

# Medical X-ray Image Hierarchical Classification Using a Merging and Splitting Scheme in Feature Space

Nooshin Jafari Fesharaki, Hossein Pourghassem

Department of Electrical Engineering, Najafabad Branch, Islamic Azad University, Isfahan, Iran

Submission: 24-01-2013 Accepted: 13-06-2013

## ABSTRACT

Due to the daily mass production and the widespread variation of medical X-ray images, it is necessary to classify these for searching and retrieving proposes, especially for content-based medical image retrieval systems. In this paper, a medical X-ray image hierarchical classification structure based on a novel merging and splitting scheme and using shape and texture features is proposed. In the first level of the proposed structure, to improve the classification performance, similar classes with regard to shape contents are grouped based on merging measures and shape features into the general overlapped classes. In the next levels of this structure, the overlapped classes split in smaller classes based on the classification performance of combination of shape and texture features or texture features only. Ultimately, in the last levels, this procedure is also continued forming all the classes, separately. Moreover, to optimize the feature vector in the proposed structure, we use orthogonal forward selection algorithm according to Mahalanobis class separability measure as a feature selection and reduction algorithm. In other words, according to the complexity and inter-class distance of each class, a sub-space of the feature space is selected in each level and then a supervised merging and splitting scheme is applied to form the hierarchical classification. The proposed structure is evaluated on a database consisting of 2158 medical X-ray images of 18 classes (IMAGECLEF 2005 database) and accuracy rate of 93.6% in the last level of the hierarchical structure for an 18-class classification problem is obtained.

**Key words:** Hierarchical classification, merging and splitting scheme, orthogonal forward selection, shape and texture features

## INTRODUCTION

Considerable development of multimedia data, including audio, video, text and image and also their extensive use in different aspects has led to increasing demands in archiving and retrieval tools.<sup>[1]</sup> The image and text databases are the most applicable databases among other multimedia databases in different applications such as medical, education, remote sensing and entertainment has led to the formation of the large databases of images.<sup>[2,3]</sup> Among different applications, extensive usage of medical images in research, training, medical education, diagnosing diseases and therapeutic plans require stronger search engines.<sup>[3]</sup> Hence, by development of data, the demand for managing and retrieval of medical images data has increased. The indexing of images used in the beginning to be done according to textual annotation of images while the techniques only based on texts provide limitations on image retrieval.<sup>[4]</sup> Manual annotation is a time consuming and subjective task due to human perception. Moreover, description of

many visual features in images such as irregular organic shapes is a very hard task contextually. Thus, retrieval of content-based medical images retrieval (CBMIR) as the response to the challenges for visual data management has been transformed in the recent years into the considerable field of research.

CBMIR systems use the classification of the image contents in the first stage as a pre-processing tool. A successful classification reduces the search space by indexing the images and by eliminating irrelevant images. In general, classification of images has an important role in searching a query image in medical databases. Automatic medical image classification is a technique for assigning a medical image to a class from among a number of pre-defined image categories that includes of three main steps: (1) representation; i.e., extraction of suitable features for describing the image contents, (2) adaptation; i.e., selecting the best subsystem from the set of features, (3) generalization; i.e., training and evaluation of classifiers.<sup>[5]</sup> Classification of X-ray images among other medical images has been

### Address for correspondence:

Dr. Hossein Pourghassem, Department of Electrical Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Isfahan, Iran.  
E-mail: h\_pourghassem@iaun.ac.ir

interested many researchers that could be due to its mass production and widespread application. For instance, more than 12,000 X-ray images were produced daily, in radiology section of Geneva University in 2002.<sup>[6]</sup>

Different algorithms have been presented for classification of the medical X-ray images. Although the variety and widespread production of medical X-ray images requires that classification is carried out with different categories, but it was limited to few categories before 2005. For example, Keyzers *et al.*<sup>[7]</sup> dealt with classifying the radiography images in six classes from Image retrieval in medical applications (IRMA) database including of 1,617 training images and 332 test images. In this research, the class labels were according to the body part examined, image modality and biological system. The classification was carried for the content-based retrieval aim, in which a kernel classifier was used with the criterion regarding distorted tangent distance and the error rate of 8% was reported. The classification of 851 medical images in eight classes with the error rate of less than 1% was also reported by Pinhas and Greenspan<sup>[8]</sup> in 2003, but considering this point is important that the low number of classes were not appropriate for medical purposes including evidence-based medicine and case-based reasoning. There has been more attention since 2005 on more details within the classes in a way that the images in each class were put in one group and thus the classifications were performed with more categories. Representing a general pattern of classification could be pointed in this regard, for categorizing 6,231 medical images into 81 groups according to the direction and modality of imaging by Lehmann *et al.*, where the accuracy rate of 85.5% was obtained.<sup>[9]</sup> In this method, encoding a large number of medical images was very time consuming. Rahman *et al.*<sup>[10]</sup> reported the accuracy rate of 81.96%, by analyzing the performance of a classification with 20 categories for 5,000 medical images in 2007. In addition, Pinhas and Greenspan reported the classification of 1,500 radiography images in 17 categories in the same year using the Gaussian Mixture Modeling-Kullback Leibler (GMM-KL) framework and 97.5% correctness was obtained.<sup>[11]</sup> This method was performed by extracting various features so that finally, a 37,500-feature vector was obtained for presenting each image. The classification problem was a complex task by using this high dimensional feature space. In this year, a classification scheme by extracting features in the local, global and pixel levels was also represented by Mueen *et al.* in which the performance of the classification was analyzed on IRMA database in 57 categories including 9,000 training images and 1,000 testing images and accuracy rate of 89% was obtained. In this method, manually labeling of the large set of training images was very time consuming and despite the reported overall accuracy rate, the accuracy rate for many categories was less than 50 or even zero.<sup>[12]</sup> In 2012, Mohammadi *et al.* was faced with a classification problem of 4,402 medical

X-ray images into 21 categories in which a combination of shape and texture features based on Gabor filter was used. The correctness of 88.7% was reported by using a support vector machine (SVM) classifier.<sup>[13]</sup> In addition, a classification scheme for categorizing of IRMA database consisting of 1169 medical X-ray images into 15 classes was represented by Ghofrani *et al.* in the same year. This method was performed by defining Gabor-based Centre Symmetric-Local Binary Pattern (GCS-LBP) features based on Gabor and using SVM classifier where the accuracy rate of 90.8% was obtained.<sup>[14]</sup>

In all mentioned works in above, only a single classifier had been used that made handling the overlapping between similar classes in the X-ray image database with difficulty and as a result, by increasing the number of classes, the complexity of the feature extraction stage and finally the proposed structure is observed. However in the recent previous years, the hierarchical techniques for medical image classification were used as an efficient method versus this problem. In 2010, Ray and Sasmal used an effective clustering method based on multi-level features consisting of global, local and pixel levels in order to cluster medical X-ray images by using a combination of the hierarchical techniques and K-means method.<sup>[15]</sup> This algorithm was evaluated on 150 X-ray images in 5 classes without using any dimension reduction technique, but its capability for large databases still has been remained as a challenge. A new hierarchical merging scheme for using in CBMIR systems was also presented by Pourghassem and Ghassemian<sup>[16]</sup> So that, in the first stage of the proposed structure, the homogenous classes were created from overlapping classes and then, this merging based classification was progressive to achieve all classes in two stages. The merging conditions of the research realized a supervised classification method and an unsupervised clustering technique. Furthermore, the conditions for selecting the used features in each group have had a lot of complexity.<sup>[16]</sup> Therefore, using single classifier in medical X-ray image classification usually goes along with complexity in some steps such as feature extraction by increasing the number of labeled classes. On the other hand, although significant works have been presented by using the hierarchical classification, but it seems the hierarchical classification methods in the previous works have also been applied for small databases or with complex structure for large databases.

