2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS)

A Hybrid Feature Modeling Approach for Content-Based Medical Image Retrieval

Karthik K
Department of Information Technology
National Institute of Technology Karnataka
Surathkal, India
2karthik.bhat@gmail.com

Sowmya Kamath S
Department of Information Technology
National Institute of Technology Karnataka
Surathkal, India
sowmyakamath@nitk.edu.in

***Abstract*—With the proliferation of various imaging based diagnostic procedures in the healthcare field, patient-specific scan images constitute huge volumes of data that needs to be well-organized and managed for supporting clinical decision support applications. One such crucial application with a significant impact on point-of-care treatment quality is a Content Based Medical Image Retrieval (CBMIR) system that can assist doctors in disease diagnosis based on similar image retrieval. Medical images are multi-dimensional and often contain manifold information, due to which efficient techniques for optimal feature extraction from large-scale image collections are the need of the day. In this paper, an efficient CBMIR model is proposed that is built on multi-level feature sets extracted from medical images. Four different feature extraction techniques are used to optimally represent images in a multi-dimensional feature space, for facilitating classification using supervised machine learning algorithms and top-k similar image retrieval. Experimental validation of proposed model on the standard ImageCLEF 2009 dataset containing 12,560 X-ray images across 116 classes showed promising results in terms of classification accuracy of 85.91%.**

***Index Terms*—Content based medical image retrieval (CBMIR), Information Retrieval, Automated Medical Diagnostics**

I. Introduction

Over the past two decades, there have been significant advancements in medical imaging technologies for disease diagnosis, resulting in various types of imaging data like X-Ray scans, MRI (Magnetic Resonance Imaging), CT Scans (Computed Tomography), PET (Positron-Emission Tomography) etc. Consequently, the volume of medical scan images is growing exponentially, given the large number of patients seeking medical attention. For improving organization and storage of medical scan images, several standards have been proposed and standardized. The DICOM standard (Digital Imaging and Communications in Medicine) [1] can be used for storing patient information and their scan images for communication. Systems like PACS (Picture Archiving and Communication Systems) [2] were created for improving access to stored DICOM files, so that higher-level applications like decision making can be supported. However, most such systems have many restrictions, like the dependency on text-based search for images, requiring keyword matching capabilities. Thus, image retrieval often suffers from limited accuracy, due to factors like ambiguous/sparse textual descriptions of images, non-existent/low quality image annotations (e.g. test description, patient profile, and other technical parameters) etc. To overcome these limitations, effective techniques called Content-Based Image Retrieval (CBIR) that use pure visual cues to retrieve relevant images from huge image collections were developed.

Content-Based Medical Image Retrieval (CBMIR) has been an area of significant research interest as its application assist doctors in efficient retrieval of similar images that helps in faster patient disease diagnoses [3]. The IRMA project (Image Retrieval in Medical Applications) project [4] focuses on building a CBMIR system for X-ray images for automated diagnosis. IRMA is built on a pipeline composed of interdependent processes like categorization, registration, feature extraction and selection, indexing and identification for enabling content-based image retrieval.

The problem of medical image classification and retrieval is also an area of active research. Scan images like X-rays, CT scans etc can encompass several internal organs, it is essential to devise automatic classification approaches to deal with the diversity. Fesharak et al [5] used shape based feature extraction techniques and used Bayesian rule classifiers for classifying these images across 28 classes. Pourghassem and Daneshvar [6] proposed a merging based classification algorithm that measured weighted Euclidean distance for retrieval of images over 57 classes. This CBMIR model was evaluated on 10,000 training and 1000 query images. Zare et al [7] proposed a X-ray classification technique that uses low level features extracted from scans for training a support vector machine classifier and reported 70 classes had a classification accuracy greater than 80%. Random barcode method by Tizhoosh [8] based on random transform, is one of the preliminary work for retrieval of medical images which achieved a good result.

