

# Image Retrieval and Classification of Carotid Plaque Ultrasound Images

C.I. Christodoulou<sup>\*1</sup>, C.S. Pattichis<sup>1</sup>, E. Kyriacou<sup>2</sup>, A. Nicolaides<sup>3</sup>

<sup>1</sup>Dep. of Computer Science, University, of Cyprus, 75 Kallipoleos Str, P.O. Box 20537, 1678 Nicosia, Cyprus

<sup>2</sup>Dep. of Computer Science and Engineering, Frederic University, Limassol, Cyprus

<sup>3</sup>Cyprus Cardiovascular Disease Educational and Research Trust, Nicosia, Cyprus

**Abstract:** The extraction of multiple features from high-resolution ultrasound images of atherosclerotic carotid plaques, characterizing the plaque morphology and structure can be used for the classification and retrieval of similar plaques and the identification of individuals with asymptomatic carotid stenosis at risk of stroke. The objective of this work was to develop an automated image retrieval and classification system for the retrieval of similar carotid plaque ultrasound images, which will assist the physician in making his diagnostic decision based on similar previous cases. The neural self-organizing map (SOM) and the statistical K-nearest neighbor (KNN) classifiers were used for the retrieval and the classification of the carotid plaques into symptomatic or asymptomatic. Twenty different feature sets including texture, shape, morphological, histogram and correlogram features were extracted from the carotid plaque images and the classification results were further combined in order to improve the success rate. The results on a dataset of 274 carotid plaque ultrasound images show that image retrieval and classification for carotid plaque image are feasible and that features like multi-region histogram or texture can be used successfully for the identification of cases with similar symptoms output.

**Keywords:** Carotid plaque, image retrieval, classification.

## INTRODUCTION

The huge amount of medical and other digital images made available in the recent years necessitate image retrieval systems in order to effectively and efficiently use the information that is intrinsically stored in these image databases. The visual characteristics of a disease as shown through medical imaging modalities carry diagnostic information and often visually similar images correspond to the same symptom output. The retrieval of similar medical images to a given query image from a database of images may be used as a 'second opinion' to the physician in making diagnostic decisions. A critical step for developing a successful image retrieval system is the automated extraction of features characterizing the image and the selection of the most appropriate features to the given task.

There was a lot of work and great interest for the development of image retrieval systems during the last decade that triggered a stronger association of weakly related fields such as, computer vision, machine learning, information retrieval, human-computer interaction, database systems, Web and data mining, information theory, statistics, and psychology [1]. In a review paper, Muller *et al.* [2] provided an extensive review of content-based image retrieval (CBIR) systems in medical applications, their clinical benefits and future directions. There were several approaches in dealing with image retrieval systems. Chbeir and Favetta in their work [3] developed a medical image management system, which translated contextual and

semantic data of medical imaging into SQL queries in order to accomplish efficient image retrieval. Liu [4] in his paper described an enhanced independent component analysis method and its application to content based face image retrieval, which had an improved performance compared to other popular face recognition methods. In [5] Nezamabadi-pour and Kabir used histogram for image retrieval whereas in [6] Laaksonen *et al.* used the SOM classifier and different feature distributions for comparing different classes and different feature representations of the data in the context of the PicSOM CBIR system. Furthermore in their work, Amores and Radeva [7] presented a CBIR system for intravascular ultrasound images using a generalization of correlograms in order to extract local, global and contextual image features.

In previous preliminary work [8] on the same problem with a different dataset comparable results were achieved with the ones presented in this study. Also compared to other previous work [9], in this study new features, beyond the traditional texture features, were investigated like morphology, multi-region histogram and corellogram. Furthermore, an enhanced dataset which was more carefully selected and labeled by the physicians was used.

Carotid endarterectomy is an operation during which a vascular surgeon removes the inner lining of the carotid artery if it has become thickened or damaged. This procedure eliminates atherosclerotic plaque from the carotid artery in order to restore blood flow. There is evidence that carotid endarterectomy in patients with asymptomatic carotid stenosis may reduce the incidence of stroke [10]. The current practice is to operate patients based on the degree of internal carotid artery stenosis of 70% to 99% as shown in X-ray angiography [11]. However, a large number of patients may

<sup>\*</sup>Address correspondence to this author at the Department of Computer Science, University, of Cyprus, 75 Kallipoleos Str, P.O. Box 20537, 1678 Nicosia, Cyprus; Tel: +35799543747; E-mail: cschr2@ucy.ac.cy

be operated unnecessarily. Therefore it is necessary to identify patients at high risk, which will be considered for carotid endarterectomy and patients at low risk, which will be spared from an unnecessary, expensive and often dangerous operation.

Furthermore, endovascular angioplasty treatment of atherosclerotic carotid artery stenosis with stenting may be an alternative to surgical endarterectomy [12]. This treatment is gaining increasing popularity as a less invasive technique in symptomatic patients with severe (>70%) carotid artery stenosis especially for those patients who are at high risk for complications after surgery. No significant difference in the major risks of treatment was found but the wide confidence intervals indicate that it is not possible to exclude a difference in favor of one treatment. Minor complication rates favor endovascular treatment [12].

There are indications that the morphology of atherosclerotic carotid plaques, obtained by high-resolution ultrasound imaging, has prognostic implications. Smooth surface, echogenicity and a homogenous texture are characteristics of stable plaques, whereas irregular surface, echolucency and a heterogeneous texture are characteristics of potentially unstable plaques [9, 13]. The extraction of features characterizing efficiently the structure of ultrasound carotid plaques is important for the correct plaque classification and retrieval.

