLCM) and histogram texture feature extraction method. The feature is selected using fuzzy rough set and Genetic Algorithm (GA). The proposed approach has high approximation capability and much faster convergence.

Keywords- Classification; Linear Discriminent Analysis (LDA); Support Vector Machine (SVM); GA, Fuzzy Rough set; GLCM; Histogram Texture feature.

I. INTRODUCTION

Image Classification is becoming more important for diagnosis process. The need of efficient research on features extracted and their role to the classification results makes researchers to select features randomly as input to their systems. Genetic algorithms have been discovered successful for best possible solution amongst a huge number of possible solutions. In addition, a mixture of genetic algorithms and fuzzy can prove to be very powerful in classification problems. Previously genetic algorithms have been used either to progress neural network topology (Stathakis 2006) [4] or to pick features ( Kavzoglou 2002 [5]) but not both at the same time.

Different methods of DICOM image feature reduction method have been used to solve for efficient classification. The method used for feature extraction are statistical methods and texture based methods [3]. Some other methods are based on fuzzy theory [11] and neural networks [2, 21, 22].

For texture classification problem GLCM, Histogram, level set, Gabor filters, and wavelet transform [6, 7, 8, 9]. These are the approaches for image analysis. The Gabor filters are poor due to their lack of orthogonality that results in redundant features, while wavelet transform is capable of representing textures at the most suitable scale, by varying the spatial resolution and there is also a wide range of choices for the wavelet function.

In medical image analysis, the determination of normal and infected brain is classified by using texture. DICOM and CT image texture proved to be useful to determine the Normal brain [10] and to detect the brain disease part [11].

There is a major problem in selecting the optimal features in medical imaging. Therefore it is important for an effective classification, feature reduction and feature selection method. The previous study classification is done by a hybrid approach [15,16].

A modified method for classifying DICOM images with lower computational requirements is proposed and selection percentage and accuracy is analyzed. The tables provide the user with all relevant information for taking efficient decision. Thus a combination of LDA and SVM is used, which gives better accuracy when compare to traditional classifiers.

This paper is organized as follows: The proposed method is presented in Section 2 and feature selection and classification approach is presented in Sections 3. Section 4
contains results and discussion while conclusions and future work are presented in Section 5.

II. THE PROPOSED METHOD

The proposed method is based on the following techniques, Gray Level Co-occurrence Matrix, Histogram Features, Genetic Algorithm, Rough set, Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM) [17, 18]. It consist of three stages,

1. Feature Extraction
2. Feature Selection
3. Classification

The proposed hybrid technique for DICOM image classification is shown in figure 1.

The average value of intensity histogram features obtained for different type of medical image is given Table 2 as follows:

A. GLCM Features

The Gray-Level Co-occurrence Matrix (GLCM) is a statistical method that considers the spatial relationship of pixels, which is also known as the gray-level spatial dependence matrix. The pixel and the adjacent pixel is consider as the spatial relationship and also another spatial relationships can be specified between these two pixels.

The following GLCM features were extracted in this paper: Autocorrelation, Contrast, Correlation, Cluster Prominence, Cluster Shade, Dissimilarity Energy, Entropy, Homogeneity, Maximum probability, Sum of squares, Sum average, Sum variance, Sum entropy, Difference variance, Difference entropy, Information measure of correlation, information measure of correlation and Inverse difference normalized.

The value obtained for the above features for a typical normal and infected DICOM image is given in the following Table 3.
III. FEATURE SELECTION AND CLASSIFICATION

To improve the prediction accuracy and minimize the computation time, feature selection is used. Feature selection occurs by reducing the feature space. This is achieved by removing irrelevant, redundant and noisy features which performs the dimensionality reduction. Popularly used feature selection algorithms are Sequential forward Selection, Sequential Backward selection, Genetic Algorithm and Particle Swarm Optimization. In this paper a combined approach of fuzzy roughest method with Genetic Algorithm is proposed to select the optimal features. The selected optimal features are considered for classification.

A. Genetic Algorithm (GA) based Feature selection:

During classification, the number of features can be large, irrelevant or redundant. So the optimal solution is not occurred. To solve this problem, feature reduction is introduced to improve the process by searching for the best features subset, from the original features.

GA is an adaptive method of global-optimization searching and simulates the behavior of the evolution process in nature. It is based on Darwin’s fittest principle, which states that an initial population of individuals evolves through natural selection in such a way that the fittest individuals have a higher chance of survival.

The total features extracted are 40 from the extracted features using GA and the selected features are 8.

