Automated Classification of Magnetic Resonance Brain Images Using Wavelet Genetic Algorithm and Support Vector Machine

Ahmed Kharrat
National Engineering School of Sfax
Computer & Embedded Systems Laboratory (CES)
B.P 1173, Sfax 3038, Tunisia
Ahmed.Kharrat@ieee.org

Karim Gasmi
National Engineering School of Sfax
Computer & Embedded Systems Laboratory (CES)
B.P 1173, Sfax 3038, Tunisia
Gasmikarim@yahoo.fr

Mohamed Ben Messaoud
National Engineering School of Sfax
Computer & Embedded Systems Laboratory (CES)
B.P 1173, Sfax 3038, Tunisia
M.BenMessaoud@enis.rnu.tn

Nacéra Benamrane
Department of Computer Science Faculty of Science
Vision and Medical Imagery Laboratory U.S.T.O.
B.P 1505, EL-Mnaouer, Oran, 31000, Algeria
nabenamrane@yahoo.com

Mohamed Abid
National Engineering School of Sfax
Computer & Embedded Systems Laboratory (CES)
B.P 1173, Sfax 3038, Tunisia
Mohamed.Abid@enis.rnu.tn

Abstract—**In this paper we propose a new approach for automated diagnosis and classification of Magnetic Resonance (MR) human brain images, using Wavelets Transform (WT) as input to Genetic Algorithm (GA) and Support Vector Machine (SVM). The proposed method segregates MR brain images into normal and abnormal. Our contribution employs genetic algorithm for feature selection witch requires much lighter computational burden. An excellent classification rate of 100% could be achieved using the support vector machine. We observe that our results are significantly better than the results reported in a previous research work employing Wavelet Transform and Support Vector Machine.**

Keywords: Wavelets Transform (WT); Genetic Algorithm (GA); Support Vector Machine (SVM); Magnetic Resonance Imaging (MRI).

I. INTRODUCTION

Magnetic resonance (MR) imaging is currently an indispensable diagnostic imaging technique in the study of the human brain [1]. It's a non-invasive technique that provides fairly good contrast resolution for different tissues and generates an extensive information pool about the condition of the brain. Such information has dramatically improved the quality of brain pathology diagnosis and treatment. However this big amount of data makes manual interpretation impossible and necessitates the development of automated image analysis tools.

There is a variety of automated diagnostic tools that are developed by applying sophisticated signal/image processing techniques utilizing transforms and, may be, subsequently applying some computational intelligent techniques. In one possible methodology, the process of automatic segregation of normal/abnormal subjects, based on brain MRIs, is illustrated as a three-step process: feature extraction, feature selection and nonlinear classification.

To extract features from the MR brain images several image analysis methods are used: e.g. Gabor filters, Independent Component Analysis (ICA) [2], techniques employing statistical feature extraction (like mean, median, mode, quartiles, standard deviation, kurtosis, skewness, etc.) [3], Fourier Transform (FT) based techniques [4], Wavelet Transform (WT) based techniques [5, 6], etc. while Fourier Transform provides only frequency analysis of signals, Wavelet Transforms provide time-frequency analysis, which makes it a useful tool for time-space-frequency analysis and particularly for pattern recognition.

We use Genetic Algorithm (GA) to find minimum features subset giving optimum discrimination between extracted features. GA proves to be the most efficient compared with classical algorithms [7] including sequential forward selection (SFS), sequential backward selection (SBS), sequential floating forward selection (SFFS) and sequential floating backward selection (SFBS).

We apply machine learning algorithms to obtain the classification of images under two categories, either normal or abnormal [8, 9, 10]. Support Vector Machines (SVMs) area

Proc. 9th IEEE Int. Conf. on Cognitive Informatics (ICCI'10)
F. Sun, Y. Wang, J. Lu, B. Zhang, W. Kinsner & L.A. Zadeh (Eds.)
978-1-4244-8040-1/10/$26.00 ©2010 IEEE

widely used for classification tasks due to their appealing generalization properties and their computational efficiency.

The rest of the paper is organised as follows. Section 2 presents the Wavelet transform for feature extraction. Section 3 is devoted for feature selection employed for Genetic Algorithm. Image Classification is presented in Section 4. The performance evaluation is presented in Section 5. Finally, the section 6 presents our conclusions.