In this paper, we try to develop a way for categorizing the large medical X-ray image databases in more classes by using the existing approaches and reducing the complexities surrounding this issue. For this purpose, first, by employing tools such as local adaptive histogram, edge detection algorithm together with threshold and also morphological operations, in addition to image enhancement, the boundary of the main object in each image is extracted. Then only by the shape features

and also using the orthogonal forward selection (OFS) method according to Mahalanobis separability measure, a subset of the extracted shape features is obtained for constructing the main classification in the first level of the proposed hierarchical classification. In other words, we use two introduced accuracy rate and miss-classified ratio measures in<sup>[16]</sup> and the tertiary difficult and unsuitable measure, i.e., dissimilarity measure in<sup>[16]</sup> is discarded due to computational complexity and its dependency of each class to optimal threshold determination. Instead, we used a powerful feature selection method (i.e., OFS method according to Mahalanobis separability measure) instead of used common feature selection method in<sup>[16]</sup> (i.e., feature forward selection method) along with a complete set of shape and texture features for each class based on its content. In the next levels, using the shape or texture features or a combination of them, each class splits into smaller classes and this procedure continues for obtaining all the primary formed classes. This scheme in the classification problem of many classes causes that to improve the classification performance; an expert classifier with a set of the selected features especially is assigned to a sub-space of the feature space. In other word, instead of assigning a single classifier to all feature space, a set of the classifiers are assigned to sub-spaces of the feature space based on the complexity or simplicity of the classification problem in each sub-space.

This paper is organized as follows. In Section 2, pre-processing stage is described. In Section 3, shape and texture feature extraction in different levels of classification are introduced. The OFS method and also the proposed structure of hierarchical classification based on merging and splitting of classes are presented in Section 4. In Section 5, experimental results are reported and then, a comparison between our proposed structure and other presented works is obtained in Section 6. Finally, Section 7 gives a conclusion to the work.

## PRE-PROCESSING

The medical X-ray images have specific features that create challenges in their processing and classification.<sup>[17]</sup> These characteristics are mainly including of intensity variations, low contrast<sup>[14]</sup> and high rate of noise in X-ray images. The important point is the brightness difference in the separation places of hard and soft-tissues that are not considerable in the medical X-ray images. Therefore, the contrast of these images is very low. On the other hand, the imaging by X-ray imposes the high rate of noise to the captured images. Thus, in the first stage of pre-processing, we should enhance the quality of these images by a combination of methods for reducing noise and improving the contrast. Then, for extracting the boundaries of bones in the images, some threshold techniques and morphologic functions are employed. These processes play a key role in the perfect and precise extraction of shape feature.

## Contrast Enhancement and Noise Reduction

A combination of techniques for reducing noise and improving contrast is applied for improving the quality of images. High frequency noise is among the characteristics of X-ray images that often low-pass filters are used for eliminating it to remain the main image data by elimination of the high frequencies. Hence, by applying a median filter for the first step, the noise resulted from digital imaging system is reduced. The median filter for elimination of noise in 2-dimensional signals is a very suitable method. To improve the contrast and uniform the overall intensity distribution, the intensity of each pixel is determined through the intensity of pixels in the neighbor of the respective pixel ( $3 \times 3$  square window). Furthermore, the grey level of all the image pixels is adjusted again by a way that the area between 0 and 255 is filled.<sup>[14]</sup> Later on, by employing tools such as adaptive histogram equalization, increasing the contrast would be possible on small areas of the image. For this purpose, first a window with definite dimensions is considered and slipped over the whole image. At each stage, the histogram of the pixels under the window is equalized and the pixels with new values are replaced with them. Equalizing the histogram could be performed in three ways: (1) Uniform distribution, (2) Rayleigh distribution and (3) exponential distribution. The best result in the present research is obtained by the respective tools and using the exponential distribution. It is important to note that equalizing of the histogram, although provides improvement of the image contrast with no destructive effects on the areas with higher contrast,<sup>[18]</sup> but it increases the noise on the image. As a result, after equalizing the histogram, a median filter (with  $5 \times 5$  square windows) is used again on the image to reduce the noise [Figure 1].

## Proposed Binarization Method

After improving the quality of images, the edge detection filters are used to highlight the separation boundary of hard and soft-tissues. Combing edge detection techniques such as Sobel and Laplacian on the images would be arrived at a



**Figure 1:** (a) The main image, (b) the result of contrast increase and noise reduction procedure (processed image)

better conclusion. Hence, by smoothing the image resulted from applying the Sobel filter and multiplying it to the processed image for reducing unwanted edges, the Laplacian filter will be applied on it and finally a coefficient ( $\alpha$ ) of the final image is added to it. The coefficient is obtained by experience for each class. This combination provides a more suitable image for the thresholding. For instance, the unwanted edges extracted from applying the Sobel edge detection are justified due to their sensitivity to noise in the combination state. After this stage, by defining an appropriate threshold, the binary image is obtained by showing the bone for the relative area. Ultimately, the obtained threshold is exercised by a proposed thresholding method. This method is according to the normalized histogram processing for each image [Figure 2]. The block-diagram of the proposed binarization algorithm is shown in Figure 3.

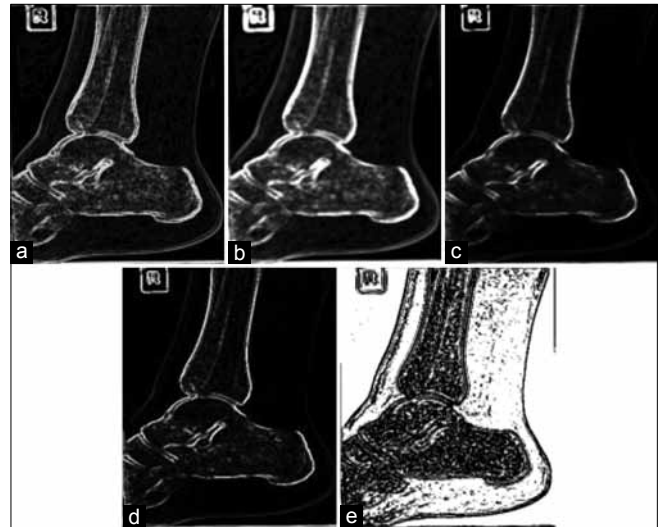
It is important to notice that by applying the thresholding method, in the most cases apart from the hard tissue, a part of the foreground including the boundary that distinguishes soft-tissues is also shown. Furthermore, some letters and symbols are used in some images to indicate the area of imaging in medical cases that has same grey levels as compared with the hard tissue and hence their boundary is revealed. Thus, after applying the thresholding method, morphologic tools are used to obtain the binary image, merely including the hard tissues.

### Morphological Operations

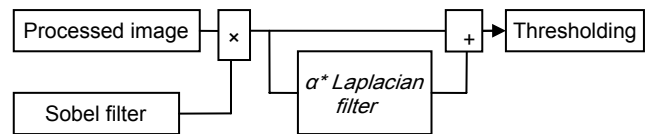
In addition to the revealed objects in the foreground during the proposed binarization method, the most of the X-ray images include symbols and letters that define the area of the imaging. The use of morphological operations provides the possibility of eliminating these symbols and objects. For this purpose, first by applying dilation operation, some pixels are added to the boundary of objects, for filling the boundary of objects [Figure 4a]. Then, the whole image is labeled according to the forming objects and the smallest bounding box in each area is computed. Since, the bones are shown as large objects in images, the most of the small objects (related to the foreground with the soft-tissue, signs and letters or the probable existing noise in the image) are eliminated by providing the threshold on the maximum length of the bounding boxes. For improving the boundary of the objects (hard tissue) and removing the redundant and small objects, the largest object is remained in the image and other objects are removed [Figure 4b and c]. Consequently, through applying a simple edge detection method, the exact boundary of the main object will be extracted [Figure 4d].

### FEATURE EXTRACTION

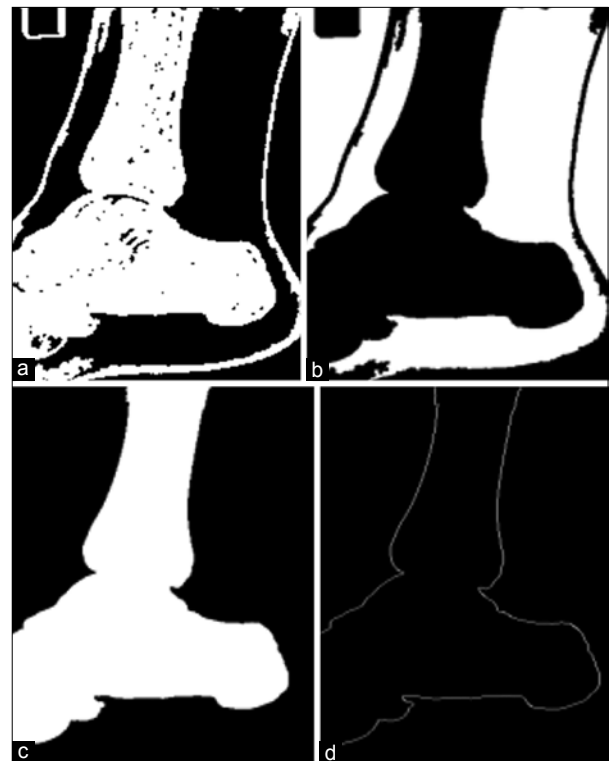
Feature extraction is referred to any science that obtains considerable and important components from images.