TABLE 4 FEATURE SELECTED BY GENETIC ALGORITHM METHOD

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Area</td>
</tr>
<tr>
<td>2</td>
<td>Centroid</td>
</tr>
<tr>
<td>3</td>
<td>Minor Axis Length</td>
</tr>
<tr>
<td>4</td>
<td>Autocorrelation</td>
</tr>
<tr>
<td>5</td>
<td>Sum Entropy</td>
</tr>
<tr>
<td>6</td>
<td>Diff. Variance</td>
</tr>
<tr>
<td>7</td>
<td>Mean</td>
</tr>
<tr>
<td>8</td>
<td>Energy</td>
</tr>
</tbody>
</table>

The above Table 4 shows the feature selected by GA method.

B. Feature selection by Rough Set

Fuzzy set involves more advanced mathematical concepts, real numbers and functions, whereas in classical set theory the notion of a set is used as a fundamental notion of whole mathematics and is used to derive any other mathematical concepts, e.g., numbers and functions [13,14].

The selected features using Rough set method are tabulated as follows

TABLE 5 FEATURE SELECTED BY ROUGH SET METHOD

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Kurtosis</td>
</tr>
<tr>
<td>2</td>
<td>Std</td>
</tr>
<tr>
<td>3</td>
<td>Sum Average</td>
</tr>
<tr>
<td>4</td>
<td>Sum Variance</td>
</tr>
</tbody>
</table>

2.2 Feature selection

In feature selection,

N = number of Features
n1 = features selected by Rough set
n2 = features selected by GA

1. N number of features is extracted by GLCM and Histogram texture features from the preprocessed Image
2. Apply roughest algorithm to select the optimal set containing n1 number of features where n1 < N
3. Apply genetic algorithm to select the best subset containing n2 number of features where n2 < N
4. Find the Union of n1 features and n2 features to form final n features
5. Use the n features where n < N for Classification.

FIGURE 2 PROPOSED APPROACH FOR FEATURE SELECTION

The above Figure 2 shows the feature selection by proposed approach. The following Table 6 gives feature selected by proposed approach.
TABLE 6 FEATURE SELECTED BY PROPOSED APPROACH

<table>
<thead>
<tr>
<th></th>
<th>Area</th>
<th>Centroid</th>
<th>Minor Axis Length</th>
<th>Autocorrelation</th>
<th>Sum Average</th>
<th>Sum Variance</th>
<th>Sum Entropy</th>
<th>Diff. Variance</th>
<th>Mean</th>
<th>Energy</th>
<th>std</th>
<th>Kurtosis</th>
</tr>
</thead>
</table>

2.3 Classification

The Selected Features are used for classification process. Classification works on the combination of MLDA and SVM [19].

A. Multi Class Linear Discriminant Analysis (MLDA)

Multi Linear discriminant analysis is valuable to structure a predictive model to group a membership based on observed features of each case. The system generates a discriminant function (or, for more than two groups, a set of discriminant functions) based on linear combinations of the predictor variables that provide the best bias between the groups. For grouping membership functions are generated from a sample of cases and also applied to the new cases [17].

All of these approaches are analogous discriminant function. This is analyzed to determine variable discriminate between two or more as expected occurring groups. LDA is a more sophisticated classification technique than correlation analysis, which does not consider the covariance structure. On the other hand, estimation of the covariance Σ requires more data, for this reason, LDA could perform sub-optimally on small data sets. MLDA overcomes this limitation by accepting large dataset.

B. Support Vector Machine

Support vector machine are a relatively new learning process influenced highly by advances in statistical learning theory and a sufficient increase in computer processing power in recent years. SVMs are very effective in a wide range of bio-informatic problems and in particular, perform well in analyzing microarray expression data and detecting remote protein homologies.

In Support Vector Machine a modelling error which occurs when a function is too closely fit to a limited set of data points that can be avoided by limiting the complexity of the models that the method may be generated. A specific approach for controlling the complexity of the models is given by the Vapnik-Chervonenkis (VC) theory and the structural risk minimization principle [19]. This is applied to the training sets of classification. Non-linearity in a data set is accounted for with kernel functions, which map the input vectors to some higher dimensional space [20]. A major advantage is that optimization problems resulting from SVMs have a global minimum solution and can be solved with standard quadratic programming tools.

Support vector machine is a learning system that uses a hypothesis space of linear functions in a high dimensional feature space, trained with a learning algorithm from the optimization theory. It attempts to minimize the upper bound on the generalization error based on the principle of Structural Risk Minimization (SRM) (28). The decision function implemented by SVM can be written as:

\[
f(x) = \text{sign} \left( \sum_{i=1}^{N} y_i a_i k(x, x_i) + b \right)
\]

Proposed Classifier

In the Proposed approach, MLDA is used for training and testing the multiple group dataset and SVM is used for Classification purpose. The multiple group is defined as Group 1 as Normal, Group 2 as medium and Group 3 as severe patients and use that variable as the dependent variable in a multiple regression analysis. The results are obtained that are analogous to those occurred via linear discriminate analysis.

