Proceedings of the 2005 IEEE
Engineering in Medicine and Biology 27th Annual Conference
Shanghai, China, September 1-4, 2005

Adaboost and Support Vector Machines for White Matter Lesion Segmentation in MR Images

Azhar Quddus, Paul Fieguth and Otman Basir *Member, IEEE*

***Abstract*— The use of two powerful classification techniques (boosting and SVM) is explored for the segmentation of white-matter lesions in the MRI scans of human brain. Simple features are generated from Proton Density (PD) scans. Radial Basis Function (RBF) based Adaboost technique and Support Vector Machines (SVM) are employed for this task. The classifiers are trained on severe, moderate and mild cases. The segmentation is performed in T1 acquisition space rather than standard space (with more slices). Hence, the proposed approach requires less time for manual verification. The results indicate that the proposed approach can handle MR field inhomogeneities quite well and is completely independent from manual selection process so that it can be run under batch mode. Segmentation performance comparison with manual detection is also provided.**

I. Introduction

The presence of *white matter hyperintensities* (WMH) lesions in *Magnetic Resonance* (MR) images give important clues regarding presence of abnormality in the human brain. The presence of lesion may not have been fully understood pathologically, nevertheless, WMH lesions are important for obvious reasons.

Many techniques have been suggested in the literature. In the following, we discuss some of the prominent works available in the literature. There are several threshold based techniques (such as [1], [2], [3]), requiring different degrees of human assistance, with improved intra- and inter-reader variability ranging from 3-10 percent for volume estimation, compared to manual tracing. More sophisticated techniques include model-based classifiers ([4]), neural networks ([5], [6]) and fuzzy connectedness ([7]) based approaches. In case of [4], simulated image data was used for the purpose of comparison. In [5], supervised artificial neural network (ANN) was used. After applying backpropagation neural networks, postprocessing was applied to clean artifacts. Finally, lesions were picked-up by manual selection processes. In [6], neural networks were used but the volumes were coregistered to Tailarach space. The results indicate that segmentation results considerably depend on coregistration process. In [7], some training points have to be selected manually to sample Gray-Matter (GM), White-Matter (WM) and Celebro-Spinal Fluid (CSF) for each brain before starting automatic segmentation. In [8], anything not segmented as GM/WM/CSF were taken as lesions hence, partial voluming, artifacts and hyperintensities may produce lots of false positives. More recently, lesions were detected as outliers [9]. They used "probabilistic brain atlas" to train the classifier. This approach requires rotations to bring volume from acquisition to brain Atlas domain. It was claimed to be fully automatic but had best spatial agreement in the range of 50 percent with human observers. Also, well designed studies such as [10], [6] have shown that there is very wide variation in lesion volumes estimated by different human observers, especially when they where trained at different institutions.

Boosting ([11], [12]) is well known technique to improve the classification performance of a weak learner. Whereas, support vector machines (SVM) [13] try to separate decision boundary in a nonlinear hyperdimensional space. These techniques were used in various applications (such as [14]) and performed very well .

In this work, Adaboost [11] and SVM [13] are employed to segment WMH lesions. The main features of this work include,

- The use of robust automatic segmentation technique based on Adaboost and SVM.
- The segmentation is performed in T1 acquisition space rather than in standardized space with large number of slices (such as [6], [9]). Hence, the proposed approach requires less time for manual verification [15].
- Both the Adaboost and SVM perform very well with high dimensional data and hence, the proposed approach can be extended with feature-set with many dimensions.

This article is organized as follows. In Section 2, we briefly present MRI Data analysis. Adaboost and SVM are introduced in section 3, followed by results and discussion in section 4.

II. MRI Data Analysis

Figure 1 shows normalized histograms of lesion and non-lesion gray levels in PD scans with severe WMH. Gaussian approximation using ML estimation is shown as dotted curves. Here the voxels corresponding to background, skull and ventricular *Celebro-Spinal Fluid* (CSF) are set to zero. This is achieved by masking PD scan using T1 segmentation. Figure 2 shows a case of mild WMH lesions where there is huge amount of overlap between the classes.

Clearly, the use only gray levels as features is not sufficient. Since WMH are only located in the white matter and mostly near the ventricles it is reasonable to include distance measure as features. Hence, we propose to use distance of pixels from the center of the mass along with PD gray-levels as features. Let p_{ijk} be the pixel gray values at location

A. Quddus and O. Basir are with the PAMI Lab, Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. ({aquddus,obasir}@uwaterloo.ca)

P. Fieguth is with the Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada. (pfieguth@uwaterloo.ca)

0-7803-8740-6/05/$20.00 ©2005 IEEE.