**Figure 2:** Preparing procedure of image for boundary extraction, (a) applying Sobel filter on processed image, (b) smoothing, (c) multiplying smoothed image and processed image, (d) the sum of obtained image in (c) and the filtered one by Laplacian filter, (e) the resulted image of thresholding method (binary image)



**Figure 3:** Block diagram for representing the combination of edge detection filters



**Figure 4:** (a) The result of applying the dilation operation on the obtained binary image. (b) and (c) The image after removing the small and redundant objects. (d) Boundary extraction of the object

In different subjects of image processing and pattern recognition, feature extraction is considered as a specific type of data reduction with the aim of finding a subset of informative variables from image data.<sup>[19]</sup> Clearly, the main key in the images classification is included within the process of feature extraction, where the calculated features are used for defining their contents.<sup>[15]</sup>

As a whole, process of feature extraction could be performed in three different levels of global, local and pixel. The features in the global level are extracted from the whole image while in local level; it is carried out by dividing the main image into smaller parts. Finally, in the pixel level, the simplest features of the image could directly be obtained according to the pixel amounts of the considered image. Moreover, the most systems use low level features for the medical X-ray image classification application,<sup>[15]</sup> i.e., the features that directly express the image contents by compressing data based on pixels amounts. These features are based on color, texture and shape. Since, the X-ray images are mainly in grey level; color could not thence be a suitable feature. Although texture could also be considered, but due to the existence of noise in the X-ray images, evaluation of the texture is not properly possible. In contrast, the forming and shape data is an important feature in the X-ray images.<sup>[4]</sup> In the first stage of the present research, the shape features are used and furthermore, with regard to the inter-class details, the texture features are also used.

## Shape Features

Efficient shape features for exhibiting the image contents should have some of the principle characteristics, including: Identifiable potentials, noise resistance, statistically-independence, reliability and translation, rotation and scale invariance.<sup>[20]</sup> In this research, the extracted shape features include simple geometric features such as axis of least inertia (ALI), eccentricity based on principle axes method and elongation regarding to minimum bounding rectangle, circularity ratio, ellipse variance. These features have quite simple calculations and are suitable to discriminate shapes with large differences, which are consisting of our purpose, especially in first levels of hierarchical classification. The other features are moments such as invariant moments (IM) and Zernike moments (ZM) that are usually concise, robust and easy to compute. The moments are invariant to scaling, rotation and translation of the object, too. Fourier descriptor (FD) is also used as a valid description tool because of its simplicity in computing, being robust to noise and compacting.<sup>[21]</sup> Then, all extracted features are reported through a feature vector. Prior to executing the feature extraction process, the enhanced images are re-sampled into  $K$  samples so that feature vectors with the same length are acquired from them.

### ALI

ALI is unique to the shape and shows the direction of the shape. This feature is defined as a line that the integral of the square distance of the points on the shape boundary is the minimum value. To obtain ALI, since it passes the center of gravity of the shape, the shape is transferred in such a way that the shape center of gravity is placed on the origin of Cartesian coordinates system. The parameter equation of ALI is shown in.<sup>[20]</sup>

### Eccentricity

This is obtained by the ratio of the major axis length to the minor axis length. The main axes of the shape are defined exclusively for each shape as two lines that intersect each other deliberately on the center of gravity of the shape and indicate the directions with zero cross correlation.<sup>[22]</sup> The shape contour in this method is observed as a sample of statistical distribution. By assuming covariance matrix for the contour, the lengths of the main two axes are considered equal to the eigenvalues of this matrix. The considered relations are presented in.<sup>[22]</sup>

### Elongation

Elongation is another concept based on eccentricity that shows the symmetrical rate of the shape, obtained by minimum bounding rectangular. This rectangle, also called the smallest bounding box, is the smallest rectangle consisting of all points on the shape. By calculating the length and width of the rectangle, the elongation criterion is calculated.<sup>[20]</sup>

### Circularity ratio

This feature, also called circularity variation, indicates how much the shape is close to a circular shape. By defining the standard deviation and the mean of the radial distance from the centroid of the shape to the boundary points, the parameter equations of circularity ratio are shown in.<sup>[23]</sup>

### Ellipse variance

Ellipse variance shows the mapping errors of shape for filling an ellipse by the covariance matrix, equal to the relevant shape.<sup>[20]</sup>

### IM

IM are also called geometric moments that include the simplest moment function. By calculating the geometric moment function, the geometric central moments are defined.<sup>[24]</sup> These moments are invariant with regards to rotations and translation, which require quite simple calculations.<sup>[24]</sup> These moments could also be normalized, for not to be subjected to variations with regards to scales. Based on the normalized central moments, a set of IM are obtained.<sup>[25]</sup> In this paper, 7 moments are applied.

### ZM

ZM are orthogonal moments. Compound ZM are resulted from orthogonal Zernike polynomials. By defining orthogonal radial

polygon, ZM are obtained for a binary image.<sup>[24]</sup> The important point in calculating ZM is that the coordinate space of the image should be transformed to the limit that the orthogonal polygon is defined. Furthermore, the calculation complexities of Zernike radial polygon are increased for larger degrees. In this paper, the real parts of 6 moments are calculated.<sup>[26]</sup>

### FD

As a whole, the FD is obtained by employing Fourier transform on a shape signature resulted from the shape boundary coordinates. The normalized coefficients of Fourier transform in this situation are called FD. The FD resulted from different signatures provide different significant performances. Centroid distance function and complex coordinate function could be defined as shape signatures for showing the boundary.

Centroid distance function is a function expressed by the distance of boundary points from the center of gravity of the shape and complex coordinates function is simply the complex number generated from the coordinates of boundary points. In this case, FD is obtained from Fourier function on centroid distance function and complex coordinates function which outperform FD derived from other shape signatures in terms of overall performance<sup>[20]</sup> and in the last, DC and the first frequency components are used to normalize the coefficients of Fourier transform.<sup>[20]</sup>

### Texture Features

Texture plays a significant role in image-related applications and is a key component for human visual perception. Accordingly, texture features have been extensively studied in the research area of image classification.<sup>[27]</sup> In this paper, despite the high rate of noise in the X-ray images, the texture features are only used for separating groups having definite texture, to smaller classes in later stages of classification. The utilized texture features include correlation, contrast, angular second moment and homogeneity extracted from grey level co-occurrence matrices (GLCM) and also statistical features (such as mean and standard deviation) of sub-bands from applying wavelet transform on the image.

### Computed Texture Measures From GLCM

GLCM method is one of the statistical techniques in analyzing the texture that was offered by Haralick<sup>[28]</sup> for estimating image properties related to second order statistics of image. In other words, GLCM is an estimate of a joint probability density function of grey level pairs in an image. GLCM could be expressed according to the following expression:

$$P_{d,\theta}(i, j), (i, j = 0, 1, \dots, N - 1) \quad (1)$$

Where  $i$  and  $j$  show the grey level of 2 pixels having the distance of  $d$  from each other and the angle  $\theta$  and  $N$  is the

number of grey levels in the image.<sup>[29-31]</sup> Once the matrices were formed, angular second moment, homogeneity, contrast and correlation are extracted from them.

