MULTI-LEVEL SVM BASED CAD TOOL FOR CLASSIFYING STRUCTURAL MRIs

Jerrin Thomas Panachakel and Jeena R.S.

Abstract The revolutionary developments in the field of supervised machine learning have paved way to the development of CAD tools for assisting doctors in diagnosis. Recently, the former has been employed in the prediction of neurological disorders such as Alzheimer's disease. We propose a CAD (Computer Aided Diagnosis tool for differentiating neural lesions caused by CVA (Cerebrovascular Accident) from the lesions caused by other neural disorders by using Non-negative Matrix Factorisation (NMF) and Haralick features for feature extraction and SVM (Support Vector Machine) for pattern recognition. We also introduce a multi-level classification system that has better classification efficiency, sensitivity and specificity when compared to systems using NMF or Haralick features alone as features for classification. Cross-validation was performed using LOOCV (Leave-One-Out Cross Validation) method and our proposed system has a classification accuracy of over 86%.

1 Introduction

Cerebrovascular accident (CVA) or cerebrovascular insult (CVI), commonly referred to as "stroke" is the leading cause of death, next to ischaemic heart disease

Jerrin Thomas Panachakel
Dept. of Electrical Engineering
Indian Institute of Science
Bangalore
e-mail: jp@ee.iisc.ernet.in

Jeena R.S.
Dept. of Electronics and Communication Engineering
College of Engineering
Trivandrum
e-mail: jeena_rs@yahoo.com

and the leading cause of adult disability worldwide [1, 2]. Globally, 15 million people suffer from stroke every year and of this, a third dies and half of the remaining struggle with permanent disabilities [3]. The statistics is no different for developing countries like India [4, 5, 6]. In Trivandrum, the capital of the Indian state of Kerala, the incidence rate of stroke is 135.0 and 138.0 per 1,00,000 inhabitants in the urban community and the rural community respectively per year[5]. Clearly, stroke has transformed from a disease pertaining to developed nations to a global hazard.

It was these facts which motivated us to develop a prediction system for CVA. As a predecessor to the proposed CAD (Computer Aided Diagnosis) tool for prediction, we have developed a CAD tool for differentiating brain MRIs of subjects who have suffered from stroke (these MRIs will be hereafter referred to as "stroke MRI") from the MRI of subjects suffering from other neural disorders (hereafter referred to as "non-stroke MRI"), which can be diagnosed from structural MRIs. To the best of knowledge of the authors, this is a maiden work in the field of stroke prediction using neuroimaging, though there are several similar works in the literature, pertaining to other neurological disorders, such as schizophrenia [7], Alzheimer's disease [8] etc.

In this work, we have used the Haralick features [9] and Non-negative Matrix Factorisation [10] for feature extraction and SVM for classification. Also, we have introduced a computationally efficient method for combining feature vectors which are linear in two different kernel spaces. For this, we have made use of the distance of the feature vector from the hyperplane as a measure of confidence value of classification, thus improving the classification efficiency obtained by using either one of the two feature vectors individually.

The rest of the paper is organized as follows: Section 2 gives the details about the database used in this work. Section 3 briefly discusses the various features used for classification. Section 4 describes multi-level classification approach. Performance metrics used is discussed in Section 5. Finally, the results are given in Section 6.

2 Database

The database used for this work is "The Whole Brain Atlas", developed by Keith Johnson, MD, and Alex Becker, PhD., with the support of the Brigham and Women's Hospital Departments of Radiology and Neurology, Harvard Medical School, the Countway Library of Medicine, and the American Academy of Neurology. The database includes the MRI images of neurological disease such as:

1. Neoplastic Disease (brain tumor)

   - Metastatic adenocarcinoma
   - Metastatic bronchogenic carcinoma
   - Meningioma
   - Sarcoma

2. Degenerative Disease

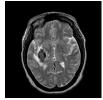- Alzheimer's disease
- Huntington's disease
- Motor neuron disease
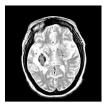- Cerebral calcinosis

3. Inflammatory or Infectious Disease

- Multiple sclerosis
- AIDS dementia
- Creutzfeld-Jakob disease
- Cerebral Toxoplasmosis

4. Cardiovascular Accident (CVA)

- Acute stroke: Speech arrest
- Acute stroke: "alexia without agraphia"
- Subacute stroke: "transcortical aphasia"
- Chronic subdural hematoma
- Hypertensive encephalopathy, and
- Cerebral hemorrhage.

