

Multi-Level of Feature Extraction and Classification for X-Ray Medical Image

M.M Abdulrazzaq¹, Imad FT Yaseen², SA Noah³, Moayad A. Fadhil⁴

^{1,2}Computer Science Department, International Islamic University Malaysia, IIUM, Malaysia

³Faculty of Information Science & Technology, Universiti Kebangsaan Malaysia, UKM, Malaysia

⁴Faculty of Information Technology, University of Philadelphia, Jordan

Article Info

Article history:

Received Jan 9, 2018

Revised Mar 2, 2018

Accepted Mar 18, 2018

Keywords:

X-ray medical image

SVM

K-NN

Feature Extraction

Classification

ABSTRACT

There has been a rise in demand for digitized medical images over the last two decades. Medical images' pivotal role in surgical planning is also an essential source of information for diseases and as medical reference as well as for the purpose of research and training. Therefore, effective techniques for medical image retrieval and classification are required to provide accurate search through substantial amount of images in a timely manner. Given the amount of images that are required to deal with, it is a non-viable practice to manually annotate these medical images. Additionally, retrieving and indexing them with image visual feature cannot capture high level of semantic concepts, which are necessary for accurate retrieval and effective classification of medical images. Therefore, an automatic mechanism is required to address these limitations. Addressing this, this study formulated an effective classification for X-ray medical images using different feature extractions and classification techniques. Specifically, this study proposed pertinent feature extraction algorithm for X-ray medical images and determined machine learning methods for automatic X-ray medical image classification. This study also evaluated different image features (chiefly global, local, and combined) and classifiers. Consequently, the obtained results from this study improved results obtained from previous related studies.

Copyright © 2018 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Mohammed Muayad Abdulrazzaq,

Computer Science Department,

Kulliyah of Information and Communication Technology (KICT),

International Islamic University Malaysia

E-mail: eng.alobaydee81@gmail.com

1. INTRODUCTION

The production and relatively straightforward management of digital visual content have been increasingly in demand over these recent years. Specifically in the medical domain for digital information, the continuous development of medical images such as X-ray, Computed Tomography (CT) scans, and Magnetic Resonance Image (MRI) scans contributed substantial amount of images daily. For example, the Department of Radiology in University Hospital of Geneva produced 12,000 to 15,000 images daily in 2002 [1]. The number of produced and stored images daily for this department continued to increase to 50,000 images in 2007 [2] and 114,000 images in 2009 [3]. Essentially, these images reveal critical information of visually inaccessible body parts, which are essential for medical diagnosis, medical education, and medical studies.

Therefore, effective techniques to navigate and search substantial amount of medical images accurately are necessary. The conventional image retrieval system depends on keyword search, in which the keywords or annotated image descriptions are manually assigned for indexing purpose. Subsequently,

relevant images are retrieved using this indexing system, which is known as Text Based Image Retrieval (TBIR). However, the TBIR method is disregarded due to the presence of thousands or even millions of image in the database. The process of entering metadata to each of these images is costly and time-consuming [4].

Consequently, rather than depending on TBIR, the Content Based Image Retrieval (CBIR) method is opted, in which the image retrieval process depends on features extracted from the image itself (the visual content of an image). Specifically, low-level features such as color, texture, and shape are considered as feature vectors, which are automatically extracted in the process of searching for specific images with respect to the query image. Accordingly, this technique is less time-consuming compared to the technique that depends on texts for the purposes of indexing and retrieving [5]. However, CBIR does not interpret data in the same way that a human does. Additionally, it is inexpedient for the system to elucidate high pixel images as how a human perceives images. Such limitation is known as semantic gap [6], which is specifically defined as the difference between how a human perceives an image based on a high-level semantic concept and how a computer classifies an image based on low-level features. Nevertheless, in practice, CBIR cannot be achieved based on only simple independent visual features. Various medical image classification methods using machine learning are developed to reduce the issues of semantic gap. With that, this study formulated an effective classification system for X-ray medical images based on multi-level feature extraction, feature reduction, and multi-classification techniques. The evaluation of this integration was performed using ImageCLEF2005 database. Attempts to utilize global or local features with either Support Vector Machine (SVM) classifier or k-Nearest Neighbor (k-NN) classifier for X-ray medical images were performed in various related studies, as summarized in Table 1 [7]-[9]. For this study, the evaluation was based on correctness rate. The correctness rate, as shown in Equation 1, is the result of dividing the number of correctly classified images by the total number of images.

$$\text{Correctness} = \frac{\text{Number of images classified correctly}}{\text{Total number of images}} \quad (1)$$

2. ANALYSIS AND PROPOSED SOLUTION

Realistically, it is a challenge to reduce the semantic gap because visual features of images do not present high-level semantic concepts and instead of utilizing the content of images, users opt for text-based query. With that, this has further instigated studies to develop effective medical image classification methods. However, the familiarization process with the semantic model in classifying images and enhancing the retrieval performance is complex.

Conversely, the results obtained from the previous studies to classify X-ray medical images, as shown in Table 1, utilizing global or local features with either SVM or k-NN classifiers are not regarded as the finest solutions to the issue of reducing the semantic gap. These results remained vary from one another. For example, referring to Table 1, RWTH-i6 team achieved error rate of 12.6% while Montreal team achieved error rate of 55.0% for the same dataset. Meanwhile, Mueen [10] combined feature extractions of global, local, and pixel for X-ray medical image classification and annotation using both SVM classifier and k-NN classifier. The resultant outcome of this combined feature extraction, consisting of 57 classes (ImageCLEF2005 database) revealed that the performance of SVM exceeded the performance of k-NN in most of the classes (specifically, 48 classes) while the performance of k-NN exceeded the performance of SVM in the remaining nine classes only. With that, SVM was considered for annotation purpose. Three hierarchical levels of image annotation were applied to reduce the semantic gap.