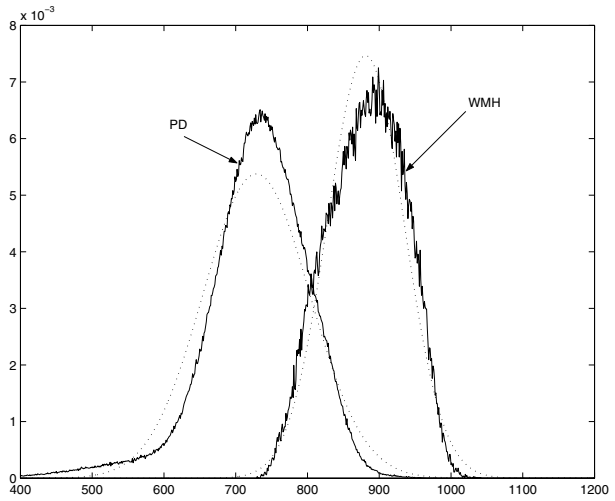

Fig. 1. Severe Case: Normalized histogram of gray level in PD for non-lesion and lesion. ML estimation with gaussian approximation (Dotted)

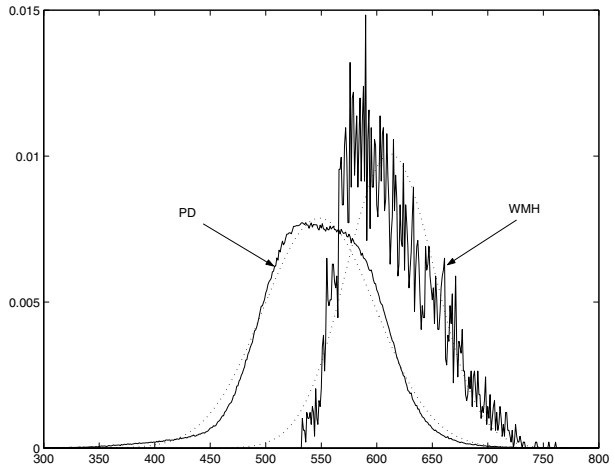

Fig. 2. Mild Case: Normalized histogram of gray level in PD for non-lesion and lesion. ML estimation with gaussian approximation (Dotted)

(i, j, k) in 3-D cartesian space. We denote location as $l_{i,j,k}$ and calculate C_o (the center of the mass) as,

$$C_o = \sum_{i,j,k} l_{i,j,k} \quad \forall p_{ijk} > 0$$

The feature set consists of,

$$\{p_{ijk}, \|C_o - l_{i,j,k}\|\} \quad \forall p_{ijk} > 0 \tag{1}$$

III. Boosting and SVM

AdaBoosting and SVM are introduced in this section.

A. AdaBoost

The AdaBoost algorithm, introduced in 1994 by Freund and Schapire [11], solved many of the practical difficulties of the earlier boosting algorithms [16]. In the following, a brief introduction of Adaboost algorithm is provided. A detailed treatment can be found in [16].

The algorithm takes as input a training set (x_1, y_1),..., (x_m, y_m) where each x_i belongs to some *domain or instance space* X, and each *label* y_i is in some label set Y. For two class problem, $Y = \{-1, +1\}$. Initially, all weights are set equally, but on each round, the weights of incorrectly classified examples are increased so that the weak learner is forced to focus on the hard examples in the training set. The weak learner's job is to find a *weak hypothesis* $h_t : X \to \{-1, +1\}$ appropriate for the distribution D_t .

$$\epsilon_t = Pr_{i\sim D_t}\left[h_t(x_i) \neq y_i\right] \sum_{i:h_t(x_i)\neq y_i} D_t(i)$$

Notice that the error is measured with respect to the distribution D_t. The pseudo code is provided as follows:

Given: (x_1, y_1),...,(x_m, y_m), where $x_i \in X$, $y_i \in Y = \{-1, +1\}$
Initialize $D_1(i) = 1/m$.
For $t = 1, ..., T$:
• Train weak learner using distribution D_t.
• Get weak hypothesis $h_t : X \to \{-1, +1\}$ with error

$$\epsilon_t = Pr_{i\sim D_t}\left[h_t(x_i) \neq y_i\right]$$

• Choose $\alpha_t = \frac{1}{2} ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right)$
• Update:

$$\begin{aligned} D_{t+1}(i) &= \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t} & h_t(x_i) = y_i \\ e^{\alpha_t} & h_t(x_i) \neq y_i \end{cases} \\ &= \frac{D_t(i) exp(-\alpha_t y_i h_t(x_i))}{Z_t} \end{aligned} \tag{2}$$

where Z_t is the normalization factor (chosen so that D_{t+1} will be a distribution).
Break if $\epsilon_t = 0$ or $\epsilon_t \geq 1/2$ and set $T = t - 1$