Angular second moment is also known as energy or uniformity, expressing the repetition of pairs of pixel of an image. Contrast is a measurement resulted from local existing variations in the image. Correlation indicates the order of grey levels. In other words, it is represented the linearity of grey levels in relation to pixel pairs. Homogeneity is inversely proportional to the contrast in constant energy.<sup>[29]</sup> In our application, the co-occurrence matrices are calculated for four directions  $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$  and for  $d = 1$ . Hence, 16 features are totally obtained for each image.<sup>[30]</sup>

### Texture features from multi-resolution space

The methods based on multi-resolution filtering could be used for analyzing image textures. One of the most popular multi-resolution methods is Wavelet Transform that provides sub-bands statistics such as average and standard deviation as the texture features. The best result is obtained by extracting statistics in approximation band Low-Low-Low (LLL) of the wavelet coefficients in the most classes and in four sub-bands LLL, Low-Low-High (LLH), Low-High-Low (LHL) and Low-High-High (LHH) of the wavelet coefficient just for one category during the implementation of the proposed hierarchical classification structure.

## FEATURE SELECTION AND CLASSIFICATION

### Feature Selection

Relevant features are not predefined in the most cases in the real world. Thus, many of the features at the beginning are irrelevant or redundant for showing the complete image. Moreover, in large databases in the most cases, the learning process could not be performed before elimination of some unwanted features. Furthermore, discarding the some unwanted features reduces the running time of the algorithm, intensely. Therefore, the feature selection deals with selecting the smallest subset of the extracted features for the classification on the reduced spaces for better performance as compared to the main space.<sup>[32]</sup>

In this paper, the feature selection is carried out based on the OFS, where the feature subsets are evaluated according to Mahalanobis class separability measure. In the orthogonal decomposition-based methods like OFS method, the features are de-correlated in the orthogonal space and they can be evaluated and selected independently. Furthermore, the main purpose of these methods is redundancy reduction of the feature subsets.<sup>[32]</sup> Gram-Schmidt orthogonal transform could be used for selecting features, since the features of physical meaningless in the Gram-Schmidt space could

be referred to their equivalent variables in the measuring space.<sup>[33-35]</sup>

Assuming, there are  $N$  samples existing in the form of  $x(i)$ ,  $i = 1, 2, \dots, N$  where each sample is represented by a  $k$ -dimensional vector. The feature vector and feature matrix are defined as follows, respectively:

$$x_i = [x_i(1), \dots, x_i(N)]^T \quad (2)$$

$$X = \begin{bmatrix} x_1(1) & \dots & x_k(1) \\ \dots & \dots & \dots \\ x_1(N) & \dots & x_k(N) \end{bmatrix} \quad (3)$$

By defining  $R$  as the triangular matrix and  $Q$  as the orthogonal matrix, the feature matrix could be decomposed as follows:

$$X = QR \quad (4)$$

By assuming  $q_i$  as the new feature vector in the orthogonal space, matrix  $Q$  is calculated as follows by using Gram-Schmidt orthogonal transform:

$$q_1 = x_1 \quad (5)$$

$$q_i = x_i - \sum_{j=1}^{i-1} \alpha_{i,j} q_j \quad (6)$$

where,

$$\alpha_{ij} = \begin{cases} \frac{q_j^T x_i}{q_j^T q_j} & j = 1, \dots, i-1 \\ 1 & j = 1 \end{cases} \quad (7)$$

Quality of the feature subset could be evaluated according to its capabilities in creating the large class separation. Hence, in this paper, Mahalanobis class separability measure is used. By defining the mean vector of the samples in class ( $i$ ), in the form of  $m_i = [m_{i1}, \dots, m_{ik}]^T$  and covariance matrices of different classes of  $i$  and  $j$  diagonally in orthogonal space, according to the following relation:

$$c_{ij} = \text{diag}[\sigma_{ij}^2, \dots, \sigma_{kij}^2] \quad (8)$$

Mahalanobis class separability measure for multi-category problems could be used as the evaluation factor for the features subset, where  $L$  represents the number of categories:

$$J = \frac{2}{L(L-1)} \sum_{k=1}^n \sum_{i=1}^{L-1} \sum_{j=i+1}^L \frac{(m_{ki} - m_{kj})^2}{\sigma_{ij}^2} \quad (9)$$

Thus, the procedure of orthogonal forward feature selection is summarized as follows: First, all the variables  $x_i$ , ( $i = 1, \dots, n$ ) in the first stage are considered and Mahalanobis class separability measure is calculated for all of them. The variable that leads to maximum class

separability is identified and added to the features subset. Second, all the remaining variables are considered in the second stage and Mahalanobis class separability measure is calculated for them and the feature that creates the most of the class separability is identified and added to the subset. Third, this Procedure continues up to when the class separation by the next best feature subset becomes less than the pre-determined threshold.

## Medical X-ray Image Classification

### Classifiers

$K$ -nearest neighbor (KNN) classifier and multi-layer perceptron (MLP) neural network are used for classifying images. The basis of decision making for KNN classifier on  $K_n$  samples closer to unknown sample vector is according to the used distance measure, such class, which has more votes on the closer  $K_n$  samples could identify the considered class.<sup>[36]</sup> In the KNN classifier, two important parameters should be set. The optimal value of  $K_n$  and distance measure. We use Euclidean distance as distance measure and also the optimal values of  $K_n$  is determined based on the classification performance in the experimental results section.

The other utilized classifier is feed forward MLP neural network including of three layers (input, hidden and output layers) with the activity functions of tansig and hardlim in the hidden and output layers, respectively. The number of neurons of the input and output layers is equal to the number of features in the feature vector and the output classes, respectively. The number of hidden layer is also determined by trial and error way. The back propagation algorithm is used to train the MLP classifier.

### The proposed hierarchical classification structure

In general, by increasing the classes in the medical X-ray image classification, many of the classes (categories) are overlapped by the use of a single classifier and limited features and no appropriate results will be obtained. Hence, combination of some classifiers in a hierarchical structure and using suitable features set in each level are proposed as a strategy for solving this problem. In general, the hierarchical classifier could be either in agglomerative or divisive forms. In agglomerative method, each object is placed in a separate group and the groups that are rather similar are merged successively for the final condition to be achieved. While in divisive technique, all objects are considered to be in a similar group and split into smaller groups, consecutively for the final condition to be obtained.<sup>[15]</sup>

In this paper, we use a combination of both methods. First, the medical X-ray images are classified by using only the shape features due to similarity of images in the shape contents. Then, the overlapping classes are merged based

on the merging measures. Therefore, in the first level of the proposed hierarchical structure, a set of main classes that are consisted of the similar classes based on their shape contents, are formed. From the first through last levels of the proposed hierarchical classification structure, each main class (group) is divided by using the shape or texture features or combination of both and it continues until all primary classes are reconstructed again. Figure 5 shows the block diagram of the proposed hierarchical classification structure.

In the first level, the overlap classes (categories) are merged together based on the merging measures, to improve the performance of the classification. For this purpose, accuracy rate and miss-classified ratio are used.<sup>[16]</sup> The accuracy rate ( $A_i$ ) for class  $i$  is considered as,

$$A_i = \frac{\text{The number of images classified correctly to class } i}{\text{The total number of assigned images to class } i} \quad (10)$$

The miss-classified ratio ( $Mc_{ij}$ ) between two classes  $i$  and  $j$  is also defined as,

$$Mc_{ij} = \frac{\text{The number of images of class } i \text{ classified to class } j}{\text{The total number of assigned images to class } i \text{ and } j} \quad (11)$$

Hence, if  $A_i$  is less than a pre-determined threshold  $\delta$  ( $A_i < \delta$ ), the class  $i$  will have the potential for merging. In order to find the candidate class  $j$  for merging with class  $i$ , the miss-classified ratio of class  $j$  and class  $i$  ( $Mc_{ij}$ ) should be the highest value. The merged classes are considered as a group. After that, a classifier evaluates the test dataset on the merged categories again. In case that the total performance is lower than the pre-determined value ( $P_{pre}$ ), this procedure will be continued. The merging procedure is summarized as,

```

P = the total accuracy rate;
F = the total number of classes;
(1) for i = 1:F
    if  $A_i < \delta$ 
        for j = 1:F, j ≠ i
            calculate  $Mc_{ij}$ ;
        end;
        if  $Mc_{ij}$  is the highest;
    
```

