International Journal of Advanced Trends in Computer Science and Engineering

Optimal Feature Subset Selection and MSVM Classification Based CBIR for Medical Images

S.Sankar Ganesh[1], S.V.Sudha[2], S.Ayyasamy[3], G.Pandiya Rajan[4], S.Chidambaram[5]
[1]Dr.N.G.P. Institute of Technology, Coimbatore, India, ssganesa@gmail.com
[2]Dr.N.G.P. Institute of Technology, Coimbatore, India, svsudha.mvenki@gmail.com
[3]Dr.N.G.P. Institute of Technology, Coimbatore, India, ayyasamyphd@gmail.com
[4]National Engineering College, Kovilpatti, India, glpandiyan86@gmail.com
[5]National Engineering College, Kovilpatti, India, chidambaramraj1@gmail.com

ABSTRACT

The main objective of this study is to design and implement a novel CBIR system for medical image retrieval system by comparing the features of the query image with the DB image. Here it is discussed about the feature selection for image comparison and retrieval. At present the entire process uses a wrapper approach using Association rules-based on feature selection and PS-classifier. As a result of this experiment, the proposed model is comparable to the other models on diabetic retinal datasets. Here it is introduced three different classifiers such as PS-Classifier, Artificial Neural Network and Association rules for selecting an optimal feature subset on diabetic retinal dataset as DRIVE to evaluate the effectiveness of proposed feature selection method. From the optimal feature subset, the Multi Class Support Vector Machine algorithm is used for classifying the abnormal condition of the medical images. The entire system is experimented in MATLAB software and the results are verified. From the results, it is observed that the feature selection based CBIR outperforms than the other existing one discussed in the literature. This paper obtained an improved accuracy in all classifiers in feature selection for any DR Dataset. Also, the result shows that the feature selection based CBIR has more efficiency in terms of relevant image retrieval. In addition, the specificity and sensitivity also improved after feature selection.

Key words: Image processing, Content Based Image Retrieval, Bag-of-Words, SIFT, SURF, Feature Extraction, Image Comparison.

1. INTRODUCTION

Content based image retrieval is a study deals with search, navigate, compare and retrieve digital images from a database. IR is one of the fast and emerging research area used for image retrieval specifically. It obtains user required image in terms of similarity measures. Images are compared and retrieved by comparing the related features extracted. Most of the IR methods are designed and developed by feature description methods. The set of all visual, geometrical and statistical properties of an image is called as features. The images are compared after encoding the image data in such a manner they can be compared efficiently. There are two different types of feature descriptions such as local and global feature descriptions. The local features can able to describe the image patches in terms of pixels, but a global feature describes the visual content of the image. In other words, a local feature can provide pixel information (hidden) whereas a global feature can provide visual features of the image. Comparing with local feature, the global features work faster than the local features. In this paper to increase the speed as well as to learn the entire information about an image for comparison both local and global feature descriptors are used([4],[5],[12],[20]).

Comparing with the global features, local features are used for efficient image retrieval with great discriminative power [15]. Also it is extracting a large amount of local features especially from large size data set. Due to that some of the existing approaches used BOVW ([14],[3]) model where it solves the scalability issues by quantizing descriptors into visual words. In order to avoid these kinds of problems this paper SSBOVW a system for image retrieval using local features using BOVW model. In this paper the author motivated to increase the accuracy in terms of image retrieval using local descriptors (SIFT [8], [9], SURF [2], [16]) based BOW model.

In terms of medical industry based CBIR, various earlier works were proposed and not able to provide high accuracy. The authors in [13] proposed an image retrieval method for medical image data based on the images content such as image shape, because, efficient content-based image Retrieval in the medical domain is still a challenging problem. In [11] the author used indexing based image retrieval system for surgical planning and medical training and patient diagnoses. It is not time effective since more manual processes are included.

The authors in ([7], [17]) developed a CBIR system names as Image Retrieval for Medical applications (IRMA) for rich image management process. The database used in this paper is available in distributed manner and global features are used for comparison. But the IRMA system found lack of ability in obtaining relevant pathological information based images. In ([1],[6],[18]) national library at USA developed a Spine Pathology and Image Retrieval System (SPIRS) based on the shape features. But SPIRS works well only for homogeneous images. The novelty of the proposed approach is the feature selection is carried out using three different efficient optimization algorithms. The best feature subset is obtained by finding the intersection among the resultant features selected by the optimization algorithms. The experiment is carried out over the bench mark dataset DIABETDB-1 and DIABETDB-2.

