Distinguishing Among 3-D Distributions for Brain Image Data Classification

A. Lazarevic, D. Pokrajac, V. Megalooikonomou & Z. Obradovic

Center for Information Science and Technology, Temple University
Philadelphia, PA 19122, USA

ABSTRACT: To facilitate the process of discovering brain structure-function associations from image and clinical data, we have developed classification tools for brain image data that are based on measures of dissimilarity between probability distributions. We propose statistical as well as non-statistical methods for classifying three dimensional probability distributions of regions of interest (ROIs) in brain images. The statistical methods are based on computing the Mahalanobis distance and Kullback-Leibler distance between a new subject and historic data sets related to each considered class. The new subject is predicted to belong to the class corresponding to the dataset that has the smaller distance from the given subject. The non-statistical methods consist of a sequence of partitioning the brain image into hyper-rectangles followed by applying supervised neural network models. Experiments performed on synthetic data representing mixtures of nine distributions as well as on realistic brain lesion distributions from a study of attention-deficit hyperactivity disorder (ADHD) after closed head injury showed that all proposed methods are capable of providing accurate classification of the subjects with the Kullback-Leibler distance being the least sensitive on the size of the training set and on information about the new subject. The proposed statistical methods provide comparable classification to neural networks with appropriately generated attributes, while requiring less computational time.

INTRODUCTION

Data mining in brain imaging is proving to be an effective methodology for providing prognosis, treatment, and a better understanding of brain functionality. The detection of relationships between human brain structures and brain functions (i.e., human brain mapping) has been recognized as one of the main goals of the Human Brain Project [1]. Development of large databases [2,3,4] for the purpose of meta-analysis of data pooled from multiple studies is now funded by several government initiatives worldwide. These databases consist of a large collection of studies that include 3-D images from different medical imaging modalities that capture structural (e.g., MRI[1], CT[2]) and functional /physiological (e.g., PET[3], fMRI[4]) information about the human brain. Traditionally, two approaches have been used in functional brain mapping. The first approach seeks associations between lesioned structures and neurological or neuropsychological deficits. The second approach seeks associations between brain activations patterns and tasks performed. Independent of the approach used, a current obstacle in human brain mapping is the lack of methods to automatically classify ROIs (i.e., lesions, brain activations, etc) and quantitatively measure their levels of similarity.

In this paper, in order to assist the process of discovering brain structure-function associations from image and clinical data and to make retrieval of similar brain scans possible, we have developed statistical and neural network methods for classification of brain image data based on measures of dissimilarity between 3-D probability distributions. Although the proposed methods can be used in classifying any type of ROIs here we apply them to lesion-deficit analysis and MR data sets. Given a clinical image of a new subject that contains a number of lesioned voxels, the goal is to determine whether it belongs to a group of subjects who did or did not develop attention-deficit hyperactivity disorder (ADHD) after closed head injury.

In brain mapping, behavioral and image data are collected from patients and analyzed in order to detect associations among spatial regions of the brain and their functions. The image data resulting from scanning of the patient are multiple layers of images that are combined into a voxel-based 3-D representation. The first step in the process is to make data comparable across subjects. In particular for image data, the ROIs are identified (segmented) and image registration is performed to bring the patient's image data into register, i.e., spatial coincidence, with a common spatial standard (i.e., a reference brain or anatomical atlas). The methods that we present are applied after these pre-processing steps are performed.

BACKGROUND AND RELATED WORK

Let $\boldsymbol{x}$ be a multivariate random variable that can assume any of the values from a multidimensional domain D. We denote by $P(\boldsymbol{x})$ the probability that $\boldsymbol{x}$ falls into a subdomain $V \subset D$. More precisely,

$$P[\boldsymbol{x} \in V] = \int_V p(\boldsymbol{x}) \cdot d\boldsymbol{x}$$

[1] MRI: Magnetic Resonance Imaging
[2] CT: Computed Tomography
[3] PET: Positron Emission Tomography
[4] fMRI: Functional MRI

where $p(\boldsymbol{x})$ is a probability density function satisfying non-negativity ($p(\boldsymbol{x}) \geq 0$) and normalization ($\int_V p(\boldsymbol{x}) \cdot d\boldsymbol{x} = 1$) conditions. The probability density function $p(\boldsymbol{x})$ uniquely determines a distribution of vectors $\boldsymbol{x}$, drawn from the distribution. Each distribution can be characterized by its histogram [5] and a parametric one can also be specified by its parameters [5, 6].

The problem stated in the introduction can be formulated as follows. Let r_{xyz} denote the value of a *voxel* (volume element) of a 3-D brain image (volume). In our study, a voxel has a value $r_{xyz} = 1$ if it belongs to a lesion (such voxels are referred to as "lesioned voxels"). Otherwise, the voxel has a zero value ($r_{xyz} = 0$). Given two sets S_Y and S_N that contain coordinates of lesioned voxels for N subjects who did or did not develop ADHD respectively, the task is to identify whether a data set s_z, that corresponds to lesioned voxels of a new subject, comes from the same distribution as the set S_Y or the set S_N. Therefore, the objective is to characterize the distribution of the new data set s_z and to compare it to two given distributions corresponding to subjects who did or did not develop ADHD. Methods for distinguishing among distributions can in general be categorized into:

- distance based methods
- maximum likelihood methods

Distance based methods rely on an appropriately defined distance measure between distributions in order to determine to which existing distribution a new distribution is closer to. Frequently used distances include Euclidean distance, Mahalanobis distance [7], Bayesian distance [6], Patrick-Fisher distance [8], Bhattacharyya distance [9] and Kullback-Leibler distance [10].

