

Classification of MR brain images using kernel support vector machine

Maske Hema Shivajirao¹, Sanjeev Dadarao Bhosale²

¹IPG Student at Electronics & Telecommunication department

²Asst. Professor at Electronics & Telecommunication department, Shri Tuljabhavani college of engineering, Tuljapur

Abstract - An automated and accurate analysis of MR brain images is really important for medical analysis and interpretation. Over the last decade multitudinous ways have before been proposed. In this paper, we presented a fully unique way to classify a given MR brain image as normal or abnormal. The proposed way first employed wave transubstantiate to yank features from images, followed by applying principle member analysis (PCA) to span back the width of features. The reduced features were submitted to a kernel support vector machine (KSVM). The strategy of K-fold stratified cross witness was used to enhance stereotype of KSVM. We chose seven common brain troubles (glioma, meningioma, Alzheimer's trouble, Alzheimer's trouble plus agnosia, Pick's trouble, sarcoma, and Huntington's trouble) as abnormal smarts. 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 transform, the transformation may be nonlinear and the transformed space high dimensional feature space, it may be nonlinear in the original input space.

Index Terms - PCA, DWT, KSVM, Kernel chose LIN, HPOL, IPOL, and RBF.

I. INTRODUCTION

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. Clinical decisions regarding the treatment of brain neoplasms rely, in part, on MRI at various stages of the treatment process. Radiological diagnosis is based on the multi- parametric imaging profile (CT, conventional MRI, advanced MRI). Tumor

characterization is difficult, because neoplastic tissue is often heterogeneous in spatial and imaging profiles [1], and for some imaging techniques often overlaps with normal tissue (especially the infiltrating part) The diagnostic values of MRI are greatly magnified by the automated and accurate classification of the MRI images [6{8]. 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 [9]. In order to reduce the feature vector dimensions and increase the discriminative power, the principal component analysis (PCA) was used [10]. PCA is appealing since it effectively reduces the dimensionality of the data and therefore reduces the computational cost of analyzing new data [11]. Then, the problem of how to classify the input data arises.

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) [12] and k-nearest neighbors (k-NN) [13]. The other category is unsupervised classification [14], including self-organization feature map (SOFM) [12] and fuzzy c-means [15]. While all these methods achieved good results, and yet the supervised classifier performs better than the unsupervised classifier in terms of classification accuracy (success classification rate). However, the classification accuracies of most existing methods were lower than 90%, so the goal of this paper is to find a more accurate method.

II. METHODOLOGY

The proposed 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. We will explain the detailed procedures of the preprocessing in the following subsections.

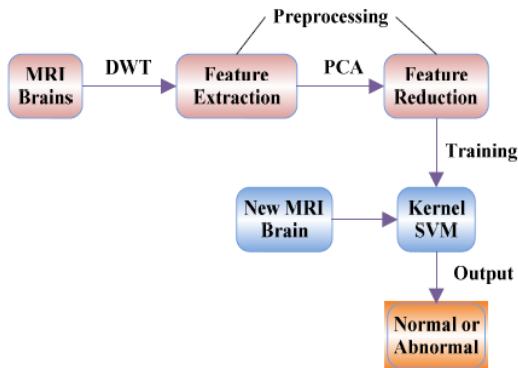


Fig 1 Methodology of proposed algorithm

A. Feature Extraction

In signal analysis the most conventional tool is Fourier transform (FT), which breaks down a time domain signal into constituent sinusoids of various frequencies, which 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, analyst cannot 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) [23]. 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. The development of signal analysis is shown in Fig. 2.

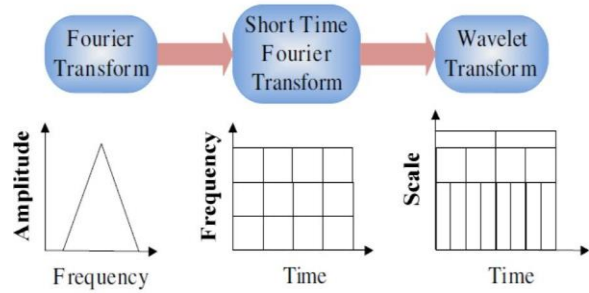


Fig. 2 The development of signal analysis

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".