Despite significant improvements in medical image retrieval, conventional retrieval models that support querying based on text/keywords fail to capture the latent visual features in an image, paving the way to CBIR systems for medical images. As medical images are inherently multi-dimensional and complex in the varied information that can be inferred from them, designing effective feature extraction mechanisms can help to improve overall retrieval accuracy and also recall. In this paper, a hybrid feature model for optimal representation of medical image information is proposed for classification and

Part No: CFP1858A-ART; ISBN No:978-1-5386-8492-4/18/$31.00 ©2018 IEEE

content-based retrieval of radiography images. Four different feature extraction methods are used for constructing the hybrid feature vector that represents the shape and texture properties of the images, which is used for effective classification of the images. The search interface is provided as a Query By Example (QBE) interface, where a representative image can be submitted and top-k similar images are retrieved using a pair-wise similarity measure. The rest of this paper is organized as follows: In Section II, the proposed methodology for radiography medical image Classification and Retrieval using texture and shape feature extraction techniques is presented. The experimental results are discussed in section III, followed by conclusion and possible directions for future work.

II. Proposed Methodology

In this section, we discuss the proposed supervised learning framework that incorporates hybrid feature modeling techniques for large-scale medical image data, to enable content-based image retrieval. For our experiments, we used the ImageCLEF 2009 dataset [4], considering 12,560 images across 116 different categories. The processes involved in the proposed CBMIR system is illustrated in Fig. 1.

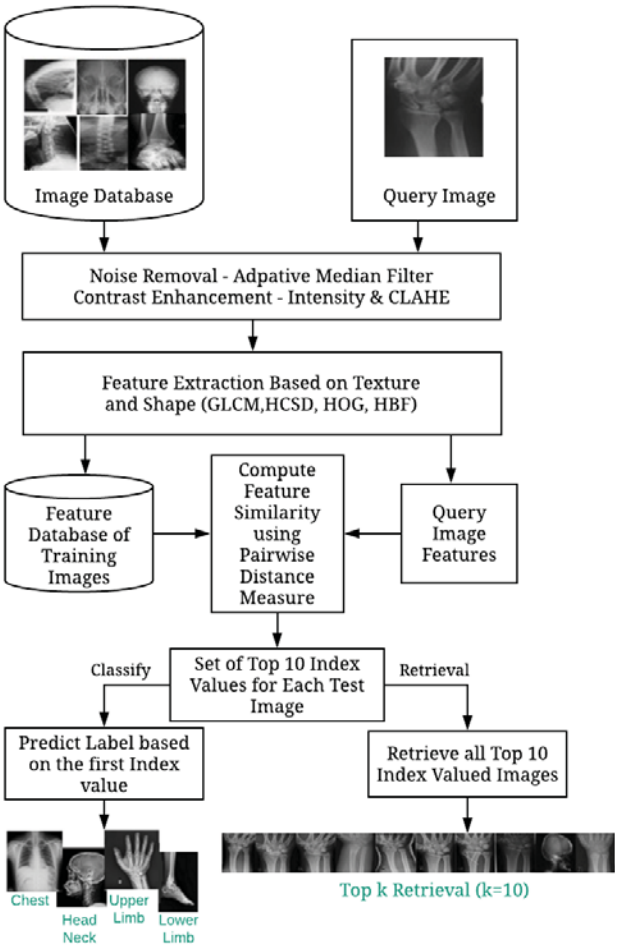


Fig. 1: Workflow of the Proposed CBMIR Model

A. Noise Removal & Contrast Enhancement

Medical images often contain some visual noise. In x-Ray images specifically, the presence of noise is commonly due to random photon distributions. As this causes an undesirable variations and also affects the visibility of objects of interest (like bones, organs etc.) in the image, it is crucial to remove this noise before the process of feature extraction. We used an adaptive median filtering technique for removing noise from the radiography images in our dataset. Adaptive median filtering [9] is an image enhancement technique that works well due to its low sensitivity to pixel value changes. It also does not affect the sharpness of the scan image, preserving even small objects-of-interest.

Next, intensity changes and contrast adjustment techniques are used to improve image quality further. The intensity value of an image is adjusted by saturating all its pixel values by 1%. Similarly, image contrast is boosted by transforming its pixel values with a technique called Contrast-Limited Adaptive Histogram Equalization (CLAHE) [10]. CLAHE operates on small regions in the image, called tiles, rather than the entire image. Each tile's contrast is enhanced individually. The contrast, especially in homogeneous areas, can be limited to avoid amplifying any noise that might be present in the image. The transformation of the image after applying these processes is shown in Fig.2. After these processes, feature modeling is performed for extracting texture and shape based features.