(a) Original MRI   (b) Preprocessed MRI

Fig. 1 An example of preprocessing the MRI image

A total of 30 T2 weighted MRI images from the database were used in the work. Out of these, 16 were non-stroke MRIs and 14 were stroke MRIs. All images were manually orientation correccted, cropped and resampled to a resolution of 256×256. Prior to feature extraction, the images were subjected to normalisation. Normalising the images with respect to the maximum pixel intensity value may introduce noise [8], so the maximum pixel intensity was considered to be the mean of the highest 0.1% pixel intensity values, as proposed in [8] and [11]. The images were further processed for reducing noise using a bilayer filter [12, 13]. Original and preprocessed images are shown in Fig. 1.

3 Feature Extraction

We have relied on two sets of features for classification: (1) textural features and (2) features extracted using NMF.

3.1 Textural features

The textural features introduced by Robert M. Haralick in the paper titled "Textural features for image classification" published in the year 1973 [9] is one of the most important feature extraction techniques for images when the texture of the images has a pivotal role.. These features are generic in nature, meaning they are not developed for a specific imaging modality but can be used for a wide range of images, including biomedical images, where texture is an important property due to the intrinsic spatial tonal variations. The features are computed from matrix referred to as the gray-tone spatial dependence matrix, given by:

$$\mathbf{G} = \begin{bmatrix} c(1,1) & c(1,2) & \cdots & c(1,N_g) \\ c(2,1) & c(2,2) & \cdots & c(2,N_g) \\ \vdots & \vdots & \ddots & \vdots \\ c(N_g,1) & c(N_g,2) & \cdots & c(N_g,N_g) \end{bmatrix} \tag{1}$$

For an image with N_g gray levels, the matrix will be a square matrix of dimension N_g. The element in location $\{i, j\}$ is the count of number of pixels with value i in the neighbourhood of a pixel with value j. This can be extended to a probability matrix by dividing with the appropriate count. As shown in Fig. 2, neighbourhood can be defined in 4 different ways *viz.* horizontal, vertical, left diagonal and right diagonal. Hence, 4 different gray-tone spatial dependence matrices are possible for each image.

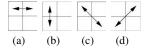
        (a)   (b)   (c)   (d)

Fig. 2 Four different directional adjacencies possible. (a) Horizontal. (b) Vertical . (c) Left diagonal (d) Right diagonal

14 different statistics are computed from each of the four gray-tone spatial dependence matrix using the following relations:

1. Angular Second Moment:

$$\sum_i \sum_j c(i,j)^2 \tag{2}$$

2. Contrast:

$$\sum_{n=0}^{N_g-1} n^2 \{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} c(i,j)\}, |i-j| = n \tag{3}$$

3. Correlation:

$$\frac{\sum_i \sum_j (ij)c(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{4}$$