Apart from that, in another study on 4,937 X-ray medical images, Fesharaki & Pourghassem [11] achieved accuracy rate of 82.8% using feature extraction of shape and Bayesian classifier. Conversely, Ghofrani [12] achieved higher accuracy rate (90.8%) using feature extraction of shape and edges as well as SVM classifier on a dataset of 1,169 X-ray medical images. The accuracy rate increased to 94.2% with the integration of feature extraction of shape and texture and SVM classifier (rather than neural classification technique) on a dataset of 4,402 X-ray medical images [13]. Zare [14] utilized feature extractions of Gray Level Co-occurrence Matrix (GLCM), Canny, pixel, BoW, and LPBd as well as SVM and k-NN classifiers, where SVM achieved higher accurate rate (90%) based on ImageCLEF2007 database.

In conclusion, there is a need to utilize an effective classification that integrates multi-level feature extraction (global and local features) and multi-classification techniques for X-ray medical image classification.

Table 1. Summaries of related work

Author	Features	Classifier	Database	Semanticgap	Results
RWTH-i6	IDM with X32 thumbnails sizes, Sobel filter	1-NN	ImageCLEF2005	Not addressed	Error Rate 12.6%
RWTH-mi	CCF and IDM	1-NN	ImageCLEF2005	Not addressed	Error Rate 13.3%
Ulg.ac.be	16X16 randomly extracted patches from images.	Decision Tree	ImageCLEF2005	Not addressed	Error Rate 14.1%
Geneva-gift	Gabor Texture Filters	5-NN	ImageCLEF2005	Not addressed	Error Rate 20.6%
Infocomm	Texture features	SVM	ImageCLEF2005	Not addressed	Error Rate 20.6%
MIRACLE	weighting function for 20-NN	20-NN	ImageCLEF2005	Not addressed	Error Rate 21.4%
NTU	Gray values	1-NN 2-NN	ImageCLEF2005	Not addressed	Error Rate 21.7%
NCTU-DBLAB	Scaling	SVM	ImageCLEF2005	Not addressed	Error Rate 24.7%
CEA	Sobel filter	3-NN	ImageCLEF2005	Not addressed	Error Rate 36.9%
Mtholyoke	Tamura texture features and Gabor	k-NN	ImageCLEF2005	Not addressed	Error Rate 37.8%
CINDI	Canny Edge detector	SVM	ImageCLEF2005	Not addressed	Error Rate 45.3%
Montreal	Fourier shape and contour descriptors	-	ImageCLEF2005	Not addressed	Error Rate 55.7%
Qiu & Xu (2005)	Gray level texture and contrast	SVM	ImageCLEF2005	Not addressed	Accuracy: 89%
Mueen (2009)	Texture, Shape, Local & global features.	k-NN SVM	ImageCLEF2005	Addressed	Accuracy: 82% for k-NN, 89% for SVM
Fesharaki & Pourghassem (2012)	Shape features	Bayesian	4937 X-ray	Not addressed	82.87%
Ghofrani et al., (2012)	edges and shape	SVM	1169 X-ray	Not addressed	90.88 %
Mohammadi et al., (2012)	Shape and texture	SVM, Euclidean distance, and neural network	4402 X-ray	Not addressed	88.77 % 94.2 %
Fesharaki & Pourghassem (2013)	Shape and texture	k-NN and neural network	2158 X-ray	Addressed	93.6%
Zare et al., (2013)	GLCM, Canny Pixel, BoW, LPB	SVM k-NN	ImageCLEF2007	Not addressed	90% for SVM 86% for k-NN

3. METHODOLOGY

This present study proposed a framework to classify X-ray medical images based on multi-level feature extraction using the ImageCLEF2005 database. In this study, the development of the proposed framework was based on feature extraction, combination and selection, and classification, which are specifically discussed in the following sections.

3.1 Feature Extraction

This study extracted, combined, and utilized various features to explore different aspects of X-ray medical images. As presented in Table 1, several feature extractions were utilized, where global feature and local feature were considered in certain studies. Meanwhile, for this study, the following feature extraction algorithms were considered: (1) global feature, (2) local feature, (3) pixel feature, and (4) speeded up robust features (SURF).

In particular, global features were extracted from each image by applying feature techniques of shape and texture, which generated 282 features. These features included 130 dimensions of shape features and 152 dimensions of texture features. The local features, on the other hand, were extracted by segmenting the input image into four non-overlapping blocks of pixels, resulting to the extraction of 282 dimensions from each patch. The pixel feature was extracted after resizing each image to 15 x 15 pixels, which generated 225 features. SURF technique subsequently extracted 150 features from each image.

3.2 Texture Feature

Essentially, texture features refer to the underlying structural arrangement of the surfaces in the input image. There are two types of texture features, which are (1) Gray Level Co-occurrence Matrix (GLCM) and (2) Wavelet Transform (WT). GLCM was firstly introduced by [15]. It is mainly utilized to compute the second-order texture characteristics in solving the issues of categorization efficiently. For $N \times N$ image, it includes pixels with gray levels of 0, 1, 2, ..., $(G - 1)$ and represented by matrix $C(i, j)$, where each matrix element stands for the joint incidence of intensity levels i and j with prospects at a certain distance, d (which refers to the related distance between each pair of pixels and a related orientation angle) [16].

In order to obtain enhanced outputs, several co-occurrence matrices must be considered; one for each related location offers various texture features or similar features at various scales. Several texture measures of GLCM could be directly calculated [15] [17] [18] [19] [20] [21] [22] [23]. Generally, θ is quantized into four different directions: 0o, 45o, 90o, and 135o.

In this study, 22 co-occurrence matrices for each of these four directions were obtained, which included (1) autocorrelation, (2) cluster prominence, (3) cluster shade, (4) contrast, (5) correlation, (6) difference entropy, (7) difference variance, (8) dissimilarity, (9) energy, (10) entropy, (11) homogeneity, (12) inverse difference, (13) inverse difference normalized, (14) inverse difference moment normalized, (15) information measures of correlation 1, (16) information measures of correlation 2, (17) maximum probability, (18) maximum probability, (19) sum average, (20) sum entropy, (21) sum of squares, and (22) sum variance. Consequently, 88 dimensions were obtained. The following section reveals the notations used to describe the various features of GLCM and equations utilized for texture statistics in images, as shown in Equations 2 – 23 [15] [17] [18] [19] [20] [21] [22] [23].