2. EXISTING WORK

The author in ([10],[12]) proposed a new approach for IR using BOW. The authors used SIFT features with BOW and hue descriptor for representing only color information in the image description. The experiment is carried out over benchmark dataset and the efficiency is verified. In order to avoid data loss, there are two different codebook is generated for quantization. More reference makes more confusion and takes more memory and CPU time.

3. PROPOSED WORK

In order to increase the efficiency in terms accuracy and applied for medical images, this paper motivated to provide a novel CBIR system with various classifiers including Association rules method. Also the novelty of this paper is

chooses the common features selected and classified from the feature labels for comparing the images with entire feature information available in the images. Nowadays medical applications are tremendously growing based on cloud, the medical information (especially medical images) are shared among the medical experts in online. To effectively retrieve relevant medical images, this paper provides an effective CBIR system especially for medical dataset. The proposed work comprises of two main stages such as training stage and testing stage. The training stage provides an optimal feature subset for efficient comparison with the testing stage. In the testing stage the query image is preprocessed, enhanced, GLCM feature extracted and features selected by Association rules algorithm. The entire process of the proposed system used in this paper is shown in Figure-1.
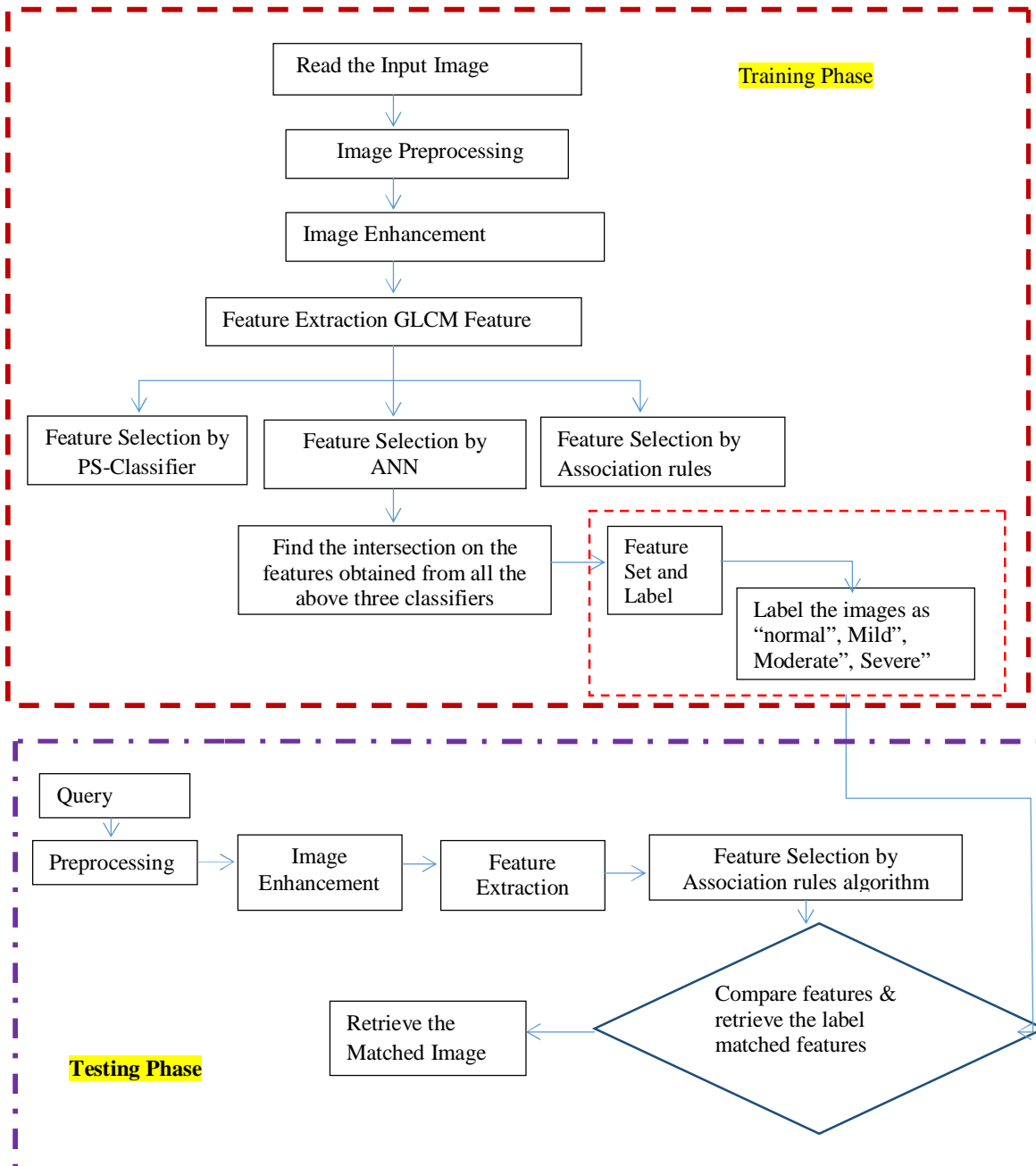