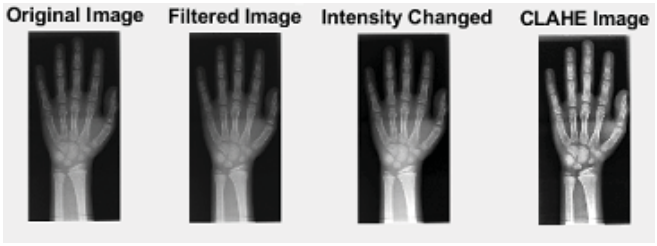


Fig. 2: Images enhancement process

B. Feature Modeling

The performance of a CBIR system is highly dependent on the inherent properties of images, represented formally to facilitate retrieval. Most diagnostic scan images are generally monochromatic (e.g. X-ray). Hence, effective local and global-level analysis of the images are critical. Also, the collection contains scans belonging to different classes, hence capturing this variation is also crucial in order to deal with the variety of images. In our work, four different feature extraction techniques were used for obtaining a well-rounded image representation. As feature extraction plays an important role in image retrieval performance, we focused on texture and shape based feature extractions.

The Gray Level Co-occurrence Matrix (GLCM) [11] is a statistical method that examines the spatial relationship between the pixels in an image for generating features. The GLCM feature vector is composed of 14 features which are primarily texture based features. These are extracted for each

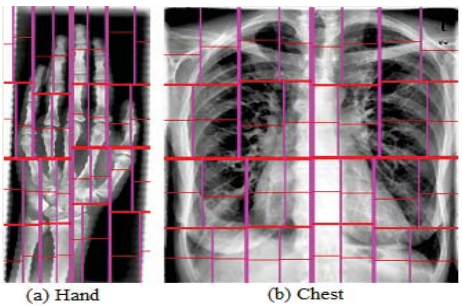
Part No: CFP1858A-ART; ISBN No:978-1-5386-8492-4/18/$31.00 ©2018 IEEE

X-ray image in our dataset. These features are related to the specific characteristics of the image such as Angular Second Moment, Contrast, Correlation, Variance and Homogeneity. These features capture the individuality of an image in the form of probability of a pixel finding its gray level intensity i at distance d, which can be formulated as $P(i, j : d)$. Here, each pixel is associated with its 8 neighboring pixels except the edge pixels.

The Hierarchical Centroid Shape Descriptor (HCSD) [12] generates a 124-dimensional feature space, and is augmented using a kd-tree technique proposed by Sexton et al [13]. HCSD is a shape feature extraction method which starts by recursively decomposing an image into sub-images. Initially, it takes image I as the input and computes its transpose I^T, calculates the centroid C(x_c, y_c) for each input with 1^{st} order moment along x and y axis, where x_c=m_{10}/m_{00} and y_c=m_{01}/m_{00}. The *order of moment* of a 2D function $f(x, y)$ is formulated as per Eq. (1). A digital image's raw moment m_{pq} with $I(i, j)$ pixel intensities is calculated as per Eq. (2). The *centroid* is calculated by dividing the image area recursively until the desired depth (here, we considered depth=7) is reached. Fig. 3 illustrates how the centroid values are calculated from the gray scale images. At each level, the axis of the coordinates is captured, after which the generated vector is normalized to [-0.5, 0.5] range, point 0 being the root level. Finally, the features extracted are concatenated to the image feature.

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx.dy \tag{1}$$

$$m_{pq} = \sum_{i=0}^{M} \sum_{j=0}^{N} i^p j^q I(i, j) \tag{2}$$



(a) Hand (b) Chest

Fig. 3: HCSD extraction from X-ray images

Histogram of Oriented Gradients (HOG) is implemented by dividing the image dimension into smaller regions and storing the histogram orientation of the pixel values of that region [14]. Each pixel of the region has a weight as per the gradient L2-norm. Here, histogram channels are performed based on the rectangular region. As overlapping can occur, regions are contributed more than once to the final feature vector. Initially, this feature extraction was used for generic images, but it is found to be well-suited for medical images, as it can be used to extract the directional change of intensity levels in the image. A total of 9 rectangular regions are used here with 9 bin histogram per region. The 9*9 feature vector is now concatenated to the image's feature vector, thus resulting in 81 dimensional feature vector.