The objective of this work was to develop a system that will facilitate the automated retrieval of similar carotid plaque ultrasound images and their classification into symptomatic or asymptomatic based on the following features: (i) texture, (ii) shape, (iii) morphology, (iv) histogram and (v) correlogram. A total number of 20 different feature sets were extracted from the carotid plaques, which were classified using the neural self-organizing map (SOM) and the statistical K-nearest neighbor (KNN) classifiers. The classification results were further combined in order to enhance the classification output using the following combining techniques: (i) majority voting, (ii) weighted averaging based on a confidence measure and (iii) considering the most common retrievals. The criterion for correct retrieval was their clinical symptom outcome (symptomatic or asymptomatic). The aim was to identify plaque images with similar structure and based on their clinical history and known symptoms to decide the course of treatment for the query plaque/subject. The ultimate task was to identify individuals with asymptomatic carotid stenosis at risk of stroke. Stroke is the third leading cause of death in the western world and the major cause of disability in adults.

## MATERIAL

A total of 274 carotid plaque ultrasound images (137 symptomatic and 137 asymptomatic) were analyzed. The carotid plaques were labelled as symptomatic after one of the following hemispheric or retinal symptoms was identified in a follow up period of 6 to 84 months: stroke, transient ischemic attack or amaurosis fugax. Patients with cardioembolic symptoms were excluded from the study. A resting 12 lead ECG and a CT-brain scan without contrast were performed and reported by two cardiologists and one neuroradiologist respectively [10]. Asymptomatic plaques

were truly asymptomatic if they had never been associated with symptoms in the past [10, 14].

The ultrasound images were collected in the Irvine Laboratory for Cardiovascular Investigation and Research, Saint Mary's Hospital, UK, using an ATL (model HDI 3000 - Advanced Technology Laboratories, Seattle, USA) duplex scanner with a 5-10 MHz multifrequency probe at 20 pixels/mm. The images were normalized manually by adjusting the image linearly so that the median gray level value of blood was 0-5, and the median gray level of adventitia (artery wall) was 180-200 [14]. The scale of the gray level of the images ranged from 0 to 255. This normalization (i.e. using blood and adventitia as reference points) was necessary in order to extract comparable measurements in case of processing images obtained by different operators or different equipment.

The plaque identification and segmentation tasks are quite difficult and were carried out manually by the expert physician and/or vascular ultrasonographer based on color flow imaging. The main difficulties are due to the fact that the plaque boundaries cannot be distinguished from blood based on brightness level difference, or using only texture features, or other measures. Also calcification and acoustic shadows make the problem more complex. Thus, acoustic shadows were excluded. Fig. (1) shows an ultrasound image of the carotid artery with the atherosclerotic carotid plaque region of interest (ROI) outlined. The plaques used in this study were manually segmented, however it should be mentioned that a semi-automated segmentation technique [15] based on snakes, provided plaque delineations that were highly comparable to the manual ones.

## FEATURE EXTRACTION

A wide range of twenty different features sets was investigated in an effort to identify the most suitable for carotid plaque classification and retrieval. The following texture [16-21], shape, morphology [22], histogram and correlogram [7] feature sets were extracted from the segmented plaque images:

### Texture Features

(i) *Statistical Features (SF)*: The following statistical features were computed: 1) Mean value, 2) Median value, 3) Standard Deviation, 4) Skewness, and 5) Kurtosis.

(ii) *Spatial Gray Level Dependence Matrices (SGLDM)*: The spatial gray level dependence matrices as proposed by Haralick *et al.* [16] are based on the estimation of the second-order joint conditional probability density functions that two pixels ( $k,l$ ) and ( $m,n$ ) with distance  $d$  in direction specified by the angle  $\theta$ , have intensities of gray level  $i$  and gray level  $j$ . Based on the probability density functions the following texture measures were computed: 1) Angular second moment, 2) Contrast, 3) Correlation, 4) Sum of squares: variance, 5) Inverse difference moment, 6) Sum average, 7) Sum variance, 8) Sum entropy, 9) Entropy, 10) Difference variance, 11) Difference entropy, and 12), 13) Information measures of correlation. For a chosen distance  $d$  (in this work  $d=1$  was used, i.e.  $3 \times 3$  matrices) and for angles  $\theta = 0^\circ, 45^\circ, 90^\circ$  and  $135^\circ$  we computed four values for each of the above 13 texture measures. In this work, the mean and the

**Fig. (1).** Ultrasound image of the carotid artery with the atherosclerotic carotid plaques outlined.

range of these four values were computed for each feature, and they were used as two different feature sets.

(iii) *Gray Level Difference Statistics (GLDS)*: The GLDS algorithm [17] uses first order statistics of local property values based on absolute differences between pairs of gray levels or of average gray levels in order to extract the following texture measures: 1) Homogeneity, 2) Contrast, 3) Angular second moment, 4) Entropy, and 5) Mean. The above features were calculated for displacements  $\delta = (0, d), (d, d), (d, 0), (d, -d)$ , where  $\delta \equiv (\Delta x, \Delta y)$ , and their mean values were taken. In this work the parameter  $d=1$  was used.

(iv) *Neighborhood Gray Tone Difference Matrix (NGTDM)*: Amadasun and King [18] proposed the Neighborhood Gray Tone Difference Matrix in order to extract textural features, which correspond to visual properties of texture. The following features were extracted, for a neighborhood size of  $(2d+1) \times (2d+1)$  where  $d=1$  was chosen: 1) Coarseness, 2) Contrast, 3) Busyness, 4) Complexity, and 5) Strength.