Output the final hypothesis:
$H(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$

B. Support Vector Machines (SVM)

Support Vector Machines (SVM) were first introduced by Vapnik [13] in 1995. A brief introduction is provided here from [17]. The SVM regression formulation uses a special loss function called *Vapnik's loss function*, which is a linear loss function with an insensitive zone:

$$L_\epsilon(y, f(x, w)) = \begin{cases} \epsilon & |y - f(x, w)| \geq \epsilon \\ |y - f(x, w)| & (otherwise) \end{cases} \tag{3}$$

Here **x** and **y** are input and output respectively. Parameter ϵ controls the width of the insensitive zone. Then the goal of SVM regression is to minimize the following functional:

$$R(w) = \frac{1}{n} \sum_{i=1}^{n} L_\epsilon(y_i - f(x_i, w)) \tag{4}$$

Subject to the constraints $\|w\|^2 < C$ (the constraint on the norm of coefficients is imposed to trade off the complexity

of the solution). By using standard Lagrange multiplier techniques, it can be shown [18] that the minimization of (4) leads to the dual optimization problem. The SVM solution is in the form of the following linear expansion of kernel functions:

$$f(x, \alpha_i, \alpha_i^*) = \sum_{j=1}^{m} (\alpha_j^* - \alpha_j) K(x, x_j) \tag{5}$$

There are several types of basis functions suggested in the literature. In this work, *Radial Basis Functions* (RBF) are used, which is defined as:

$$f(x) = \sum_{i=1}^{m} w_i \, exp\left\{ -\frac{|x - x_i|^2}{\sigma^2} \right\} \tag{6}$$

and the corresponding kernel: $K(x_1, x_2) = exp\left\{ -\frac{\|x_1 - x_2\|^2}{\sigma^2} \right\}$, where σ^2 controls the width of the RBF kernel. The following section deals with the implementation of these algorithms for the segmentation of WMH lesions in 3-D MRI scans.

IV. Results and Discussion

The PD scan consists of $256 \times 256 \times 52$, 16-bit voxels. The PD volume is coregistered and resliced in T1 acquisition space which consists of $256 \times 256 \times 124$ voxels, where each voxel is of size $0.8 \times 0.8 \times 1.2$ mm. The background, skull and ventricular CSF is removed using T1 segmentation mask. In order to have robust detection, the PD gray-level is scaled to have a fix mean value (here 700). The feature set consists of 2 dimensions as described in (1). Three sets (severe, moderate and mild cases) of training and testing data was generated from manual labeling of WMH lesions.

For Adaboost implementation, RBF neural network with 2 inputs and 5 hidden nodes is employed. Each component of input vector **x** feeds forward to 5 basis functions whose outputs are linearly combined with weights $\{w_i\}_{i=1}^{5}$ into the network output $f(x)$. The training is done through *Fletcher-Reeves Conjugate Gradient* optimization technique. Figure 3 shows the WMH lesion segmentation results obtained by *Adaboosted* RBF neural network. The right column with manual segmentation is provided for comparison.

For SVM implementation, RBF kernels with $\sigma^2 = 300$ were used because overfitting was observed with smaller values. As shown in section 2, feature set is non-separable. Hence, different costs (50 for non-lesion and 100 lesion) were used in the objective function. Details regarding costs for non separable case can be found in [19]. Figure 4 shows the WMH lesion segmentation results obtained by SVM using RBF kernel. The right column with manual segmentation is also provided for comparison. The segmentation performance metrics were calculated as follows:

$$Correct\ Detection\ Factor\ (CDF) = \frac{CD}{NT} \tag{7}$$

$$False\ Positive\ Factor\ (FPF) = \frac{FP}{NT} \tag{8}$$

$$False\ Negative\ Factor\ (FNF) = \frac{FN}{NT} \tag{9}$$

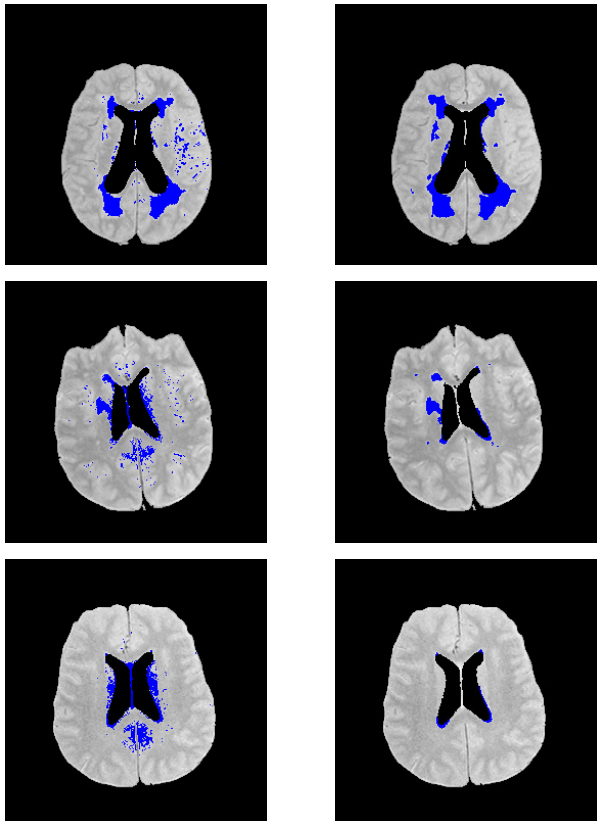

Fig. 3. Adaboost Results (left column) and manual truth (right column) for severe, moderate and mild cases (top to bottom)

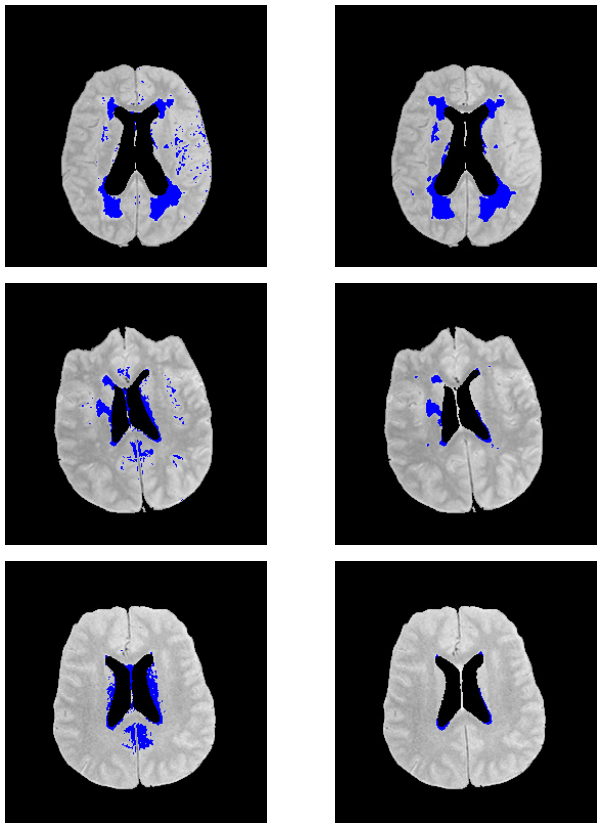

Fig. 4. SVM Results (left column) and manual truth (right column) for severe, moderate and mild cases (top to bottom)

TABLE I

TRAINING PERFORMANCE

Number of Training Data	Speed(seconds)	
	Adaboost	SVM
1000	18.8170	76.1500
2000	34.3290	164.6970
3000	43.6930	496.3430
4000	64.2420	642.5940
5000	81.0360	1096.6

TABLE II

ADABOOST SEGMENTATION PERFORMANCE

Case	CDF	FPF	FNF
Severe	0.6703	0.1718	0.3297
Moderate	0.5516	6.3089	0.4484
Mild	0.7007	18.3964	0.2993

Here *CD* is the number of correctly detected voxels, *FP* is the number of false positive voxels, *FN* is the number of false negative voxels and *NT* is the number of true lesion voxels (from manual selection). Clearly for ideal segmentation CDF is one whereas FPF and FNF are zero. Adaboost is found to be faster than the SVM as shown in Table I. However, the results in figures 3 and 4 and tables II and III show that the segmentation quality is quite similar with both the approaches. The results indicate that segmentation error (false positives) is more in mild case than the other two cases. It follows the intuition that the mild cases are more difficult than the severe and moderate cases. For objective segmentation quality assessment, postprocessing and segmentation noise filtering were not employed.

V. CONCLUSIONS

Two advanced classification techniques (Adaboost and SVM) were employed for WMH lesion segmentation in MRI images. Both the Adaboost and SVM perform very well with high dimensional data and hence, the proposed approach can easily be extended with feature-set with many dimensions. Proton Density (PD) gray level and distance (from centroid) are used as features. The training is performed separately for three cases with severe, moderate and mild WMH lesions. Results indicate that segmentation error is more in mild case than the other two cases. Also, the effect of MR field inhomogeneity is very much reduced. Results were found to be comparable with both Adaboost and SVM, however, Adaboost was found to be faster.

