

A Novel MRI Brain Images Classifier Using PCA and SVM

Geeta Palki¹, Ashwini Patil², Sandeep Kumar³, Shrivatsa D Perur⁴, Shubham Kumar⁵

Dept. of ISE, Gogte Institute of Technology,
Belagavi

Abstract - Automated and accurate classification of MR brain images is extremely important for medical analysis and interpretation. Over the last decade, numerous methods have already been proposed. In our project, we presented a novel method to classify a given MR images as normal or abnormal. The proposed method first employed the wavelet transform to extract features from images. The element abstraction is a method of representation of the image with raw data by performing the processing to extract the useful data from the image to improve the process of decision-making like the classification of different patterns. Followed by applying principle component analysis (PCA) to reduce the dimensions of features. The reduced features were submitted to a kernel support vector machine (KSVM). These methods combine the intensity and the components of the shapes and different orders with the textures of the tumor from the MRI images. SVM is automatically classified brain MRI images under two categories, either normal or abnormal. The determination of normal and abnormal brain image is based on symmetry which is exhibited in the axial and coronal images. Using feature vector gained from the MRI images. SVM classifier is used to classify the images. We choose seven common brain diseases (glioma, meningioma, Alzheimer's disease, Pick's disease, sarcoma and Huntington's disease) as abnormal brains, and collected 160 MR brain images (20 normal and 140 abnormal) from Harvard Medical School website. We performed our proposed method with four different kernels and found that linear kernel achieves the best classification accuracy.

Keywords- *Magnetic Resonance Imaging, Wavelet Transform, Principle Component Analysis, Support Vector Machine, Classification.*

I. INTRODUCTION

Automatic Magnetic resonance imaging (MRI) is a noninvasive medical imaging technique used in Computer aided diagnosis (CAD) to visualize detailed internal structure and limited functions of the body such as brain diseases, Alzheimer disease or movement's disorders such as Parkinson. The diagnostic values of MRI are greatly magnified by the automatic and accurate classification of the MR images. Different algorithms are exposed in each step on automatic brain tumor detection; this process has its appropriate methods. Wavelet transform is an effective tool for feature extraction from MR brain images, because it allows analysis of images at various levels of resolution due to its multi-resolution analytic property. However, this technique requires large storage

and is computationally expensive. In order to reduce the feature vector dimensions and increase the discriminative power, the principal component analysis (PCA) was used. PCA is appealing since it effectively reduces the dimensionality of the data and therefore reduces the computational cost of analyzing new data. In recent years, researchers have proposed a lot of approaches for this goal, which fall into two categories. One category is supervised classification, including support vector machine (SVM) and k -nearest neighbors (k -NN). The other category is unsupervised classification, including self-organization feature map (SOFM) and fuzzy c -means. While all these methods achieved good results, and yet the supervised classifier performs better than unsupervised classifier in terms of classification accuracy (success classification rate). However, the classification accuracies of most existing methods were lower than 95%, so the goal of this paper is to find a more accurate method. Among supervised classification methods, the SVMs are state-of-the-art classification methods based on machine learning theory. Compared with other methods such as artificial neural network, decision tree, and Bayesian network, SVMs have significant advantages of high accuracy, elegant mathematical tractability, and direct geometric interpretation. Besides, it does not need a large number of training samples to avoid overfitting. Original SVMs are linear classifiers. In this paper, we introduced the kernel SVMs (KSVMs), which extends original linear SVMs to nonlinear SVM classifiers by applying the kernel function to replace the dot product form in the original SVMs. The KSVMs allow us to fit the maximum-margin hyperplane in a transformed feature space. The transformation may be nonlinear and the transformed space high dimensional; thus though the classifier is a hyperplane in the high-dimensional feature space, it may be nonlinear in the original input space. The structure of the rest of this paper is organized as follows. Next gives the detailed procedures of preprocessing, including the discrete wavelet transform (DWT) and principle component analysis (PCA). It introduces the motivation and principle of SVM, and then turns to the kernel SVM. And also introduces protecting the classifier from overfitting. Experiments used totally 160 images as the dataset, showing the results of feature extraction and reduction. Afterwards, we compare our method with different kernels to the latest methods in the decade.

