# Classification-Driven Pathological Neuroimage Retrieval Using Statistical Asymmetry Measures

Y. Liu, F. Dellaert, W. E. Rothfus*, A. Moore, J. Schneider, and T. Kanade

The Robotics Institute, Carnegie Mellon University, Pittsburgh 15213, USA **
yanxi,dellaert,awm,schneide,tk@cs.cmu.edu
* University of Pittsburgh Medical Center, Pittsburgh, PA

**Abstract.** This paper reports our methodology and initial results on volumetric pathological neuroimage retrieval. A set of novel image features are computed to quantify the statistical distributions of approximate bilateral asymmetry of normal and pathological human brains. We apply memory-based learning method to find the most-discriminative feature subset through image classification according to predefined semantic categories. Finally, this selected feature subset is used as indexing features to retrieve medically similar images under a semantic-based image retrieval framework. Quantitative evaluations are provided.

## 1 Motivation

Medical images form an essential and inseparable component of diagnosis, intervention and patient follow-ups. In this work, we use a patient's image as an index to retrieve medically similar and relevant patient cases from a large multimedia database. Common practice in the image retrieval and pattern recognition community is to map each image into a set of numerical or symbolic attributes called *image indexing features*. Thus each image corresponds to a point in a multidimensional image feature space. Existing "content-based" image retrieval (CBIR) systems [6, 13] depend on general visual properties such as color and texture to classify diverse, two-dimensional (2D) images. However, these general visual cues often fail to be effective discriminators for image sets taken within a single domain, where images have subtle, domain-specific differences. Furthermore, these global statistical color and texture measures do not necessarily reflect or have proven correspondence to the meaning of an image, i.e. the image semantics, nor are they suitable for handling three-dimensional (3D) volumetric images. Our objective is to go beyond the ill-defined, subjective visual feature indexing practiced in many current CBIR systems. Our approach is based on statistical learning, and contains the following components:

1. **Feature extraction** maps each volumetric image into a multi-dimensional image feature space;
2. **Feature weighting and image similarity construction** imposes relative weights on the image feature space;
3. **Adaptive image retrieval** chooses image similarity best suited for the user intention.

** This research is supported in part by an NIST grant #70NANB5H1183 and in part by the NIH/NCI research contract # N01-CO-07119.

The kind of medical database accessing capability we are developing will have many potential applications in clinical practice and medical education, including on-line consultation, differential diagnosis, surgical planning, recovery/outcome evaluation, and tele-medicine. As a realistic test-bed of our methodology, we choose a neuroimage database composed of volumetric CT image sets of hemorrhage (acute blood), bland infarct (stroke) and normal brains. Justifications for this endeavor are two-fold: first of all, a database composed of volumetric images and collateral information in a particular medical domain provides objective, semantically well-defined training sets and quantifiable results; second, due to the limitation of the popular color and texture image features used by many existing CBIR systems, finding novel image features and most discriminating feature subsets for medical image characterization become crucial.

As pointed out in [5]: "To date, all too often image analysis algorithm development ignores the analysis of different abnormal, pathological or disease states". Work on neuroimage analysis has been concentrated more on morphological variations of normal brains [2, 7] or brains with mental disorders [15] using high resolution, high density MRI data. The image data used in this work, in contrast, is CT images obtained from clinical practice. The sparseness and incompleteness of the CT image data set pose more challenges in image understanding and indexing than does the complete, dense data used in other reported work. However, the database images we use more realistically reflect the type of query images that an end user will supply in practice.

Existing work on pathological medical image database retrieval deals mostly with 2D images and has not paid much attention to a systematic approach for image feature weighting. In the work of a dental radiography image database retrieval [17], the authors use a deformable shape contour selected by the system designer as the primary feature for image indexing. By selecting different modes in the finite element representation and eigen-decomposition of the contours (hand drawn by an expert dentist), the authors achieve classification rates between 87% (for normals) and 62% (for pathologies). What is missing in the image retrieval practice is an objective, quantitative evaluation of the extracted image features **before** they are used for image retrieval. The method we propose here directly addresses this issue.