II. FEATURE EXTRACTION USING WAVELET TRANSFORM

For the feature extraction there is a wide variety of multiresolution approaches mainly Fourier transform (FT) and wavelet transform (WT). Wavelets are mathematical tools for analysis of complex datasets. These mathematical functions decompose data into different frequency components and then study each component with a resolution matched to its scale. Compared with Fourier transform, wavelet transform seems as an efficient tool in many ways. The Fourier Transform suffers from the limitation that the provided image representation is based only on its frequency content and is not localized in time. Another problem is that the Fourier Transform cannot provide time evolving effects of frequencies in non stationary signals whereas wavelet transform functions provides a hierarchy of scales ranging from the coarsest scale in stationary or in non-stationary signals. Hence wavelet transform has received much attention as a promising tool for feature extraction from images because it can represent an image at various resolutions and because there is a wide range of choices for the wavelet functions.

The mother wavelet is the basis of a wavelet transform. As the pixel intensity values vary smoothly, we choose Daubechies-2 [5, 6] for efficient representation of smoothly changing signals. Although Daubechies-2 is expensive to compute, it is better than Haar wavelet and can render excellent classification accuracy. Daubechies-2 level 1 wavelet approximation coefficient of the MR brain images are extracted and used as feature vector for optimisation.

III. FEATURE SELECTION VIA GENETIC ALGORITHM

Genetic algorithms are stochastic global adaptive search techniques based on the mechanisms of natural selection. GAs comprise a subset of Darwinian evolution-based optimisation techniques focusing on the application of selection, mutation, and recombination to a population of competing problem solutions. Recently, GAs have been recognized as parallel, iterative optimizers and efficient techniques to solve optimization problems [11], including many pattern recognition and classification tasks. Compared with other optimization techniques, GAs start with a random initial population containing a number of chromosomes where each one represents a solution of the problem which performance is evaluated by a fitness function (1). They operate in cycles called generations; the population undergoes reproduction in a number of iterations.

$$fitness = W_A \times Accuracy + W_{nb} \times \frac{1}{N} \qquad (1)$$

Where W_A is the weight of accuracy and W_{nb} is the weight of N feature participated in classification where $N \neq 0$.

The GA maintains a population of competing feature transformation matrices. To evaluate each matrix in this population, the input patterns are multiplied by the matrix, producing a set of transformed patterns which are then sent to a classifier. The classifier typically divides the patterns into a training set, used to train the classifier, and a testing set, used to evaluate classification accuracy. The accuracy obtained is then returned to the GA as a measure of the quality of the transformation matrix used to obtain the set of transformed patterns. Using this information, the GA searches for a transformation that minimizes the dimensionality of the transformed patterns, while maximizing classification accuracy.

Basically, GA consists of three main stages: Selection, Crossover and Mutation. At each step, the Genetic algorithm selects individuals from the current population to be parents and uses them to produce the children for the next generation. The parents which are subject to genetic operators produce offspring. The offspring which may be better than their parents are inserted into the population. Candidate solutions are usually represented as strings of fixed length, called chromosomes. A fitness or objective function is used to reflect the goodness of each member of the population and to measure the fitness of a chromosome. Chromosomes of low fitness are eliminated and the ones of high fitness are kept and moved to the next generation. The application of these three basic operations is repeated for many generations and finally stops when reaching individuals that represent the optimum solution to the problem.

GA can be applied to the tuning of brain MRIs in clinical medicines to ensure the selection of optimal feature set. The block diagram for the entire system is given below "Fig. 1".

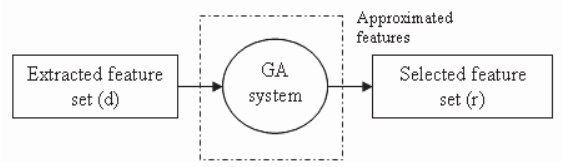


Figure 1. Block diagram of the entire system

The goal of GA System is to find a subset of size *r* among *d* variables (*r*<<*d*), which optimizes the performance of the classifier.