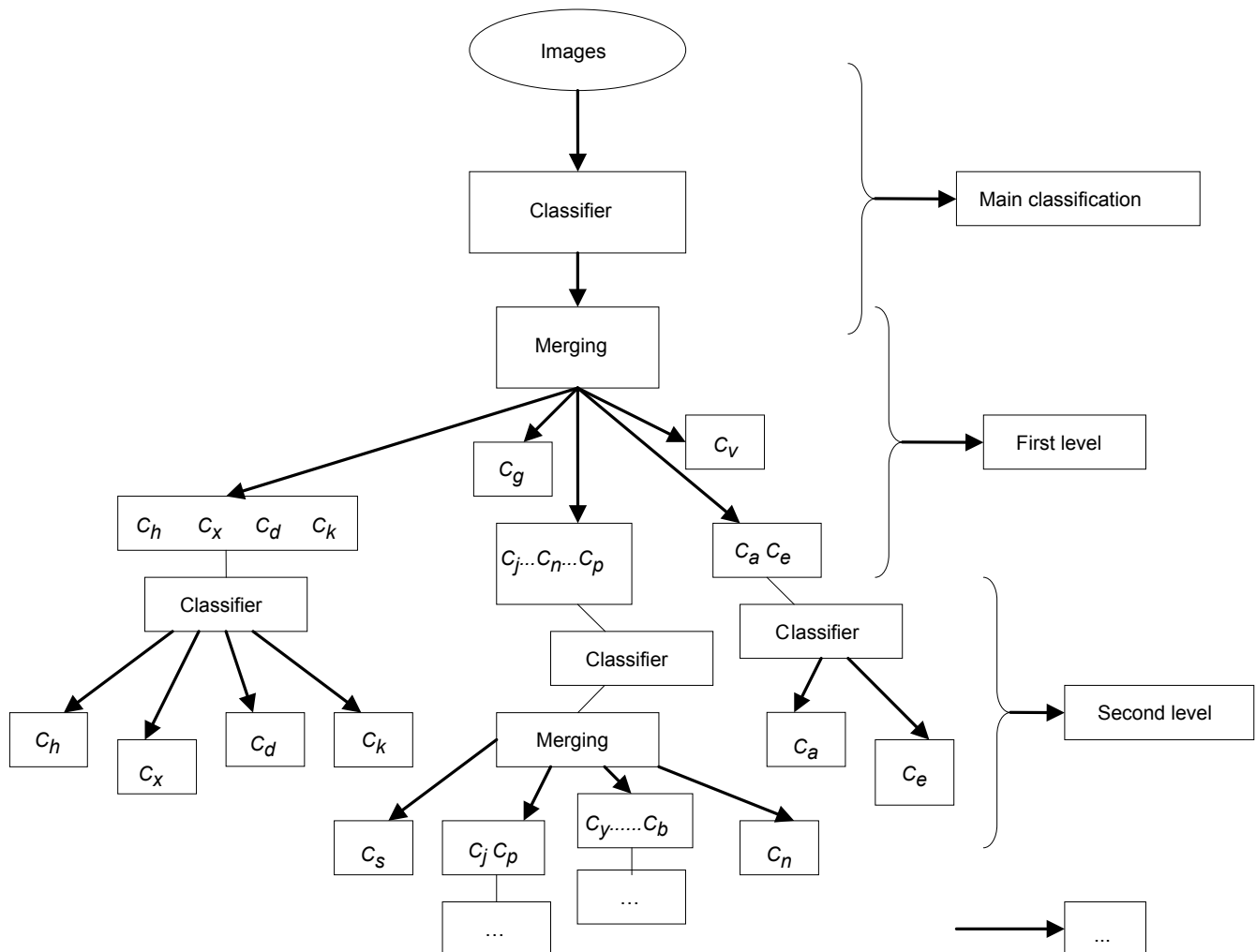


Figure 5: Block diagram for representing the proposed hierarchical classification structure



```

class i is merged with class j;
consider class i and j as a group;
end;
end;
end;
(2) if  $P < P_{pre}$ 
     $F =$  the total number of groups;
    go back to step 1;
else
    break
end;
end;

```

Then in the next levels, with regard to more attention to the inter-class details and using texture and shape features, each class is split into smaller categories. This procedure is carried out reversely; i.e., all the merged classes in a group are considered separately and the merging scheme is performed by determining the appropriate threshold again. Thus, by stopping the merging scheme, each group is split into smaller classes. This procedure continues inside each group for all the primary classes as splitting scheme.

## EXPERIMENTAL RESULTS

### Medical X-ray Images Database

A subset of IRMA dataset,<sup>[37]</sup> comprising 2,158 X-ray images in 18 different classes is used to evaluate our proposed structure. The classification of the images was performed according to the imaging orientation, biological system and anatomy region. The classes of X-ray images are shown in Table 1, together with details. We took two-third of the images as training dataset and one-third of the images as test dataset. Therefore, a total of 1,439 images for training and 719 images for the test are considered to evaluate our structure.

### Main Classification Based on Shape Features

At first, all images are resized to  $512 \times 512$ . Then after pre-processing the images, the extracted shape features provide a feature vector for each image and hence the dimensions of representation are reduced. The feature vector includes 786 features for each image [Table 2] that a subset of features is selected by applying OFS method according to Mahalanobis class separability measure, which includes 48 features for each image. Figure 6 shows class separability measure for different features. As shown in Figure 6, after an index in x-axis, the variation of the class separability measure changes negligibly. The experiences show this index can be defined as the pre-defined threshold for OFS method. Some of the feature indexes have also represented in Figure 6. Although the highest class separability is obtained by IM feature ( $F_7$ ), among

Table 1: Medical X-ray image classes

Class number	Anatomical region	Imaging orientation	System
1	Cranium	Coronal	Musculoskeletal
2	Spine	Coronal	Musculoskeletal
3	Hand	Coronal	Musculoskeletal
4	Radial carpal joint	Coronal	Musculoskeletal
5	Elbow	Coronal	Musculoskeletal
6	Foot	Coronal	Musculoskeletal
7	Ankle joint	Coronal	Musculoskeletal
8	Lower leg	Coronal	Musculoskeletal
9	Knee	Coronal	Musculoskeletal
10	Neuro cranium	Sagittal	Musculoskeletal
11	radial carpal joint	Sagittal	Musculoskeletal
12	Forearm	Sagittal	Musculoskeletal
13	Elbow	Sagittal	Musculoskeletal
14	Ankle joint	Sagittal	Musculoskeletal
15	Lower leg	Sagittal	Musculoskeletal
16	Right breast	Axial	Reproductive
17	Left breast	Axial	Reproductive
18	Knee	Axial	Musculoskeletal

Table 2: The extracted features in different levels of the hierarchical classification structure

Feature number	Feature name	Feature length
1	Axis of least inertia	1
2	Eccentricity	1
3	Elongation	1
4	Circularity ratio	1
5	Ellipse variance	1
6	Invariant moments	7
7	Zernike moments	6
8	Fourier descriptor with complex coordinate	498
9	Representation fourier descriptor with centroid distance representation	250
10	Computed texture features from GLCM matrices	16
11	Texture features from wavelet transform	2 (LLL), 6 (LLH, LHL, LHH)

GLCM – Grey level co-occurrence matrices; LLL – Low-Low-Low; LLH – Low-Low-High; LHL – Low-High-Low; LHH – Low-High-High

the different features, the simple geometric features such as ALI, Eccentricity etc., and ZM form the feature subset with higher class separability than the others ( $F_{14}$ ,  $F_{21}$ ,  $F_{15}$ ,  $F_{19}$ ,  $F_1, \dots$ ) and then IM and FD, especially Fourier Descriptor with complex coordinate representation (FCC) ( $F_8$ ,  $F_{12}$ ,  $F_{22}$ ,  $F_{24}, \dots$ ) are used to form the rest of the feature subset. After selecting the optimal features, the MLP classifier is trained and the accuracy rate of 44.73% is obtained on the test dataset. This performance is obtained by using 34 neurons in the hidden layer of MLP. Figure 7 shows the results of classification on 18 categories, using MLP classifier. Furthermore, k-nearest-neighbor classifier is used for classifying images in 18 classes by using shape features and the best result (the accuracy rate of 49.07%) is obtained by selecting  $k_n = 1$  [Figure 8].