## 2 Our Approach

A basic framework for classification-driven semantic based image retrieval is demonstrated in Figure 1, containing two stages with five essential components. The two stages are an off-line stage for classification-driven image similarity metric learning (feature weighting), and an on-line stage for image retrieval. The five essential components are (1) image preprocessing, (2) image feature extraction, (3) feature subset weighting via image classification, (4) image retrieval, and (5) quantitative evaluation. Though the two stages share some common components, the goals and constraints differ. In the off-line stage the goal is to find the best and smallest subset of image features that capture image semantics. It requires an explicitly labeled image data set, and sufficient computer memory space to store a large feature attribute matrix and to support extensive search on this matrix. High computational speed is a plus but not necessary. During the on-line stage, on the other hand, the demands are fast retrieval speed and visualization of retrieved images given similarity metric and a query image.

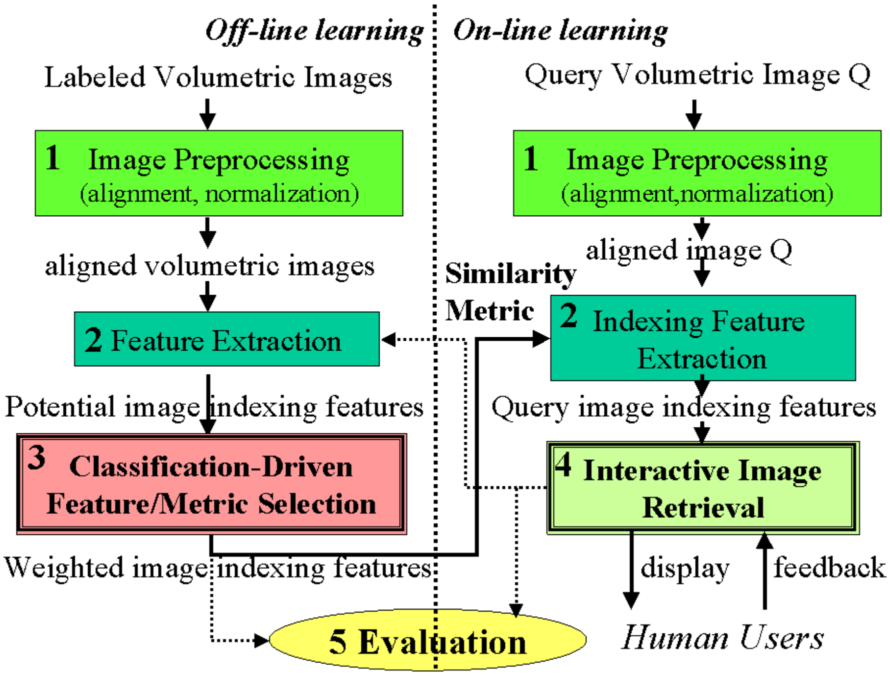


**Fig. 1.** Basic components for classification-driven semantic image retrieval

### 2.1 Novel Neuroimage Features

Human brains are approximately bilaterally symmetrical. In this work we have experimented with quantifying human brain asymmetry as a departure from an ideal midsagittal plane (Section 3.2).

### 2.2 Classification-Driven Feature Selection

We consider semantic-based image retrieval as a process of image classification by a $k$ nearest neighbor classifier (kNN) [4]. As is well-known, the advantage of a kNN classifier is its non-linear, non-parametric nature and tolerance to arbitrary data distributions without the assumption that the forms of the underlying densities are known. It can therefore bypass probability estimation and go directly to decision functions [4]. However, a kNN classifier is suboptimal compared to Bayes error, **and** is sensitive to irrelevant, redundant, and noisy features. It can be proven that adding more features can increase the error rate for a kNN classifier [3]. To achieve a reasonable performance from a kNN classifier, therefore, it is *mandatory* to first go through a feature selection process. Given a query image and its $k$ nearest neighbors, the goal of image retrieval under this context is to maximize the number of retrieved images that have the same semantic content as the query image. This is achieved by varying relative feature weights (dimension), and the distance function (metric).

In this work, feature selection is defined as a mapping from a potential feature set $F_1 = \{f_1, f_2, ..., f_n\}$ to another feature set $F_2 = \{w_1 f_1, w_2 f_2, ..., w_n f_n\}$ where $0 \leq w_i \leq 1$. Since some of the $w_i$s may take the value 0, $|F_2| \leq |F_1|$. Two types of features are expected to be removed from $F_1$: irrelevant features and redundant features. As a result, for any image semantic class $c_i$ the posterior probabilities $P(c_i|F_1)$ and $P(c_i|F_2)$ are equivalent. The values of the $w_i$s are the result from our learning algorithm. When many $w_i$s are zeros (as is the case in this work) $F_2$ becomes a much lower-dimensional space than $F_1$, and computation cost is reduced greatly.