$c(i, j) \equiv (i, j)^{th}$: it represents the “entry in a normalized GLCM”

$N_g \equiv$ it represents the gray levels number

$$c_x(i) \equiv \sum_{j=1}^{N_g} c(i, j) \quad c_y(j) \equiv \sum_{i=1}^{N_g} c(i, j)$$

$$c_{x+y}(k) \equiv \sum_{i, j | i+j=k} c(i, j) \quad \text{for } k=2,3,\dots,2N_g$$

$$c_{x-y}(k) \equiv \sum_{i, j | i-j=k} c(i, j) \quad \text{for } k=0,1,\dots,N_g - 1$$

$\mu \equiv$ Mean of $c(i, j)$

$\mu_x, \mu_y \equiv$ Mean of p_x and p_y respectively

$$\mu_{x-y} \equiv \sum_{i=0}^{N_g-1} i p_{x-y}(i)$$

$\sigma_x, \sigma_y \equiv$ They represent the “standard deviations of p_x and p_y ”, respectively”

$MH, MY \equiv$ They represent “the entropies of p_x and p_y respectively”

$$HXY1 \equiv -\sum_{ij} c(i, j) \log(c_x(i) c_y(j))$$

$$HXY2 \equiv -\sum_{ij} c_x(i) c_y(j) \log(c_x(i) c_y(j))$$

The following equations were utilized to compute the presented twenty-two texture statistics:

$$\text{Energy} = \sum_i \sum_j c^2(i, j) \tag{2}$$

$$\text{Entropy} = \sum_i \sum_j c(i, j) \log c(i, j) \tag{3}$$

$$\text{Dissimilarity} = \sum_i \sum_j |i - j| \cdot c(i, j) \tag{4}$$

$$\text{Contrast} = \sum_i \sum_j (i-j)^2 \cdot c(i, j) \quad (5)$$

$$\text{Correlation} = \frac{\sum_i \sum_j (i-\mu_x)(j-\mu_y) c(i, j)}{\sigma_x \sigma_y} \quad (6)$$

$$\text{Homogeneity} = \sum_i \sum_j \frac{c(i, j)}{1+|i-j|} \quad (7)$$

$$\text{Autocorrelation} = \sum_i \sum_j (ij) c(i, j) \quad (8)$$

$$\text{Cluster Shade} = \sum_i \sum_j (i+j-\mu_x-\mu_y)^3 c(i, j) \quad (9)$$

$$\text{Cluster Prominence} = \sum_i \sum_j (i+j-\mu_x-\mu_y)^4 c(i, j) \quad (10)$$

$$\text{Maximum Probability} = \max(i, j) c(i, j) \quad (11)$$

$$\text{Sum of Squares} = \sum_i \sum_j (i-\mu)^2 c(i, j) \quad (12)$$

$$\text{Sum Average} = \sum_{i=0}^{2G-2} i c_{x+y}(i) \quad (13)$$

$$\text{Sum Variance} = \sum_{i=0}^{2N_g-2} (i - \text{Sum Average})^2 c_{x+y}(i) \quad (14)$$

$$\text{Sum Entropy} = \sum_{i=0}^{2G-2} c_{x+y}(i) \log(c_{x+y}(i)) \quad (15)$$

$$\text{Difference variance} = \sum_{i=0} (i - \mu_{x-y})^2 c_{x-y}(i) \quad (16)$$

$$\text{Difference Entropy} = - \sum_i c_{x+y}(i) \log(c_{x+y}(i)) \quad (17)$$

$$\text{Information Measures of Correlation 1} = \frac{\text{Entropy} - HXY1}{\max\{HX, HY\}} \quad (18)$$

$$\text{Information Measures of Correlation 2} = \sqrt{1 - \exp[-2(HXY2 - \text{Entropy})]} \quad (19)$$

$$\text{Maximum Correlation Coefficient} = (\text{second largest eigen value of } Q) 0.5 \quad (20)$$

$$\text{Where } Q(i, j) = \sum_k \frac{c(i, k) c(j, k)}{c_x(i) c_y(k)} \quad (21)$$

$$\text{Inverse Difference Normalized} = \sum_i \sum_j \frac{c(i, j)}{1+|i-j|} \quad (22)$$

$$\text{Inverse Difference Moment Normalized} = \sum_i \sum_j \frac{c(i, j)}{1+(i-j)^2} \tag{23}$$

Meanwhile, one of the most commonly used methods for multi-resolution image description and analysis is the WT. It specifically offers an efficient set of tools for various applications such as compression of images or signals, detection of objects, improvement of images, and noise removal. Wavelets are functions, satisfying a linear combination of various conversion and scaling processes of a wave function. It utilizes wavelet transform, specifically the Haar wavelet, to extract texture feature. This first known wavelet is considered as the simplest wavelet basis, which was utilized for orthonormal wavelet transform with compact support [24]. Equation 24 represents the Haar function equation using a step function, ψ unc.

$$\psi \text{ as } \begin{cases} 1 & 0 \leq t < \frac{1}{2}, \\ -1 & \frac{1}{2} \leq t < 1, \\ 0 & \text{otherwisw} \end{cases} \tag{24}$$

The Haar wavelet was applied in this study since it is the most efficient technique to calculate the feature vector [25]. This was performed by applying the Haar wavelet for four times in order to divide the input image into 16 sub-images, as illustrated in Figure 2.