(v) *Statistical Feature Matrix (SFM)*: The statistical feature matrix [19] measures the statistical properties of pixel pairs at several distances within an image, which are used for statistical analysis. Based on the SFM the following texture features were computed: 1) Coarseness, 2) Contrast, 3) Periodicity, and 4) Roughness. The constants  $L_r, L_c$  which determine the maximum intersample spacing distance were set  $L_r=L_c=4$ .

(vi) *Laws Texture Energy Measures (TEM)*: For the Laws TEM extraction [20, 21], vectors of length  $l=7, L=(1, 6, 15, 20, 15, 6, 1), E=(-1,-4,-5, 0, 5, 4, 1)$  and  $S=(-1,-2, 1, 4, 1,-2,-1)$  were used, where  $L$  performs local averaging,  $E$  acts as edge detector and  $S$  acts as spot detector. If we multiply the column vectors of length  $l$  by row vectors of the same length, we obtain Laws  $l \times l$  masks. In order to extract texture features from an image, these masks are convoluted with the image and the statistics (e.g. energy) of the resulting image are used to describe texture. The following texture features were extracted: 1) LL - texture energy from LL kernel, 2) EE - texture energy from EE kernel, 3) SS - texture energy from SS kernel, 4) LE - average texture energy from LE and EL ker-

nels, 5) ES - average texture energy from ES and SE kernels, and 6) LS - average texture energy from LS and SL kernels.

(vii) *Fractal Dimension Texture Analysis (FDTA)*: The Hurst coefficient  $H^{(k)}$  [21] was computed for image resolutions  $k=1, 2, 3, 4$ . A smooth surface is described by a large value of the parameter  $H$  whereas the reverse applies for a rough surface.

(viii) *Fourier Power Spectrum (FPS)*: The radial sum and the angular sum of the discrete Fourier transform [21] were computed in order to describe texture.

### Shape Features

The following shape features were calculated from the plaque images [9]: 1) X - coordinate maximum length of the plaque image frame, 2) Y - coordinate maximum length of the plaque image frame, 3) Area of ROI, 4) Perimeter of ROI, and 5)  $\text{Perimeter}^2/\text{Area}$ . The idea was to investigate whether the size and complexity of the shape of the segmented plaque had any diagnostic value.

### Morphology Features

Morphological image processing allows the detection of the presence of specific patterns, called structural elements, at different scales. The simplest structural element for near-isotropic detection is the cross '+' consisting of five image pixels. Using the cross '+' as a structural element, pattern spectra were computed for each plaque image as defined in [22]. The mean cumulative distribution functions (CDF) and the mean probability density functions (PDF) were computed as two different morphological feature sets.

### Histogram Features

(i) *Histogram*: The grey level histogram of the ROI of the plaque image was computed for 32 equal width bins and was used as an additional feature set. Histogram despite its simplicity provides a good description of the plaque structure.

(ii) *Multi-region Histogram*: Three equidistant ROIs were identified by eroding the plaque image outline by a factor based on the plaque size. The histogram was com-

puted for each one of the three regions as described above and the 96 values comprised the new feature vector. This feature was computed in order to investigate whether the distribution of the plaque structure in equidistant ROIs has a diagnostic value and more specifically if the structure of the outer region of the plaque is critical whether the plaque will rupture or not.

### Correlogram Features

Correlograms are histograms, which measure not only statistics about the features of the image, but also take into account the spatial distribution of these features [7]. In this work two correlograms were implemented for the ROI of the plaque image: (i) based on the distance of the distribution of the pixels' gray level values from the center of the image, and (ii) based on their angle of distribution. For each pixel the distance and the angle from the image center was calculated and for all pixels with the same distance or angle their histograms were computed. In order to make the comparison between images of different sizes feasible, the distance correlograms were normalized into 32 possible distances from the center by dividing the calculated distances with *maximum\_distance/32*. The angle of the correlogram was allowed to vary among 32 possible values starting from the left middle of the image and moving clockwise. The resulting correlograms were matrices 32x32 (gray level values over 32 were set to be the white area surrounding the region of interest and were not consider for the calculation of the features).

The histogram and correlogram features were used for classification with their original values whereas the rest of the feature sets were normalized before use by subtracting their mean value and dividing with their standard deviation.

### Feature Selection

In order to enhance the classification score, feature selection was implemented by computing the distance between the two classes for each feature as [9]:

$$dis = \frac{|m_1 - m_2|}{\sqrt{\sigma_1^2 + \sigma_2^2}} \quad (1)$$

where  $m_1$  and  $m_2$  are the mean values and  $\sigma_1$  and  $\sigma_2$  are the standard deviations of the two classes for each feature. The features with the highest discriminatory power were considered to be the ones with the greatest distance. The 15 best features were selected from the 56 texture features and were used as a separate feature vector. Furthermore, for the morphology feature sets principal component analysis (PCA) was applied as another feature selection method.

## IMAGE RETRIEVAL AND CLASSIFICATION

For the retrieval of similar plaque images the neural self-organizing feature map (SOM) classifier and the statistical K-nearest neighbor (KNN) classifier were used. The SOM and the KNN classifiers are well suited for image retrieval because they perform their classification task on the basis of identifying similar patterns. The SOM identifies similar patterns assigned to the same or near to a specific output node, whereas the KNN by inherently searching for the nearest neighbors of the evaluation pattern.

For training and evaluating the classifiers, the leave-one-out method was implemented where each time a plaque was

used for evaluation and the remaining plaques of the dataset were used for training. The leave-one-out method made the classification results independent of bootstrap sets (i.e. dividing the data in training and evaluation sets) and therefore more robust and reliable.