4. Sum of Squares: Variation:

$$\sum_i \sum_j (i-\mu)^2 c(i,j) \tag{5}$$

5. Inverse Difference Moment:

$$\sum_i \sum_j \frac{1}{1+(i-j)^2} c(i,j) \tag{6}$$

6. Sum Average:

$$\sum_{i=2}^{2N_g} i p_{x+y}(i) \tag{7}$$

7. Sum Variance:

$$\sum_{i=2}^{2} N_g (i - f_s)^2 p_{x+y}(i) \tag{8}$$

8. Sum Entropy:

$$-\sum_{i=2}^{N_g} p_{x+y}(i) \log\{p_{x+y}(i)\} \tag{9}$$

9. Entropy:

$$-\sum_i \sum_j c(i,j) \log(c(i,j)) \tag{10}$$

10. Difference Variance:

$$\sum_{i=0}^{N_g-1} i^2 p_{x-y}(i) \tag{11}$$

11. Difference Entropy:

$$-\sum_{i=0}^{N_g-1} p_{x-y}(i) \log\{p_{x-y}(i)\} \tag{12}$$

12. Information Measure of Correlation 1:

$$\frac{HXY - HXY1}{\max\{HX, HY\}} \tag{13}$$

13. Information Measure of Correlation 2:

$$(1 - \exp[-2(HXY2 - HXY)])^{0.5} \tag{14}$$

14. Maximal Correlation Coefficient:

$$\text{second largest eigen value of } Q^{0.5} \tag{15}$$

where μ_x, μ_y, σ_x and σ_y are the mean and standard deviations of p_x and p_y, the probability density functions. $Q(i, j)$ is given by the following relation:

$$Q(i, j) = \sum_k \frac{c(i, j)c(j, k)}{p_x(i)p_y(k)}. \tag{16}$$

Since each gray-tone spatial dependence matrix corresponds to a specific directions, the features obtained may vary with the change in the orientation of the image. To have some level of orientation independence, we took the mean and range of the features, hence giving only a total of 28 features per image instead of 56.

3.2 Non-negative Matrix Factorisation

One of the major drawbacks with feature extraction techniques such as PCA, is that negative values of the basis vectors are quite difficult for being interpreted in many practical applications. One of remedies for this is to have a representation where non-negativity is imposed by some means. Given a non-negative matrix A, the task is to decompose the matrix into two non-negative matrices V and H subjected to the condition that the Forbenius norm between the given matrix A and the product of the two vectors VH is minimum.

$$\min_{V \geq 0, H \geq 0} ||A - VH||_F^2 \tag{17}$$

This decomposition is known as Non-Negative Matrix Factorization (NMF). The matrix V is called the basis matrix or the mixing matrix and H represents unknown or hidden sources. The beauty of (17) lies in the perspective of viewing each column of A as a linear combination of columns of V with the weights given by the components of each column in H [14, 15, 16]. The number of columns in V is very much less than the number of rows in A. Though NMF seems to be a computationally complex operation, several algorithms have been developed in the last decade for the computationally efficient implementation of NMF.

The maiden work in the field of NMF (Non-negative Matrix Factorisation) can be traced back to a 1994 paper by Paatero and Tapper [10] in which they performed factor analysis on environmental data [17]. Their aim was to find the common latent features or latent variables that explained the given set of observation vectors. Some elementary variables combine together positively to give each of the factor. A factor can either be present, in which case it has a positive effect or the factor can be absent, in which case the factor has null influence. Clearly, there is no room for a "negative" influence and hence it often makes sense to constrain the factors to be non-negative.

4 Multilevel SVM Classification

SVM is a widely used supervised machine learning algorithm for classification developed by Vladimir N. Vapnik and the current standard incarnation (soft margin) was proposed by Vapnik and Corinna Cortes in 1995 [18]. An in-depth discussion on SVM classification can be found in [19]. In our proposed multi-level classification, two support vector models are created using NMF features and Haralick features. Now, given a feature vector $\Phi(\overrightarrow{x})$, in the feature space $\Phi(.)$, we compute a score based on its distance from the decision boundary hyperplane $f(x)$ as,[20]

$$\text{d}ist(\Phi(x), f(x)) = \frac{f(x)}{\sum\limits_{i \in SV} |y_i \alpha_i \Phi(x)|^2} \tag{18}$$

where $y_i \in \{-1, 1\}$, α_i : vector weights for support vectors.

The scores provide an estimate of how good the classification is. Larger the score, larger will be the distance from the hyperplane and hence higher will be probability of the sample to lie in that class [20]. In our work, for a given test sample, we compute the scores for both the support vector models. The model which gives the highest absolute value for the score is assumed to have classified the sample correctly. This approach improved the accuracy of the classification.

5 Performance Metrics

Three performance metrics were used in this work:

1. Sensitivity (*SN*), which is the measure of the system's ability to identify stroke MRIs.
2. Specificity (*SP*), which is a measure of the system's ability to identify non-stroke MRIs.
3. Accuracy (*AC*), which is a measure of the system's net classification efficiency.

