International Journal of Latest Research in Engineering and Technology (IJLRET)
ISSN: 2454-5031
www.ijlret.com || Volume 03 - Issue 10 || October 2017 || PP. 60-69

Classification of brain tumors(Based)OnNoise-reduced MRI by Non Local PCA

Ahmad Mahmoudi shourabi[a], Mohammad.H Shafi Abadi[b], Fatemeh Safara[c]
[a]*Master's student of Computer, Islamic Azad University of Eslamshar*
[b,c]*Faculty of Computer, Islamic Azad University of Eslamshar*

Abstract: The field of medical imaging gains its importance with increase in the need of automated and efficient diagnosis in a short period of time. Magnetic resonance imaging (MRI) is an imaging technique that has played an important role in neuroscience research for studying brain images. Denoistion is an important part before the segmention and also Classification is an important stage in retrieval system in order to distinguish between normal patients and those who have the possibility of having abnormalities or tumor. Over the last decade numerous methods have already been proposed.In this study, we presented a novel method to classify a givenMR brain image as normal or abnormal. The proposed method first employed Non local pca denoising(NLPCA) to reduce noise exist in the images and then wavelet transform to extract features from images,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). The strategy of Kfold stratified cross validation was used to enhance generalization of KSVM. We chose seven common brain diseases (glioma, meningioma, Alzheimer's disease, Alzheimer's disease plus visual agnosia, 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 methodwith the RBF kernel (Radial Basis Function) achieves one of the highest classification accuracy.
Keyword: MRI, Non local PCA, Denoising, Discrete Wavelet Transform, kernel support vector machine, K-FOLD cross validation

1. Introduction

Noise reduction is an important *image* processing method which has wide applications in different fields. The key to noise reduction *is to* reduce the noise. Magnetic resonance (MR) imaging has very important role on current medical and research procedures. However, these images are inherently noisy and thus filtering methods are required to improve the data quality. This denoising process is usually performed as a preprocessing step in many image processing and analysis tasks such as registration or segmentation. There is a large amount of bibliography related to the denoising topic that highlights the relevance of this issue for the scientific community. A large review of MRI denoising methods can be found at Mohan et al. (2014). Currently, most denoising methods can be classified on those that use the intrinsic pattern redundancy of the data and those exploiting their sparseness properties. On the first class, the well known non-local means (NLM) filter (Buades et al., 2005) is maybe the most representative method. This method reduces the noise by exploiting the self-similarity of the image patterns by averaging similar image patterns. In MRI, early works using the NLM method are from Coupé et al. (2008) and Manjn et al. (2008). The bibliography related to this method is quite extensive (Trist Un-Vega et al., 2012; Coupé et al., 2012;Manjn et al., 2009, 2010, 2012; Wiest-Daesslé et al., 2008; Heand Greenshields, 2009; Rajan et al., 2012, 2014). On the other hand, 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 [1,5]. 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 on 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 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 [16,18]. 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 over-fitting [19].Original SVMs are linear classifiers. In this study, 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 [20]. The KSVMs allow us to fit the maximum-margin hyper-plane in a transformed feature space. The transformation may be nonlinear and the transformed space high dimensional; thus though the classifier is a hyper-plane in the high-dimensional feature space, it may be nonlinear in the original input space [21].The structure of the rest of this article is organized as follows. Next Section 2 gives the detailed procedures of proposed work. Section 3 including Noise reduction by Nonlocal PCA, Segmentation, discrete wavelet transform (DWT) and Feature Reduction using Principal Component Analysis, Section 4 introduces the motivation and principles of linear SVM, and then turns to the kernel SVM. Section 5 introduces the K-fold cross validation, that protecting the classifier from over-fitting. Section 6 Case study & dataset. Final Section 7 is devoted entirely to conclusions.

2. PROPOSED WORK

In total, our method consists of three stages:
Step 1. Preprocessing (including **Noise reduction,** Segmentation and 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 [22]. We will explain the detailed procedures of the preprocessing in the following subsections.