We employ *kernel regression*, an efficient memory based learning (MBL) technique, to evaluate feature subsets, their relative weights and their distance function through classification. The error metric we seek to minimize is *cross entropy* defined as $E = -\sum_i \sum_C \delta_{iC} ln \hat{P}(c|\mathbf{x}_i)$ where $\delta_{iC}$ represents the 1-of-m multiple class membership encoding, and $\hat{P}(c|\mathbf{x}_i)$ is the approximation of the posterior probability $P(c|\mathbf{x})$ of a class $c$ given a feature vector $\mathbf{x}$ via Bayes law. Minimizing this function will yield a metric and a smoothing parameter $\sigma$ for which kernel regression best approximates the posterior probabilities $P(c|\mathbf{x})$, and is thus optimally suited for classification [1]. To look for a good *similarity metric* or *classifier* in this feature space, off-line, combinatorial search over the space of potential features is performed to minimize the classification error [14].

### 2.3 Image Retrieval and Evaluation

Once a locally optimal classifier is found, the associated image feature weights provide a proper scaling of the original feature space. The extracted image features form an $N$ dimensional vector space, the overall distance function $D$ of two images $a, b$ is defined as $D(a,b) = \sqrt{(\boldsymbol{A}-\boldsymbol{B})^T \Sigma^{-1} (\boldsymbol{A}-\boldsymbol{B})}$ where $\boldsymbol{A} = f(a), \boldsymbol{B} = f(b)$ are the $N$ dimensional image feature vectors of images $a$ and $b$ respectively, $\Sigma$ is an $N \times N$ covariance matrix. When $\Sigma$ is a diagonal matrix, $D$ is a non-uniformly scaled Euclidean distance. Otherwise $D$ is the Mahalanobis distance. Image retrieval is done using kNN with Euclidean distance. Evaluation of the retrieved images is measured by the *precision* rate $\mathbf{R}_{\mathbf{P_k}} = N_{rp}/K$ and *recall* rate $\mathbf{R}_{\mathbf{R_k}} = N_{rp}/N_p$ in the top $K$ retrieved images, where $N_{rp}$ is the number of correctly retrieved images with pathology class $p$ in the top $K$ retrieved images for the query, and $N_p$ is the total number of instances in the database that have pathology class $p$ and should be retrieved. When $K < N_p$, recall rate is identical to the precision rate.

## 3 Experiments

A study was performed using a data set of 48 volumetric CT brain scans containing normal (26), stroke (14) and blood cases (8). These are clinical CT images collected directly from a local hospital emergency room.

### 3.1 Image Alignment

The first step is to align all 3D images such that the ideal midsagittal planes (iMSP) of different brain images are parallel and pitch angle is zero. 3D image

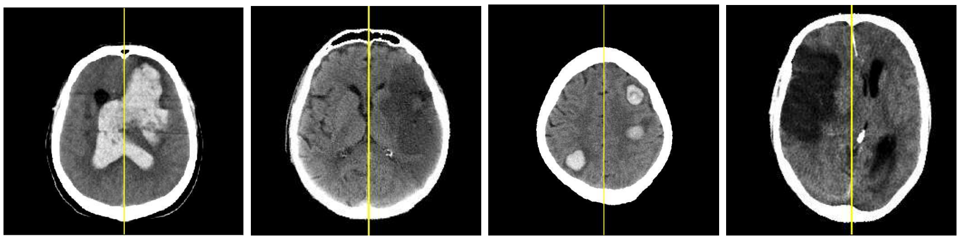

**Fig. 2.** The ideal symmetry axis is extracted as the intersection of a 3D brain image and its ideal midsagittal plane.