## HIERARCHICAL CLASSIFICATION RESULTS

In the previous stage, by considering 18 different classes, the classification was applied on the optimal feature space. In this stage, the classification results of category merging in the first level of hierarchical classification structure are used. The threshold of accuracy rate measure is set to  $\delta = 60\%$  and then the class with the highest overlap with the considered class is selected to merge. For this purpose, by sorting the values of the miss-classified ratio, each class that has the highest value of the miss-classified ratio with the considered class is merged. The merging scheme continues until the total accuracy level reaches to 85%. in this research, the threshold of 60% for accuracy rate of each class is acceptable to be prevented from merging of a lot of classes and as a result, defining another features for progressing of the classification structure in more levels with more complexity by choosing a higher value and on the other hand, making it hard to reach to total accuracy by choosing a lower amount. According to the error rate of each level and the repeat of that in the other levels, the threshold of 85% for total accuracy level is also desirable. After applying the merging process on the classification

results, 4 classes with an accuracy rate of 88.2% are obtained by using the MLP classifier and 17 neurons in the hidden layer [Figure 9]. By applying the merging process, using KNN classifier with  $k_n = 3$  and by choosing the class 13 in different category, the less accuracy rate than 85% for the 4-category classification is resulted. The merging measures are obtained for the 3-category classification with the accuracy rate of 85.21%, which is less than the 4-category classification by using MLP classifier. The results of using KNN classifier are shown in Figure 10.

The merging results of classes in the first level of hierarchical classification are used for the next steps of structure. in the second level, with regard to more details inside the classes, 18 texture features including contrast, correlation, homogeneity and angular second moment by using GLCM matrices in four directions and for  $d = 1$  and also the mean and variance of one sub-band (LLL) (and 4 sub-bands (LLL, LHL, LHH, LLH in the lower level) of Wavelet coefficients are used for splitting each class to smaller categories by MLP classifiers. First, according to the results of the classification, all the merged ones are considered separately in one group and the merging process is implemented on them by determining the

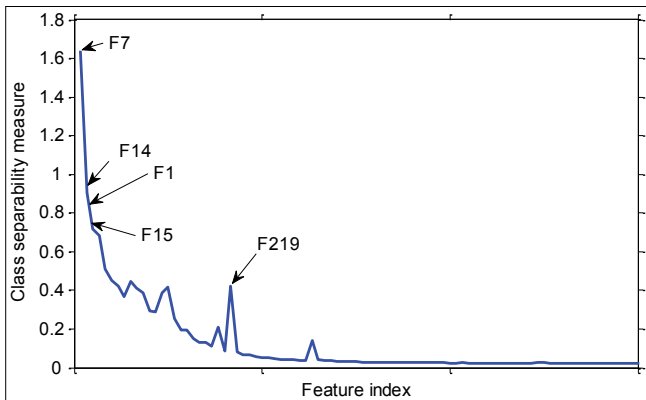


Figure 6: Class separability measure for different features (only some of the first feature indexes are shown accidentally)

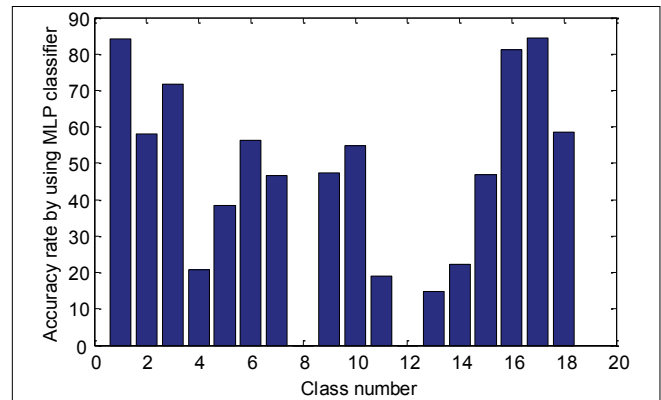


Figure 7: The results of classification by Multi-Layer Perceptron classifier (total accuracy rate of 44.73%)

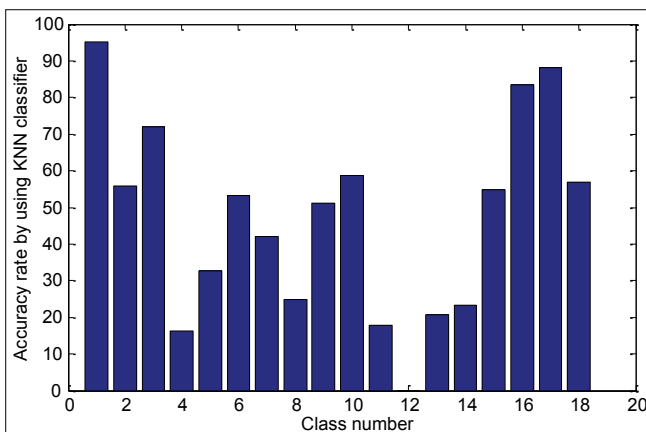


Figure 8: The results of classification by K-nearest neighbor classifier (total accuracy rate of 49.07%)

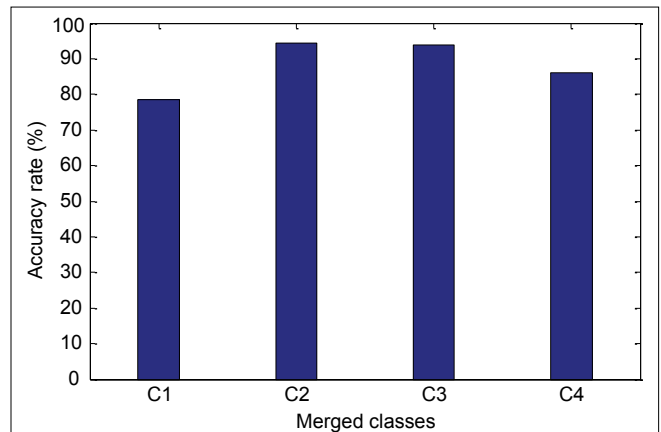


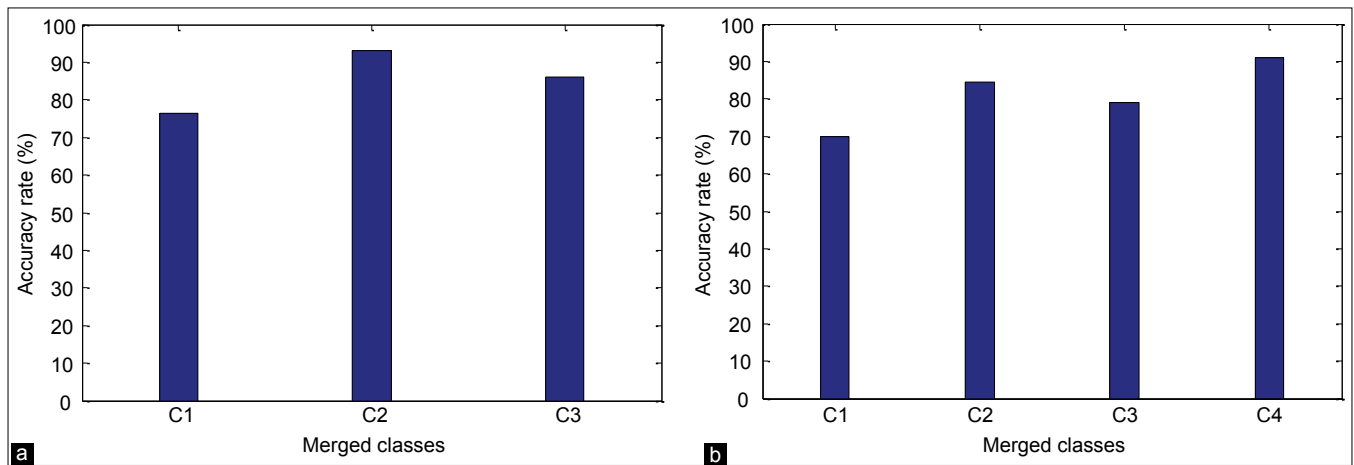
Figure 9: The result of the first level of hierarchical classification structure by Multi-Layer Perceptron classifier

appropriate threshold. Hence, by stopping merging process, each group is split into smaller classes. This procedure is continued up to form all the classes. Table 3 shows the results of inter-class merges, the number of used features and classification results.