Before defining these metrics mathematically, we introduce the following terms:

- TP: True Positive, stroke MRI identified as stroke MRI.
- TN: True Negative, non-stroke MRI identified as non-stroke MRI.
- FP: False Positive, non-stroke MRI identified as stroke MRI.
- FN: False Negative, stoke MRI identified as non-stroke MRI.

Now, we can mathematically define sensitivity, specificity and accuracy as:

$$SN = \frac{TP}{TP + FN} \tag{19}$$

$$SP = \frac{TN}{TN + FP} \tag{20}$$

$$AC = \frac{TP + TN}{TN + TP + FP + FN} \qquad (21)$$

The confusion matrix used in this work is shown in Fig. 3.

Fig. 3 Sample confusion matrix

6 Results

LOOCV (Leave-One-Out Cross Validation), which is a special case of LpOVC (Leave-p-Out Cross Validation) was used for cross validation. The steps adopted were,

1. **Step 1:** Out of the N samples with ground truth, $N-1$ were chosen as the training set and remaining one was chosen as the test data. Using this set of training data, both classification models were generated and using these models, the test data was classified based on the SVM score. Depending on the classification output and the ground truth, one of TP, TN, FP or FN was incremented.
2. **Step 2:** In the next iteration, a different sample was chosen as the test data and Step 1 is repeated.
3. **Step 3:** This process is continued until all the samples have been used as a test data once.

Table 1 Classification results when Haralick features alone is used

	Linear	MLP	RBF
Sensitivity	71.43	78.57	71.00
Specificity	87.50	81.25	81.25
Accuracy	80.00	80.00	76.67

TABLE 1 and II shows the result of using NMF and Haralick features independently for classification using different kernel functions. For MLP, the parameter

(a) Confusion matrix when NMF weight vector alone is used as the features

(b) Confusion matrix when Haralick features alone is used

Fig. 4 Confusion matrices for various features considered individually.

Table 2 Classification results when NMF weight vector alone is used as the features
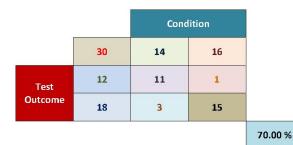
	Linear	MLP	RBF
Sensitivity	78.57	64.29	71.43
Specificity	87.50	87.50	81.25
Accuracy	83.33	76.67	76.67

Table 3 Comparison of multi-level SVM with simple SVM

	Haralick	NMF	Concatenated	Multi-Level
Sensitivity	78.57	78.57	64.29	**78.57**
Specificity	81.25	87.50	81.25	**93.75**
Accuracy	80.00	83.33	73.33	**86.67**

		Condition	
	30	**14**	**16**
	12	**9**	**3**
Test Outcome			
	18	**5**	**13**
			73.33 %

(a) Confusion matrix obtained when concatenated Haralick feaures and NMF weight vector is used as the feature vector

		Condition	
	30	**14**	**16**
	12	**11**	**1**
Test Outcome			
	18	**3**	**15**
			70.00 %

(b) Confusion matrix when multi-level SVM is used

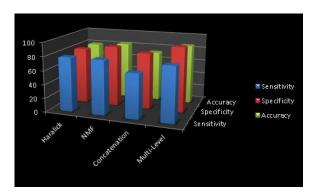**Fig. 5** Confusion matrices for various features considered simultaneously.

Fig. 6 Comparison of multi-level SVM with simple SVM

was $[1-9]$ and for RBF, the sigma value was 40. These values were found to give the best results.

By using the proposed multi-level classification, we could improve the accuracy, sensitivity and specificity rather than by just concatenating the features. The results are shown in TABLE III, Fig. 5 and Fig. 6.

7 Conclusion

In this paper, we have proposed a novel CAD tool that can classify the structural MRI images of patients who had stroke from the images of patients suffering from other neurological disorders. By incorporating a multi-level classification system, we have achieved a classification efficiency of more than 83% by using Non-negative matrix factorisation and Haralick features as the features and SVM as the classifier. This can be further improved by incorporating other features. We believe that this can be extended to a CAD systems for predicting CVAs.