intra- and inter-subject, multi-modality rigid and deformable registrations of normal brains have achieved excellent results [16, 11]. However, no existing technology is readily available for registration of single or multi-modality, inter-subject, pathological brain images [8]. Matching two 3D pathological brains remains a difficult and controversial problem. This situation poses challenges to any 3D pathological neuroimage databases, especially for CT images. In our work, we have taken advantage of the approximate bilateral structure of human brains to develop a robust midsagittal plane (MSP) computation algorithm [9] for pathological brain alignment and comparison. The effect of the ideal midsagittal plane (iMSP) extraction algorithm [9] is not to find where the midsagittal plane is, but where it is supposed to be. This is especially useful for pathology brains since the anatomic midsagittal plane is often distorted (shifted or bent) due to large lesions. Figure 2 shows some 2D sample results after the midsagittal plane is extracted. The iMSP algorithm is robust under adverse conditions and has been evaluated extensively [9]. Axial and coronal slices are used simultaneously to handle large out-of-plane rotation angles effectively. No statistically significant difference is found between the iMSP algorithm and two experienced neuroradiologists.

### 3.2 Potential Indexing Feature Extraction

Though lesions are usually obvious to a trained eye, automatic image segmentation is a very hard problem in medical image analysis [5]. We take an alternative approach for image feature extraction which does not require precise segmentation of the image: human brains present an approximate bilateral symmetry, from which pathological brains often depart. Even the brains where pathology appears symmetrically present a different kind of symmetry from normal brains. Our intention is to quantify and capture the statistical distribution difference of various brain asymmetries. A set of relatively simple and computationally inexpensive statistical image features is collected (Figure 3). After the iMSP is aligned in the middle of each 3D volumetric image (left most image in Figure 3), three types of asymmetry features are computed: 1) global statistical properties, 2) measures of asymmetry of halved and quartered brains, and 3) local asymmetrical region-based properties. These features are extracted from the original image (Fig. 3, image I) with its iMSP aligned, the difference image (Fig. 3, image D) of the original image and its mirror reflection with respect to iMSP, the

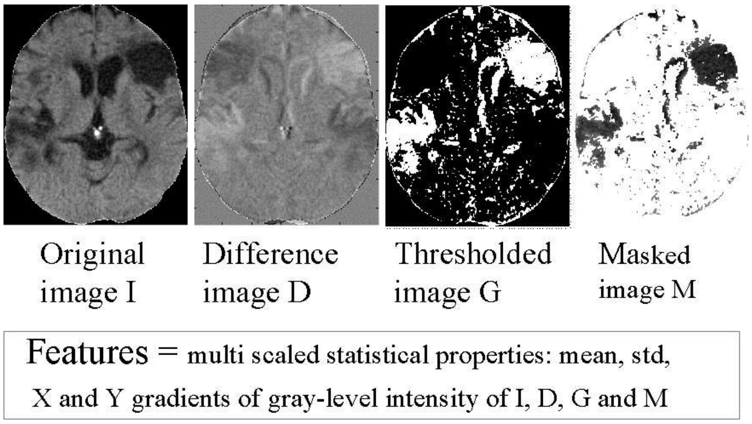


**Fig. 3. Image I**: original image where its 3D MSP aligned and centered vertically. **Image D**: the intensity difference image between the original image I and its own vertical reflection. **Image G**: a thresholded image of image D. **Image M**: the product of image G with original image I (a masked image).

thresholded difference image (image G[1]), and the original image masked by the thresholded binary image (Figure 3, image M). The image features include: the means, variances, X and Y gradients of the intensity images at different regions and under various Gaussian smoothing, scaling and thresholding. A total of 48 image features are computed from each image.

### 3.3 Classification-Driven Feature Subset Evaluation

Now each image $i$ becomes a vector $V_i$, and the length $N = |V_i|$ is the total dimension of the potential image feature space computed above. All the image data in the database form an $M \times N$ 2D sheet with $M$ image points in the $N$ dimensional feature space. A labeling, given by an experienced neuroradiologist, is added to the end of each row indicating the semantic class of the image, either simply the pathology type or pathology plus the anatomical location of the lesion(s).

We use a CMU proprietary combinatorial search engine called "Vizier" [14], to simultaneously find the image feature sub-dimensions and a proper kernel regression classifier that minimizes the leave-one-out cross-validation [12] on cross-entropy error of the training data given the classifier. The "Vizier" engine searches through a large set of possible classifiers within user specifications, and stops either when all possible choices are exhausted or a time limit given by the user is reached. Typically, we run a search between 20 minutes to a full hour on a standard PC. The output from Vizier is a specification of the best similarity metric found so far, consisting of the weight on each input feature attribute and a smoothing parameter $\sigma$.