The Euclidean distance between two vectors depends on the sum of squared diffrences of their components. Therefore, given two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, the Euclidean distance between them is computed as $d_E = \sqrt{(\boldsymbol{x}-\boldsymbol{y})^T \cdot (\boldsymbol{x}-\boldsymbol{y})}$. Multivariate data with normal distribution tend to cluster about the mean vector $\boldsymbol{\mu}$, falling in an ellipsoidally shaped cloud whose principal axes are eigenvectors of the covariance matrix [10]. When computing Mahalanobis distance, this fact is considered by including a covariance matrix Σ into calculation. Therefore, Mahalanobis distance between two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, is measured as $d_M = \sqrt{(\boldsymbol{x}-\boldsymbol{y})^T \cdot \boldsymbol{\Sigma}^{-1} \cdot (\boldsymbol{x}-\boldsymbol{y})}$. The Mahalanobis distance equals to Euclidean distance only when the covariance matrix $\boldsymbol{\Sigma}$ is an identity matrix. The Bayesian distance [6] differs from the Mahalanobis distance by incorporating information regarding the size of the distribution, as well as the *a priori* probability of the class. The former accommodates the normalization requirement, while the latter merely offsets the distance according to the relative frequency that one class occurs compared to another. Due to the offsets, the Bayesian "distance" does not satisfy the non-negativity requirement for metrics [6].

The Patrick-Fisher distance [8] between two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ is measured as $d_{PF} = \left(\frac{\boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_y}{2}\right)^{-1}(\boldsymbol{x}-\boldsymbol{y})$, where $\boldsymbol{\Sigma}_x$ and $\boldsymbol{\Sigma}_y$ are the corresponding covariance matrices.

The Bhattacharyya distance between two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined as $d_B = \frac{1}{8}(x-y)^T\left(\frac{\boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_y}{2}\right)^{-1}(x-y) + \frac{1}{2}\ln\frac{\left|(\boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_y)/2\right|}{\sqrt{\left|\boldsymbol{\Sigma}_x\right| \cdot \left|\boldsymbol{\Sigma}_y\right|}}$, where $\boldsymbol{\Sigma}_x$ and $\boldsymbol{\Sigma}_y$ have again the same meaning [9].

Unlike the previously mentioned distances, the Kullback-Leibler (KL) distance $d_{KL}(p(\cdot), q(\cdot))$ is defined as a measure of similarity between two distributions and is equivalent to relative entropy. Although nonnegative, and equal to zero only between the same vectors, the Kullback-Leibler distance is not a true metric, since it is not necessarily symmetric ($d_{KL}(p(\cdot), q(\cdot)) \neq d_{KL}(q(\cdot), p(\cdot))$) and does not satisfy the triangle inequality.

In maximum likelihood methods, given a new distribution (new subject in our application) and estimated probability densities of existing distributions, a likelihood that a new distribution is same as one of existing distributions is computed. To perform a maximum likelihood technique, probability densities of the distributions should be estimated, which can be performed using parametric, non-parametric or semi-parametric techniques [6].

Various statistical and neural network techniques have been applied to different problems in brain image data analysis from image segmentation and registration [11] to detection of electromagnetic field sources [12,13,15] and analysis of fMRI activations. The divergence between probability distributions based on the Kullback-Leibler distance has been used in the analysis of the fMRI signal in order to construct a brain activation map [15]. In other studies, using statistical tests, the likelihood for particular voxels to exhibit significant changes between conditions is estimated [16]. Statistical methods such as SPM (statistical parametric mapping) [17,18,19] are of great value in the analysis of fMRI activations but they do not automatically classify or compare activations. Data mining methods have been recently applied to brain images in order to discover associations between lesions and deficits [20] (the interested reader can see [21] for a complete treatment). However, little work has been done in brain image data classification and efficient discovery of associations between structures and functions.

METHODOLOGY

In this paper, we present statistical and neural network methods for classifying three-dimensional probability distributions of regions of interest (ROIs) in brain images. The statistical methods are based on computing the Mahalanobis and Kullback-Leibler distances. The distances are computed between a new sample (subject) and data sets related to each considered class (distribution). In the neural network method, the brain images are partitioned into three-dimensional hyper-rectangles and the neural networks are then applied on the obtained hyper-rectangles.

Statistical Distance Based Methods

Given two data sets S_Y and S_N containing lesioned voxels respectively from two classes with subjects who did and did not develop ADHD, the task is to classify a new subject to one of these two classes. The new subject is specified through a data set s_z containing a number of lesioned voxels. Therefore, the

new sample s_z is predicted to belong to the class that corresponds to one of the datasets S_Y or S_N, which has the smaller distance from the given subject.

The Mahalanobis distance and the Kullback-Leibler (KL) distance are considered in this paper. Given a new data set s_z, the Mahalanobis distance between the new subject s_z and an existing data set S (S_Y or S_N) is computed as:

$$d_M = \sqrt{(\mu_{s_z} - \mu_S)^T \cdot \Sigma^{-1} \cdot (\mu_{s_z} - \mu_S)}$$

where μ_{s_z} and μ_S are mean vectors of the data sets s_z and S respectively, and Σ is the pooled sample covariance matrix [5] given as:

$$\Sigma = \frac{(m_z - 1) \cdot \Sigma_{s_z} + (m - 1) \cdot \Sigma_S}{(m_z + m - 2)}$$

with Σ_{s_z} and Σ_S denoting covariance matrices of data sets s_z and S, respectively.