### DISCUSSION AND COMPARISON WITH OTHER WORKS

The similarity of medical X-ray images in different classes and also great inter-class variability in some of the classes [Figure 11] cause the complexity of the used tools in some parts of the usual classification methods. Therefore, the hierarchical classification seems to be a good solution for this problem. The proposed merging and splitting scheme in the proposed hierarchical

classification structure and the used features and classifiers in different levels have a major effect on improvement of the reported classification performance. In the medical X-ray image classification application, the shape features can be very powerful. But, for employing them in the main classification, finding the exact boundary of bones in all images plays a significant role in improving the performance of the shape feature extraction. For this purpose, we tried to extract the boundary of bones by employing an appropriate pre-processing consisting of the contrast enhancement, the noise reduction, the precise binarization and also the morphological operations, exactly. Among of different extracted shape features, the most parts of the feature subsets are formed from the simple geometric features and ZM and then IM and FCC, respectively.



**Figure 10:** (a) The 3-category classification with the total accuracy rate of 85.21% by K-nearest neighbor (KNN) classifier and (b) the 4-category classification with the total accuracy rate of 81.15% by KNN classifier

**Table 3:** The hierarchical classification results

1 <sup>st</sup> level Class	2 <sup>nd</sup> level			3 <sup>rd</sup> level			4 <sup>th</sup> level			5 <sup>th</sup> level		
	Feature	Class	Accuracy (%)	Feature	Class	Accuracy (%)	Feature	Class	Accuracy (%)	Feature	Class	Accuracy (%)
1	43 shape features	1, 10	95.2	GLCM	1	100	GLCM	2	100			
2		2, 4, 13	88	W	10	97.4	W (LLL)	4	96			
4				(LLL)	13	79.3						
10				34 shape features	2,4	98.7						
13												
3	GLCM	3	98									
6	W (LLL)	6	97.5									
12		12	66.7									
5	48 shape features	8, 18	95.13	GLCM	8	100	GLCM	7	98.4	GLCM	5	95.45
7		5, 7, 9, 11, 14, 15	99	W (LLL)	18	100	W (LLL)	11	75	W (LLL)	14	82.6
8				GLCM	7, 11	84.27	GLCM	5, 14	89.84	GLCM	9	98.57
9				W (LLL)	5, 9, 14, 15	97.52	W	9, 15	90.06	W (LLL)	15	100
11												
14												
15												
18												
16	GLCM	16	100									
17	W (LLL)	17	100									

GLCM – Grey level co-occurrence matrices; LLL – Low-Low-Low

Although merging the classes by only the shape features and applying the defined merging measures decreases the total performance of classification in the first level of the hierarchical classification structure, but, this process is performed with less computational complexity in comparison with the past same works. Separating the global features in different levels of classification is another preference of our proposed structure, i.e., in the primary merging levels which the similarity of the images is considered, the shape features are only used and in the superior levels, the texture features become more highlight. Figure 12 shows the similarity of images in different categories including of the overlapped classes.

Furthermore, based on some items such as the desirable accuracy rate of user and the computational complexity level, some changes are applied to the feature sub-space and the number of levels of the hierarchical classification structure according to the selected classification parameters. Therefore, if the higher values of performance are desirable, the more computational complexity will add in the next levels and ultimately the hierarchical classification will be performed with more levels.

An exact comparison across schemes suggested in the literature is a complex task (especially as no standard dataset is available at this time). However, to evaluate results of our proposed structure, we obtain a comparison between our proposed structure and some other presented classification techniques in the literature. Several state-of-the-art classification techniques can be discussed at the present time. For example, Keyzers *et al.*<sup>[7]</sup> evaluated the IRMA database including 1617 training images and 332 images as test dataset into six classes by the aim of content-based image retrieval in 2003. In this research, a kerner classifier was used with the measure according to distorted tangent distance in which the accuracy rate of 92% was reported. However, the capability of this work for more labeled classes has remained as a challenge.<sup>[7]</sup> In 2007, Lehmann *et al.* represented a general pattern of classification for categorizing 6,231 medical images into 81 groups according to the direction and modality of imaging. In this method; although the accuracy rate of 85.5% was obtained, encoding a large number of medical images was very time consuming.<sup>[9]</sup> In the same year, Pinhas and Greenspan<sup>[8]</sup> dealt with a classification problem on 1,500 X-ray images into 17 classes. Although by employing a GMM-KL framework and using a multidimensional feature



**Figure 11:** Inter-class variability for three different classes, (a) elbow, (b) radial carpal joint and (c) hand



**Figure 12:** The similarity between the merged images in a, b and c categories

Table 4: Comparison between our work and other presented works

Algorithm	Database	Number of classes	Accuracy rate (%)
Keyser et al. 2003 <sup>[7]</sup>	IRMA (1617 training/332 test)	6	92
Lehmann et al. 2007 <sup>[9]</sup>	6231	80	85.5
Pinhas and Greenspan. 2007 <sup>[8]</sup>	1500	17	97.5
Mueen et al. 2007 <sup>[12]</sup>	IRMA (9000 training/1000 test)	57	89 (SVM), 82 (KNN)
Pourghassem and Ghassemian 2008 <sup>[16]</sup>	IRMA (9100)	40	90.83 (25 classes in 1 <sup>st</sup> level) and 94 (17 classes in 2 <sup>nd</sup> level)
Ray and Sasmal 2010 <sup>[15]</sup>	150	5	C1=72, C2=89, C3=87, C4=92, C5=81
Ghofrani et al. 2012 <sup>[14]</sup>	IRMA (1169)	15	90.88
Mohammadi et al. 2012 <sup>[13]</sup>	IRMA (4402)	21	88.77
Proposed structure	IRMA (2158)	18	93.6 (last level)

KNN – K-nearest neighbour; IRMA – Image retrieval in medical applications; SVM – Support vector machine

space, the performance of 97.5% was obtained, the high dimensional feature space and using no reduction technique so that a 37,500-feature vector for each image was used increased the complexity.<sup>[11]</sup> In this year, the work of Mueen *et al.*<sup>[12]</sup> can be also mentioned in which the multilevel features, namely, global, local and pixel were used SVM. They analyzed the performance of this classification scheme on the IRMA database with 57 categories consisting 9,000 training images and 1,000 test images and the correctness of 89% was obtained. In their work, the manual labeling of a large number of training images was time-consuming and although the high total accuracy rate, no correctness rate or less than 50% was reported for lots of classes.<sup>[12]</sup> Pourghassem and Ghassemian reported a hierarchical classification for using in CBMIR systems in 2008. The classification was progressive in two stages by using a supervised classification method and an unsupervised clustering technique in which the conditions for selecting the used features in each group were with complexity.<sup>[16]</sup> They obtained the accuracy rate of 90.83% on 25 merged classes in the first level and 94% on 17 merged classes on the second level for classifying 9,100 images in 40 groups. In 2010, Sasmal and Ray<sup>[15]</sup> represented a classification scheme according to a combination of K-means and hierarchical methods and by using multilevel features. In this research, only 150 radiography images in five classes had been used, which has remained the capability of this classification technique for a large database as a challenge.<sup>[15]</sup> In 2012, Mohammadi *et al.*<sup>[13]</sup> dealt with a classification problem consisting 4402 medical X-ray image in 21 classes in which a combination of shape-texture features based on Gabor filter was used. They reported the classification accuracy rate of 88.7% by using SVM classifier.<sup>[13]</sup> In the same year, Ghofrani *et al.*<sup>[14]</sup> presented a classification scheme for categorizing an IRMA database including 1,169 medical X-ray images into 15 classes. In this scheme, Gabor-based Centre Symmetric -Local Binary Pattern (GCS-LBP) features with SVM classifier had been used and the total correctness of 90.8% was obtained [Table 4].

## CONCLUSIONS

In this paper, a novel medical X-ray image hierarchical classification structure based on merging and splitting measures and by separating the feature space based on

shape and texture features was proposed. This structure was performed on a database consisting 2158 medical X-ray images in 18 classes. By the aim of improving the total classification correctness, we used only the shape features in the first level of our proposed structure to form the overlapped classes. In other word, according to the shape similarity of classes and by applying the supervised merging measures, the similar classes were grouped into the overlapped classes. Then, in the next levels, the combination of shape and texture features or texture features only were used to split the overlapped classes in smaller classes consecutively for forming all the classes, separately. We also used OFS algorithm according to Mahalanobis class separability measure as a feature selection method. By applying this algorithm, better results was obtained for classification performance. Eventually, the accuracy rate of 93.6% in the last level of the hierarchical structure for the 18-class classification problem was obtained. In the other word, in this work instead of assigning a single classifier to all feature space, a set of the classifiers were assigned to the sub-spaces of the feature space based on the complexity or simplicity of the classification problem in each sub-space. The provided results showed the effectiveness of the proposed structure.