[1] In our implementation, the absolute value of image G is used, whose value quantifies how asymmetrical the corresponding voxel regions are.

Most physicians prefer using a 2D slice as a query to access digital databases. Though our framework and system design are for 3D volumetric images, 2D slices can be treated as basic image units as well. In this study, we have used three pathology types: blood,infarct and normal brain images, and used 2D slices as basic units. The total 3D image set $S_{3D}$ is randomly divided (within each class) into a training set containing two thirds of $S_{3D}$ (31) and a hold-out test set containing one third of $S_{3D}$ (17), amounting to a total of 1250 2D slices. Care is taken to separate the test set and training set in such a way that there are no 2D slices from the same 3D brain belong to both training and testing sets.

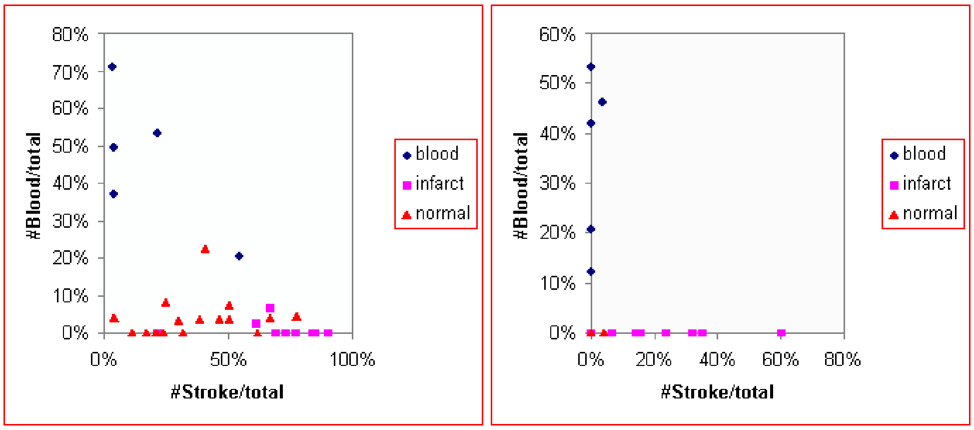


**Fig. 4.** The 3D image points are clustering in the two-ratio feature space while increasing the false negative penalty value of the 2D three-class classifier. The X-axis is $r_S$ and Y-axis is $r_B$. The relative weights learned by Vizier are 0.5 and 1 respectively.

### 3.4 Hierarchical Classifiers

An image classifier is simply a by-product of the process of feature subset evaluation through classification. The performance of the classifier may predict how well the selected feature subset will behave during image retrieval. For example, we can use a naive Bayes classifier as a 3D image classifier built on top of the 2D image classifiers, by using the ratios $r_S = n_S/n$ and $r_B = n_B/n$ as two 3D image features, where $n_N, n_S, n_B$ are the numbers of predicted normal, stroke and blood 2D images from the same 3D image, and $n = n_N + n_S + n_B$. The image distribution in this two-dimensional feature space can be observed in Figure 4. From these feature spaces, one can observe that normal and blood cases are better separated than the normal and stroke cases. Using Bayes law, this probability can be expressed as $P(c|[r_S, r_B]) = \frac{P([r_S,r_B]|c)P(c)}{P([r_S,r_B])}$. A *cost matrix* can be imposed to bias the classifier. For example, a *false negative penalty* $w > 1$ is incurred whenever a pathological image is classified as normal, whereas a unit cost or zero cost is incurred when a normal image is classified as pathological or when a class chosen is the correct class, respectively. The performance of such a 3D image classifier (when $w = 4$) is shown in the left of Figure 5.