The Kullback-Leibler distance for distinguishing between the new subject s_z and an existing data set S (S_Y or S_N) is defined as:

$$d_{KL}(s_z, S) = \int_D p_z(x) \ln \frac{p_z(x)}{p(x)} \cdot dx$$

where $p_z(x)$ and $p(x)$ are probability densities corresponding to the distributions from which data sets s_z and S are drawn respectively. Although $d_{KL}(s_z, S) + d_{KL}(S, s_z)$ is a true distance metric, we consider only $d_{KL}(s_z, S)$ thus consistently measuring the divergence between s_z and S. However, in order to compute this Kullback-Leibler divergence, these distributions need to be estimated beforehand. In the following text, we present the algorithm for this estimation.

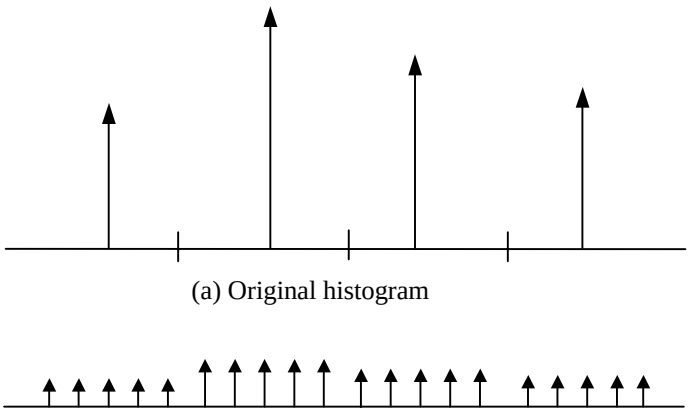
Since the data sets S_Y, S_N and s_z obtained from medical imaging or simulation contain discrete values for lesioned voxels, here Kullback-Leibler divergence is computed using the following discrete approximation:

$$d_{KL}(S_Z, S) = \sum_D p_z(x) \ln \frac{p_z(x)}{p(x)} \cdot \Delta x$$

where $p_z(x)$ and $p(x)$ are estimated discrete probability densities, and Δx is the product of discretization intervals in each dimension.

The estimation of distribution histograms is performed using the following procedure.

1. *Discretization*. The brain image data set is given with lesioned voxels inside the domain that is discretized into a $N_x * N_y * N_z$ three-dimensional grid. For each of the three dimensions, the interval $[1,N_i]$ (i = x, y, z) is divided into k equal intervals thus resulting in k^3 equal three-dimensional hyper-rectangles. The initial histogram is obtained by approximating the distribution in each 3-D hyper-rectangle by the ratio of lesioned voxels that fall inside the hyper-rectangle and the total number of lesioned voxels inside the brain volume [6].
2. *Histogram padding*. After performing the *discretization* step, the number of histogram hyper-rectangles is smaller than the number of possible discrete location of voxels. In order to match the resolution of original data, the histogram resolution is increased such that the representative value for each 3-D cube is repeated k times. The histogram is then scaled such that the sum of all values still equals to one (Figure 1). After performing this operation, the number of histogram values is equal to the number of hyper-rectangles in an original $N_x * N_y * N_z$ three-dimensional grid.



(a) Original histogram

(b) Histogram after padding

Figure 1. Illustration of histogram padding for 1-D histograms. Each of the values (represented as arrows) is converted into r = 5 equal values that correspond to new smaller bins.

3. *Histogram smoothing*. To avoid the problem of discontinuity between histogram values at boundaries of initial k^3 bins [6], padded histograms are smoothed using a 3-D low-pass filter with a specified window of size w^3 discretization intervals. A simple filter whose output is the average of the inputs within the window with size w (Figure 2) is applied.

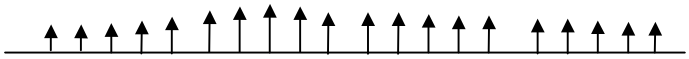


Figure 2. Smoothed histogram after applying filtering with w=2 to padded histogram from Fig 1b.

4. *Histogram modification*. Due to the presence of the finite number of voxels in the brain image data set, it is possible that some 3-D bins in the estimated histogram do not contain any examples [6], even after the histogram smoothing performed in *Step* 3. Hence, in that case, the estimated density corresponding to empty 3-D cube would be zero although the true density might be not. This may cause problems in computing the Kullback-Leibler distance, since some of the values $p_1(\mathbf{x}) \cdot \log(p_1(\mathbf{x}) / p_2(\mathbf{x}))$ from a sum will be computed as infinity. To avoid this problem, a small positive value is added to estimated densities, such that the normalization condition (sum of densities = 1) remains satisfied.

By estimating the histograms using the procedure explained above, we avoid averaging of histograms for a comparatively small number of samples that could occur if the histogram is directly estimated on 3-D hyper-rectangles corresponding to the original discretization intervals.

Neural Network Method

An optional method for classifying new subjects based on neural networks is proposed, since such universal approximators

were often reported to outperform the alternatives for classification of real life non-linear phenomena [14].
In order to apply a neural network model to the problem of classification between classes of subjects who did and did not develop ADHD, the value of each voxel can be treated as an attribute. However, the number of voxels in a brain image is usually in the order of 10^7 and training a neural network model with such large number of input attributes is infeasible in many realistic applications, due to the curse of dimensionality effect as well as computational issues of non-linear optimization. Therefore, there is a need to reduce the number of attributes that are used for constructing a neural network model.