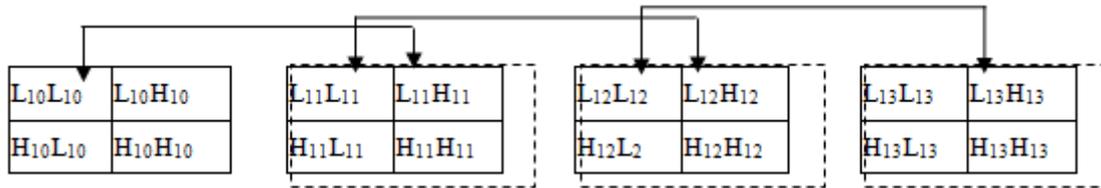


Figure 2. Applying Haar Wavelet Four Times

Each image I of the size of $r \times r$ was initially resized into 100×100 pixels. The Haar wavelet was subsequently applied to each image before dividing it into four sub-images, where each image has $(r \times r)/2$ size - $L10L10$, $L10H10$, $H10L10$, and $H10H10$. In the sub-image of $L10L10$, low frequencies were present in both horizontal direction and vertical direction. Specifically, low frequencies were present in the horizontal direction while high frequencies were present in the vertical direction. However, in the sub-image of $H10L10$, high frequencies were present in the horizontal direction while low frequencies were present in the vertical direction. and in the $H10H10$ sub-image, there are high frequencies in both directions. Following that, the Haar wavelet was applied on the image of $L10L10$ with the size of $(r \times r)/2$ to obtain four new sub-images, where each image has $(r \times r)/4$ size - $L11L11$, $L11H11$, $H11L11$, and $H11H11$. Similar process was repeated twice to obtain sub-images of $(r \times r)/8$ and $(r \times r)/16$, respectively, as illustrated in Figure 2. Additionally, four features were computed for each of the presented four procedures, which are (1) entropy, (2) energy, (3) mean, and (4) standard deviation. With that, there were 64 features computed from all sub-images. Figure 3 illustrates a sample image used as an input for the Haar WT, which was obtained from the ImageCLEF2005 database.



Figure 3. Image sample from ImageCLEF2005

The obtained sub-images after applying the second, third, and fourth Haar wavelet are depicted in Figure 4, based on the aforementioned. Haar wavelet was applied four times in order to divide the input image into 16 sub-images to get the most information about the image, applying Haar wavelet for the five time gets the sub images L13L13 equal to zero, therefore was applied only four times.

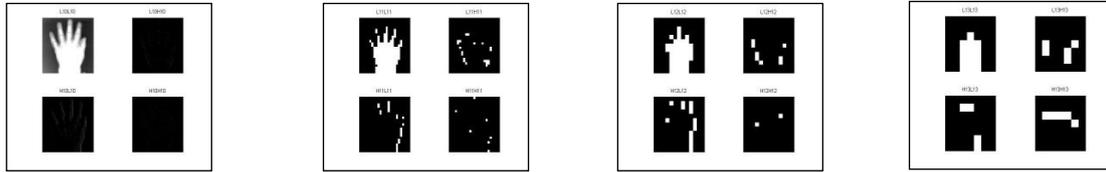


Figure 4. The obtained sub-images after applying the second, third, and fourth Haar wavelet

3.3 Shape Feature

The shape feature offers geometrical information concerning an image object, which does not vary with the variations in the orientation, scale, and location. For this process, the shape information of an image was explored based on edges. Thus, the histogram of edge techniques and SURF technique were applied in this study to extract the shape feature of images. Histogram of edge was utilized to explore the shape feature for each image. In particular, both gradient histogram and edge orientation histogram were applied. The first edge histogram technique was utilized to extract 50 features from each image while the second edge histogram technique was utilized with a Canny filter to extract 80 features from each image [26].

The SURF technique has a scale and rotation invariance property, which facilitates object identification with no regards to the image's resize or representation of rotation around a certain axis [27]. Realistically, variance occurs because not all information could be captured from a specific recording. Invariance is an essential property of image since the similarity measurement is probable based on the feature between two images that cannot be duplicated. Thus, the SURF technique was applied to extract 150 features from each image.

3.4 Combination and Selection

Combined feature refers to the combination of global feature, local feature, pixel feature, and SURF into one vector. Figure 5 depicts the overall process of feature extraction as well as combination and selection. In order to extract pixel features, images were resized to 15 x 15, which contributed a vector of 225 pixel features. The global features refer to the features of shape and texture, which were extracted from the whole image; thus the resultant outcome of this combined vector was 282 features, specifically 130 features from the edge histogram, 64 features from the WT and 88 features from the GLCM. Conversely, the local features were extracted by segmenting the image into four non-overlap patches, which shared similar 282 features. This led to 1,128 features, combined in one local feature vector. Meanwhile, 150 features were obtained for the SURF.

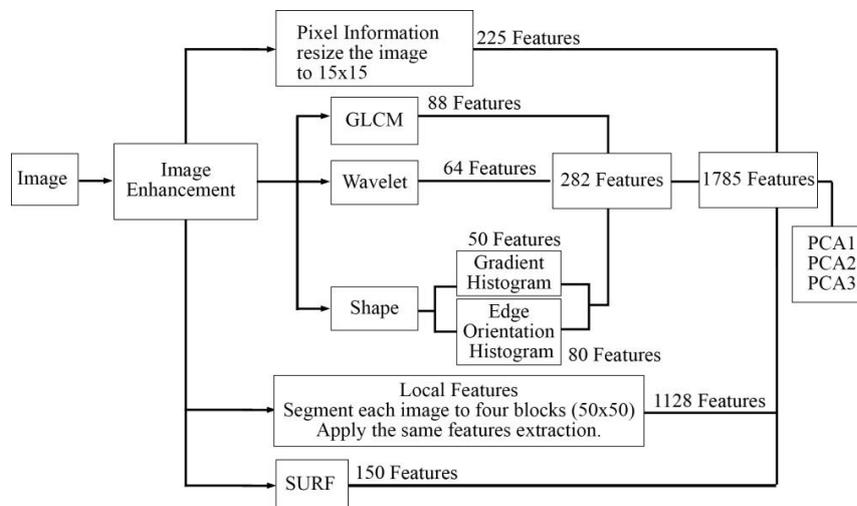


Figure 5. Feature extraction, combination and selection

As a result, the overall feature vector dimensionality for each image equals to 1,785 feature vectors. Given the substantial number of feature vectors involved, a certain dimensionality reduction technique must be performed to decrease the feature vectors. The most commonly used dimensionality reduction technique is the principal component analysis (PCA) [28]. This simple technique effectively decreases the dimensionality of data. With the application of this technique, the feature vectors were reduced from 1,785 into 25, 50, and 100 one to study and choose the optimal precision outcomes.

3.5 Classification

The classification of image is presented as the main aspect of this present study with respect to the objectives of this study. Four distinct features were initially extracted from the input image, which were global feature, local feature, pixel feature, and SURF. Following that, these extracted features were combined into one feature vector. PCA was subsequently performed to decrease the dimensionality of feature vectors. The developed image classification system from this study was evaluated using the ImageCLEF2005 database [29]. This database was segmented into training set and testing set. The training set was categorized into 57 known classes, which were pre-defined.