IV. SUPPORT VECTOR MACHINE FOR CLASSIFICATION

A. Review of support vector machine learning

A support vector machine, introduced by Vapnik, is a supervised, multivariate classification method that takes as input labeled data from two classes and outputs a model file for classifying new unlabeled/labeled data into one of two classes. The method has previously been applied to neuroimaging data [8, 9, 10, 12]. It yields successful classification results mainly making binary classification and

solving linear and non linear classification problems. The image data doesn't need to satisfy the assumptions of random Gaussian field theory so that image smoothing is unnecessary. The use of SVM, involves two basic steps namely training and testing. Training an SVM involves feeding known data to the SVM, to form a finite training set. The training set allows SVM to get its intelligence to classify unknown data. SVMs are related to other multivariate methods such as canonical variate analysis, a method successfully applied to fatty acid images of patients with Alzheimer's disease [13]. SVM is based on the structural risk minimization principle from the statistical learning theory. It is applied basically for the binary classification and then extended to the multiclass case [14]. Suppose we have a training set composed of N samples $X=\{X_i\}_{i\le N}$, $X_i \in \Re^n$. Let scalar y denote its class label that is, $y = \pm 1$. Let $\{(x_i, y_i), i = 1, 2, ..., l\}$ denote a given set of l training samples.

1) Linear separation: It is the simplest case where the input patterns are linearly separated by a hyper-plane defined in (2),

$$F(x) = WT\,x + b = 0 \qquad (2)$$

Where W is an adjustable weight vector, and b is the bias term. For each training example x_i the $f(x) \ge 0$ for $y_i = +1$ and $f(x) \le 0$ for $y_i = -1$. If y is "1", it means that the input example is normal. If y is "-1", the input example is abnormal. In "Fig. 2", the margin between two hyper-planes H_1: $W^T x_1 + b = 1$ and H_2: $W^T x_1 + b = -1$ is $\frac{2}{\|w\|}$, and the hyper-plane that maximizes the margin is the optimal separating hyper-plane. Thus, the optimization is now a convex quadratic programming problem.

2) Non linear separation: It is the case in which the linear hyper-plane could not be found to separate data even with the use of relaxation variable. It uses a non-linear operator $\Phi(.)$ to map the input pattern x into higher-dimensional space. The non-linear classifier so obtained is defined as in (3),

$$F(x) = WT\ \Phi(x) + b \qquad (3)$$

Which is linear in terms of the transformed data $\Phi(x)$ but non linear in terms of the original data $x \in \Re^n$. Following non-linear transformation, the parameters of the decision function $f(x)$ are determined by the following minimization criteria,

$$MinJ(W,\xi) = \frac{1}{2}\|W\|^2 + C\sum \xi_I \qquad (4)$$

Subject to

$$y_i(W^T\Phi(x_i) + b) \ge 1 - \xi_i, \quad \xi_i \ge 0 \qquad (5)$$

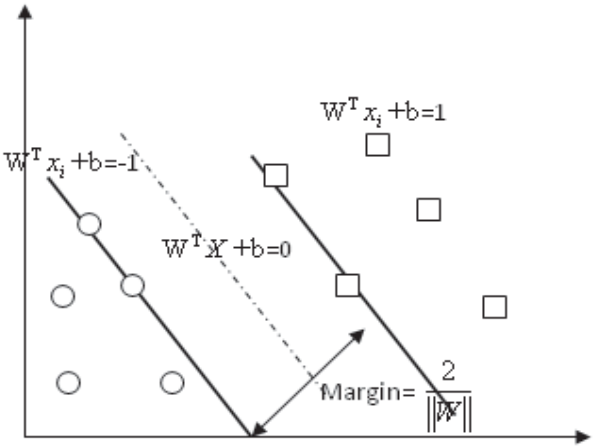


Figure 2. Separating hyper-plane between two classes

B. Support vector machine kernel functions

The kernel function in an SVM has an important role that consists in implicitly mapping the input vector (through an inner product) onto a high-dimensional feature space. It aims at controlling the empirical risk and classification capacity in order to maximize the margin between the classes and to minimize the true costs. When choosing a kernel function, it is necessary to check whether the set is linearly or non-linearly separable. When the set is linearly separable, $K(Xi,X)$ is kernel function and means inner product $\langle X_i, X \rangle$. When the set is non-linearly separable, $K(Xi, X)$ is kernel function, and it must satisfy the Mercer condition. Mercer's theorem states that a non-linear mapping underlies a kernel $K(Xi,X)$ provided that $K(Xi,X)$ is a positive integral operator [12]; that is, for every square integrable function $g(.)$ defined on the kernel $K(Xi,X)$, the kernel satisfies the following condition,

$$\iint K(x,y)g(x)g(y)dxdy \ge 0 \qquad (6)$$

There are several types of kernel learning methods that satisfy Mercer's condition such as polynomial and RBF. These are among the most commonly used kernels in SVM research.