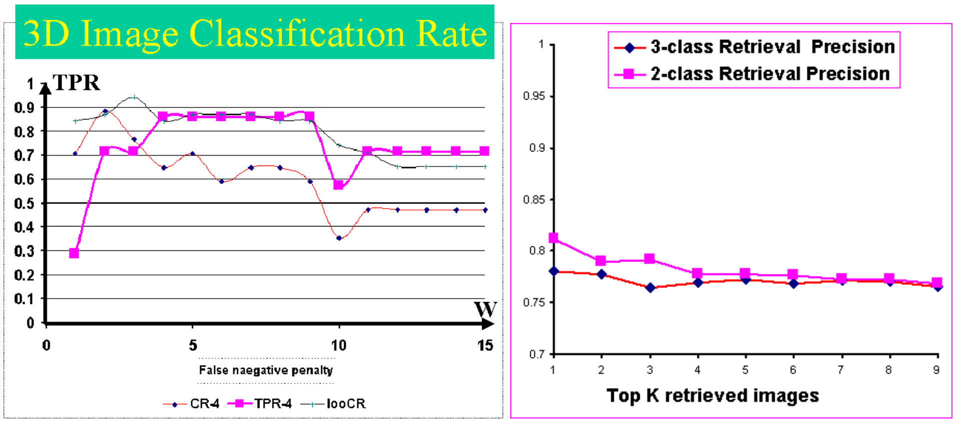


**Fig. 5.** Left (image classification, Bayes classifier): performance of a 3D image classifier on a randomly chosen, hold-out test-set when its false negative penalty $w = 4$. TPR-4: true positive rate. CR-4: classification rate. The horizontal axis is the 2D classifier false negative penalty value $w$. Right (image retrieval, kNN classifier): the mean value of retrieval precision as a function of the top $K$ retrieved images. 2-class: normal, pathology. 3-class: normal, infarct, blood.

### 3.5 Image Retrieval and Evaluation

Our approach is ideally suited to tuning an image similarity metric to the specific type of query that a user may submit to the system. As an example, one may simply want to know whether the query image is normal. In this case, the feature subset that is selected and optimally weighted for discriminating three different pathology types may not be optimal for a binary normal/abnormal discrimination. Thus we need to find an alternative feature subset. When we tried this, the locally optimal classifier found a different metric than the one from the ternary classification case. In both cases, Vizier finds a most discriminating feature subset containing 9-5 image features — a reduction in indexing feature dimension by nearly 5 to 10 fold. The similarity metric found in image classification can now serve as an image indexing vector for retrieving images in the reduced feature space. The right side of Figure 5 shows the mean retrieval precision rate for the hold out test images, one for the three-class case (normal, blood, infarct) and one for the two-class (normal, abnormal) case. One can observe a slightly better performance for the 2-class than for the 3-class image set. This is to be expected since 2-class classification leaves less space for errors to be made than the 3-class classification problem. However, in this problem the difficult class separation is between normals and strokes, which can be observed from the original image feature values and 2D classifier output (Figure 4), thus the improvement in 2-class is minimal.

Given near 80% precision rate on average for image retrieval, this result implies that in average 8 out of 10 top ranked retrieved images have the same pathology as the query image. Figure 6 shows two retrieval results for pathology cases: top – blood, bottom – infarct.

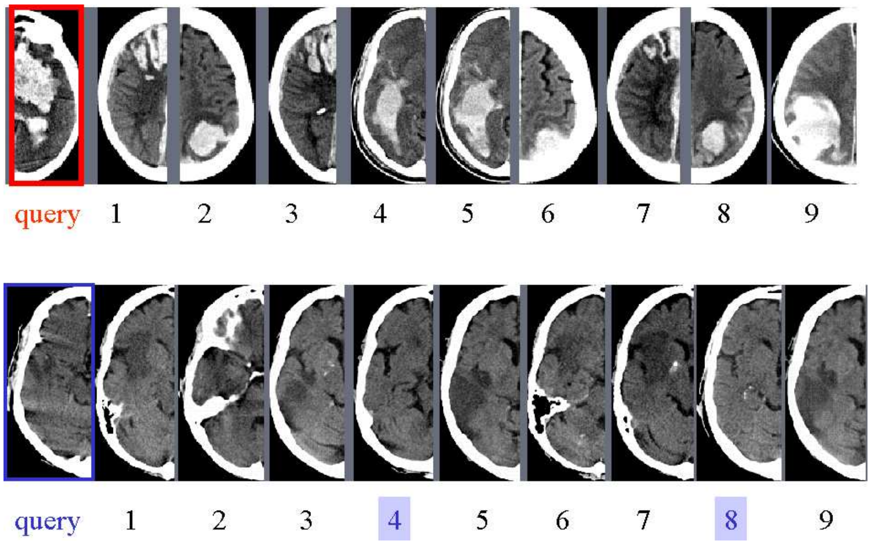


**Fig. 6.** Top left most: query image with acute blood, bottom left most: query image with bland infarct. The first nine retrieved half slices follow, from left to right in descending order of similarity. The pathologies on the top row retrieved images are all acute blood. The pathologies on the bottom row retrieved images are: infarct, infarct, infarct, normal, infarct, infarct, infarct, normal, infarct.