Figure 1: Overall Architecture of the Proposed System

3.1 Dataset

The images used in this paper are diabetic retinal images having various diseases like exudates, hemorrhages, Microaneurysms and other damages on the surface of the retinal images. The entire dataset is taken from benchmark dataset DRIVE which is publically available in http://www.isi.uu.nl/Research/Databases/DRIVE/download. php. The images in the DRIVE dataset are taken from a screening program in Netherlands. It comprises of 400 images with 7 signs. The patient's age group is from 25 years to 90 years. All the images are acquired using a CANON CR5 NON-MYDRIATIC 3 CCD camera. It has high resolution of pixels. Hence the image processing results can be perfect.

3.2 Training Stage

There are 30% of the images are taken for training purpose. The extracted features form GLCM feature extraction method is feed as input into PS-classifier, ANN-Classifier and Association rules -Classifier for selecting an optimal set of features. From the three obtained optimal feature set the common features are selected as the final feature subset for comparing and retrieval. The final feature subset is classified using a multi class SVM classifier with four classes such as "Normal", "Mild", "Moderate", and "Sever". This optimal feature subset and the four classes are used to compare with the test images and the relevant image is retrieved.

3.3 Testing Stage

In the testing stage, the remaining 70% of the images in the DRIVE dataset is considered as the test images/query images. Any image can be selected as the query image and the relevant images are retrieved from the database according to the feature subset and classes. Here a two-fold comparison is applied for improving the efficiency of the CBIR. Comparing the optimal feature subset is considered as the first fold comparison and comparing the classes is considered as the second-fold comparison. Since the final feature set is optimal obtained using various feature selection method, classes obtained from MSVM method the final retrieved image in accordance to the query image is more accurate.

3.4 Image Preprocessing

It is the first step of the process to acquire a digital image from the image database. The image database consists of the collection of n number of images depends on the user range and choice. It is the process of improving the image in its appearance and efficient representation. It involves extraction, analysis, and recognition of image coding, filtering, normalization, segmentation, and object identification. Image segmentation is the process of dividing an image into multiple parts. The output of this stage is a set of significant regions and objects, [19]. The noise occurred in the image I is removed by the filter parameter "h" whereas this parameter is estimated based on the noise variation σ^2. The output images are noise removed image which can be represented by

$$DIhv + n(DIh, v)$$

where, denoised image is represented as DIh, and $n(DIh, v)$ is noise predicted and v is the decomposed image.

3.5 Image Enhancement

It is more essential that enhancing the images for effective image processing. The poor quality of the image degrades the output of the image processing steps. Hence various image processing methods like Histogram equalization, contrast equalization and etc., are used. One of the main drawback in histogram equalization is it modifies the brightness of the image and it affect the original look of the images and it does not suit for all kind of images.

In order to overcome this issue, this paper used Contrast Limited Adaptive Histogram Equalization (CLAHE) method for enhancing the image. Initially wiener filter is used to sharpen the image and minimize the variations among the input image and the processed image using the mean value of brightness. Hence CLAHE control the level of contrast in the image which is enhanced image.