4. EVALUATION

A series of experiments was conducted to evaluate the performance of the proposed method in this study. In particular, this is to validate the proposed method and its significance for X-ray medical image. The implementation of the proposed method included feature extraction, feature combination and reduction, and X-ray medical image classification using SMV and k-NN classifiers, which were evaluated to determine its performance based on the results of accuracy rate. As a result, four experiments were conducted. The specific methods and settings of these experiments are described and obtained results in this study are presented and discussed in the following sub-sections.

Essentially, the ImageCLEF2005 database was utilized in this study [29], which contained 10,000 X-ray images, divided into 9,000 training images and 1,000 testing images. These images were in gray scale with different resolutions, which were obtained using different imaging techniques. There were 57 classes, containing different number of sample images. During the evaluation stage, the training dataset was randomly partitioned into two sets. The first dataset was divided into 80% of training images and 20% of testing images while the second dataset was divided into 90% of training images and 10% of testing images. This ensured that each class contained training images and corresponding test images. The first dataset was selected to compare with the obtained results from previous related studies while the second dataset was selected in that specific ratio that it was similar to the ImageCLEF2005 database, where it contained similar percentages of both training set (90%) and testing set (10%). The number of training images and test images for each class in these two datasets was tabled in Table 3 (80:20) and Table 4 (90:10) (mentioned in columns with fine dashed style). It should be noted that the sequence refers to the class number.

Table 3. Number of images, 80% training and 20% testing

Class	No. Of Images	80%	20%	Class	No. Of Images	80%	20%	Class	No. Of Images	80%	20%
1	336	269	67	20	31	25	6	39	38	38	8
2	32	26	6	21	194	155	39	40	51	41	10
3	215	172	43	22	48	38	10	41	65	52	13
4	102	82	20	23	79	63	16	42	74	59	15
5	225	180	45	24	17	14	3	43	98	78	20
6	576	461	115	25	284	230	57	44	193	154	39
7	77	62	15	26	170	136	34	45	35	28	7
8	48	38	10	27	109	87	22	46	30	24	6
9	69	55	14	28	228	182	46	47	147	118	29
10	32	26	6	29	86	69	17	48	79	63	16
11	108	86	22	30	59	47	12	49	78	62	16
12	2563	2050	513	31	60	48	12	50	91	73	18
13	93	74	19	32	78	62	16	51	9	7	2
14	152	122	30	33	62	50	12	52	9	7	2
15	15	12	3	34	880	704	176	53	15	12	3
16	23	18	5	35	18	14	4	54	46	37	9
17	217	174	43	36	94	75	19	55	10	8	2
18	205	164	41	37	22	18	4	56	15	12	3
19	137	110	27	38	116	93	23	57	57	46	11

Table 4. Number of images, 90% training and 10% testing

Class	No. Of Images	90%	10%	Class	No. Of Images	90%	10%	Class	No. Of Images	90%	10%
1	336	302	34	20	31	28	3	39	38	34	4
2	32	29	3	21	194	175	19	40	51	46	5
3	215	193	22	22	48	43	5	41	65	58	7
4	102	92	10	23	79	71	8	42	74	67	7
5	225	202	23	24	17	15	2	43	98	88	10
6	576	518	58	25	284	256	28	44	193	174	19
7	77	69	8	26	170	153	17	45	35	31	4
8	48	43	5	27	109	98	11	46	30	27	3
9	69	62	7	28	228	205	23	47	147	132	15
10	32	29	3	29	86	77	9	48	79	71	8
11	108	97	11	30	59	53	6	49	78	70	8
12	2563	2307	256	31	60	54	6	50	91	82	9
13	93	84	9	32	78	70	8	51	9	8	1
14	152	137	15	33	62	56	6	52	9	8	1
15	15	13	2	34	880	792	88	53	15	13	2
16	23	21	2	35	18	16	2	54	46	41	5
17	217	195	22	36	94	85	9	55	10	9	1
18	205	184	21	37	22	20	2	56	15	13	2
19	137	123	14	38	116	104	12	57	57	51	6

4.1 Experiment 1 - Feature Reduction

This experiment was conducted to evaluate and investigate the performance of the our proposed system after reducing the number of features using PCA, which was essential to determine the accuracy rate of the system. Feature vectors were reduced from 1,785 features into 25 features, 50 features, and 100 features, which was termed as PC1, PC2 and PC3, respectively. The results of accuracy rate with optimal feature reduction were obtained with and without a threshold for both datasets containing specific ratio of training set and testing set (80:20; 90:10) and were subsequently compared among these obtained values.

The PCA is considered competent and effective in reducing the dimensionality of data. Both SVM (with RBF kernel) and k-NN (with $k = 1$) were employed for each evaluation stage in this experiment. Table 5 reveals the obtained results using k-NN classifier while Table 6 reveals the obtained results using SVM classifier. Consequently, PC2 (50 features) achieved the highest accuracy rate. Specifically, PC2 obtained the highest percentages, with and without the threshold using both classifiers. Based on this experiment, PC2 was considered for the subsequent experiments. It should be noted that undeniably, it is essential to have sufficient number of features for discrimination and for high accuracy rate. Having few features might lead to low accuracy rate and inadequate number of features subsequently affects the discrimination among the features of other images. Nevertheless, high accuracy rate is not warranted with high number of features due to the high occurrence of common features, which affects the discrimination among the features of other images as well. Consequently, PC2 was proven to achieve the highest accuracy rate, rather than PC1 (25 features) and PC3 (100 features).