For attribute reduction, our proposed method first partitions the brain volume into a number of 3-D hyper-rectangles. In order to be able to compare the obtained classification results to the results achieved by statistical methods, as a partitioning algorithm we use the same discretization procedure explained at Step 1 in the previous section. The voxels inside the small 3-D hyper-rectangles are averaged over the total number of voxels inside 3-D hyper-rectangles and these averaged values of voxels are treated as new attributes for training with the neural network classification model. For determining the minimal number of hyper-rectangles sufficient for successful classification using the neural network, discretization is incrementally increased until satisfactory classification accuracy is achieved.

We trained multilayer (2-layered) feedforward neural network classification models with the number of hidden neurons equal to the number of input attributes, although experiments with a fixed number of hidden neurons were also performed. The neural network classification models had the number of output nodes equal to the number of classes, where the predicted class was from the output with the largest response. We used two learning algorithms: resilient propagation [22] and Levenberg-Marquardt [23].

EXPERIMENTAL RESULTS

Our experiments were tested on synthetic data and on realistic brain lesion distributions generated using a lesion-deficit simulator [24].

Experiments on synthetic data

Synthetic data used in our experiments contained samples from two mixtures of nine normal distributions. We were varying the parameters (means and variances) of mixture components, thus constructing different mixtures of distributions (see Figure 3).

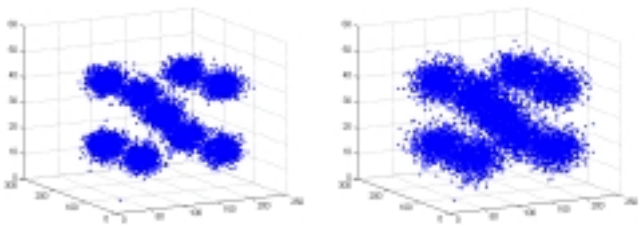
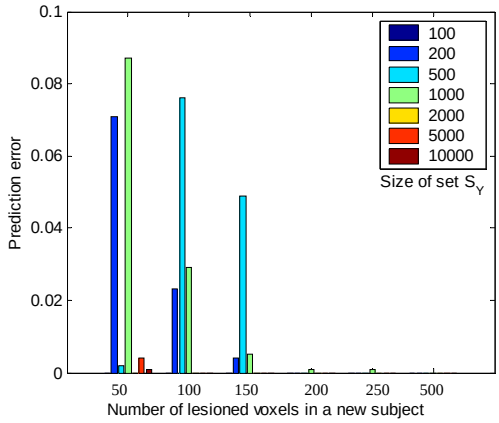


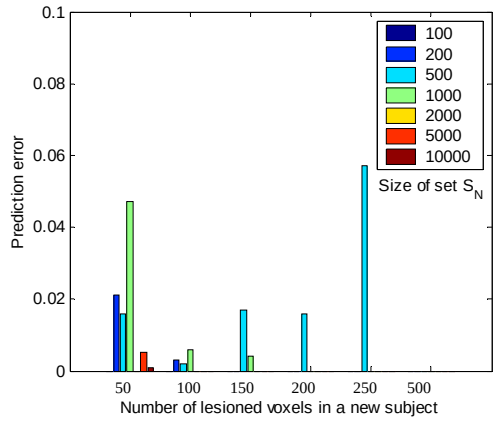
Figure 3. Two mixtures of distributions that differ only in the variance of the distribution components

In the first series of experiments, the distribution components had the same variances but different means for each class. We have repeated the experiments through 200 rounds, and each round consisted of random drawing of a new subject from one of the classes. The classification performance was monitored by measuring accuracy rate as the ratio of the number of rounds when a new subject was correctly classified and the total number of rounds. The subjects contained the number of lesioned voxels that varied from 50 to 500.

When using the Mahalanobis distance, we were able to adequately classify a new set of samples that belonged to one of two mixtures in 90% to 99% of cases, depending on the size of sets S_Y, S_N and the number of lesioned voxels in a new subject (Figure 4). Analyzing the charts from Figure 4, it can be noticed that the prediction error of our classification methods decreased when the size of sets S_Y, S_N increased and when the number of lesioned voxels increased too.



a) Subjects who belong to the first distribution



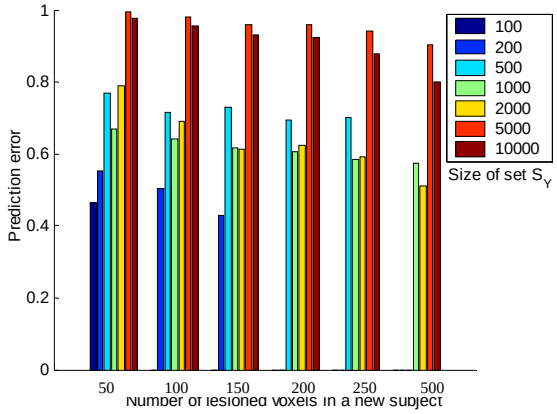
b) Subjects who belong to the second distribution

Figure 4. The prediction error when classifying new subjects from two distributions with different means using Mahalanobis distance. Mean difference was 0.6.

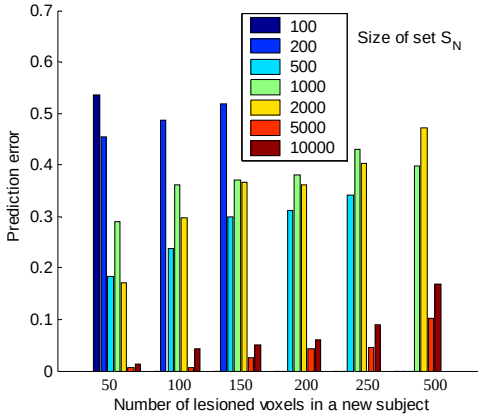
Unlike the method with Mahalanobis distance, statistical method based on computing Kullback-Leibler (KL) distance and neural network classification method, achieved almost perfect classification, for all considered sizes of data sets and numbers of lesioned voxels in a new subject (the prediction error was less then 1%).