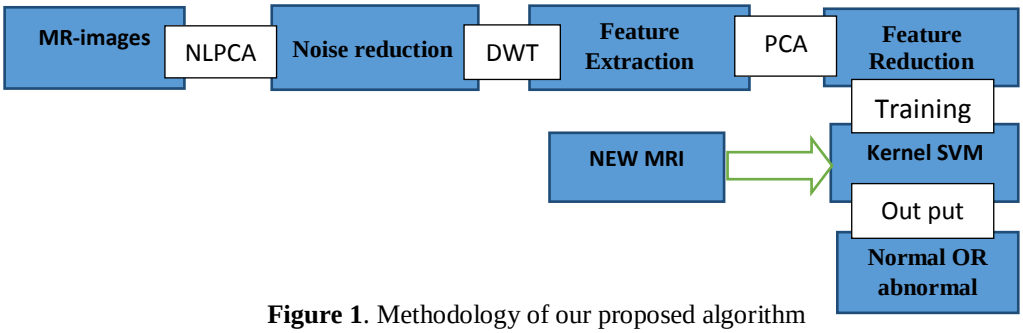

Figure 1. Methodology of our proposed algorithm

3. Preprocessing

3.1.Noise reduction

Brain images mostly contain noise, in homogeneity and sometimes deviation. However, the process of denoising of these images is very important. The NLM filter is a neighbourhood filter [23] which achieves denoising by averaging similar image pixels according to their intensity similarity. The main difference between the NLM and previous related filters is that the similarity between pixels has been made more robust to the noise level by using region comparison rather than pixel comparison; furthermore, pattern redundancy has been not restricted to be local (non-local). That is, pixels far from the pixel being filtered are not penalized due to its distance to the current pixel, as for example happens in the bilateral filter[24].In this paper we present a novel method related with the filter proposed [26]

3.2. Non local PCA

In NL-PCA, A set of N similar patches are selected to create a matrix X [25].This matrix is then transformed using PCA and the least significant eigenvectors are removed by hard thresholding. Finally, the filtered matrix X is obtained by inverting the PCA decomposition, then performed and the less significant components are erased using a hard thresholding rule, **Fig. 2** (i.e., eigenvectors with a standard deviation lower

than a threshold s are set to zero). Finally, since each cluster has contributions from different patches all estimates are combined using a uniform averaging rule. We will refer this method as non-local PCA (NL-PCA) filter. In this work we mainly focus on the two highlighted points clustering in the context of very photon-limited data, and specific denoising method for each cluster [26].

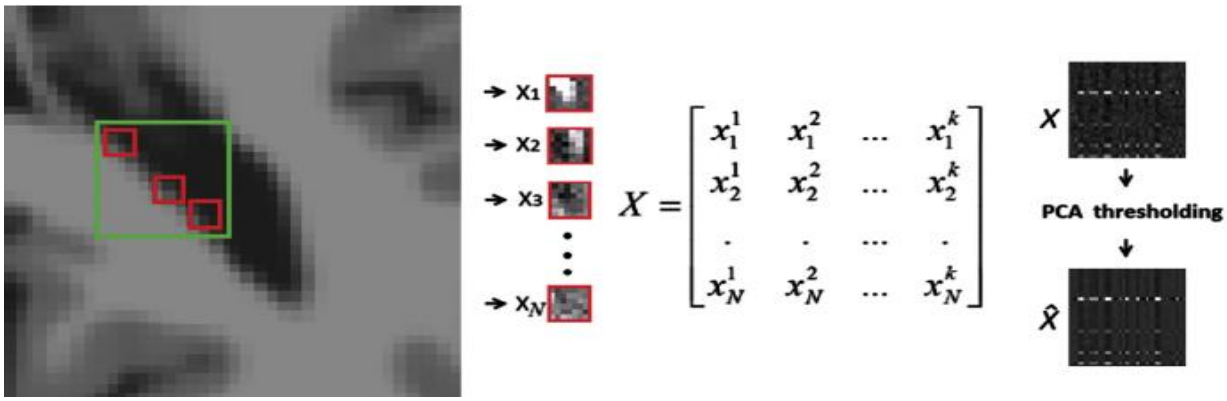

Fig. 2. Overview of NL-PCA scheme.