## 4 Conclusion and Future Work

In this work, we have demonstrated quantitatively the discriminating power of the statistical measurements of human brain asymmetry. One novelty of our approach in comparison to others in the medical image retrieval domain is to let the computer learn an image similarity metric suited for the given image semantics, instead of imposing such a metric by a human system designer subjectively. The main computational tools used in our study include memory based learning, Bayesian classification, and an effective search engine Vizier [14]. We have extended the basic framework of semantic-based image retrieval to image datasets other than pathology neuroimages, for example, multispectrum biology images [10].

Future work includes the study of a more comprehensive feature set, different weighted feature subsets for different query contents, and dynamic switching among them according to user feedback. We would also like to combine image features with collateral information such as age, gender and medical history of the patient to obtain a better retrieval rate. The basic framework presented here has provided us with such an *information fusion* capability.

## References


1. C. M. Bishop. *Neural Networks for Pattern Recognition.* Clarendon Press, 1995. ISBN:0198538499.
2. C. Davatzikos, M. Vaillant, S. Resnick, J.L. Prince, S. Letovsky, and R.N. Bryan. A computerized approach for morphological analysis of the corpus callosum. *Comp. Ass. Tomography.*, 20:88–97, Jan./Feb. 1996.
3. L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition.* Springer-Verlag, London, Paris, Tokyo, 1997.
4. R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification.* John Wiley & Sons, New York, 2001.
5. J.S. Duncan and N. Ayache. Medical image analysis: Progress over two decades and the challenges ahead. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):85,106, 2000.
6. D. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, and W. Equitz. Efficient and effective querying by image content. *Journal of Intelligent Information Systems*, 1994.
7. R. Guillemaud, M. Sakuma, P. Marais, J. Feldmar, R. Crow, L. DeLisi, A. Zisserman, and M. Brady. Cerebral symmetry analysis from mri scans. *Submitted to Psychiatry Research Neuroimaging*, 1998.
8. D. L. G. Hill, P.G. Batchelor, M. Holden, and D. J. Hawkes. Medical image registration. *Physics in Medicine and Biology*, 46:1–45, 2001.
9. Y. Liu, R.T. Collins, and W.E. Rothfus. Robust Midsagittal Plane Extraction from Normal and Pathological 3D Neuroradiology Images. *IEEE Transactions on Medical Imaging*, 20(3), March 2001.
10. Y. Liu, T. Zhao, E.S. Wachman, D.L. Farkas, and T. Kanade. Learning discriminant features in multispectrual biological images. In *Mathematical Methods in Biomedical Image Analysis.* submitted, 2001.
11. F. Maes, A. Collignon, D. Vandermeulun, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187,198, 1997.
12. A. W. Moore, D. J. Hill, and M. P. Johnson. An Empirical Investigation of Brute Force to choose Features, Smoothers and Function Approximators. In S. Hanson, S. Judd, and T. Petsche, editors, *Computational Learning Theory and Natural Learning Systems, Volume 3.* MIT Press, 1994.
13. A. Pentland, R.W. Picard, and S. Sclaroff. Photobook: Content-based manipulation of image databases. *IJCV*, 18(3):233–254, June 1996.
14. J. Schneider and A.W. Moore. A locally weighted learning tutorial using vizier 1.0. *tech. report CMU-RI-TR-00-18, Robotics Institute, Carnegie Mellon*, 1997.
15. P.M. Thompson, C. Schwartz, and A.W. Toga. High-resolution random mesh algorithms for creating a probabilistic 3d surface atlas of the human brain. *Neuroimage*, 3(1):19–34, Feb 1996.
16. J. West, M. Fitzpatrick, and et al. Comparison and evaluation of retrospective intermodality brain image registration techniques. *Journal of Computer Assisted Tomography*, 21:554–566, July/August 1997.
17. W. Zhang, S. Sclaroff, S. Dickinson, J. Feldman, and S. Dunn. Shape-based indexing in a medical image database. *Workshop on Biomedical Image Analysis*, pages 221–230, June 1998.