Another group of experiments on synthetic data involved mixtures that had the same component means but different component variances for each one of the classes. In this case, classification was typically more challenging since the mixture of distributions with smaller variances is overshadowed by the mixture with larger variances. When using Mahalanobis distance in this scenario, the achieved classification accuracy was very low when predicting the mixture with smaller variances (from 0% to 50%), and significantly higher when

predicting the mixture with larger variances (from 50% to 99%) (Figure 5).



a) Subjects who belong to the distribution with smaller variance



b) Subjects who belong to the distribution with larger variance

Figure 5. The prediction error when classifying new subjects from two distributions with different component variances using Mahalanobis distance. The variance of distributions was 0.01 and 0.1.

The method based on computing the KL distance was more successful in predicting new subjects when they belonged to one of the distributions. When predicting the mixture of distributions with smaller variance, the accuracy varied from 14% for the small size of the set S_Y to 99% for the larger size of set S_Y (Figure 6), which is much better than using Mahalanobis distance. When predicting the mixture of distributions with larger variances, the method with KL distance was able to perform almost perfect classification in all cases (error less than 1%).

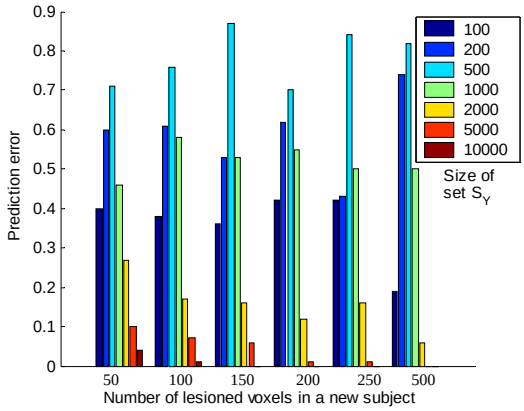


Figure 6. The prediction error when classifying new subjects from the distribution with smaller component variances using KL distance. The variance of one distributions was 0.01 and 0.1.

By applying the proposed neural networks method to the mixture of distributions when they differ only in variance components, we failed to achieve good classification when using only 2^3 3-D hyper-rectangles (accuracy was around 50%). However, when we used 3^3 3-D hyper-rectangles, the classification accuracy drastically improved (from 55% for small number of lesioned voxels and small sizes of set S_Y to 99% for large size S_Y and large number of lesioned voxels in a new subject). The achieved accuracy for both distributions was similar, and only the accuracy for the distribution with smaller variance is reported in Figure 7. The reason for poor performance of neural network classification models when using only 8 3-D hyper-rectangles was in totally overlapping distributions in obtained hyper-rectangles, and similar averaged number of lesioned voxels in those hyper-rectangles, such that the discriminative attributes were not relevant.

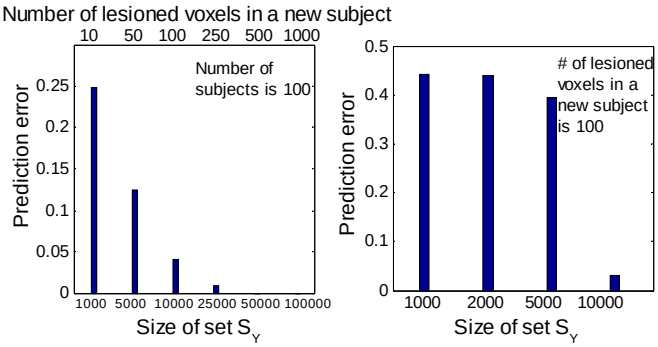


Figure 7. The prediction error when classifying new subjects by partitioning into $\mathbf{3^3}$ 3-D hyper-rectangles and then applying neural networks

Analyzing the Figures 5-7, it is evident that the method with Mahalanobis distance was inferior to the method of computing KL distance and to the neural network model. The more detailed comparison of all proposed methods for chosen size of set S_Y and the number of lesioned voxels is shown in Table 1, which also indicates the poor quality of method with Mahalonobis distance. This was probably due to the fact that the Mahalanobis distance is based on the assumption that compared distributions are normal ones.

Table 1. The comparison of prediction errors among proposed methods for classifying a new subject (The number of subjects was 5, and the number of lesioned voxels per subject was 200)

		Distribution components differ in	
Method		Mean (0.6)	Variance (1. 0.01, 2. 0.1)
Mahalanobis		1.9	64.0
Kullback-Leibler		0	57.0
Neural Network	2^3 rectangles	3.0	50.0
	3^3 rectangles	0.3	44.3

Experiments on realistic data

The segmentation of ROIs in the study was performed manually by a neuroradiologist using thresholding. A nonlinear method based on a 3-D elastically deformable model [24] was used to register the ROIs to the Tailarach anatomical atlas [25]. After a normalization of image data to a common coordinate system, we applied the proposed methods to lesion-deficit analysis and magnetic resonance imaging data sets. We performed classification of realistic brain lesion distributions that were generated using a lesion-deficit simulator [26] with the spatial statistical model conforming to the Frontal Lobe Injury in

Childhood (FLIC) study [27]. The subjects were classified into two classes according to subsequent development of ADHD after closed head injury. Therefore, there were two distributions corresponding to subjects who developed ADHD ("yes ADHD" class) and did not develop ADHD ("no ADHD" class) (Figure 8). Given a new subject with a set of lesioned voxels, the goal was to determine the more plausible class. The subjects contained the number of lesioned voxels that varied from 50 to 500, although in the specific FLIC study [26] approximately 200 lesioned voxels is present on average per 3-D brain image (i.e., per subject).