In case of Histogram Based Features (HBF), we used five different features, namely – mean, variance, standard deviation, median absolute and RMS values. Let I be the variable indicating image gray levels, $p(z_i)$ be the histogram where $i = 0, 1, 2, ...L-1$, where L is the number of discrete gray levels. The average gray level of each region in the image (mean) is computed as per Eq. (3). Variance is the amount of the difference in the gray level (Eq. (4)).

$$\mu = \frac{1}{N} \sum_{i=1}^{N} A_i \tag{3}$$

$$S = \frac{1}{N-1} \sum_{i=1}^{N} |A_i - \mu|^2 \tag{4}$$

Standard deviation captures the amount of dispersion from the mean gray level of an image (Eq. (5)). A low value of standard deviation indicates that it is very close to the mean value, on the other hand a high value means that the data are spread over a range of values. The Median absolute is the measure of the variability in the image and it is computed as per Eq. (6). Root Mean Square (RMS) is used to compute the overall contrast of the image as per Eq. (7).

$$S = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} |A_i - \mu|^2} \tag{5}$$

$$M = \mu |X - \mu(X)| \tag{6}$$

$$I_{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} |I_i|^2} \tag{7}$$

The feature values obtained after using the four feature extraction techniques described are fused to form a hybrid feature vector, which optimally represents each X-ray image. The hybrid feature vector is a 224-dimensional vector, composed of features as shown in Table I. Thus, at this stage, all the images are processed and represented by their 224-dimensional feature vector. These are used for identifying how close the feature vectors of any two given images are, when a query image feature vector is submitted to the retrieval system.

C. Pair-wise Similarity & Class Label Prediction

Once the training and testing image feature vectors are generated, a pairwise distance calculation technique is used on them for determining the most relevant image index values for each of the testing image. We used the Standard Euclidean pairwise distance method, which gave the best result, i.e., it gave the best nearest image index values for the testing image feature vectors to the training feature vectors. As we aimed for top-k retrieval, where k=10, we have generated 10 image index values from the training set for each of the testing image.

TABLE I: Summary of Feature Modeling processes

Method	Description	Total features
GLCM	Texture properties of the image	14
HCSD	Recursively decomposes the image into sub-images	124
HOG	9 rectangular regions with 9 bin histogram per region	81
HBF	Mean, variance, standard deviation, median absolute and RMS values of the input image	5
Hybrid	Combination of GLCM, HCSD, HOG and HBF features	224

From the top-10 index values, the first index value is considered for the actual classification of the test image. The label of all images at the first index value is fetched from the training set which will be the predicted label for the classification task. For classification and label prediction, we used the k-nearest Neighbors (kNN) algorithm [15]. kNN classifies an unknown object into its category among the training data and it uses the nearest neighbors between the two set of vectors i.e., between the test vector and the training vectors. kNN is widely used machine learning algorithm due to its simplicity and the results that were obtained when this technique is applied was found to be excellent in many applications. Hence, we incorporated the kNN classifier for label prediction. We experimented with different variations of the kNN algorithms, by varying the number of neighbors used. We used six kNN variants that are based on different distance measures. The kNN models, Fine, Medium and coarse are based on the nearest number distance calculations set to 1, 10 and 100 respectively. In cosine, cubic and weighted kNNs, the cosine, cubic and weighting distance metric is used for comparing two dimensional vectors. The classification task was validated by using 10-fold cross validation.

D. Content-based Image Retrieval

After the pairwise distance calculation for all the testing images with the training image set, top 10 image index values were obtained, and their labels are also predicted. These image index values are now used for the content-based image retrieval. For each test image, its pairwise distance calculated with reference to the image index value is read from the training image set. Next, the test image and the read top 10 indexed images are displayed as the retrieval results, and the performance of the retrieval is observed.

III. Experimental Results

For the experimental validation of the proposed methodology, we used the ImageCLEF2009 dataset, considering 12,560 labeled X-ray images of different body regions like face, nose area, shoulder, elbow, forearm, chest, arm, hand, wrist, finger, knee, leg, foot, ankle etc. Each category of images was divided into training and testing with a ratio of 70:30. Some sample images from the dataset are shown in Fig. 4.