3.6 CLGM Feature Extraction

GLCM is a method to examine texture of spatial image. It compares and coordinates the pair of pixels with defined values. A texture property includes contrast between the pixels stored in matrix, correlation, homogeneity and energy. For an Image G, matrix P is defined as,

$$P(i,j) = \sum_{x=1}^{R}\sum_{y=1}^{C} \begin{cases} 1 & if\ I(x,y) = i\ and\ I(x+\Delta x, y+\Delta y) = j \\ 0 & else \end{cases} \tag{1}$$

$$Energy = \sum_{i=1}^{R}\sum_{j=1}^{C} P(i,j)^2 \tag{2}$$

Contrast of texture is given as,

$$Contrast = \frac{1}{N^2}\sum_{i=1}^{R}\sum_{j=1}^{C} P(i,j)(i-j)^2 \tag{3}$$

Correlation deals with the grey level of surrounding pixels and defined as,

$$correlation = \frac{\sum_{m=1}^{R}\sum_{n=1}^{C} nmp(m,n) - \mu_x \times \mu_y}{\sigma_x \sigma_y} \tag{4}$$

GLCM is a statistical approach which quantizes the matrix levels in different samples. Here Image description consists all the features described above, and based on intensity levels the matrix is adjusted. GLCM updates relationship from i to j in the matrix to consider the probability of relationship. By using proper mean function the mean intensity obtained.

4. CLASSIFICTION
4.1 Classification Based On PS

A classification model considers a set of attribute set as an input and provides class label as an output. Function f, is a learning function which maps to each attribute say x to the one of class label y. A classification model retrieves learning algorithm into learning model by applying certain functionalities and deducing to class label. Pruned set method based classification considers P-Infrequent label sets from the trained data set and removes examples in that to make generic label sets, and leads to P-Frequent label sets followed by classifier method. The main focus is to reduce unique class values.

Consider the feature set F,

$$F = \{f_1, f_2, \ldots, f_n\} \qquad (5)$$

```
Algorithm_PS_Classifier( )
{
Begin
        Pruned set = Random sizable
        Pruned set = Technical Handlers
        for I in 1.. n
        loop
                P=infrequent (label set)
                label-set-new  -> label-set-old
                p = frequent (label-set-new)
        end loop
End PS
}
```

4.2 Classification Based on Association rules

The set of association rules triggers to identify the best set of attributes from the selected dataset. It improves accuracy and reduces the training time.

$$Si < (s, a, Action(S), Result(s, a), Cost(s, a)) > \qquad (6)$$

where, S, gives the set of all possible states and A is set of possible actions. The performance metric accuracy is defined as number of correct predictions with total number of predictions. Here the realization obtained by analyzing 500 total images in the dataset. For an unknown features, the predictions to be identified to get maximum confidence score.

```
Algorithm_Rule_Based_Algorithm ( )
Begin
        for i in 1.. n
        loop
        rule-t -> Rule(x, attribute, y)
        rule(i) -> rule-t
        end loop
endRule_Based_Algorthm
```

4.3 Classification based on ANN

ANN is considered good for classification algorithm, due to high generalization performance and less interference. Image is considered as model that learns to generate combination of labels and images. ANN back-prorogation is considered to classify the medical images because it is non-linear. ANN is preferable because mapping complex class distributions is easy. ANN yields intelligent analysis of medical imaging area compared to other conventional methods. Back-Propagation classifier takes feature vector recognizing the data to provide class identifier. Back-propagation considers all the input and output neurons to provide final classification output. Input neurons are considered as features, Output neurons are considered as classes and intermediate/hidden neurons are called as experiments starts from input node to output node.

$$ANN_i = \begin{cases} features \ \forall \ Neuronset \ t \\ classes \ \forall \ Neuronset \ O \\ \exp \forall \ Neuronset \ I \end{cases} \qquad (7)$$

4.4 Adjudicator Unit

All three classifier outputs are given to the intermediate unit called adjudicator which is shown in Figure-2. It will remove the duplications and integrate the obtained feature set (f$_s$) from the entire three classifier algorithm.
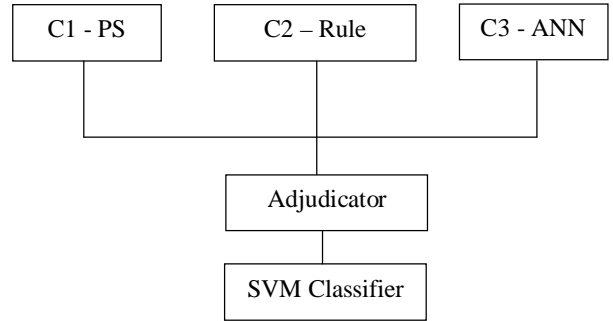

Figure 2: Adjudicator Unit for Combining Feature Selection

4.5 Multi Class Support Vector Machine Based Feature Selection

Support Vector Machine classifies the entire input data into two different classes under the label of 0 and 1. In other words it can be said as positive or negative or lower or upper classes. But, in order to classify the dataset under more than two classes, it is essential to use Multi-Class Support Vector Machine algorithm. SVM and Multi-Class SVM are belongs to machine learning approaches, it is able to learn any kind of data by itself without any external technical supports.