In experiments, we varied both the size of data sets for the classes and the number of lesioned voxels belonging to a new subject. For each combination of these parameters, we performed the experiments through a specified number of rounds (200 in our experiments). Each round consisted of random drawing of a new subject from one of the classes. The classification performance was again monitored by measuring accuracy rate computed as for synthetic data.

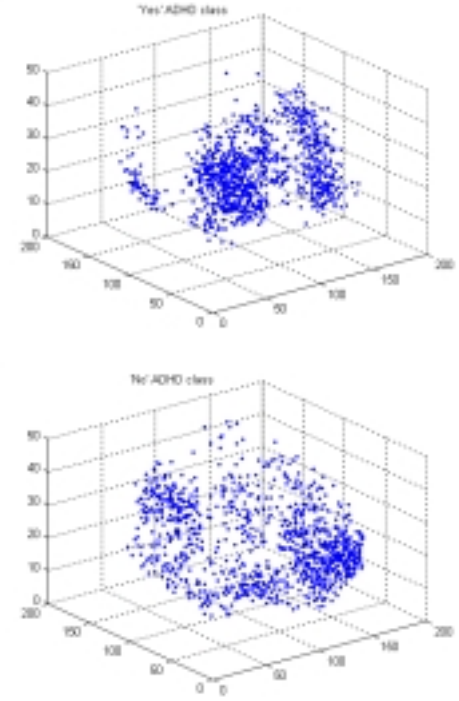
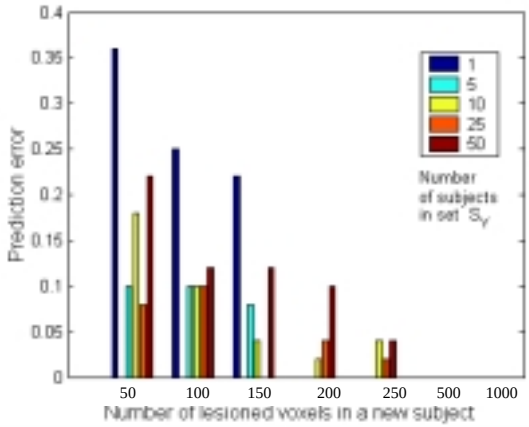


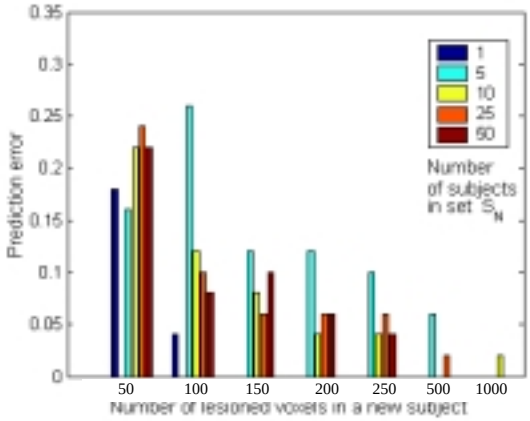
Figure 8. Distributions for "yes ADHD" and "no ADHD" class

Experiments on realistic brain lesion distributions showed that the proposed method based on Mahalanobis distance could provide more reliable and more accurate classification between the subjects regarding the development of ADHD than when classifying the subjects from synthetic data. Figure 9 demonstrates that the classification with error less than 10% was possible both for the subjects who did and who did not develop ADHD, when a sufficient knowledge of the distribution corresponding to the subject was available. This was apparent especially when 150 or more lesioned voxels are available for a new subject. The prediction was perfect (0% error) when the number of lesioned voxels in a new subject was larger than 1000. It is interesting to notice that the classification accuracy was slightly better when predicting subjects in "yes" ADHD class than in "no ADHD class" (Figure 9).

The method based on Kullback-Leibler (KL) distance was even more successful in classification of new subjects, especially when classifying subjects with a small number of lesioned voxels comparable to the size of data sets for the classes (Figure 10). When the number of lesioned voxels per a subject was 50 or 100, the prediction error was less than 2% and when more than 200, the prediction was always perfect (0% prediction error).



a) Subjects who developed ADHD



b) Subjects who did not develop ADHD

Figure 9. The prediction error when classifying new subjects using Mahalanobis distance

When performing the proposed method of partitioning the brain image into 3-D hyper-rectangles and applying neural network models, the prediction results were again comparable to KL method. For small number of lesioned voxels in a new subject (10, 50, 100), prediction accuracy achieved by neural network was better that using statistical methods. However, when this number of lesioned voxels increased, there was no significant increase in prediction accuracy achieved by neural networks. Our experiments have shown that the number of 3-D hyper-rectangles sufficient for satisfactory prediction was fairly small (only 2^3 and 3^3 3-D hyper-rectangles).