Table 5. Accuracy results for PCA by using k-NN

	80%-20%	80%-20%	90%-10%	90%-10%
	With threshold	Without threshold	With threshold	Without threshold
PCA 1	60.952	85.901	84.853	91.620
PCA 2	61.118	87.011	87.881	92.138
PCA 3	60.788	86.503	86.036	91.731

Table 6. Accuracy results for PCA by using SVM

	80%-20%	80%-20%	90%-10%	90%-10%
	With threshold	Without threshold	With threshold	Without threshold
PCA 1	90.047	91.810	89.899	94.201
PCA 2	90.450	92.202	91.924	95.368
PCA 3	90.290	92.066	90.761	95.122

4.2 Experiment 2 - Feature Combination

This experiment aimed to investigate the performance of a single feature extraction from each of the four features (global feature, local feature, pixel feature, and SURF) and the combination of these four feature extractions. The resultant outcome of this experiment was crucial in terms of accuracy rate and indexing. These results were compared with those of related previous studies in term of feature sets. Most of

the medical content-based image retrieval systems utilize global features. The main advantage of global features is the computation speed where the feature extraction and matching similarity are computationally faster. However, they may fail to identify pertinent visual characteristics. The classification process of global features includes two phases, which are training and testing. In the training phase of this study, global features were extracted from all training images and the classifier was subsequently trained on these extracted features to create a model. In order to classify the test images, features were initially extracted in the same way as in the training phase. The model was then utilized to classify test images.

However, local features are inherently robust against translation. In this experiment, local features were extracted from four square images, which were taken from original ones after dividing the image into four blocks. Similar classification process that was applied for global features was subsequently applied for local features, except that local features were extracted from each sub-image. The pixel value comparison is also an effective approach to seek similar images in the database. For most applications, this approach is not feasible because the difference between the pixels of one image to another is not evident. However, it is feasible for the pixel value comparison to identify only one specific object of equal size and located at similar position (similar row and column of an image matrix) between images with small resolutions. For this study, Experiment 2 also utilized pixel information.

The SURF, a descriptor feature, is also a scale and rotation invariant detector. The scale and rotation invariance denotes that an object could be identified even when it is scaled in size or rotated. The SURF was applied in this experiment as well, but it was not utilized as one of the local features given that the extraction of these features is a time-consuming process. In the training phase, all images were resized to 100 x 100 pixels, where the resultant large feature vector containing 1,785 features was reduced to 25 features, 50 features, and 100 features using PCA. Referring to the obtained result of Experiment 1, PC2 (50 features) was considered for Experiment 2.

For the generation of model, both SVM classifier and k-NN classifier were compared. The SVM is widely used for statistical learning and classification. Primarily, the SVM deals with binary classification issues. There are presently two multiple classification approaches in use, specifically one-against-one approach and one-against-all approach. The one-against-all approach was specifically considered for this experiment because it is computationally faster than the other approach. Accordingly, the RBF kernel was applied with $g = 0.0625$, and a trade-off between the training error and margin, $c = 8$. It should be noted that these values were obtained from an empirical study. The second most widely used classification method is the k-NN ($k = 1$), which was used for further comparisons (details on the parameters of SVM and k-NN are further discussed for Experiment 4). Results were calculated after performing random sampling on the dataset for 10 times in order to produce reliable results.

The results shown in Table 7 and Table 8 refer to the correctness rate of different feature sets using both SVM classifier and k-NN classifier, respectively. It could be observed in Table 8 that in the XMIAR prototype, the combined features of all four features using the SVM classifier achieved the highest accuracy rate (95.368%) by applying the second set of evaluation (90% of training images and 10% of testing images) without a threshold. The combined features of all features contained pixel information, global features (features of shape and texture), local features (features of shape and texture), and SURF. Therefore, the application of the SVM using combined features outperformed the other applications using each of the feature sets separately as follows: (1) global feature set, (2) local feature set, (3) pixel value set, and (4) SURF set. The comparison of these distinct features also revealed that the use of pixel features outperformed the uses of both global features and local features for all evaluation sets while the local features provided results of higher accuracy rate than the global features for all evaluation sets.

Table 7. Accuracy results of extracted features by using k-NN

	80%-20% with threshold	80%-20% without threshold	90%-10% with threshold	90%-10% without threshold
SURF	54.276	84.532	80.652	86.823
Global	61.111	82.751	86.317	88.313
Local	61.564	85.590	87.593	88.868
Pixel	63.887	87.596	86.679	91.293
All features	69.381	89.327	87.885	92.712

Table 8. Accuracy results of extracted features by using SVM

	80%-20% with threshold	80%-20% without threshold	90%-10% with threshold	90%-10% without threshold
SURF	82.800	84.963	83.409	87.261
Global	86.792	88.799	87.950	91.430
Local	88.792	90.165	88.887	92.046
Pixel	89.197	91.116	91.859	93.672
All features	90.450	92.202	91.924	95.368

Meanwhile, the combined features achieved the highest correctness rate (95.368%) by applying the SVM and a slightly lower correctness rate of 92.202% by applying the k-NN (90:10) without a threshold. In practice, different features of images reflect different attributes, which explained why the combined features provided results of higher correctness rate. Figure 6 shows the accuracy rate for each class using both SVM classifier and k-NN classifier with the evaluation set of 90:10 without a threshold. It could be observed that the SVM classified images more efficiently than the k-NN for various classes, such as classes 15, 23, 29, 37, and 51 while the k-NN outperformed the SVM for other classes such as classes 21 and 44. On the other hand, both classifiers achieved almost similar accuracy rate for classes 50 and 52. For the remaining classes, both SVM and k-NN classifiers provided convergent results. There were certain classes with substantial amount of training images such as class 12 while there were also other classes with few training images such as classes 51, 52, and 55, with only eight samples, as shown in Table 4. The k-NN classifier performed more efficiently when the objects in images were distinctly clear in contrast with the backgrounds and when all gray pixels were in one part of the images.

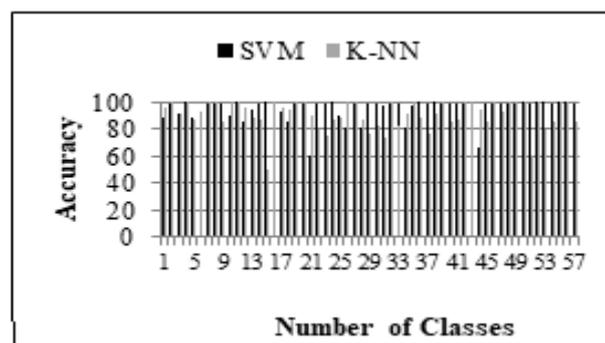


Figure 6. Accuracy results of k-NN and SVM for combinations of features

The results from this study revealed improvement in comparison with the results obtained in previous related studies using the same dataset. In fact, the proposed method in this study provided higher accuracy rate in comparison to the winner teams of the ImageCLEFmed2005 task; RWTH -i6. Additionally, the proposed method provided higher accuracy rate in comparison to the previous related studies [28], [10]- [14].