Both approaches have in-built mechanisms to verify the similarities, distance and matching score among two different objects such as trained object and test object. But the Multi-class SVM classifies the data under various classes required by the user. The classification method used in MSVM is illustrated in Figure-3.
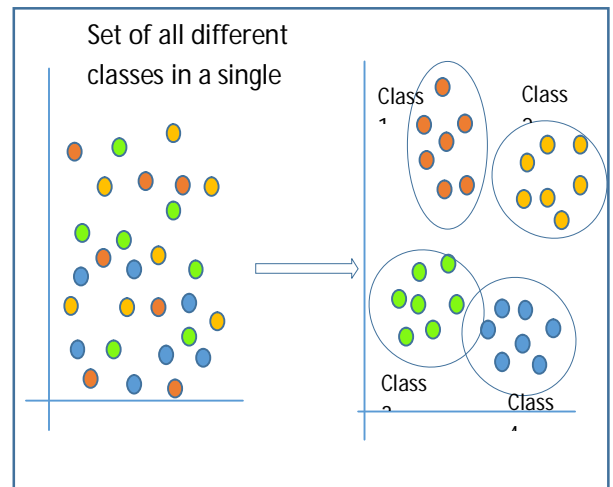

Figure 3: Multi Class SVM

MSVM classification is used frequently while handling medical images; it works on maximization of margin around the hyper plane. Feature vectors and class labels are considered for sampling the classes. Proper classifier gives the feature selection hints Apriori for the given sets which is already trained. The effectiveness is further more extracting from the adjudicator and assigns the class label to normal, mild, moderate and severe.

$$\text{Vectors } X_i = (V_{i1}, V_{i2}, \ldots, V_{in}) \qquad (8)$$

where n is the dimensionality

$$SVM \rightarrow w + W^T X \qquad (9)$$

where$w$ is the weight vector $(V_{t1}, V_{t2}, ..., V_{tm})$

$$f(x) = sgn\{\sum_{i=1}^{setvector} \lambda^* y_i (X_t \cdot x) + b^k\} \quad (10)$$

The predictor specifies positive and negative objects from the hyper plane, and b is normal to the hyper plane. Trained SVM model based on the feature Subset. Here 8 datasets belong to 4 classes is considered. Based on the improvised feature selection algorithm final cluster set is obtained. The four classes are:

$$c1 = Normal, \quad c2 = Mild,$$
$$c3 = Moderate, \quad c4 = Severe \quad (11)$$

The entire process of the proposed approach is written in the form of step-wise procedure as:

Proposed Algorithm ( )
{
- Image preprocessing
- Image Enhancement
- GLCM Based Feature Extraction
- Feature Classification using PS-Classifier obtain the result as **R1**
- Feature Classification using ANN-Classifier obtain the result as **R2**
- Feature Classification using GA-Classifier obtain the result as **R3**
- Get final classified results **R={R1∩ R2∩ R3 }**
- Read the Query Image
- Image preprocessing
- Image Enhancement

- GLCM Based Feature Extraction
- Feature Classification using GA-Classifier obtain **Rt**
- Compare **Rt** with R and retrieve the matched image as the resultant Image
}

In order to implement the proposed approach and verify the result the entire process of the proposed approach is given the form of an algorithm. The algorithm is implemented in MATLAB software and the results are verified. From the experimental results the accuracy is verified. The performance of the proposed approach is evaluated by comparing the obtained results with the existing results. To find out the accuracy the performance metrics such as True positive rate, false positive rate, True negative rate, false negative rate, sensitivity and specificity.

5. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this paper, the proposed approach is programmed in MATLAB software and the results are verified. The following Figure-4 shows the resultant images obtained in each stage of the proposed approach like input image read from the dataset, noise removed, gray scale converted and CLAHE based enhanced image. The input image can be selected randomly from the testing dataset (70%) other than the trained images. In order to differentiate the training images and testing images are stored separately in two different folders in the experimental database.
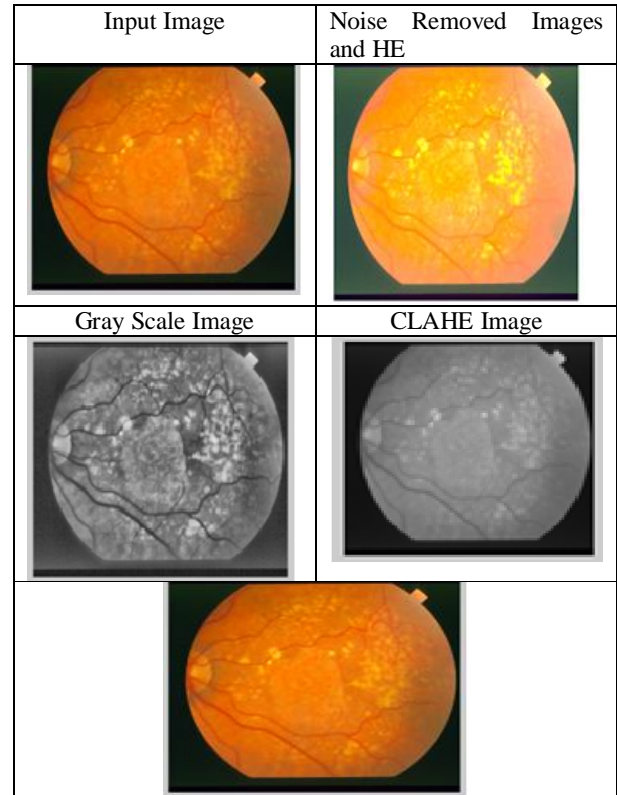

Figure 4: Resultant Image obtained after noise Removed, Gray scaled converted and CLAHE applied

The matrix shows the GLCM matrix attained from the DR image where all these values in matrix characterize the features of the image in Figure-5. Gray level, contrast, correlation and energy are the textural features are calculated by the GLCM. The retrieved image obtained from the experiment carried out on the proposed approach is given in the last row in Figure-4.

$$\begin{bmatrix} 888 & 147 & 9 & 0 & 0 & 0 & 0 & 0 \\ 137 & 2897 & 59 & 50 & 3 & 0 & 0 & 0 \\ 19 & 32 & 65 & 7 & 8 & 1 & 0 & 0 \\ 0 & 59 & 56 & 4262 & 47 & 1 & 0 & 0 \\ 0 & 11 & 13 & 463 & 6263 & 88 & 0 & 0 \\ 0 & 0 & 0 & 3 & 86 & 78 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Figure 5: GLCM Matrix Obtained from the Experiment

After feature extraction all the features are feed into PS, ANN and Association rule based methods for selecting the optimal features. From the optimal features the test image features are compared and the relevant images are retrieved. It is very difficult in developing CBIR for Diabetic retinopathy based medical images due to the heavy similarity among all the images. Manually it is well known that the similarities of the medical images are higher than the dissimilarities. In accordance to the query image the relevant images retrieved by the proposed approach is shown in Table-1. The first column shows the input images used as query images. Other columns show the relevant images retrieved from the experiment. From Table-1, the efficiency can be noticed. Even though the efficacy can be computed by the precision recall values calculated in the experiment.

Few images are taken from the testing set and used as the query image. In accordance to the query image the retrieved relevant images is given in column 2 and in column 3. In order to make simplicity only two resultant images are taken from the retrieved images and given in column 2 and in 3. The calculated energy, entropy and contrast values for the query images are given in column 3 to column 5. The relevancy of the query image is computed among the optimal feature vector of the query image and the trained images. It is obtained by computing the Euclidian distance among the optimal feature vector. For comparison the threshold values for energy, entropy and contrast are referred from the below Table-2. In order to classify the entire feature, set and label the classes the threshold values given in Table-2 is compared with the obtained feature values.