## ACKNOWLEDGMENT

The authors would like to thank the IRMA Group, Aachen, Germany, for making the database available for the experiments.

## REFERENCES

1. Kyuheon K, Seyoon J, Byung TC, Jae Yeon L, Younglae B. Efficient video images retrieval by using local co-occurrence matrix texture features and normalized correlation. Proceeding of the IEEE Region 10 Conference, TENCON 99. Vol 2. 1999. p. 934-7.
2. Feng DD. Biomedical Information Technology. Massachusetts: Academic Press; 2008.
3. Smeulder M, Worring S, Santini A, Gupta A, Jain R. Content-based image retrieval at the end of the early years. IEEE Trans Pattern Anal Mach Intell 2000;22:1349-80.
4. Aggarwal P, Sardana HK, Jindal G. Content based medical image retrieval: Theory, gaps and future directions. Int J Graph Vision Image Process (ICGST-GVIP) 2009;9:1687-398X.
5. Jain K, Duin RP, Mao J. Statistical pattern recognition: A review. IEEE Trans Pattern Anal Mach Intell 2000;22:4-36.
6. Müller H, Michoux N, Bandon D, Geissbuhler A. A review of content-based image retrieval systems in medical applications-clinical

- benefits and future directions. *Int J Med Inform* 2004;73:1-23.
7. Keyzers D, Dahmen J, Ney H, Wein BB, Lehmann TM. Statistical framework for model-based image retrieval in medical applications. *J Electron Imaging* 2003;12:59-68.
  8. Pinhas A, Greenspan H. A continuous and probabilistic framework for medical image representation and categorization. *Proceeding of SPIE, PACS and Imaging Informatics 2003*;5371:230-8.
  9. Lehmann TM, Güld MO, Deselaers T, Keyzers D, Schubert H, Spitzer K, *et al.* Automatic categorization of medical images for content-based retrieval and data mining. *Comput Med Imaging Graph* 2005;29:143-55.
  10. Rahman MM, Bhattacharya P, Desai BC. A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback. *IEEE Trans Inf Technol Biomed* 2007;11:58-69.
  11. Greenspan H, Pinhas AT. Medical image categorization and retrieval for PACS using the GMM-KL framework. *IEEE Trans Inf Technol Biomed* 2007;11:190-202.
  12. Mueen, Baba MS, Zainuddin R. Multilevel feature extraction and X-ray image classification. *J Appl Sci* 2007;7:1224-9.
  13. Mohammadi M, Helfroush MS, Kazemi K. Novel shape-texture feature extraction for medical X-ray image classification. *Int J Innov Comput Inf Control* 2012;8:659-76.
  14. Ghofrani F, Helfroush M, Danyali H, Karimi K. Medical X-ray image classification using gabor-based CS-local binary patterns. *Int Conf Electron Biomed Eng Appl (ICEBEA)* 2012;284:8.
  15. Ray Ch, Sasmal K. A new approach for clustering of X-ray images. *J Comput Sci Issues* 2010;7:1694-0784.
  16. Pourghassem H, Ghassemian H. Content-based medical image classification using a new hierarchical merging scheme. *Comput Med Imaging Graph* 2008;32:651-61.
  17. Rui Y, Huang TS, Chang SF. Image retrieval: Current techniques, promising directions, and open issues. *J Vis Commun Image Represent* 1999;10:39-62.
  18. Sharifi S, Zaroug SA, Chester EG, Owen JP, Lee EJ. Bone edge detection in hand radiographic images. *Proceedings of the 16<sup>th</sup> IEEE International Conference on Engineering in Medicine and Biology Society, Vol 1.* 1994. p. 514-5.
  19. Egmont-Petersen M, de Redder D, Handels H. Image processing with neural networks – A review. *Pattern Recognit* 2002;35:2279-301.
  20. Yang M, Kpalma K, Ronsin J. A survey of shape feature extraction techniques. *Pattern Recognit* 2008;8:43-90.
  21. Tsai M, Chen MF. Object recognition by a linear weight classifier. *Pattern Recognition Lett* 1995;16:591-600.
  22. Peura M, Iivarinen J. Efficiency of simple shape descriptors. *Proc 3<sup>rd</sup> Int Workshop on Visual Form (IWVF3)*; 1997.
  23. Zhang D, Lu G. Review of shape representation and description techniques. *J Pattern Recognit* 2004;37:1-19.
  24. Celebi ME, Aslandogan YA. A comparative study of three moment-based shape descriptors. *Proc. International Conference Information Technology: Coding and Computing*; 2005. p. 788-93.
  25. Hu MK. Visual pattern recognition by moment invariants. *IRE Trans Inf Theory* 1962;8:179-87.
  26. Fesharaki NJ, Pourghassem H. Medical X-ray images classification based on shape features and Bayesian rule. *Fourth International Conference on Computational Intelligence and Communication Networks (CICN 2012)*. Mathura, Uttar Pradesh, India: IEEE; 2012. p. 369-73.
  27. Jian M, Liu L, Guo F. Texture image classification using perceptual texture features and Gabor wavelet features. *Conf Asia Pac Conf Inf Process* 2009;2:55-8.
  28. Haralick RM, Shamugan K, Dinstein I. Textural features for image classification. *IEEE Trans* 1973;3:610-21.
  29. Xian GM. An identification method of malignant and benign liver tumors from ultrasonography based on GLCM texture features and fuzzy SVM. *Int J Expert Syst Appl* 2010;37:6737-41.
  30. Fesharaki NJ, Pourghassem H. Medical X-ray image clustering using a new Gabor function-based image representation. *Int Rev Comput Softw* 2012;7:143-8.
  31. McNitt-Gray MF, Wyckoff N, Sayre JW, Goldin JG, Aberle DR. The effects of co-occurrence matrix based texture parameters on the classification of solitary pulmonary nodules imaged on computed tomography. *Comput Med Imaging Graph* 1999;23:339-48.
  32. Dash M, Liu H. Feature selection for classification. *J Intell Data Anal* 1997;1:131-56.
  33. Mao KZ. Orthogonal forward selection and backward elimination algorithms for feature subset selection. *IEEE Trans Syst Man Cybern B Cybern* 2004;34:629-34.
  34. Mao KZ. Fast orthogonal forward selection algorithm for feature subset selection. *IEEE Trans Neural Netw* 2002;13:1218-24.
  35. Colak S, Isik C. Feature subset selection for blood pressure classification using orthogonal forward selection. *29<sup>th</sup> IEEE Bioengineering Conference*; 2003. p. 122-3.
  36. Fukunaga K. *Introduction to Statistical Pattern Recognition*. 2<sup>nd</sup> ed. New York: Academic Press; 1990.
  37. IRMA Group, Available from: <http://www.irma-project.org>.

**How to cite this article:** Fesharaki NJ, Pourghassem H. Medical X-ray image hierarchical classification using a merging and splitting scheme in feature space. *J Med Sign Sens* 2012;3:150-63.

**Source of Support:** Nil, **Conflict of Interest:** None declared

## BIOGRAPHIES



**Nooshin Jafari Fesharaki** was born in 1986 in Isfahan, Iran. She received her B.Sc. degree in communications engineering from Department of Electrical Engineering of Najafabad Branch, Islamic Azad University, Isfahan, Iran, 2008, where she

is now a M.Sc. student of Communications Engineering. Her research interests are digital image processing especially medical ones, pattern recognition and content-based image retrieval.

**E-mail:** nj.fesharaki@yahoo.com



**Hossein Pourghassem** received his PhD in Biomedical Engineering from Tarbiat Modares University (TMU) in 2008, in Tehran, Iran. Since 2008, he has been with Department of Electrical Engineering, Najafabad Branch, Islamic Azad University

in Isfahan, Iran, where he is now an Assistant Professor at IAUN. His teaching and research interests are content-based image retrieval, biometrics, and pattern recognition, digital image processing and neural networks and has yet published many papers in international conferences and journals. He is a member of the machine vision and image processing (MVIP) society of Iran.

**E-mail:** h\_pourghassem@iaun.ac.ir