4.3 Classification and Parameters

The main two classifiers in this study were the SVM with RBF kernel and the k-NN with Euclidean distance metrics to locate the nearest neighbor. Thus, this subsequent experiment was conducted to compare the SVM classifier with the k-NN classifier using different parameters. Additionally, optimal parameters for each classifier were identified with its respective accuracy rate. For this comparison experiment, the k-NN was considered due to its popularity and classification performance shown in previous related studies. Moreover, compared to SVM, the implementation of k-NN is simpler because there is no offline training. For this study, the SVM was applied from the Library for Support Vector Machine (LIBSVM) [29]. The LIBSVM is generally defined as incorporated software to support vector classification and to sustain multi-class classification. Its main features included effective multi-class classification, cross validation for model choice, and various kernels. For this experiment, the k-NN was examined with different values of kernel (k). In addition, a comparison was conducted for the SVM using RBF kernel of different values.

It should be noted that this experiment was an empirical study of trial and error to select the optimal kernel function. Hence, this empirical study revealed that the values obtained using k-NN (k = 1) and SVM (-

$t = 2$, $-c = 8$, $-g = 0.0600$) were optimal based on the classification of ImageCLEF2005 database. For the parameters of SVM, $-t$ represents the type of kernel, $-c$ refers to the tradeoff between the training error and margin, and $-g$ denotes gamma, which refers to how far the influence of a single training example could be achieved. In this context, low values of $-g$ indicated far while high values of $-g$ indicated close.

According to the previous experiment, the dataset was divided into 90:10 to calculate the results obtained in this particular experiment. Similar feature vectors of global, local, pixel, and SURF were utilized with the application of PCA.

As illustrated in Figure 6, results revealed that when c equals to 8, improved classification was achieved for the SVM classifier. Moreover, the default value of gamma (g) is obtained using the following: $1 / \text{num_features}$. Given that both training images and testing images contained 50 features per image, the value obtained equals to 0.02. Conducted empirical study revealed that increased value of gamma provided higher stability and accuracy rate.

The results using k-NN revealed that the highest accuracy rate was achieved when $k = 1$, which is, in fact, the default value of k . A comparison between this value and other values ($k = 2$, $k = 3$) was shown in Figure 7 and 8. One drawback of using the SVM is the time required for offline training. For this study, using the processor of Intel(R) i7-4500U with 8GB RAM and MATLAB 2012a version for coding specifically, the SVM took approximately five hours while for the k-NN, it performed rather instantaneously.

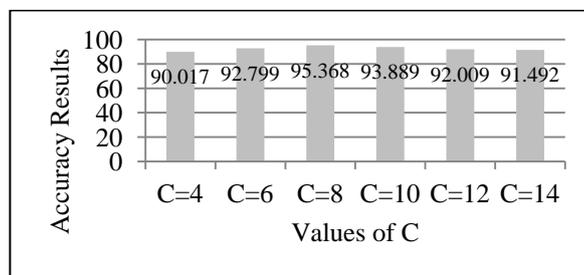


Figure 7. Accuracy results based on different values of C

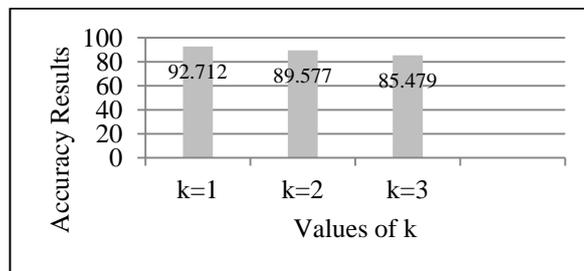


Figure 8. Accuracy results based on different values of k

5. DISCUSSION

These experiments validated the significance of classifying the X-ray medical image for meaningful image retrieval. Experiment 1 was specifically conducted to obtain optimal number of features for subsequent experiment using PCA. The resultant outcome of Experiment 1 is that PC2 (50 features) achieved the highest accuracy rate for both classifiers. Meanwhile, the obtained results from Experiment 2 distinctly revealed that combined features yielded higher accuracy rate compared to the application of single feature. Given the complexity of medical images, this study revealed that utilizing all available features was the optimal approach to enhance the performance of retrieval and classification. Typically, the classification of an image depends on low-level features while the annotation depends on the accuracy rate of classification.

Additionally, in Experiment 3, the classification techniques and its parameters were conducted. Based on the optimal performances of both SVM classifier and k-NN classifier, these two classifiers were utilized in the proposed system, where the SVM and k-NN are used for better accuracy results.