It is not sure that the GLCM feature values are very accurate, since the texture properties of the DR images are merely similar in terms of color, shape and textures. In order to evaluate the performance of the proposed approach the experiment is carried out in the entire dataset and the results are verified. According to the result the performance metrics such as precision and recall are calculated and presented here.
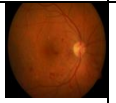
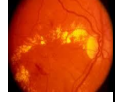**Table 1 :** Results Obtained from the Experiment
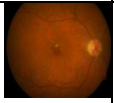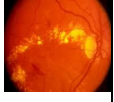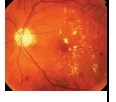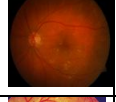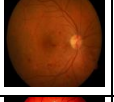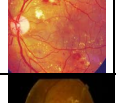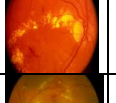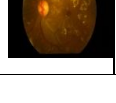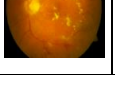
Query Image	R1	R2	Energy	Entropy	Contrast
			0.032	4.321	0.192
			0.065	7.983	0.828
			0.035	6.574	0.268
			0.054	8.020	0.872
			0.033	7.843	0.645

Table 2 : Feature Values denote four different classes

Features	Normal	Abnormal	Moderate	Severe
Energy	0.1 to 0.2	0.2 and 0.6	0.6 to 0.9	0.9 and above
Contrast	0.0015 to 0.003	0.003 to 0.005	0.005 to 0.075	0.075 and above
Entropy	4 to 5.5	5.5 to 6	6 to 7	7 and above

In CBIR systems the precision and recall represents the accuracy of the system in terms of relevancy for the query images. Precision is calculated as:

$$Precision = \frac{TP}{TP+FP} \qquad (12)$$

and recall is calculated as:

$$Recall = \frac{TN}{TP+FN} \qquad (13)$$

TP is calculated as the number of correctly retrieved out of all the images

FP is calculated as the number of incorrectly retrieved out of all the images

TN is calculated as the number of incorrectly retrieved from the incorrect images

FN is calculated as the number of incorrectly retrieved from the correct images

The experiment is carried out in over 300 images whereas 100 images belongs to normal class, 50 images belongs to mild class, 75 images belongs to moderate class and 75 images belongs to severe class. The number classified images are given in Table 3. The performance of the proposed approach is evaluated by comparing the results among the classifiers individually. The classification by PS-classifier with MSVM, ANN-classifier with MSVM and AR-Classifier with MSVM are calculated and compared. The obtained results using the individual classifiers are given in Table-3. From the obtained results it is very clear and noticed that AS-classifier with MSVM is better than the other classifiers. From the entire results it is noticed and clear the proposed approach obtained more classification accuracy than the other individual classifiers.

Table 3 : Classes Obtained from the Experiment

Images	Normal	Mild	Moderate	Severe
Database Images	100	50	75	75
PS-classifier with MSVM	98	50	73	72
ANN-classifier with MSVM	99	49	74	73
Association Rule Based Classifier with MSVM	99	50	74	74
Proposed Approach	99	50	75	74

TP = 298/300 =0.993333
FP =2/300 =0.006667
TN =1/150 =0.006667
FN =1/100 =0.01

From the obtained TP, FP, TN and FN the precision, recall, sensitivity, specificity and accuracy are calculated as:

$$Precision = \frac{TP}{TP+FP} = 0.993333$$

$$Recall = \frac{TP}{TP+FN} = 0.990033$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = 0.999891$$

From the above calculation, in terms of percentage, the obtained precision, recall and accuracy are 0.99333, 0.990033 and 0.999891 respectively. From the obtained results it is very clear and noticed that the proposed approach outperforms in retrieving relevant images.

6. CONCLUSION

The main objective of this study is to design and implement a novel CBIR system for medical industry. Medical experts want to retrieve relevant images in terms of various classes

under normal and abnormal. Most of the images are very similar in terms of structure, texture, shape other external features. Since, the accuracy of the existing approaches is less, it is essential to design and develop an accurate CBIR for medical industry. In this paper diabetic retinal images are considered as the medical images used in the experiment. To do that there are three different algorithms are used for optimal feature selection and they are classified using Multi-class SVM classifier. Due to the optimal feature selection the accuracy of the image retrieval is high. From the experiment it is noticed that the proposed approach obtained precision, recall and accuracy are 0.99333, 0.990033 and 0.999891 respectively. Comparing with the other approaches it is high.