B. Discrete Wavelet Transform

The discrete wavelet transform (DWT) is a powerful implementation of the WT using the dyadic scales and positions [24]. The fundamentals of DWT are introduced as follows. Suppose $x(t)$ is a square-integrable function, then the continuous WT of $x(t)$ relative to a given wavelet $\Psi(t)$ is defined as

$$W_{\Psi}(a, b) = \int_{-\infty}^{\infty} x(t) \Psi_{a,b}(t) dt \tag{1}$$

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}} \Psi\left(t - \frac{a}{b}\right) \tag{2}$$

Here, the wavelet $\Psi_{a,b}(t)$ is calculated from the mother wavelet $\Psi(t)$ by translation and dilation: a is the dilation factor and b the translation parameter (both real positive numbers). There are several different kinds of wavelets which have gained popularity throughout the development of wavelet analysis. The most important wavelet is the Harr wavelet, which is the simplest one and often the preferred wavelet in a lot of applications.

Equation (1) can be discretised by restraining a and b to a discrete lattice ($a = 2^j$ & $a > 0$) to give the DWT, which can be expressed as follows.

$$\begin{aligned} ca_{j,k}(n) &= DS \left[\sum_n x(n) g_j^*(n - 2^j k) \right] \\ cd_{j,k}(n) &= DS \left[\sum_n x(n) h_j^*(n - 2^j k) \right] \end{aligned} \tag{3}$$

Here $ca_{j,k}$ and $cd_{j,k}$ refer to the coefficients of the approximation components and the detail components, respectively. $g(n)$ and $h(n)$ denote for the low-pass filter and high-pass filter, respectively. j and k represent the wavelet scale and translation factors, respectively. DS operator means the down sampling. Equation (3) is the fundamental of wavelet decomposes. It decomposes signal $x(n)$ into two signals, the approximation coefficients $ca(n)$ and the detail components $cd(n)$.

This procedure is called one-level decompose. The above decomposition process can be iterated with successive approximations being decomposed in turn, so that one signal is broken down into various levels of resolution. The whole process is called wavelet decomposition tree, shown in Fig. 3.

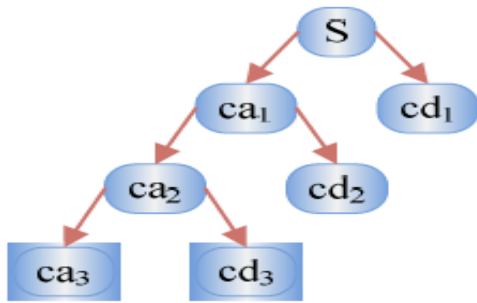


Figure 3. A 3-level wavelet decomposition tree.

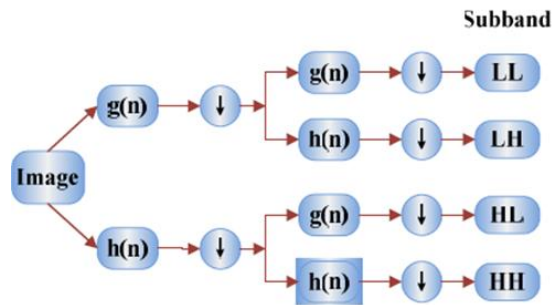


Figure 4. Schematic diagram of 2D DWT.

The LL subband are often considered the approximation component of the image, while the LH, HL, and HH subbands are often considered 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 may be a technique issue associated with digital filter which is usually utilized

in the DWT. As we filter the image, the mask will extend beyond the image at the sides, therefore the solution is to pad the pixels outside the pictures. In our algorithm, symmetric padding method was utilized to calculate the boundary value.

B. 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 amount of features.

PCA is an efficient tool to reduce the dimension of a data set consisting of an outsized number of interrelated variables while retaining most of the variations. It is achieved by transforming the info set to a new set of ordered variables consistent with their variances or importance.

This technique has three effects: it orthogonalizes the components of the input vectors in order that uncorrelated with one another, it orders the resulting orthogonal components in order that those with the most important variation come first and eliminates those components contributing the least to the variation within 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.