Acknowledgement

References

[1] Saver, J.L., Starkman, S., Eckstein, M., Stratton, S., Pratt, F., Hamilton, S., Conwit, R., Liebeskind, D.S., Sung, G., Sanossian, N.: Methodology of the field administration of stroke therapy–magnesium (FAST-MAG) phase 3 trial: Part 1–rationale and general methods. International Journal of Stroke **9** (2014) 215–219

[2] Organization, W.H., et al.: The top 10 causes of death. July 2013. Available at: who. int/mediacentre/factsheets/fs310/en/.[Last accessed: October 2013] (2014)

[3] Nagi, D., Aamar, A., Rahim, I., Sadaf, M.I., Butt, Z., Ahsan, I., Abbas, F., Wahab, A.: Clinical audit of patients with cerebrovascular accident and transient ischemic attack in Our Lady of Lourdes Hospital Drogheda, Ireland. Annals of King Edward Medical University **18** (2014)

[4] Prasad, K., Vibha, D., et al.: Cerebrovascular disease in South Asia–Part I: A burning problem. JRSM cardiovascular disease **1** (2012) 20

[5] Sridharan, S.E., Unnikrishnan, J., Sukumaran, S., Sylaja, P., Nayak, S.D., Sarma, P.S., Radhakrishnan, K.: Incidence, types, risk factors, and outcome

of stroke in a developing country the Trivandrum stroke registry. Stroke **40** (2009) 1212–1218

[6] Chauhan, S., Aeri, B.T.: Prevalence of cardiovascular disease in India and it is economic impact-A review. International Journal of Scientific and Research Publications (2013) 212

[7] Potluru, V.K., Calhoun, V.D.: Group learning using contrast NMF: Application to functional and structural MRI of schizophrenia. In: Circuits and Systems, 2008. ISCAS 2008. IEEE International Symposium on, IEEE (2008) 1336–1339

[8] Padilla, P., López, M., Górriz, J.M., Ramirez, J., Salas-Gonzalez, D., Álvarez, I.: NMF-SVM based CAD tool applied to functional brain images for the diagnosis of alzheimer's disease. Medical Imaging, IEEE Transactions on **31** (2012) 207–216

[9] Haralick, R.M., Shanmugam, K., Dinstein, I.H.: Textural features for image classification. Systems, Man and Cybernetics, IEEE Transactions on (1973) 610–621

[10] Paatero, P., Tapper, U.: Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. Environmetrics **5** (1994) 111–126

[11] Illán, I., Górriz, J., Ramírez, J., Salas-Gonzalez, D., López, M., Segovia, F., Chaves, R., Gómez-Rio, M., Puntonet, C.G.: 18F-FDG PET imaging analysis for computer aided alzheimer's diagnosis. Information Sciences **181** (2011) 903–916

[12] Panachakel, J.T.: Contourlet transform and iterative noise free filtering based bilayer filter for enhancing echocardiogram. In: Green Technologies (ICGT), 2012 International Conference on, IEEE (2012) 301–306

[13] Panachakel, J.T., George, M.: Bilayer denoising of echocardiography images using contourlet transform and iterative noise-free filtering. International Journal of Applied Engineering Research **7** (2012)

[14] Lee, D.D., Seung, H.S.: Algorithms for non-negative matrix factorization. In: Advances in Neural Information Processing systems. (2001) 556–562

[15] Donoho, D., Stodden, V.: When does non-negative matrix factorization give a correct decomposition into parts? In: Advances in neural information processing systems. (2003) None

[16] Berry, M.W., Browne, M., Langville, A.N., Pauca, V.P., Plemmons, R.J.: Algorithms and applications for approximate nonnegative matrix factorization. Computational Statistics & Data Analysis **52** (2007) 155–173

[17] Tropp, J.A.: Literature survey: Nonnegative matrix factorization. University of Texas at Asutin (2003)

[18] Cortes, C., Vapnik, V.: Support-vector networks. Machine learning **20** (1995) 273–297

[19] Burges, C.J.: A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery **2** (1998) 121–167

[20] Shin, H., Cho, S.: How to deal with large dataset, class imbalance and binary output in svm based response model. In: Proceedings of the Korean Data Mining Conference. (2003) 93–107