REFERENCES

1. Antani S, Long LR, and Thoma GR. **Content-based Image retrieval for large biomedical image archives**, *Studies in health technology and informatics* Vol 107 No 2 pp. 829-833, 2004.
2. Bay H, Tuytelaars T, and Van Gool L. **Surf: Speeded up robust features**, *Computer Vision – ECCV Springer* pp. 404–417, 2006. https://doi.org/10.1007/11744023_32
3. Csurka G, Dance C, Fan L, et al. **Visual categorization with bags of keypoints**, *Work Stat Learn Comput Vision - ECCV* Vol 1 pp. 1-16, 2004.
4. Datta R, Joshi D, Li J, et al. **Image retrieval: Ideas, influences, and trends of the new age,** *ACM Computing Surveys (CSUR)* Vol 40 No 2 pp. 1-60, 2008.
5. Halawani A, Teynor A, Setia L, et al. **Fundamentals and Applications of Image Retrieval: An Overview,** *Datenbank-Spektrum* Vol 18 pp. 14–23, 2006.
6. Lai H, Yan P, Shu X, et al. **Instance-aware hashing for multi-label image retrieval,** *IEEE Transaction on Image Processing* Vol 25 No 6 pp. 2469-2479, 2016. https://doi.org/10.1109/TIP.2016.2545300
7. Lehmann TM, Guld MO, Thies C, et al. **Content-based image retrieval in medical applications,** *Methods of Information in Medicine* Vol 43 No 4 pp. 354-361, 2004.
8. Saravana Balaji B, Krishna Kumar V, Ahmed Najat Ahmed, **Semantically enriched Tag clustering and image feature based image retrieval,** *International Journal of Advanced Trends in Computer Science and Engineering* Vol 8 No 1.2 pp. 138-141, 2019.
9. Lowe DG. **Distinctive image features from scale-invariant keypoints**, *International Journal of Computer Vision* Vol 60 pp. 91–110, 2004.
10. Mohammed Alkhawlani, Mohammed Elmogy, and HazemElbakry. **Content-Based Image Retrieval using Local Features Descriptors and Bag-of-Visual Words**, *International Journal of Advanced Computer Science and Applications* Vol 6 No 9 pp. 212-219, 2015.
11. Muller H, Michoux N, Bandon D, et al. **A review of content based image retrieval systems in medical applications-clinical benefits and future directions**, *International Journal of Medical Informatics*, Vol 73 No 1 pp. 1-23, 2004.
12. Odrigues J, Cristo M, and Colonna JG. **Deep hashing for multi-label image retrieval: a survey,** *Artificial Intelligence Review, 2020* https://doi.org/10.1007/s10462-020-09820-x.
13. Ramamurthy B, and Chandran KR. **Content based Image Retrieval for Medical Images using Canny Edge Detection Algorithm**, *International Journal of Computer Applications,* Vol 17 No 6 pp. 32-37, 2011.
14. Sankar Ganesh S, and Ramar K. **Content Based Image Retrieval using Heuristic Particle Swarm Optimization**, *Asian Journal of Information Technology*, Vol 15 No 12 pp. 1929-1935, 2016.
15. Sankar Ganesh S, Ramar K, Manimegalai D, Sivakumar M. **Image Retrieval Using Heuristic Approach and Genetic Algorithm**, *International Journal of Computational Information Systems,* Vol 8 No 4 pp. 1563-1571, 2012.
16. Song G and Tan X. **Learning multilevel semantic similarity for large-scale multi-label image retrieval**, *In: Proceedings of the 2018 ACM on international conference on multimedia retrieval* pp. 64-72, 2018. https://doi.org/10.1145/3206025.3206027
17. Thies C Guld MO Fischer B et al. **Content based queries on the CasImage database with in the IRMA framework**, *Lecture Notes in Computer Science* Vol 3491 pp. 781-792, 2004.
18. Thoma GR Long LR and Antani SK. **Biomedical Imaging research and development: knowledge from images in the medical enterprise**, *Technical Report The Lister Hill National Centre for Biomedical Communications US National Library of Medicine* pp. 1-69, 2006.
19. Zhu Y Li Y and Wang S. **Unsupervised deep hashing with adaptive feature learning for image retrieval**, *IEEE Signal Processing Letters* Vol 26 No 3 pp. 395–399, 2019. https://doi.org/10.1109/LSP.2019.2892233
20. Nadiah Yusof, Amirah Ismail, Nazatul Aini Abd Majid. **Near-Duplicate Images Detection and Clustering in Large-Scale Image Dataset: A Systematic Literature Review**, *International Journal of Advanced Trends in Computer Science and Engineering* Vol 9 No 1.1 pp. 446-460,2020.