III. 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. 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. Hence, solutions are global and usually unique, thus avoiding the convergence to local minima exhibited by other statistical learning systems, such as neural networks. Traditional SVMs constructed a hyperplane to classify data, so they cannot deal with classification problem of which the different types of data located at different sides of a hyper surface, the kernel strategy is applied to SVMs . The resulting algorithm is formally similar, except that every dot product is replaced by a nonlinear kernel function. The kernel is related to the transform (xi) by the equation $k(x_i, x_j) = \varphi(x_i)\varphi(x_j)$. The value w is also in the transformed space, with $w = \sum_i \alpha_i y_i \varphi(x_i)$. Dot products with w for classification can be computed by $w \cdot \varphi(x) = \sum_i \alpha_i y_i \varphi(x_i, x)$.

In another point of view, the KSVMs allow to fit the maximum-margin hyperplane in a transformed feature space. The transformation may be nonlinear and the transformed space higher dimensional; thus, though the classifiers a hyperplane in the higher-dimensional feature space, it may be nonlinear in the original input space.

A. Classification Using SVM:

In machine learning, support vector machines (SVMs), also support vector networks are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

The models are trained using `svmtrain ()` and classified using `svmclassify()`. The kernels used are:

- Linear

- Quadratic
- Polynomial
- MLP
- RBF

The corresponding accuracies are also predicted.

Images	RBF	Linear	Polynomial	Average Accuracy
Image 1	90	80	90	87%
Image 2	80	90	70	80%
Image 3	90	90	70	84%
Image 4	90	90	80	87%
Image 5	90	90	90	90%

Table. 1 Accuracies of different KSVMs.

IV. CONCLUSION

In this study, we have developed a novel DWT+PCA+KSVM method to distinguish between benign and malignant MRIs of the brain. For a given training dataset segmentation of the tumor part was performed using a combination of thresholding and morphological operations, features were extracted using DWT, then the high dimensionality of these features were reduced using PCA. These features were stored as train set features in a .mat file. For test MRI images same set of features were extracted and classified using Kernel support vector machines. The highest accuracy was found with RBF kernel to 90% with holdout cross validation. As an enhancement it was decided to predict the age group of the test images. Volumetric analysis was performed to find out the age group i.e as an infant or an adult.

REFERENCES

[1] Zhang, Y., L. Wu, and S. Wang, Magnetic resonance brain image classification by an improved artificial bee colony algorithm," Progress in Electromagnetics Research, Vol. 116, 65{79, 2011.

[2] Mohsin, S. A., N. M. Sheikh, and U. Saeed, MRI induced heating of deep brain stimulation leads: Effect of the air-tissue interface," Progress in Electromagnetics Research, Vol. 83, 81{91, 2008.