### The SOM Classifier

The SOM was chosen because it is an unsupervised learning algorithm where the input patterns are freely distributed over the output node matrix [23]. The weights are adapted without supervision in such a way, so that the density distribution of the input data is preserved and represented on the output nodes. This mapping of similar input patterns to output nodes, which are close to each other, represents a discretisation of the input space, allowing a visualization of the distribution of the input data. The output nodes are usually ordered in a two dimensional grid and at the end of the training phase, the output nodes are labeled with the class of the majority of the input patterns of the training set, assigned to each node.

In the evaluation phase, a new input pattern was assigned to the winning output node with the weight vector closest to the new input vector. In order to classify the new input pattern, the majority of the labels of the output nodes in an  $R \times R$  neighborhood window centered at the winning node, were considered. The number of the input patterns in the neighborhood window for the two classes  $m=\{1, 2\}$ , ( $1$ =symptomatic,  $2$ =asymptomatic), was computed as:

$$SN_m = \sum_{i=1}^L W_i N_{mi} \quad (2)$$

where  $L$  is the number of the output nodes in the  $R \times R$  neighborhood window with  $L=R^2$  (e.g.  $L=9$  using a  $3 \times 3$  window), and  $N_{mi}$  is the number of the training patterns of the class  $m$  assigned to the output node  $i$ .  $W_i=1/(2 d_i)$ , is a weighting factor based on the distance  $d_i$  of the output node  $i$  to the winning output node.  $W_i$  gives the output nodes near to the winning output node a greater weight than the ones farther away (e.g. in a  $3 \times 3$  window, for the winning node  $W_i=1$ , for the four nodes perpendicular to the winning node  $W_i=0.5$  and for the four nodes diagonally located  $W_i=0.3536$ , etc). The evaluation input pattern was classified to the class  $m$  of the  $SN_m$  with the greatest value, as symptomatic or asymptomatic. The correct classification score was computed for six SOM window sizes:  $1 \times 1$ ,  $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ ,  $9 \times 9$  and  $11 \times 11$ . The SOM classifier was implemented with a  $12 \times 12$  output node architecture and it was trained for 5000 learning epochs. In order to verify the correctness of the classification results three different runs were made and their average was taken.

### The KNN Classifier

The statistical k-nearest neighbor (KNN) classifier [24] was also used for the classification of the carotid plaques. In the KNN algorithm, in order to classify a new input pattern, its  $k$  nearest neighbors from the training set were identified. The new pattern was classified to the most frequent class among its neighbors based on a similarity measure that is usually the Euclidean distance. In this work the KNN system was implemented for different values of  $k$  ( $k=1, 3, 5, 7, 9, 11$  and  $13$ ) and it was tested using for input the different feature sets.

## The Classifier Combiner

In the case of difficult pattern recognition problems, the combination of the outputs of multiple classifiers using for input multiple feature sets extracted from the raw data, can improve the overall classification performance [25]. In the case of noisy or of a limited amount of data, different classifiers often provide different generalizations by realizing different decision boundaries. Also, different feature sets provide different representations of the input patterns, containing different classification information. The combination of the results of the different features and the different classifiers increases the probability that the errors of the individual features or classifiers may be compensated by the correct results of the rest. The use of a confidence measure on how reliable the classification result is, can further improve the overall performance by weighting the individual classification results before combining.

In this work, the 20 classification results of the SOM and KNN classifiers were combined using: (i) Majority voting, (ii) Weighted averaging based on a confidence measure and (iii) Considering the most common retrievals.

### Majority Voting

In majority voting, the input plaque under evaluation was classified as symptomatic or asymptomatic by the 20 classifiers using as input the 20 different feature sets. The plaque was assigned to the majority of the symptomatic or asymptomatic votes of the 20 classification results, obtained by the SOM and KNN classifiers.

### Weighted Averaging Based on a Confidence Measure

The implementation steps for combining using weighted averaging were the following:

*Step 1: Compute confidence measure as*

$$conf = 2(\max\{SN_1, SN_2\} / (SN_1 + SN_2)) - 1 \quad (3)$$

If an input plaque pattern was classified as symptomatic then its confidence measure is multiplied with -1, whereas the asymptomatic plaques retain their positive values.

*Step 2: Calculate the average confidence.* Calculate the average of the  $n$  confidence measures which is the final output of the system combiner, as

$$\overline{conf} = \frac{1}{n} \sum_{j=1}^n conf_j \quad (4)$$

*Step 3: Classify plaque.* If  $\overline{conf} < 0$  then the plaque is classified as symptomatic, else if  $\overline{conf} > 0$  then the plaque is classified as asymptomatic.

The final output of the system combiner is the average confidence  $\overline{conf}$ , and its values are ranging from -1 to 1. Values of  $\overline{conf}$  close to zero mean low confidence of the correctness of the final classification result, whereas values close to -1 or 1 indicate a high confidence.

### Considering The most Common Retrievals

In this combining method, the plaques were considered which were most commonly retrieved among the 20 different

feature sets. In the case of the KNN classifier, from the  $k \times 20$  retrieved plaques the  $k$  plaques were identified with the maximum common occurrences. The input reference plaque was classified by using majority voting based on the labels of the  $k$  identified plaques (symptomatic or asymptomatic). In the case of the SOM classifier, in a similar approach, for the  $1 \times 1$  window the  $k=1$  common plaque occurrences were considered, for  $3 \times 3$   $k=3$ , for  $5 \times 5$   $k=5$ , etc. It will be further investigated in future work if a more elaborate selection of the  $k$  common plaque occurrences may lead to better results.