II. LITERATURE SURVEY

In this paper Karat et al. presented that Magnetic resonance imaging (MRI) is an imaging technique that produces high-quality images of the anatomical structures of the human body, especially in the brain and provides rich information for clinical diagnosis and biomedical research. The diagnostic values of MRI are greatly magnified by the automated and accurate classification of the MRI images. A wavelet transform is an effective tool for feature extraction from MR brain images because it allows analysis of images at various levels of resolution due to its multi-resolution analytic property. R. M. Haralick et al. proposed that, using 2D Wavelet transform (Daubechies-two of level one), the brain image is decomposed into four sub-bands.

The clearest appearance of the changes between the different textures represented by the sub-band whose histogram has the maximum variance. J. Huang, J. Lu, C., X. Ling et al. explained the process of the wavelet transforms. One of the most powerful methods for extraction is Wavelet transform. This is an effective tool for 2D image feature extraction because it allows for the analysis of images at various levels of resolution. The main advantage of the wavelet is that it affords localized frequency information about the function of a signal, which is particularly beneficial for classification.

III. METHODOLOGY

In total, our method consists of three stages:

Step 1: Preprocessing (including feature extraction and feature reduction);

Step 2: Training the kernel SVM;

Step 3: Submit new MRI brains to the trained kernel SVM, and output the prediction.

As shown in Fig. 1, this flowchart is a canonical and standard classification method which has already been proven as the best classification method.

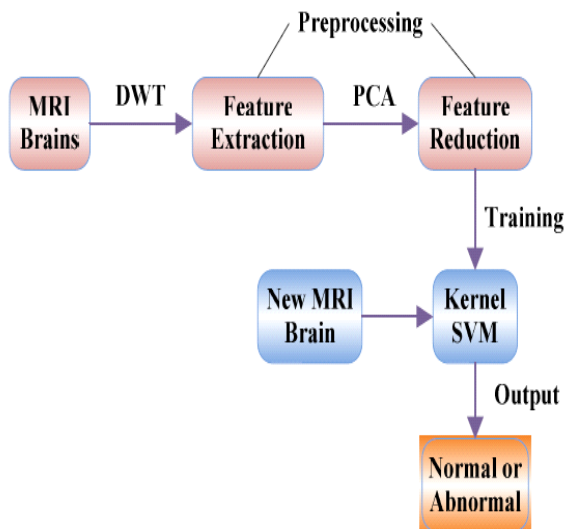


Fig 1. Methodology of our proposed algorithm.

Feature Extraction

The most conventional tool of signal analysis is Fourier transform (FT), which breaks down a time domain signal into constituent sinusoids of different frequencies, thus, transforming the signal from time domain to frequency domain. However, FT has a serious drawback as discarding the time information of the signal. For example, an analyst can not tell when a particular event took place from a Fourier spectrum. Thus, the quality of the classification decreases as time information is lost. Gabor adapted the FT to analyze only a small section of the signal at a time. The technique is called windowing or short time Fourier transform (STFT). It adds a window of particular shape to the signal. STFT can be regarded as a compromise between the time information and frequency information. It provides some information about both time and frequency domain. However, the precision of the information is limited by the size of the window. Wavelet transform (WT) represents the next logical step: a windowing technique with variable size. Thus, it preserves both time and frequency information of the signal. Another advantage of WT is that it adopts "scale" instead of "traditional frequency", namely, it does not produce a time-frequency view but a time-scale view of the signal. The time-scale view is a different way to view data, but it is a more natural and powerful way, because compared to "frequency", "scale" is commonly used in daily life. Meanwhile, "in large/small scale" is easily understood than "in high/low frequency".

2D DWT

In case of 2D images, the DWT is applied to each dimension separately. Fig. 3 illustrates the schematic diagram of 2D DWT. As a result, there are 4 sub-band (LL, LH, HH, and HL) images at each scale. The sub-band LL is used for next 2D DWT.