1) Polynomial learning machine: The polynomial kernel is defined as follows,

$$K(x, y) = (x^T y + 1)^P \qquad (7)$$

Where p, the order of a kernel, is a positive constant.
To construct polynomial decision rules of degree'd', one can use the following function for convolution of the inner product,

$$K(x, x_i) = [(x \times x_i) + 1]^d \qquad (8)$$

The decision function becomes,

$$F(x) = sign\Big(\sum_{Support} y_i \alpha_i [(x_i \times x) + 1]^d - b\Big) \qquad (9)$$

Which is a factorization of the d-dimensional polynomials in n-dimensional input space.

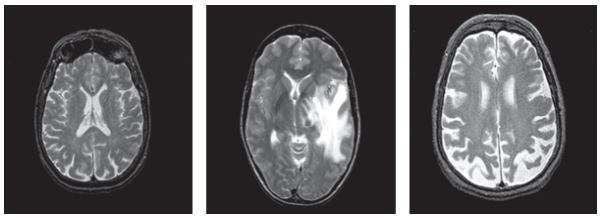
2) Radial Basis Function machines: Classical radial basis function machine uses the following set of decision rules,

$$F(x) = sign(\sum_{i=1}^{N} \alpha_i y_i K_y(|x - x_i| - b) \qquad (10)$$

Where N is the number of support vectors, γ the width parameter of the kernel function, $K_\gamma(|x - x_i|)$ depends on the distance $|x - x_i|$ between two vectors.

V. PERFORMANCE EVALUATION

The images used in this work, are some of the benchmark images downloaded from the Harvard Medical School webpage, freely available in public domain [16]. The images belong to the whole brain atlas, where the brain image datasets are acquired using several imaging technologies. We have tested our classification algorithm for several MR images, some of which belong to normal brain and others belong to pathological brain "Fig.3". All these normal and pathological benchmark images are axial, T2-weighted, MR images of 256×256 sizes. These images are acquired at several positions of the transaxial planes. By convention, for all images, the subject's left is at the right of the image. For each image available, the location of the image in the whole brain dataset is shown in the side view, i.e. in the sagittal image. For our case study, we have considered a total of 83 transaxial images (29 belonging to normal brain and 54 belonging to pathological brain, suffering from a low grade glioma, Meningioma, bronchogenic carcinoma, Glioblastoma multiforme, Sarcoma and Grade IV tumors) in several brain locations. For these pathological brains, suffering from tumors, we have included images acquired at different time instants. The main objective of our algorithm is to segregate normal brain MR images from pathological brain MR images. We have considered that all images belonging to seven persons (four men and three women). Their ages vary between 22 and 81 years.



(a) (b) (c)

Figure 3. (a) A sample MR image of normal brain, (b) A sample MR image of abnormal brain, (c) Alzheimer's didease

The proposed methodology of classifying MR images of human brain is shown "in Fig.4". The method uses the steps of feature extraction, feature selection and classification.

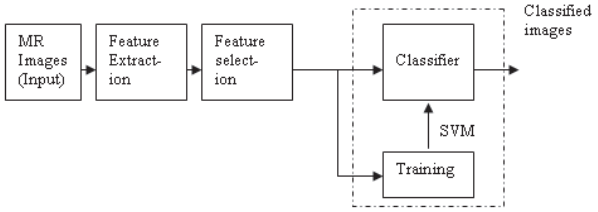


Figure 4. MR Images Classification

TABLE I. PARAMETERS OF GA

GA property	Value/Method
Size of generation	100
Initial population size	30
Performance index/fitness function	fitness (1)
Selection method	Tournoi
Probability of selection	0.05
Crossover method	Arithmetic crossover
Crossover Probability	0.9
Number of crossover points	1
Mutation method	Uniform mutation
Mutation Probability	0.01

For each image, we implement Wavelet transform and we extract five features from these outputs. As described before we applied the genetic algorithm parameters to reduce the number of extracted features. The Genetic algorithm parameters chosen as described "in table 1" prove to be more useful and accurate as they give better selection results.