VI. ACKNOWLEDGMENTS

Authors acknowledge the invaluable support in the form of research data by Dr. S. E. Black at *Linda C. Campbell Cognitive Neurology Unit*, at Sunnybrook & Women's College Health Sciences Centre (SWCHSC) in Toronto. The computing facilities at PAMI-Lab at the University of Waterloo is also acknowledged.

TABLE III

SVM SEGMENTATION PERFORMANCE

Case	CDF	FPF	FNF
Severe	0.6784	0.1547	0.3216
Moderate	0.5977	6.0426	0.4023
Mild	0.7525	18.0410	0.2475

REFERENCES

[1] I. Kapouleas. Automatic detection of white matter lesions in magnetic resonance brain images. *Comput. Programs, Meth. Biomed*, 1:17–35, 1990.

[2] D. Wicks, P. Tofts, D. Miller, C. du Boulay, A. Feinstein, and R. Sacares. Volume measurement of multiple sclerosis lesions with magnetic resonance images: A preliminary study. *Neuroradiology*, 34:475–479, 1992.

[3] F. Pannizzo, M. Stallmeyer, J. Friedman, R. Jennis, J. Zabriskie, and C. Plank. Quantitative mri studies for assesment of multiple sclerosis. *Magn. Reson. Med.*, 24:90–99, 1992.

[4] M. Kamber, R. Shinghal, D. L. Collins, G. S. Francis, and A.C. Evans. Model-based 3-d segmentation of multiple sclerosis lesions in magnetic resonance brain images. *IEEE Transaction on Medical Imaging*, 14(3):442–453, Sep 1995.

[5] A.P. Zijdenbos, B.M. Dawant, R.A. Margolin, , and A.C. Palmer. Morphometric analysis of white matter lesions in mr images: Method and validation. *IEEE Transaction on Medical Imaging*, 13(4):716–724, Dec 1994.

[6] A.P. Zijdenbos, R. Forghani, and A.C. Evans. Automatic pipeline analysis of 3-d mri data for clinical trailsapplication to multiple sclerosis. *IEEE Trans. Medical Imaging*, 21(10):1280–1291, October 2002.

[7] J.K. Udupa, S. Samarasekera, Y. Miki, M.A. Van Bucham, and R.I. Grossman. Multiple sclerosis lesion quantification using fuzzy-connectedness principles. *IEEE Transactions on Medical Imaging*, 16(5):598–609, Oct 1997.

[8] J.K. Udupa and S. Samarasekera. Fuzzy connectedness and object definition: theory, algorithms and applications in image segmentation. *Graphical Models and Image Processing*, 58(3):246–261, May 1996.

[9] K. V. Leemput, F. Maes, D. Vandermeulen, A. Colchester, and P. Suetens. Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE Transaction on Medical Imaging*, 20(8):677–688, Aug 2001.

[10] A.P. Zijdenbos, R. Forghani, and A. Evans. Automatic quantification of ms lesions in 3d mri brain data set: Validation of insect. *Proceedings of Medical Image Computing and Computer-Assisted Intervention-MICCAI'98*, 1496:439–448, 1998.

[11] Y. Freund and R. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *Proc. EuroCOLT'94: European Conference on Computational Learning Theory, LNCS*, 1994.

[12] G. Ratsch, T. Onoda, and K.-R. Muller. Soft margin for adaboost. *Machine Learning*, 42:287–320, 2001.

[13] C. Cortes and V. Vapnik. Support vector machines. *Machine Learning*, 20:273–297, 1995.

[14] Y. LeCun, L. Jackel, L. Bottou, C. Cortes, J. Denker, H. Drucker, I. Guyon, U. Muller, E. Sackinger, P. Simard, and V. Vapnik. Learning algorithms for classification: A comparison on handwritten digit recognition. *Neural Networks*, pages 261–276, 1995.

[15] A. Quddus, N. Lobaugh, B. Levine, A. Fienstien, and S.E. Black. Robust protocol for the segmentation of subcortical hyperintensities on mri scans. VASCOG, Goteborg, Sweden, 2003.

[16] R. E. Schapire. A brief introduction to boosting. *Proceedings of the Sixteenth InternationalJoint Conference on Artificial Intelligence*, 1999.

[17] X. Shao and V. Cherkassky. Multi-resolution support vector machine. *IJCNN '99. International Joint Conference on Neural Networks*, 2:1065–1070, July 1999.

[18] V. Vapnik. *The nature of Statistical Learning Theory*. Springer, 1995.

[19] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2:121–167, 1998.