## The Retrieval System

For the SOM system, the retrieved images were assigned to the same output node with the query plaque or in the neighborhood window  $R \times R$  of the winning node as explained above. In the KNN system the  $k$  most similar images using the Euclidean distance as the similarity measure were retrieved. The evaluation of how successful was the retrieval, was based on the classification of the retrieved plaques into two types: (i) symptomatic, or (ii) asymptomatic. If the type of the query reference plaque was the same with the type of the majority of the retrieved images, then the retrieval was considered successful. Figs. (2) and (3) provide examples of successful image retrievals.

## RESULTS

Table 1 tabulates the success rates of correct retrievals for all cases for the SOM system, whereas Table 2 for the KNN system. In general feature sets, which in average performed well in both systems, proved to be the multi-region histogram, the selected with Eq. 1 best 15 texture features, the NGTDM, the FDTA, and the SGLDM (range) texture features. When combined the success rate was significantly improved in both systems especially in the cases of combining with weighted averaging and when considering the most common retrievals. The morphology and the FP feature sets performed poorer in both systems.

More specific for the SOM system, best feature set in average was the multi-region histogram with 71.8% correct retrievals, followed by the NGTDM with 69.3%, the best 15 texture features with 68.2%, the GLDS with 68.1%, and the FDTA with 67.8%, and both the SGLDM (range) and the correlogram (distance) with 66.3%. When combined with weighted averaging, the SOM system yielded in average 69.2% and when combined with common retrievals 66.9%. Individually, the highest correct retrieval rate was 72.6% with the multi-region histogram window sizes  $1 \times 1$  and  $9 \times 9$ , followed by 71.2% for the NGTDM feature set with the SOM window size  $9 \times 9$ , and 70.8% when combined with weighted averaging and window sizes  $1 \times 1$  and  $3 \times 3$ .

In the KNN system, best feature set in average proved again to be the multi-region histogram with 71.6% correct retrievals, followed by all 56 textures used as a feature vector with 68.7% correct retrievals, followed by the best 15 texture features with 68.0%, the NGTDM with 66.6%, and the SGLDM (range) and the FDTA with 65.2%. When combining the classification results of the KNN system, the correct retrieval rate improved significantly and reached in average 70.9% when combined with weighted averaging, and 69.7% when combined with common retrievals. Individually the highest correct retrieval rate for the KNN system was 73.7%

**Fig. (2).** Image retrieval using the KNN classifier and the best 15 texture features as selected with Eq. (1).

**Fig. (3).** Image retrieval using the KNN classifier and the multi-region histogram feature set.

when combined with common retrievals and with the  $k=11$ , followed by the best 15 textures and  $k=13$  with 73.4% and the multi-region histogram and  $k=7$  with 73.0%. When combined with weighted averaging and  $k=7$  yielded 72.3% and when using all 56 texture and  $k=11$  yielded 71.2%.

Varying the values of  $k$  showed that bigger  $k$  values yielded in general better results as illustrated in the last row of Table 2. This was partly true also for bigger window sizes of the SOM system, as shown in Table 1.

In Fig. (2) an example is given with the five more similar plaques to the query reference plaque shown at the top left side of the figure. Below each plaque its label as symptomatic or asymptomatic is given. In Fig. (2) the KNN classifier was used with the best 15 texture features used as input. As seen from the labels of the retrieved images, 5 out of 5 of the retrieved images are labeled as asymptomatic and their label coincides with the label of the reference plaque. Fig. (3) show the retrieved images for the same reference plaque using the KNN with the multi-region histogram. As seen in the figure, again all of the retrieved images are labeled as asymptomatic while two plaques are common retrievals with the previous Fig. (2).

Receiver operating characteristic (ROC) analysis is a standard approach to evaluate the sensitivity and specificity

of diagnostic procedures [26]. ROC analysis estimates a curve of the true positive rate (sensitivity) Vs the false positive rate ( $1 - \text{specificity}$ ), which describes the inherent trade off between sensitivity and specificity of a diagnostic system. The area below the curve indicates the discrimination capability of the specific system. Fig. (4) shows the ROC curves for the KNN classifier ( $k=13$ ) and four different feature sets. In the figure caption the areas below each curve are given.

Table 3 tabulates the evaluation measures  $\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FN} + \text{FP})$ ,  $\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$ ,  $\text{Sensitivity or Recall} = \text{TP} / (\text{TP} + \text{FN})$ ,  $\text{Specificity} = \text{TN} / (\text{TN} + \text{FP})$  and the area below the ROC curve for the proposed system, which should use the multi-region histogram feature set which gave the best results, for the SOM classifier with window size  $9 \times 9$  and for the KNN classifier with  $k=7$ .

## DISCUSSION

The advancement of medical imaging modalities in recent years made available to the clinicians a vast amount of medical images that is almost impossible to be processed manually. Limitations in the human eye-brain visual system, the presence of overlapping structures in images, the vast

**Table 1. The Success Rate of Correct Retrievals in % for the Leave-One-out SOM System, Using as Input the 20 Different Feature Sets. The Results are Given for Different SOM Window Sizes and in Average. In the Third Column  $N$  the size of the Input Feature Vector is Given. In Bold are Highlighted the Best Results Per Feature Set and in Average**