The implementation was carried out with an Intel Xeon Workstation @3.31 GHz and 16 GB of memory running

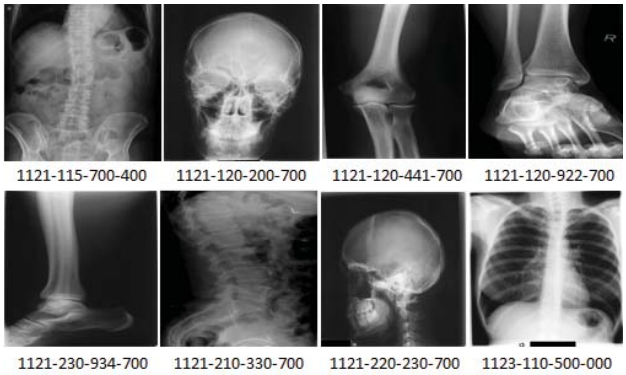


Fig. 4: Sample images from IRMA dataset and their IRMA codes [4]

Matlab v.2017. We used standard Euclidean distance metric for pairwise distance measurement between the two image vectors, and the classification results are tabulated in Table II. It can be seen that this achieved good accuracy of 85.91%, with significantly reduced false positive rate. With respect to this the performance of the proposed model with state of the art methods compared is reported in Table III.The hybrid feature vector was fed to six variants of kNN classifiers, with 10-fold cross validation. From Table IV, it can be seen that the kNN classifier model using Cosine Measure as the similarity metric achieved the best accuracy. To validate these results, the Receiver Operating Characteristic curves (ROC) were plotted for one versus other classes. The ROC curves for the class *hand* and *leg ankle joint* versus other classes plotted using cosine kNN classifier is shown in Fig. 5.

TABLE II: Results for Standard Euclidean Pairwise distance

Evaluation Metrics	Values (Before Enhancement)	Values (After Enhancement)
Accuracy	82.97	85.91
Error	17.03	14.09
Precision	60.34	63.47
Recall	55.22	58.73
Specificity	99.64	99.70
False Positive Rate (FPR)	0.36	0.30
F1 Score	57.67	61.09

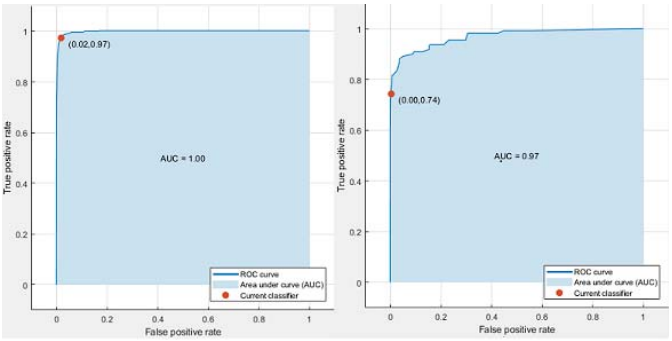
We observed that most classes had a good retrieval performance while for some classes the retrieval was average. This

TABLE III: Comparison of the proposed approach with existing works

Approach	No of Classes	Accuracy
Proposed Hybrid Approach	**116**	**85.91%**
Shape Features and Bayesian Rule [5]	28	82.87%
Merging Scheme [6]	57	90.23%
Combined Features [7]	116	Acc $\geq$ 80 (70 classes)

TABLE IV: Classification accuracy obtained for various classifiers

Classifier Model	Accuracy (%) (Before Enhancement)	Accuracy (%) (After Enhancement)
Fine kNN	79.1	81.6
Weighted kNN	79.3	81.4
Cosine kNN	78.5	81.3
Medium kNN	77.6	79.8
Cubic kNN	76.2	78.7
Coarse kNN	66.7	68.5



(a) Class Hand (b) Class Leg Ankle Joint

Fig. 5: Area under ROC Curve - one versus all other classes using cosine kNN classifier

is because, there is high class imbalance in the ImageCLEF dataset, i.e., several classes have more than 200 images, while some classes have only 10-15 images. Thus, the classifier was trained on less data, due to which label prediction accuracy and retrieval performance was low. Top-k retrieval results for some sample classes for which high accuracy was observed are shown in Fig. 6 and those classes where average accuracy was obtained are shown in Fig. 7. As per our observations, out of 116 classes in ImageCLEF, label prediction accuracy was $>$ 90% for 38 image classes, $>$ 60% for 70 classes and the remaining classes showed $<$ 40%. In intend to develop models that can address this disparity in the dataset is our future work. To evaluate retrieval performance, we used two popular IR metrics, *precision@k (p@k)* and *Mean Average Precision (MAP)*. Precision@k is given by the ratio of images that are retrieved in top k set that are actually relevant. It can be computed as per Eq. (8). The MAP of a set of testing images is defined as the mean of the average precision scores for each query image (Eq. (9)). MAP@k can also be computed accordingly, by considering precision scores at k.

$$p@k = \frac{\text{No. of retrieved images @k that are relevant}}{\text{Total images retrieved @k}} \quad (8)$$

$$MAP = \frac{\sum_{q=1}^{Q} AvgP(q)}{Q} \quad (9)$$

where, Q = Number of query/test images, q = 1,2,3, ... Q.