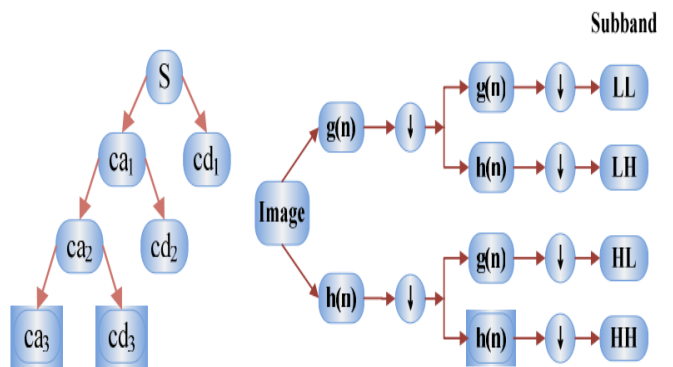


Fig 2. A 3-level wavelet Decomposition tree. Fig 3. Schematic diagram of 2D DWT

The LL subband can be regarded as the approximation component of the image, while the LH, HL, and HH subbands can be regarded as the detailed components of the image. As the level of decomposition increased, compacter but coarser approximation component was obtained. Thus, wavelets provide a simple hierarchical framework for interpreting the image information. In our algorithm, level-3 decomposition via Harr wavelet was utilized to extract features. The border distortion is a technique issue related to digital filter which is commonly used in the DWT. As we filter the image, the mask will extend beyond the image at the edges, so the solution is to pad the pixels outside the images. In our algorithm, symmetric padding method was utilized to calculate the boundary value.

Feature Reduction

Excessive features increase computation times and storage memory. Furthermore, they sometimes make classification more complicated, which is called the curse of dimensionality. It is required to reduce the number of features. PCA is an efficient tool to reduce the dimension of a data set consisting of a large number of interrelated variables while retaining most of the variations. It is achieved by transforming the data set to a new set of ordered variables according to their variances or importance. This technique has three effects: it orthogonalizes the components of the input vectors so that uncorrelated with each other, it orders the resulting orthogonal components so that those with the largest variation come first, and eliminates those components contributing the least to the variation in the data set. It should be noted that the input vectors be normalized to have zero mean and unity variance before performing PCA. The normalization is a standard procedure.

KERNEL SVM

The introduction of support vector machine (SVM) is a landmark in the field of machine learning. The advantages of SVMs include high accuracy, elegant mathematical tractability, and direct geometric interpretation. Recently, multiple improved SVMs have grown rapidly, among which the kernel SVMs are the most popular and effective. Kernel SVMs have the following advantages: (1) work very well in practice and have been remarkably successful in such diverse fields as natural language categorization, bioinformatics and computer vision; (2) have few tunable parameters; and (3) training often involves convex quadratic optimization [31]. Hence, solutions are global and usually unique, thus avoiding the convergence to local minima exhibited by other statistical learning systems, such as neural networks.

Suppose some prescribed data points each belong to one of two classes, and the goal is to classify which class a new data point will be located in. Here a data point is viewed as a p -dimensional vector, and our task is to create a $(p+1)$ -dimensional hyperplane. There are many possible hyperplanes that might classify the data successfully. One reasonable choice as the best hyperplane is the one that represents the largest separation, or margin, between the two classes, since we could expect better behavior in

response to unseen data during training, i.e., better generalization performance. Therefore, we choose the hyperplane so that the distance from it to the nearest data point on each side is maximized. Fig. 5 shows the geometric interpolation of linear SVMs, here H1, H2, H3 are three hyperplanes which can classify the two classes successfully, however, H2 and H3 does not have the largest margin, so they will not perform well to new test data. The H1 has the maximum margin to the support vectors (S11, S12, S13, S21, S22, and S23), so it is chosen as the best classification hyperplane.

Database

The datasets consists of 83 brain MR images, each of whose size is 256x256, which were downloaded from the Harvard Medical School website (<http://med.harvard.edu/AANLIB/>). The setting of the training images and validation images can be shown due to the stratified 5-fold cross validation. Therefore, we ran 5 trials with each 22 (12 normal and 10 abnormal) are used for training and the left 61 (17 normal and 44 abnormal) are used for test. A typical representative MR image of normal, benign and malignant tumor is shown in Fig 4.