Classical LDA is used for optimal transformation and minimize the class distance to obtain the maximum discrimination. The optimal transformation can be obtained by computing Eigen-decomposition value. This method is more robust than estimating the distribution of the data. The missing values in the limited information can also be retrieved using LDA [24].

1. Obtain the DICOM images using Medical Instrument.
2. About 75 DICOM images are taken. These images are equally divided into three categories as normal, medium and severe infected disease.
3. Pre-processing algorithm is applied for image enhancement and noise reduction for the captured DICOM images.
4. After pre-processing, the image is simplified using segmentation algorithm.
5. 40 features are extracted for 75 images.
6. Then Feature selection method is applied for effective computation.
7. Finally the Hybrid approach with the Combination of MLDA and SVM is used.

FIGURE 3 PROPOSED CLASSIFIER

MLDA was applied to the selected features to develop a discrimination model. Then the model is applied to the validation set for the prediction of the index of each features and the misclassification ratio was calculated. As a second process the SVM is applied to the selected features. The acquired result is applied for validation set. Then the misclassification ratio of MLDA/SVM is calculated.
IV. RESULTS AND EXPERIMENTS

The two methods are tested based on the different datasets such as Normal, Medium and Severe infected image. For having a comparison between different method, Feature selection by GA and Rough set is used. Here nearly 75 image data’s are taken and 40 features are extracted for classification. By using GA and Rough set the 40 features are reduced to 12 features for efficient classification. So that, the computation time will be reduced and improve the class prediction.

The percentage of reduction by GA method is 80%. 75 % of reduction is done by rough set method. The selected features are used for classification which improves the prediction accuracy. The proposed approach selects feature space of DICOM images which is reduced by 95%. The following Table 7 gives the results of the proposed method. The proposed approach proves efficient for DICOM image analysis.

<table>
<thead>
<tr>
<th>TABLE 7 RESULTS OBTAINED USING PROPOSED FEATURE SELECTION METHOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA method</td>
</tr>
<tr>
<td>Rough set Method</td>
</tr>
<tr>
<td>Proposed method</td>
</tr>
</tbody>
</table>

The effectiveness of the proposed method has been estimated using the following measures:

\[
\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)}
\]
\[
\text{Sensitivity} = \frac{TP}{(TP+FN)}
\]
\[
\text{Specificity} = \frac{TN}{(TN+FP)}
\]

where, True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) are the number of True Positive cases (anomalous cases correctly classified), the number of True Negatives (normal cases correctly classified), the number of False Positives (normal cases classified as anomalous), and the number of False Negatives (anomalous cases classified as normal) respectively. Accuracy is the proportion of correctly diagnosed cases from the total number of cases. Sensitivity measures the ability of the proposed method to identify anomalous cases. Specificity measures the ability of the method to identify normal cases [22, 23].

<table>
<thead>
<tr>
<th>TABLE 8 PARAMETRIC RESULTS FOR CLASSIFICATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>S.No.</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>1.</td>
</tr>
<tr>
<td>2.</td>
</tr>
<tr>
<td>3.</td>
</tr>
</tbody>
</table>

The above table 8 shows the results parametric results for classification. The proposed approach shows the better results in terms of accuracy, sensitivity and specificity.

V. CONCLUSION

DICOM image classification is a most important stage for disease analysis. But it is not a simple way for finest identification and early detection of diseases. Generally the DICOM image is a valuable and most reliable method in early detection. DICOM is used for brain scans and it is very useful and effective technique to detect the dissimilarity in brain images. In this paper a hybrid approach is proposed for DICOM image classification. The approach consists of feature extraction, selection and classification. The classification consists of Multi Linear Discriminant Analysis (MLDA) and Support Vector Machine (SVM). Classification is done on the base of parameter extracted by GLCM and histogram texture feature extraction method. The feature is selected using fuzzy rough set and genetic algorithm. The proposed approach has high approximation capability and much faster convergence.

REFERENCES


AUTHORS PROFILE

Ms. J. Umamaheswari, Research Scholar in Computer Science, Dr. GRD college, Coimbatore. She has 5 years of teaching experience and two years in Research. Her areas of interest include Image Processing, Multimedia and communication. She has more than 3 publications at International level. She is a life member of professional organization IAENG.

Dr. G. Radhamani, Director in Computer Science, Dr. GRD College, Coimbatore. She has more than 5 years of teaching and research experience. She has volume of publications at International level. Her areas of interest include Mobile computing, e-internet and communication. She is a member of IEEE.