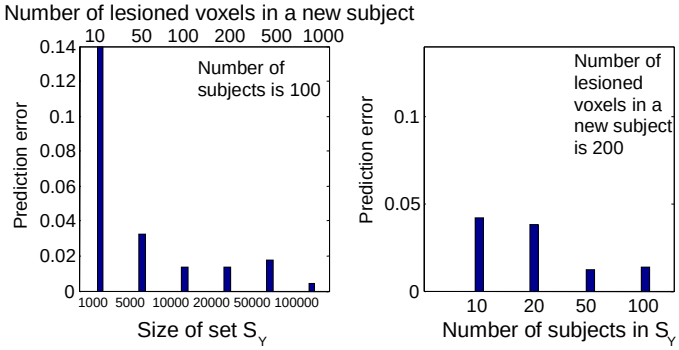


Figure 10. The prediction error when classifying new subjects by partitioning into $\mathbf{2^3}$ hyper-rectangles and then applying neural networks

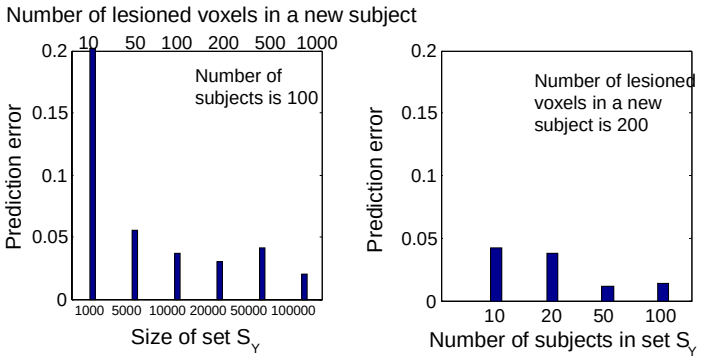


Figure 11. The prediction error when classifying new subjects by partitioning into $\mathbf{3^3}$ hyper-rectangles and then applying neural networks

The more comprehensive evaluation of all proposed methods applied on realistic data is shown in Table 2. It is evident again that the method with Mahalanobis distance is inferior when comparing to the method with KL distance. On the other side, neural network models were not so successful as for synthetic data, but they are still able of achieving comparable accuracy to the method with computing KL distance.

Table 2. The comparison of prediction errors among proposed methods for classifying a new subject (The number of subjects was 20 and 50, while the number of lesioned voxels per subject was fixed to average number of voxels per brain volume – 200)

Prediction Error (%)		(# of subjects, # of lesioned voxels)	
Method		(20, 200)	(50, 200)
Mahalanobis		10.0	10.0
Kullback-Leibler		0	0
Neural Network	2^3 rectangles	3.8	1.3
	3^3 rectangles	4.0	2.0

DISCUSSION

We proposed several methods for distinguishing between the distributions of lesioned voxels on MRI images for subjects who did and did not develop ADHD. There are several assumptions on which our methods are based. The proposed techniques operate within the accuracy of the segmentation of registration procedures used. Finally, the experiments are performed under the assumption that all the subjects in training and test sets have the same number of lesioned voxels (equal to the average number obtained from FLIC study). However, more realistic results could be obtained using data sets where the number of lesions per subject and their size follow corresponding distributions observed on real-life data [26].

In this paper, we propose non-parametric methods for modeling densities of data distributions. While more versatile, such methods typically require a higher number of data examples (total number of lesioned voxels) compared to parametric ones. Work in progress includes the examination of parametric methods for learning distributions, such as expectation-maximization algorithm [10] and clustering algorithms for partitioning distributions into distinct regions [28,29].

The proposed technique for the histogram computation involved the initial estimation on a coarse grid, followed by interpolation and smoothing in order to obtain histograms on the same grid as the underlying data. Since the voxel density is uniform in all three dimensions, during the estimation of initial histograms, we maintained the same number of discretization intervals in each direction. However, it is possible to avoid the interpolation phase by direct estimation of histograms on the original grid (where the number of discretization intervals for histogram estimation is equal to the data resolution) followed by three-dimensional filtering to smooth estimated values. Therefore, future work is necessary to determine the real necessity for this interpolation. However, the smoothing phase seems to be necessary to ensure statistically significant estimation, since the number of data points (lesioned voxels) is typically much smaller than the number of bins that correspond to the original grid.

The proposed neural network methods use attributes generated by averaging the lesioned voxels in closed spatial subregions. These methods are successful in classification if the number of lesioned voxels in these subregions differs significantly for two considered distributions. If the distributions have sufficiently different means, this condition is satisfied for a small number of subdomains. However, when distinguishing distributions with the same means and similar variances, results comparable to those obtained using statistical methods can be achieved only for a sufficiently large number of subdomains (hyper-rectangles), which was evident from the results on synthetic data.

In this paper, the subregions are of equal size to the hyper-rectangles used for histogram estimation and we did not attempt to optimize the number and the shape of these subregions. Work in progress includes the development of an adaptive procedure for determining appropriate partitioning of the three-dimensional domain into subdomains for optimal attribute generation.

Results on synthetic data suggest that all proposed methods provide almost perfect classification if the underlying distributions differ significantly in their means. However, when distributions differ slightly in variances of mixture components, the method based on Mahalanobis distance gradually ceased to provide useful classification, since these methods implicitly assume the normality of data distribution. In contrast, the method based on Kullback-Leibler distance provided good results, assuming that the size of training sets were large enough so that histograms of underlying distributions could be properly estimated.

Results on realistic data generated using a lesion-deficit simulator suggest that the proposed techniques are applicable for classification on subjects that have the number of lesioned voxels close to real-life cases (the average of 200 per 3-D brain image in the specific FLIC study [26]). Higher number of available subjects for model training usually resulted in higher accuracy. Correct classification in 95% cases was achieved when the number of lesioned voxels in a new subject was 100-150 or more. Among all examined methods, the Kullback-Leibler method was the least sensitive on the size of training set and the number of lesion voxels. The obtained results suggest that the proposed statistical methods provide comparable classification accuracy to neural networks with appropriately generated attributes, while requiring less computational time.