3.3. Segmentation by Thresholding

Brain image segmentation is one of the most important parts of clinical diagnostic tools. Therefore, accurate segmentation of brain images is a very difficult task. Image segmentation is the process of dividing an image into multiple parts. Image segmentation is one of the most important tasks in medical image analysis and is often the first and the most critical step in many clinical applications. In brain MRI analysis, and also for Features Extraction from (brain) images. in addition, image segmentation is commonly used for measuring and visualizing the brain's anatomical structures, for analyzing brain changes, for delineating pathological regions, and for surgical planning and image-guided interventions.

One of the most prominent issues in brain tumor segmentation is that tumor pixels could have Dissimilarin comparison to normal pixels.Therefore, thresholding method is frequently used for image segmentation. This is simple and effective segmentation method for images with different intensities. The technique basically attempts for finding a threshold value, which enables the classification of pixels into different categories. As shown in fig3, In this study the morphological operations are applied on the image after converting it into binary form. The basic purpose of the operations is to show only that part of the image which has the tumor that is the part of the image having more intensity.

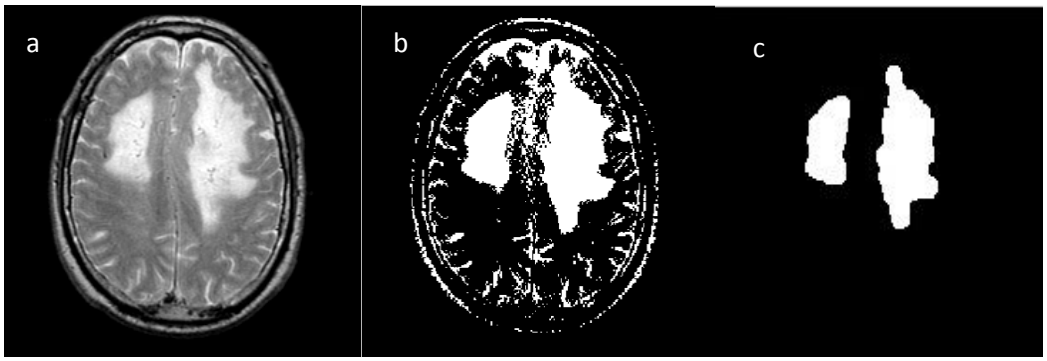

Fig3.Sample of brain MRI (Meningioma): (a) Greyscale image of brain;(b)Binary regions(c)Tumor morphology

3.4. Feature Extraction

In developing a SVM, the first important step is feature selection (new features are selected from the original inputs) or feature extraction (new features are transformed from the original inputs). In the modeling, all available indicators can be used as the inputs of SVM, but irrelevant features or correlated features could deteriorate the generalization performance of SVM. In the framework of SVM, several approaches for feature selection are also available. In [27], The Fourier transform (FT) is the most conventional tool of signal analysis is which breaks down a time domain signal into constituent sinusoids of different frequencies, thus, transforming the signal from time domain to frequency domain. In this paper, we are devoted to the use of

discrete wavelet transform (DWT)for feature extraction as an alternative to the commonly used Fourier transform (DFT).

3.5.Discrete Wavelet Transform

The discrete wavelet transform (DWT) is a powerful implementation ofthe WT using the dyadic scales and positions [28]. 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

$$w_{\psi(a,b)=}\int_{-\infty}^{\infty} x(t)\psi_{a,b}(t)dt \qquad (1)$$

Where

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-a}{b}) \qquad (2)$$

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 [29,31].Equation (1) can be discretized by restraining a and b to a discrete lattice $(\boldsymbol{a = 2^b \& a > 0})$ to give the DWT, which can be expressed as follows.

$$a_{j,k}(\mathrm{n}) = \mathrm{DS}\ [\textstyle\sum_n x(n) g_j^*\ (n - 2^j)\ \mathrm{k})]$$

$$d_{j,k}\ (\mathrm{n}) = \mathrm{DS}\ [\textstyle\sum_n x(n) h_j^*\ (n - 2^j)\ \mathrm{k})] \qquad (3)$$

Here, $a_{j,k}$and $d_{j,k}$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 a(n) and the detail components d(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. 4.