TABLE V: Evaluation of retrieval with precision@k for k=3,5,10

Test Image	k Value	Relevant images@ k	precision@k
Sample 1 *(Class-Cranium)*	k=3	3	100%
	k=5	5	100%
	k=10	10	100%
Sample 2 *(Class-Arm)*	k=3	3	100%
	k=5	5	100%
	k=10	9	90%
Sample 3 *(Class-Right Leg)*	k=3	3	100%
	k=5	5	100%
	k=10	9	90%
Sample 7 *(Class-Nose Area)*	k=3	3	100%
	k=5	5	100%
	k=10	8	80%
Sample 8 *(Class-Umbar Spine)*	k=3	3	100%
	k=5	5	100%
	k=10	8	80%

The image retrieval process was evaluated for various values of k, in order to conclusively evaluate the performance. We used k values of 3, 5 and 10 for this purpose. In case of real-world medical diagnostics applications, high precision during top-3 and top-5 retrieval is really important, as the doctor can get clear information w.r.t to the submitted image, hence we chose these values for a comprehensive evaluation. The results of our experiments for some sample image classes is presented in Table V and the MAP@k computed for all 116 classes is shown in Table VI respectively. From the results, it can be clearly observed that the top-3 and top-5 retrieval performance of the proposed approach is almost equal. This can also be seen in the MAP@k performance, as MAP@k was 87.06% for k=3, even at k=5, MAP@5 was about 86.01%, which means the performance of the proposed approach was excellent and indicates a well-balance performance. As k increases, it can be seen that the class imbalance problem comes to play, due to which performance degrades.

TABLE VI: Retrieval Evaluation for all 116 classes

k value	Observed MAP@k
k=3	87.06%
k=5	86.01%
k=10	83.91%

IV. Conclusion and Future Work

In this paper, a hybrid feature modeling based approach for content-based medical image retrieval was proposed. Most CBMIR models have restricted to a particular class or modality, we considered a large-scale ImageCLEF 2009 dataset consisting of x-ray images spanning 116 classes, which has a class imbalance issues. The experimental results showed that the proposed approach was very suitable for real-world medical image retrieval applications used for disease diagnosis and decision support, due to its excellent top-3 and top-5 retrieval performance. As part of future work, we plan to improve the hybrid feature representation further to deal with

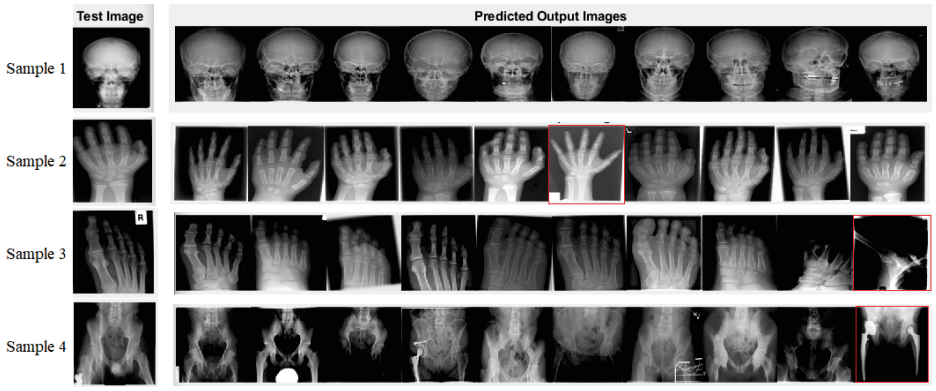


Fig. 6: Observed retrieval results (classes with high accuracy)