In general, all proposed methods have been shown capable of providing accurate classification of the subjects regarding the development of ADHD. In addition to lesion-deficit analysis,

the proposed techniques are applicable not only to the discussed domain but also to a much wider class of problems involving task-activation analysis and classification of 3D-probabilistic activation maps, such as those generated by statistical parametric maps (SPM).

ACKNOWLEDGMENTS

This work was partially supported by the National Science Foundation under Grant No. ISI-0083423.

REFERENCES

1. Koslow, S.H., Huerta, M.F., *NeuroInformatics: an Overview of the Human Brain project*, Mahway, NJ, Lawrence Erlbaum, (1997).
2. Arya, H., Cody, W., Faloutsos, C., Richardson, J. and Toga, A., A 3D medical image Database Management System, *International J of Computerized Medical Imaging and Graphics*, 20, 269-284 (1996).
3. Letovsky, S. I., Whitehead, S. H., Paik, C. H., Miller, G. A., Gerber, J., Herskovits, E. H., Fulton, T. K. and Bryan, R. N., A brain image database for structure/function analysis, *American J of Neuroradiology*, 19, 1869–1877, (1998).
4. Fox, P., Functional brain mapping with positron emission tomography, *Semin. Neurol.*, 9, 323-329, (1989).
5. Flury, B., *A First Course In Multivariate Statistics*, New York, Springer-Verlag, (1997).
6. Bishop, C., *Neural Networks for Pattern Recognition*, Oxford University Press, (1995).
7. Fukunaga, K., *Introduction to Statistical Pattern Recognition*, Academic Press, San Diego, CA, (1990).
8. Aladjem, M., Nonparametric discriminant analysis via recursive optimization of Patrick-Fisher distance, IEEE Trans. on Syst. , Man, Cybern, 28B, 2, 292-299, (1998).
9. Bhattacharyya, A., On a measure of divergence between two statisticl populations defined by their probability distributions, Bulletin of the Calcutta Mathematical Society, 35, 99-110, (1943).
10. Duda, R., Hart, P., Stork, D, *Pattern Classification*, John Wiley and Sons, (2000).
11. Maintz, A., Meijering, E. H., Viergever, M. A., General Multimodal Elastic Registration Based on Mutual Information, In Hanson, K. M., (Eds), Medical Imaging, 3338, 144-154, (1998).
12. Vigario, R., E. Oja, Independence: a new criterion for the analysis of the electromagnetic fields in the global brain?, *Neural networks*, 13, 8-9, 891-907, (2000).
13. Ropero Peaez, J., Towards a neural network based therapy for hallucinatory disorders, *Neural networks*, 13, 8-9, 1047-1061, (2000).
14. Haykin, S., Neural Networks, A Comprehensive Foundation, Prentice Hall, (1999).
15. Zhao, Q., Principe, J., Bradley, M., Lang, P., fMRI Analysis: Distribution Divergence Measure based on Quadratic Entropy, *Proc. On Human Brain Mapping*, (2000).
16. Friston, K.J., Holmes, A.P., Price, C.J., Buchel. C. and Worsley, K.J., Multisubject fMRI Studies and Conjunction Analyses, *NeuroImage*, 10, 4, 385—396, (1999).
17. Missimer, J., Knorr, U., Maguire, R. P., Herzog, H., Seitz, R. J., Tellman, L. and Leenders, K. L., On two methods of statistical image analysis, *Human Brain Mapping*, 8, 245–58, (1999).
18. Friston, K.J., Holmes, A.P., Worsley, K. J., Poline, J. P. , Frith, C. D. and Frackowiak, R. S. J., Statistical Parametric Maps in Functional Imaging: A General Linear Approach, *Human Brain Mapping*, 2, 189–210, (1995)
19. Turner, R. and Friston, K., SPM Course 1997 Notes, Wellcome Department of Cognitive Neurology, (1997).
20. Megalooikonomou, V., Davatzikos, C. and Herskovits, E., Mining lesion-deficit associations in a brain image database, *Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Diego, CA, (1999).
21. Megalooikonomou, V., Ford, J., Shen, L., Makedon, F. and Saykin, A., Data mining in brain imaging, *Statistical Methods in Medical Research*, 9, 4, 359-394, (2000).
22. Riedmiller, M., Braun, H., A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm, *Proc. of the IEEE International Conference on Neural Networks*, 586–591,(1993).
23. Hagan, M., Menhaj, M.B., Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks* 5, 989-993, (1994).
24. Davatzikos, C., Spatial transformation and registration of brain images using elastically deformable models, *Comp. Vision and Image Understand.*, 66, 2, 207-222, (1997).
25. Tailarach, J., and Tournoux, P., *Co-planar Stereotaxic Atlas of the Human Brain*, Thieme, Stuttgart, (1988).
26. Megalooikonomou, V., Davatzikos, C., Herskovits, E. A simulator for evaluating methods for the detection of lesion-deficit associations. *Human Brain Mapping*, 10, 2, 61-73, (2000).
27. Gerring, J. P., Brady, K. D., et al., Premorbid prevalence of attention-deficit hyperactivity disorder an development of secondary attention-deficit hyperactivity disorder after closed-head injury. *J Am Acad Child Adolesc Psychiatry*, 37, 647-654, (1998).
28. Lloyd, S. P., Least squares quantization in PCM, *IEEE Trans. Information Theory*, 28, 2, 129-137 (1982).
29. Sander J., Ester M., Kriegel H-P., Xu X., Density-Based Clustering in Spatial Databases: The Algorithm GDBSCAN and its Applications, *Proc. On Data Mining and Knowledge Discovery*, 2, 2, 169-194, (1998).