- [3] Golestanirad, L., A. P. Izquierdo, S. J. Graham, J. R. Mosig, and C. Pollo, Effect of realistic modeling of deep brain stimulation on the prediction of volume of activated tissue," *Progress in Electromagnetics Research*, Vol. 126, 1{16, 2012.
- [4] Mohsin, S. A., Concentration of the specific absorption rate around deep brain stimulation electrodes during MRI," *Progress in Electromagnetics Research*, Vol. 121, 469{484, 2011.
- [5] Oikonomou, A., I. S. Karanasiou, and N. K. Uzunoglu, Phased-array near field radiometry for brain intracranial applications," *Progress in Electromagnetics Research*, Vol. 109, 345{360, 2010.
- [6] Scapaticci, R., L. Di Donato, I. Catapano, and L. Crocco, A feasibility study on microwave imaging for brain stroke monitoring," *Progress in Electromagnetics Research B*, Vol. 40, 305{324, 2012.
- [7] Asimakis, N. P., I. S. Karanasiou, P. K. Gkonis, and N. K. Uzunoglu, Theoretical analysis of a passive acoustic brain monitoring system," *Progress in Electromagnetics Research B*, Vol. 23, 165{180, 2010.
- [8] Chaturvedi, C. M., V. P. Singh, P. Singh, P. Basu, M. Singaravel, R. K. Shukla, A. Dhawan, A. K. Pati, R. K. Gangwar, and S. P. Singh. 2.45 GHz (CW) microwave irradiation alters circadian organization, spatial memory, DNA structure in the brain cells and blood cell counts of male mice, *mus musculus*," *Progress in Electromagnetics Research B*, Vol. 29, 23{42, 2011.
- [9] Emin Tagluk, M., M. Akin, and N. Sezgin, \Classification of sleep apnea by using wavelet transform and artificial neural networks," *Expert Systems with Applications*, Vol. 37, No. 2, 1600{1607, 2010.
- [10] Zhang, Y., L. Wu, and G. Wei, A new classifier for polarimetric SAR images," *Progress in Electromagnetics Research*, Vol. 94, 83{104, 2009.
- [11] Camacho, J., J. Pico, and A. Ferrer, Corrigendum to "The best approaches in the on-line monitoring of batch processes based on PCA: Does the modelling structure matter?" [*Anal. Chim. Acta* Volume 642 (2009) 59-68]," *Analytica Chimica Acta*, Vol. 658, 386 Zhang and Wu No. 1, 106{106, 2010.
- [12] Chaplot, S., L. M. Patnaik, and N. R. Jagannathan, Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network," *Biomedical Signal Processing and Control*, Vol. 1, No. 1, 86{92, 2006.
- [13] Cocosco, C. A., A. P. Zijdenbos, and A. C. Evans, A fully automatic and robust brain MRI tissue classification method," *Medical Image Analysis*, Vol. 7, No. 4, 513{527, 2003.
- [14] Zhang, Y. and L. Wu, Weights optimization of neural network via improved BCO approach," *Progress in Electromagnetics Research*, Vol. 83, 185{198, 2008.
- [15] Yeh, J.-Y. and J. C. Fu, A hierarchical genetic algorithm for segmentation of multi-spectral human-brain MRI," *Expert Systems with Applications*, Vol. 34, No. 2, 1285{1295, 2008.
- [16] Patil, N. S., et al., Regression models using pattern search assisted least square support vector machines," *Chemical Engineering Research and Design*, Vol. 83, No. 8, 1030{1037, 2005.
- [17] Wang, F.-F. and Y.-R. Zhang, The support vector machine for dielectric target detection through a wall," *Progress in Electromagnetics Research Letters*, Vol. 23, 119{128, 2011.
- [18] Xu, Y., Y. Guo, L. Xia, and Y. Wu, A support vector regression based nonlinear modeling method for Sic mesfet," *Progress in Electromagnetics Research Letters*, Vol. 2, 103{114, 2008.
- [19] Li, D., W. Yang, and S. Wang, Classification of foreign fibers in cotton lint using machine vision and multi-class support vector machine," *Computers and Electronics in Agriculture*, Vol. 4, No. 2, 274{279, 2010.
- [20] Gomes, T. A. F., et al., Combining meta-learning and search techniques to select parameters for support vector machines," *Neurocomputing*, Vol. 75, No. 1, 3{13, 2012.
- [21] Hable, R., Asymptotic normality of support vector machine variants and other regularized kernel methods," *Journal of Multivariate Analysis*, Vol. 106, 92{117, 2012.

- [22] Ghosh, A., B. Uma Shankar, and S. K. Meher, A novel approach to neuro-fuzzy classification," Neural Networks, Vol. 22, No. 1,100{109, 2009.
- [23] Gabor, D., Theory of communication. Part 1: The analysis of information," Journal of the Institution of Electrical Engineers Part III: Radio and Communication Engineering, Vol. 93, No. 26,429{441, 1946.
- [24] Zhang, Y. and L. Wu, Crop classification by forward neural Progress in Electromagnetics Research, Vol. 130, 2012 387.network with adaptive chaotic particle swarm optimization," Sensors, Vol. 11, No. 5, 4721{4743, 2011.
- [25] Zhang, Y., S. Wang, and L. Wu, A novel method for magnetic resonance brain image classification based on adaptive chaotic PSO," Progress in Electromagnetics Research, Vol. 109, 325{343,2010.
- [26] Ala, G., E. Francomano, and F. Viola, A wavelet operator on the interval in solving Maxwell's equations," Progress in Electromagnetics Research Letters, Vol. 27, 133{140, 2011.