No	Feature Set	N	1x1	3x3	5x5	7x7	9x9	11x11	Average
1	SF	5	58.8	59.5	60.9	64.6	<b>65.7</b>	65.3	62.5
2	SGLDM (mean)	13	63.5	<b>64.6</b>	64.2	63.9	64.2	64.2	64.1
3	SGLDM (range)	13	62.8	66.8	<b>67.5</b>	66.1	67.2	67.5	<b>66.3</b>
4	GLDS	4	65.0	65.7	69.7	<b>70.1</b>	69.3	68.6	<b>68.1</b>
5	NGTDM	5	65.3	68.6	70.1	70.8	<b>71.2</b>	69.7	<b>69.3</b>
6	SFM	4	62.4	<b>65.7</b>	64.2	60.2	62.8	65.0	63.4
7	TEM	6	62.8	61.7	63.9	65.7	66.4	<b>67.5</b>	64.7
8	FDTA	4	67.5	67.2	<b>68.6</b>	67.5	67.5	68.2	<b>67.8</b>
9	FP	2	51.1	52.9	52.9	<b>56.6</b>	54.7	56.6	54.1
10	All texture	56	<b>67.5</b>	65.3	65.7	65.3	63.5	64.6	65.3
11	Best 15 texture	15	66.8	67.5	67.5	68.6	<b>69.7</b>	69.0	<b>68.2</b>
12	Shape	5	60.2	61.3	64.6	65.7	<b>66.1</b>	64.6	63.7
13	Morphology CDF	57	53.3	50.7	52.2	54.7	56.9	<b>58.4</b>	54.4
14	Morph. CDF (PCA)	7	<b>59.5</b>	58.4	54.4	56.2	53.3	50.7	55.4
15	Morphology PDF	57	50.4	53.3	53.6	54.7	<b>58.0</b>	57.3	54.6
16	Morph. PDF (PCA)	20	52.2	56.9	56.2	57.3	56.2	<b>58.0</b>	56.1
17	Histogram	32	60.9	65.3	65.3	65.7	<b>66.1</b>	65.7	64.8
18	Multi-Region Hist.	96	<b>72.6</b>	72.3	70.8	71.5	<b>72.6</b>	70.8	<b>71.8</b>
19	Correlogram (dist.)	32x32	62.0	65.0	65.7	67.9	68.2	<b>69.0</b>	66.3
20	Correlogram (angle)	32x32	56.6	57.3	64.6	<b>65.0</b>	63.5	62.0	61.5
21	Combine Maj. Voting		65.7	65.0	66.4	<b>68.2</b>	66.4	66.8	<b>66.4</b>
22	Combine Weight. Avg.		<b>70.8</b>	70.8	68.6	68.6	67.9	68.6	<b>69.2</b>
23	Comb. Common Retr.		62.8	68.4	<b>69.0</b>	67.4	67.4	66.3	<b>66.9</b>
	Average		61.8	63.0	63.8	64.5	<b>64.6</b>	64.5	63.7

**Table 2. The Same as Table 1 for the Leave-One-Out KNN System and Different Values of  $k$** 

No	Feature Set	N	k=1	k=3	k=5	k=7	k=9	k=11	k=13	Average
1	SF	5	54.7	55.5	56.9	59.5	<b>60.6</b>	58.4	58.8	57.8
2	SGLDM (mean)	13	62.0	60.9	<b>65.0</b>	65.0	65.0	63.9	62.8	63.5
3	SGLDM (range)	13	60.6	60.6	63.9	<b>68.2</b>	67.9	67.2	67.9	<b>65.2</b>
4	GLDS	4	62.8	61.7	63.1	61.3	66.8	65.3	<b>68.2</b>	64.2
5	NGTDM	5	61.7	60.2	67.5	<b>70.1</b>	69.0	69.7	68.2	<b>66.6</b>
6	SFM	4	55.1	62.8	63.9	66.1	<b>68.6</b>	67.9	67.2	64.5
7	TEM	6	58.0	60.9	63.1	<b>68.2</b>	67.5	66.4	65.3	64.2
8	FDTA	4	55.1	62.4	67.2	67.2	68.2	67.5	<b>68.6</b>	<b>65.2</b>

Tabel 2. Contd....

No	Feature Set	N	k=1	k=3	k=5	k=7	k=9	k=11	k=13	Average
9	FP	2	54.0	49.6	53.6	53.6	53.3	<b>56.2</b>	54.0	53.5
10	All texture	56	63.5	68.2	69.0	70.8	69.0	<b>71.2</b>	69.0	<b>68.7</b>
11	Best 15 texture	15	64.2	64.6	68.6	66.8	67.9	70.8	<b>73.4</b>	<b>68.0</b>
12	Shape	5	59.9	61.3	<b>62.0</b>	59.5	60.6	59.1	60.6	60.4
13	Morphology CDF	57	46.4	54.0	55.5	51.8	53.6	54.7	<b>56.6</b>	53.2
14	Morph. CDF (PCA)	7	54.7	47.8	51.8	52.9	54.4	54.4	<b>58.4</b>	53.5
15	Morphology PDF	57	50.7	51.5	50.7	54.0	<b>55.1</b>	52.6	53.6	52.6
16	Morph. PDF (PCA)	20	51.5	50.0	48.9	<b>52.2</b>	51.5	50.4	51.8	50.9
17	Histogram	32	58.0	59.9	63.9	65.7	<b>66.8</b>	63.9	63.9	63.1
18	Multi-Region Hist.	96	67.5	71.9	72.6	<b>73.0</b>	71.9	72.6	71.5	<b>71.6</b>
19	Correlogram (dist.)	32x32	57.3	64.6	63.5	63.5	62.8	<b>65.7</b>	65.3	63.2
20	Correlogram (angle)	32x32	62.0	62.0	62.8	64.2	64.2	65.7	<b>66.1</b>	63.9
21	Combine Maj. Voting		62.0	63.9	69.4	67.9	<b>68.6</b>	68.6	68.2	<b>66.5</b>
22	Combine Weight. Avg.		70.1	70.1	71.2	<b>72.3</b>	70.8	71.2	70.4	<b>70.9</b>
23	Comb. Common Retr.		61.7	66.8	67.5	71.2	71.5	<b>73.7</b>	71.5	<b>69.7</b>
	Average		58.9	60.5	62.6	63.7	64.2	64.2	<b>64.4</b>	62.6

**Fig. (4).** ROC analysis based on the distributions of the average confidence measures as calculated in Eq. 3 for the KNN classifier and  $k=13$  and the following feature sets: 1) Multi-region histogram (Area below curve 79.5%), 2) Best 15 texture features (Area below curve 78.2%), 3) Correlogram (angle) (Area below curve 71.7%), and 4) Morphology CDF with PCA (Area below curve 56.0%).