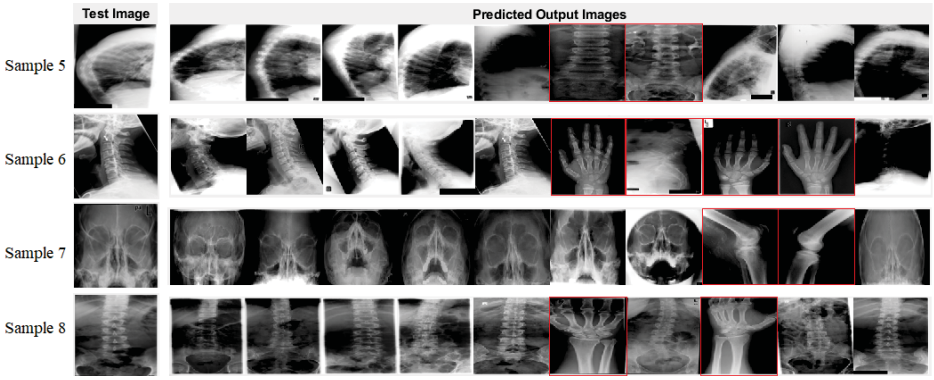


Fig. 7: Observed retrieval results (classes with average accuracy)

the class imbalance problem. Also, we intend to explore the suitability of deep neural networks for improving classification and retrieval.

ACKNOWLEDGMENT

The authors would like to gratefully acknowledge the Science and Engineering Research Board, Department of Science and Technology, Government of India for its financial support through Early Career Research Grant (ECR/2017/001056) and the facilities at Department of Information Technology, NITK Surathkal. We also thank Dr.T.M.Deserno, Department of Medical Informatics, RWTH Aachen, Germany for providing the dataset used in our experiments.

REFERENCES

[1] P. Mildenberger, M. Eichelberg, and E. Martin, "Introduction to the dicom standard," *European radiology*, vol. 12, no. 4, pp. 920–927, 2002.

[2] R. H. Choplin, "Picture archiving and communication systems: an overview." *Radiographics*, vol. 12, no. 1, pp. 127–129, 1992.

[3] J. Kim, W. Cai, D. Feng, and H. Wu, "Multidimensional medical data management: Volume of interest (voi)-based retrieval of medical images with visual and functional features," *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 3, pp. 598–607, 2006.

[4] T. Lehmann *et al.*, "The IRMA project: A state of the art report on content-based image retrieval in medical applications," in *Korea-Germany Workshop on Advanced Medical Image*, 2003, pp. 161–171.

[5] N. J. Fesharaki and H. Pourghassem, "Medical x-ray images classification based on shape features and bayesian rule," in *Computational Intelligence and Communication Networks (CICN), 2012 Fourth International Conference on*. IEEE, 2012, pp. 369–373.

[6] H. Pourghassem and S. Daneshvar, "A framework for medical image retrieval using merging-based classification with dependency probability-based relevance feedback," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 21, no. 3, pp. 882–896, 2013.

[7] M. Zare *et al.*, "Combined feature extraction on medical x-ray images," in *Computational Intelligence, Communication Systems and Networks (CICSyN), 2011 Intl. Conf. on*. IEEE, 2011, pp. 264–268.

[8] H. R. Tizhoosh, "Barcode annotations for medical image retrieval: A preliminary investigation," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 818–822.

[9] H. Hwang and R. A. Haddad, "Adaptive median filters: new algorithms and results," *IEEE Transactions on image processing*, vol. 4, no. 4, pp. 499–502, 1995.

[10] K. Zuiderveld, "Contrast limited adaptive histogram equalization," *Graphics gems*, pp. 474–485, 1994.

[11] R. Haralick *et al.*, "Textural features for image classification," *IEEE Transactions on systems, man & cybernetics*, no. 6, pp. 610–621, 1973.

[12] E. Ilunga-Mbuyamba *et al.*, "Automatic brain tumor tissue detection based on hierarchical centroid shape descriptor in tl-weighted mr images," in *Electronics, Communications and Computers, Intl. Conf. on*. IEEE, 2016, pp. 62–67.

[13] A. Sexton, A. Todman, and K. Woodward, "Font recognition using shape-based quad-tree and kd-tree decomposition," in *3rd International Conference on Computer Vision, Pattern Recognition and Image Processing*, 2000, pp. 212–215.

[14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, CVPR 2005*, vol. 1. IEEE, 2005, pp. 886–893.

[15] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.