"Table 2" presents the best chromosomes found by the algorithm during the execution. The classification performance of 100% is obtained with 5 of the whole available features. The feature size is reduced by 88.63%. Therefore it is possible to classify the normal brain and pathological brain with minimum number of features. Thus the cost of classifier can be reduced.

TABLE II. RESULTS OF FEATURE SELECTION PERFORMED BY GA FOR WAVELET FEATURES

Feature selection	Feature set	Classifier accuracy
GA	Mean of Correlation, mean of Maximum probability, mean of Difference variance, mean of Information measure of correlation I, mean of Inverse difference moment normalized, range of Contrast, range of Homogeneity	100%
	Mean of contrast, mean of Information measure of correlation I, mean of homogeneity, mean of Inverse difference moment normalized, mean of homogeneity, range of autocorrelation	100%
	Mean of contrast, mean of homogeneity, mean of sum average, mean of sum variance, range of autocorrelation	100%

The feature vectors and output labels, for all images form a complete dataset are divided into two subsets: a training

dataset and a testing dataset. We use 12 normal brain images and 20 abnormal images in the training phase. Whereas in the testing phase, we use 29 normal brain images and 54 abnormal images. The SVM classifier is trained utilizing the training dataset. Then the SVM is implemented in testing phase. In testing phase, each feature vector, corresponding to a test image, is individually input to the SVM classifier, which produces a continuous output. If the continuous output is positive, then this continuous output is assigned to the output class k_{class} =+1 (belonging to normal brain). Conversely, if the continuous output is negative, then it is assigned to the output class k_{class} =-1 (belonging to abnormal brain). To determine whether the test image is correctly classified or not we compare the output class with the corresponding k_i (which is already known before hand for the test image). This process is repeated for each exemplar in testing dataset, i.e. each test image. Finally, the testing classification accuracy of the algorithm is reported on the basis of the classification performance for the entire testing dataset.

Cross validation method with five folders is used and we find the values of C (width of the radial basis function) and γ (the error/trade-off parameter that adjusts the importance of the separation error in the creation of the separation surface) 8 and 2, respectively as the best parameters to apply in our implementation.

The linear kernel, the RBF and polynomial functions are used for SVM training and testing. The accuracy of classification is high in RBF kernel (100%) in comparison with the linear and polynomial kernels.

TABLE III. CLASSIFICATION PERFORMANCE COMPARISON FOR BRAIN MR IMAGES

Algorithm	No. of features extracted	Classification accuracy
DWT-SOM [8]	4761	94%
DWT-SVM with linear kernel [8]	4761	96.15%
DWT-SVM with polynomial kernel [8]	4761	98%
DWT-SVM with radial basis function based kernel [8]	4761	98%
Our proposed WT-GA-SVM based classifier	5	100%

"Table 3" presents the performance comparison of our proposed method, compared to recently reported brain MR classification results in S. Chaplot's manuscript [8]. In this reference, the same image data base is analysed. They proposed two methods (self-organizing maps and support vector machine) for this classification and they achieved classification accuracy of the order of 94 and 98%, respectively. To achieve these accuracies, they were compelled to utilize huge sizes of feature vectors. They utilized 4761 features extracted from DWT. In comparison with these methods, our system requires only 5 features extracted from WT to be input to the GA for feature optimisation and then for classification. The feature size is reduced by 88.63%. The implementation of our contribution requires much lighter computational burden, which is an important factor while implementing these tools in real time. Hence our proposed system could satisfy two competing requirements simultaneously. They could achieve higher classification accuracy and this could be achieved with a very small size of feature vector. In this context, we would also like to mention that the results in S. Chaplot's manuscript [8] were reported considering a total of 52 image slices (including 6 of normal brain and 46 of abnormal brain). On the other hand, our results are presented considering a total of 83 images (including 29 of normal brain and 54 of abnormal brain).

All experiments were carried out using an Intel core 2 duo machine, with 4GO RAM and a processor speed of 2GHz, run under Windows XP environment. The average CPU time consumed for extracting features, for each image, was approximately 0.07s. For all images the average is 5.249s. In the implementation phase, the classifier consumed an average time of 4.469 ms. In comparison with our method Multilayer Preceptron (MLP) requires 89×10^3 ms.