**Fig. (5).** Average of all plaques of the multi-region histogram ( $3 \times 32 = 96$  bins) of asymptomatic Vs symptomatic plaques (red dashed line). The symptomatic plaques tend towards lower bin values i.e. they look darker in all three regions.

number of normal cases in screening programs and the overload of image data which may result to oversight errors, make necessary the use of automated CBIR systems in order to effectively and efficiently use the information that is intrinsically stored in these image databases [2]. The consultation of medical CBIR systems in the frame of a computer-aided diagnosis aim not to replace the physician but to assist him in making a diagnosis by providing similar images from similar cases and helping him to consider other options and treatments used in other similar cases and not anticipated by the first examination. In previous work in image retrieval systems [2-7] several new features were proposed for different applications, which in some cases were problem dependent [4], or as in [3] a medical image management system was developed which was more based on semantics rather than image content for retrieval. Also in some of the previous studies, partly similar features to the ones implemented in this work were used, like histograms [5] or correlograms [7], or similar classifiers like SOM in [6] but to the best of our knowledge no other image retrieval system has been implemented for the retrieval and classification of carotid plaques.

In this work an image retrieval and classification system for carotid plaque ultrasound images is presented using as input multiple feature sets and based on the neural SOM and the statistical KNN classifiers and combining techniques. Best results were obtained when the multi-region histogram was used or when the classification results were combined or when the best texture features were selected. The success of features like the multi-region histogram and partly of the correlogram features show that the layout of the white hard parts of the plaque in the ultrasound image, compared to the black soft parts of the image has a diagnostic value and affects whether a plaque will be symptomatic or asymptomatic. In particular the multi-region histogram outperformed in average the common histogram by more than 7%. Fig. (5) shows the average of the multi-region histogram of asymptomatic Vs symptomatic plaques of the 274 plaques. The symptomatic plaques tend towards lower bin values i.e. they

look darker in all three regions, which is in agreement with the findings of previous work [9]. The highest individual correct classification score was 73.7% for the KNN system when combined with common retrievals and with the  $k=11$ . Traditional texture features like NGDTM, SGLDM, FDA and GLDS performed well and their results were similar to the results of previous work [8, 9] on the same problem with different datasets. More complex features like morphology failed to perform up to the expectations for the specific problem.

The performances of the KNN and the SOM classifiers were comparable. The SOM performed in average better than the KNN with 63.7% compared to 62.6% but this can be partly attributed to the fact that in the KNN-results-table the values for  $k=1$  and  $k=3$  were included which performed lower than the other values of  $k$ . However the simpler statistical KNN classifier made it easier to track back and repeat the results than the more complex neural SOM classifier, which required separate training and evaluation phases. Also the SOM required optimization of the training process by selecting the correct network architecture and was computationally relatively heavier especially in the case of the large feature vectors.

Combining the outputs of the classifiers using as input the 20 different feature sets significantly improved the final correct retrievals/classification score in both classifiers. This was especially true in the case of the KNN classifier where combining with weighted average and when combining with common retrievals improved the average retrieval rate by 7-8%. Combining with majority voting showed a lower improvement in the range of 4%. For the SOM system best combining technique proved to be the weighted averaging with an improvement from the average by 5.5% where combining with common retrievals improved the results only by 3%. This can be attributed to the fact that the selection of the  $k$  common retrievals as described in section Image Retrieval and Classification subsection C(iii) was rather not optimal

**Table 3. Accuracy, Precision, Sensitivity, Specificity and the Area Below the ROC Curve in % for the Proposed Systems which should Use the Multi-Region Histogram Feature Set which Gave the Best Results, and Either the SOM Classifier with Window size 9x9 or the KNN Classifier with  $k=7$**

Classifier	Accuracy	Precision	Sensitivity	Specificity	ROC Area
SOM 9x9	72.6	72.6	72.8	72.5	77.8
KNN $k=7$	73.0	71.7	75.9	70.1	81.8

for the SOM classifier especially for SOM window size 1x1. The results in this work demonstrate the value of combining techniques in making an image retrieval and classification system more robust and accurate. In addition, the leave-one-out method was implemented for training and evaluating the system, making the classification results independent of bootstrap sets and therefore more reliable.

Based on the results as tabulated in Tables 1 and 2, a proposed system should use the multi-region histogram feature set with either the SOM or the KNN classifiers as shown in Table 3. The specific feature set is simple and easy to compute and proved to give in average the best results.

### CONCLUDING REMARKS

In conclusion the results in this work show that image retrieval and classification for carotid plaque images is feasible and that features like multi-region histogram or texture along with combining techniques can be used successfully for the identification of cases with similar symptoms output. The novelty of this analysis lies in the investigation of multiple features and the selection of the most appropriate for image retrieval using different classifiers and furthermore in the application of combining techniques in order to improve the classification outcome.

In future work the whole classification system will be incorporated into an automated content-based image retrieval system, which through a user-friendly interface will provide to the physician not only the labels of the retrieved plaques but also all the information about the original ultrasound images, the clinical data and the history of the similar cases. Hence, an assessment could be made based on the observation of similar cases. Such a system will integrate imaging and clinical databases and will help the physician to decide the course of treatment whether surgical endarterectomy, endovascular angioplasty or pharmaceutical. This will contribute towards the implementation of the most effective strategy to minimize cardiovascular death, avoiding unnecessary operations and offering a better service to the patient.