REFERENCES

- [1] Muller, H., & Deserno, T. M. 2011. Content-based medical image retrieval. In *Biomedical Image Processing, Biological and Medical Physics, Biomedical Engineering*, pp. 471-494. Springer Berlin Heidelberg.
- [2] Datta, R., Joshi, D., Li, J., & Wang, J. Z. 2008. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys (CSUR)* **40**(2): 1-60.
- [3] Akgul, C. B., Rubin, D. L., Napel, S., Beaulieu, C. F., Greenspan, H., & Acar, B. 2011. Content-based image retrieval in radiology: current status and future directions. *Journal of Digital Imaging* **24**(2): 208-222.
- [4] Zhang, D., Islam, M. M., & Lu, G. 2012. A review on automatic image annotation techniques. *Pattern Recognition* **45**(1): 346-362.
- [5] Clough, P., Müller, H., Deselaers, T., Grubinger, M., Lehmann, T. M., Jensen, J., & Hersh, W. 2006. The CLEF 2005 cross-language image retrieval track. Springer Berlin Heidelberg. *Accessing Multilingual Information Repositories*. 4022: 535-557.
- [6] Deselaers, T., Müller, H., Clough, P., Ney, H., & Lehmann, T. M. 2007. The CLEF 2005 automatic medical image annotation task. *International Journal of Computer Vision* **74**(1): 51-58.
- [7] Amaral, I. F. A. 2010. Content-Based Image Retrieval for Medical Applications. *Master Thesis*. Faculty of Science, University of Porto. <http://www.inescporto.pt/~jsc/students/2010IgorAmaral/2010relatorioIgorAmaral.pdf>. (10 March 2014).
- [8] Mueen, A. 2009. Multi-level automatic medical image annotation. Doctoral thesis. Faculty of computer science and information technology, University of Malaya. <http://dspace.fsktm.um.edu.my/handle/1812/451>. (30 June 2013).
- [9] Fesharaki, N. J., & Pourghassem, H. 2012, November. 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*. pp. 369-373.
- [10] Ghofrani, F., Helfroush, M. S., Danyali, H., & Kazemi, K. 2012. Medical X-ray image classification using Gabor-based CS-local binary patterns. In *Proceedings International Conference on Electronics, Biomedical Engineering and its Applications, Dubai*, pp. 284 - 288.
- [11] Mohammadi, S. M., Helfroush, M. S., & Kazemi, K. 2012. Novel shape-texture feature extraction for medical X-ray image classification. *International journal of innovative computing, information & control: IJICIC (INT J INNOV COMPUT I)*. **8**(1): 658-676.
- [12] Zare, M. R., Seng, W. C., & Mueen, A. 2013. Automatic Classification of medical X-ray Images. *Malaysian Journal of Computer Science* **26**(1): 9-22.
- [13] Haralick, R. M., Shanmugam, K., & Dinstein, I. H. 1973. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on SMC-3*(6): 610-621.
- [14] Acharya, T., & Ray, A. K. 2005. *Image processing: principles and applications*. John Wiley & Sons.
- [15] Baraldi, A., & Parmiggiani, F. 1995. An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters. *Geoscience and Remote Sensing, IEEE Transactions on* **33**(2): 293-304.
- [16] Soh, L. K., & Tsatsoulis, C. 1999. Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. *Geoscience and Remote Sensing, IEEE Transactions on* **37**(2): 780-795.
- [17] Gadkari, D. 2004. Image quality analysis using GLCM. *Master thesis*. College of Arts and Sciences, University of Central Florida Orlando, Florida. http://portfolio.dhanashree.net/Gadkari_Dhanashree_MS_Thesis.pdf. (14 December 2013).
- [18] Nithya, R., & Santhi, B. 2011. Mammogram classification using maximum difference feature selection method. *Journal of Theoretical and Applied Information Technology* **33**(2): 197-204
- [19] Yang, X., Tridandapani, S., Beitler, J. J., David, S. Y., Yoshida, E. J., Curran, W. J., & Liu, T. 2012. Ultrasound GLCM texture analysis of radiation-induced parotid-gland injury in head-and-neck cancer radiotherapy: an in vivo study of late toxicity. *Medical physics*, **39**(9): 5732-5739.
- [20] Gebejes, A., & Huertas, R. 2013. Texture Characterization Based on Grey-Level Co-Occurrence Matrix. In *Proceedings Of Informatics And Management Sciences, ICTIC*, pp: 375-378.
- [21] Sathyamoorthy, D. 2014. Analysis of surface textures of physiographic features extracted from multi scale digital elevation models via grey level co-occurrence matrix. *Journal of Geodesy and Geoinformation*. **2**(1): 19-28.
- [22] Chang, P., & Piau, P. 2008. Simple procedure for the designation of Haar wavelet matrices for differential equations. *Proceedings of the International MultiConference of Engineers and Computer Scientists*. Hong Kong. pp. 2010-2013.
- [23] Rao, N. G., Kumar, V. V., & Krishna, V. V. 2009. Texture based image indexing and retrieval. *IJCSNS International Journal of Computer Science and Network Security* **9**(5): 206-210.
- [24] Canny, J. 1986. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on PAMI-8*(6): 679-698.
- [25] Bay, H., Tuytelaars, T., & Van Gool, L. 2006. Surf: Speeded up robust features. *Computer vision-ECCV 2006*. Springer Berlin Heidelberg. 3951: 404-417.
- [26] Hotelling, H. 1933. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology* **24**(6): 417-441.
- [27] Lehmann, T.M., Schubert, H., Ott, B., & Leisten, M. 2005b. ImageCLEF2005 library available at: <http://ganymed.imib.rwth-aachen.de/irma/>. (30 July 2013).
- [28] Qiu, B., Tian, Q., & Xu, C. S. 2005. Report on the annotation task in ImageCLEFmed 2005. In *Cross Language Evaluation Forum 2005 Workshop*. Vienna, Austria.

- [29] Chang, C. C., & Lin, C. J. 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2(3): 1-27.

BIOGRAPHIES OF AUTHORS



Dr. Mohammed Muayad Abdulrazzaq has received his B.Sc. in Electronic and Electrical Engineering and M.Sc. in Computer Science, PhD Information System in 2004, 2007 and 2016, respectively. His research interest includes Medical Image Retrieval, Annotation and Classification.



Prof. Dr. Imad Fakhri Taha is at the Department of Computer Science and the Head of Research at the Kulliyah of Information and Communications Technology, the International Islamic University Malaysia (IIUM) since 1 st November 2013. He is IEEE senior member, and he is the editor in chief of JACSTR (international Journal on Advanced Computer Science and Technology Research) since 2011 till now and IJPCC international journal since 2015, and the general chair of the international conference on Advanced Computer Science Applications and Technologies) since 2012 till now.



Prof. Dr. Shahrul Azman Noah is currently a professor at the Faculty of Information Science & Technology, Universiti Kebangsaan Malaysia (UKM). He received his MSc and PhD degrees from the University of Sheffield. Currently he is heading the Knowledge Technology research group at UKM. His research work is focused on semantic computing with special emphasis on information retrieval, semantic searching and ontology creation.



Dr. Moayad Al-Athami

Associative-Professor, Head of Software Engineering Department, Faculty of Information Technology, University of Philadelphia, Jordan.

Research áreas:

- Expert Systems
- Intelligent Systems
- Search Engine
- Computational Linguistics
- Computer Security
- Machine Learning