VI. CONCLUSION

In this paper a new approach for automatic classification of MR Images as normal or abnormal using WT, GA and SVM classifier is proposed. The performance of our contribution in terms of classification accuracy is interpreted. The results show that the proposed method gives better results in comparison with the methods presented in the literature. It suggests that our three-step algorithm is a promising for image classification in a medical imaging application. This automated analysis system, which requires much lighter computational time, could be further used for classification of image with different pathological condition, types and disease status.

ACKNOWLEDGMENTS

We would like to thank Mrs Ines Kallel for her helpful review of the manuscript.

Dr. B. Khalil Chtourou, Biophysics and Nuclear Medicine from the CHU Habib Bourguiba, Department of Nuclear Medicine, Tunisia-Sfax, is acknowledged for providing several clarifications on the medical aspects of the work.

REFERENCES

[1] Neeraj Sharma, Lalit M. Aggarwal, "Automated medical image segmentation techniques", *J. Med. Physics*. Vol. 35, 3-14, 2010.

[2] C.H. Moritz, V.M. Haughton, D.Cordes, M.Quigley, M.E. Meyerand, "Whole-brain functional MR imaging activation from finger tapping task examined with independent component analysis", *Am. J. Neuroradiol*. Vol. 21, 1629-1635, 2000.

[3] R.K. Begg, M. Palaniswami, B. Owen, "Support vector machines for automated gait classification", *J. IEEE Biomed*. Vol. 52, pp. 828-838, 2005.

[4] R.N. Bracewell: *The Fourier Transform and its Applications*, New York, 1999.

[5] S.G. Mallat, "A theory of multiresolution signal decomposition: the wavelet representation". *IEEE Trans. Pattern Analysis and Machine Intelligence*, pp. 674-693. PAMI, 1989.

[6] Ahmed Kharrat, Mohamed Ben Messaoud, Nacéra Benamrane and Mohamed Abid, "Detection of Brain Tumor in Medical Images". *In: 3rd IEEE International Conference on Signals, Circuits & Systems*, pp. 1-6. IEEE SCS, doi: 10.1109/ICSCS.2009.5412577, Tunisia, 2009.

[7] Siedlecki, W. and Sklanky, J., "A note on genetic algorithms for large-scale feature selection". *J. Pattern Recognition letters*. Vol. 10, pp. 335-347, 1989.

[8] S. Chaplot, L.M. Patnaik, N.R. Jagannathan, "Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network", *J. Biomed. Signal Proc*. Cont. 1, pp. 86-92, 2006.

[9] E. A. El-Dahshan, A.-B. M. Salem, T. H. Younis, "A Hybrid Technique for Automatic MRI Brain Images Classification", *Studia Univ. Babes-Bolyai, Informatica*. Vol. 1, pp. 55-67, 2009.

[10] Zacharaki EI, Wang S, Chawla S, Soo Yoo D, Wolf R, Melhem ER, Davatzikos C, "Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme", *Magnetic Resonance in Medicine*. Vol. 62, pp. 1609-1618, 2009.

[11] C.-L. Huang and C.-J. Wang. "A GA-based feature selection and parameters optimization for support vector machine". *J. Expert systems with application*. Vol. 31, pp. 231-240, 2006.

[12] Fan Y, Shen D, Davatzikos C. "Classification of structural images via high-dimensional image warping, robust feature extraction, and SVM". *In: 8th International Conference on International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 1-8. MICCAI, Berlin Heidelberg, 2005.

[13] B. Magnin;L. Mesrob;S. Kinkingnéhun;M. Pélégrini-Issac;O. Colliot;M. Sarazin;B. Dubois;S. Lehéricy;H. Benali, "Support-Vector-Machine based classification of Alzheimer's disease from whole brain anatomical MRI". *J. Neuroradiology*. Vol. 51, pp. 73-83, 2009.

[14] Yan Guermeur, SVM Multiclasss, Théorie et Applications, Nancy I, 2007.

[15] B. Scholkopf: Advances in Kernel Methods: Support Vector Learning, Cambridge, MA, 1999.

[16] HTML file (Harvard Medical School, *http://med.harvard.edu/AANLIB/*).