### ACKNOWLEDGEMENT

This work was partly supported through the project *Integrated System for the Evaluation of Ultrasound Imaging of the Carotid Artery (TALOS)* of the Research Promotion Foundation of Cyprus.

### REFERENCES

- [1] Datta R, Joshi D, Li J, Wang JZ. Image retrieval: Ideas, influences, and trends of the new Age. *ACM Comput Surv* 2008; 40(2): 1-60.
- [2] Muller 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): 1-23.
- [3] Chbeir R, Favetta F. A global description of medical imaging with high precision. *IEEE Trans Syst Man Cybern B Cybern* 2003; 33(5): 752-7.
- [4] Liu C. Enhanced independent component analysis and its application to content based face image retrieval. *IEEE Trans Syst Man Cybern B Cybern* 2004; 34(2): 1117-27.
- [5] Nezamabadi-Pour H, Kabir E. Image retrieval using histograms of uni-color and bicolor blocks and directional changes in intensity gradient. *Pattern Recogn Lett* 2004; 25(14): 1547-57.
- [6] Laaksonen J, Koskela M, Oja E. Class Distributions on SOM surfaces for feature extraction and object retrieval. *IEEE Trans Neural Netw* 2004; 17(8-9): 1121-33.
- [7] Amores J, Radeva P. In: Suri JS, Ed. *Medical image retrieval based on plaque appearance and image registration Plaque Imaging: Pixel to Molecular Level*. New Jersey: IOS Press 2005.
- [8] Christodoulou CI, Pattichis CS, Kyriacou E, Nicolaides A. Content-Based Image Retrieval For Carotid Plaque Ultrasound Images, in *Proc. of EMBE'05, 3rd European Medical and Biological Engineering Conference*. Prague, Czech Republic. November 20- 25, 2005.
- [9] Christodoulou CI, Pattichis CS, Pantziaris M, Nicolaides A. Texture based classification of atherosclerotic carotid plaques. *IEEE Trans Med Imaging* 2003; 22: 902-12.
- [10] Nicolaides A, Sabetai M, Kakkos SM, *et al*. The Asymptomatic Carotid Stenosis and Risk of Stroke (ACSRS) study. Aims and results of quality control. *Int Angiol* 2003; 22(3): 263-72.
- [11] Geroulakos G, Domjan J, Nicolaides A, *et al*. Ultrasonic carotid artery plaque structure and the risk of cerebral infarction on computed tomography. *J Vasc Surg* 1994; 20(2): 263-6.
- [12] Coward LJ, Featherstone RL, Brown MM. Safety and efficacy of endovascular treatment of carotid artery stenosis compared with carotid endarterectomy: a cochrane systematic review of the randomized evidence. *Stroke* 2005; 36(4): 905-11.
- [13] Wilhjelm JE, Gronholdt LM, Wiebe B, Jespersen SK, Hansen LK, Sillesen H. Quantitative analysis of ultrasound B-mode images of carotid atherosclerotic plaque: correlation with visual classification and histological examination. *IEEE Trans Med Imaging* 1998; 17(6): 910-22.
- [14] Nicolaides AN, Kakkos SM, Griffin M, *et al*. Effect of image normalization on carotid plaque classification and the risk of ipsilateral hemispheric ischemic events: results from the asymptomatic carotid stenosis and risk of stroke study. *Vascular* 2005; 13(4): 211-21.
- [15] Loizou C, Pattichis C, Istepanian R, Pantziaris M, Nicolaides A. Atherosclerotic Carotid Plaque Segmentation, 26<sup>th</sup> Annual Int. Conf. IEEE Eng. in Med. Biol. (EMBS), San Francisco, California, USA 2004.
- [16] Haralick RM, Shanmugam K, Dinstein I. Texture features for image classification. *IEEE Trans Syst Man Cybern* 1973; 3: 610-21.
- [17] Weszka JS, Dyer CR, Rosenfield A. A comparative study of texture measures for terrain classification. *IEEE Trans Syst Man Cybern* 1976; 6: 421-6.
- [18] Amadasun M, King R. Textural features corresponding to textural properties. *IEEE Trans Syst Man Cybern* 1989; 19(5): 1264-74.
- [19] Chung-Ming W, Yung-Chang C. Statistical feature matrix for texture analysis. *CVGIP: Graph Models Image Process* 1992; 54(5): 407-19.
- [20] Laws KI. Rapid texture identification. *Proc SPIE* 1980; 238: 376-80.
- [21] Chung-Ming W, Yung-Chang C, Kai-Sheng H. Texture features for classification of ultrasonic liver images. *IEEE Trans Med Imaging* 1992; 11(2): 141-52.

- [22] Dougherty ER. An Introduction to Morphological Image Processing. Belingham, Washington: SPIE Optical Engineering Press 1992.
- [23] Kohonen T. The self-organizing map. Proc IEEE 1990; 78(9): 1464-80.
- [24] Tou JT, Gonzalez RC. Pattern Recognition Principles. Addison-Wesley Publishing Company, Inc. 1974.
- [25] Kittler J, Hatef M, Duin R, Matas J. On combining classifiers. IEEE Trans Pattern Anal Machine Intell 1998; 20(3): 226-39.
- [26] Swets JA, Pickett RM. Evaluation of Diagnostic Systems: Methods from Signal Detection Theory. New York: Academic Press 1992.

---

Received: February 01, 2010

Revised: April 19, 2010

Accepted: July 27, 2010

© Christodoulou *et al.*; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.