Figure

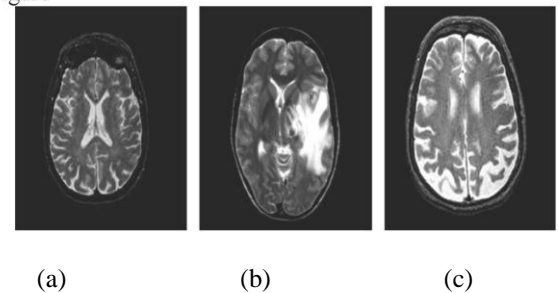


Fig. 4. Sample of brain MRIs:
(a) Normal brain; (b) meningioma; (c) Alzheimer's disease

Classification of accuracy

We tested four SVMs with different kernels (LIN, HPOL, IPOL, and GRB). In the case of using linear kernel, the KSVM degrades to original linear SVM. We computed hundreds of simulations in order to estimate the optimal parameters of the kernel functions, such as the order d in HPOL and IPOL kernel, and the scaling factor σ in GRB kernel. The element of i th row and j th column represents the classification accuracy belonging to class i are assigned to class j after the supervised classification. The results showed that the proposed DWT+PCA+KSVM method obtains quite excellent results on both training and validation images. For LIN kernel, the whole classification accuracy was $(17+135)/160 = 95\%$; for HPOL kernel, was $(19+136)/160 = 96.88\%$; for IPOL kernel, was $(18+139)/160 = 98.12\%$; and for the GRB kernel, was $(20+139)/160 = 99.38\%$. Obviously, the GRB kernel SVM outperformed the other three kernel SVMs.

CONCLUSIONS AND DISCUSSIONS

In this study, we have developed a novel DWT+PCA+KSVM method to distinguish between normal and abnormal MRIs of the brain. We picked up four different kernels as LIN, HPOL, IPOL and GRB. The experiments demonstrate that the GRB kernel SVM

obtained 99.38% classification accuracy on the 160 MR images, higher than HPOL, IPOL and GRB kernels, and other popular methods in recent literatures. Future work should focus on the following four aspects: First, the proposed SVM based method could be employed for MR images with other contrast mechanisms such as T1-weighted, Proton Density weighted, and diffusion weighted images. Second, the computation time could be accelerated by using advanced wavelet transforms such as the lift-up wavelet. Third, Multi-classification, which focuses on specific disorders studied using brain MRI, can also be explored. Forth, novel kernels will be tested to increase the classification accuracy. The DWT can efficiently extract the information from original MR images with little loss. The advantage of DWT over Fourier Transforms is the spatial resolution. In the future, we will focus on investigating the performance of these algorithms. The proposed DWT+PCA+KSVM with GRB kernel method shows superiority to the LIN, HPOL, and IPOL kernels SVMs.

REFERENCES

- [1] An Mr Brain Images Classifier Via Principalcomponent Analysis And Kernel Support Vector Machine by Y. Zhang* and L. Wu
- [2] MRI Brain Tumor Classification using Support Vector Machines and Meta-Heuristic Method by Ahmed Kharratl,2, MohamedBenHelima.
- [3] J. Huang, J. Lu, C., X. Ling, "Comparing Naive Bayes, Decision Trees, and SVM with AUC and Accuracy". Proceedings of the Third IEEE International Conference on Data Mining. IEEE Computer Society Press, pp. 553-556,2003.
- [4] S. Kirkpatrick, C. D. Gelatt et M. P. Vecchi, "Optimization by Simulated Annealing", American Association for the Advancement of Science, vol. 220, no. 4598, pp. 671-680. , 1983.
- [5] R. M. Haralick, K. Shanmugam and I. Dinstein, "Textural Features of Image Classification", In Proceeding of TEEE Transactions on Systems, Man and Cybernetics, pp. 610-621, 1973.
- [6] A Kharrat, N. Benamrane, M. Ben Messaoud and M. Abid, "Evolutionary Support Vector Machine For Parameters Optimization Applied To Medical Diagnostic", International Joint Conference on Computer Vision Theory and Applications (VISAPP 2011), pp. 201-204, 2011.
- [7] Bermejo, S., B. Monegal, and J. Cabestany, \Fish age categorization from otolith images using multi-class support vector machines," *Fisheries Research*, Vol. 84, No. 2, 247{253, 2007.