3.6. 2D Discrete Wavelet Transform

In Case of 2D images, the DWT is applied to each dimension separately. Fig. 5 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.

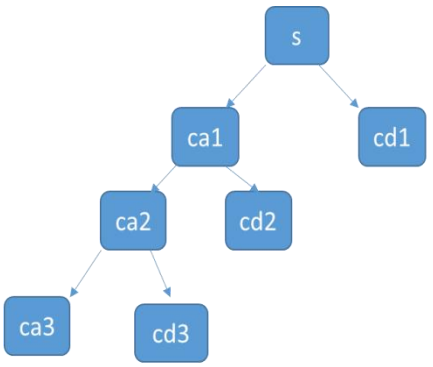

Figure 4. A 3-level wavelet decomposition tree

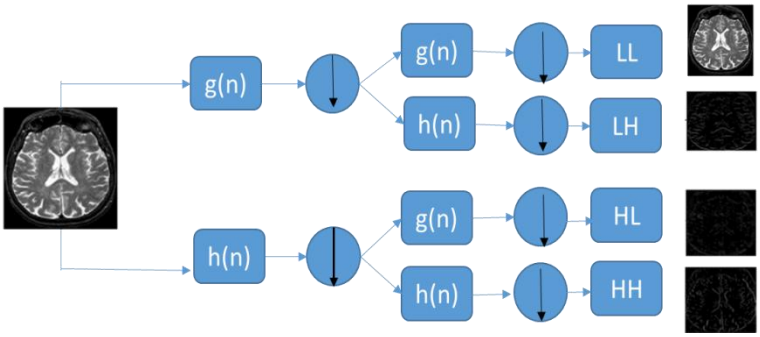

Figure 5. Schematic diagram of 2D DWT

The LL sub-band can be regarded as the approximation component of the image, while the LH, HL, and HH sub-bands 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 [32] was utilized to calculate the boundary value.

3.7. Feature reduction scheme using PCA

In this paper, PCA is used as a feature reduction, The Principal component analysis (PCA) is a well-known method for feature reduction. By calculating the eigenvectors of the covariance matrix of the original inputs, PCA linearly transforms a high-dimensional input vector into a low-dimensional one whose components are uncorrelated. In PCA, the input feature space is transformed into a lower-dimensional feature space using the largest eigenvectors of the correlation matrix. PCA also is the most widely used subspace projection technique. These methods provide suboptimal solution with a low computational cost and Computational complexity. 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. Details about PCA could be seen in Ref. [10].

4. Support vector machine (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 [33]. SVM Unlike traditional methods which minimize the empirical training error, SVM aims at minimizing an upper bound of the generalization error through maximizing the margin between these parating hyper-plane and the data. This can be regarded as an approximate implementation of the Structure Risk Minimization principle. What makes SVM attractive is the property of condensing information in the training data and providing a sparse representation by using a very small number of data points Hence, solutionsare global and usually unique, thus avoiding the convergence to local minima exhibited by other statistical learning systems, such as neural networks.

4.1. Motivation

Let us consider data points each belong to one of two classes, and the aim 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 goal is to create a (p-1)-dimensional hyper-plane. There are many possible hyper-planes that might classify the data successfully. One reasonable choice as the best hyper-plane 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 extension performance. Therefore, we choose the hyper-plane so that the distance from it to the nearest data point on each side is maximized [34]. Fig. 6 shows the geometric interpolation of linear SVMs, here H1, H2, H3 are three hyper-planes 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 ($\mathbf{s_{11}, s_{12}, s_{13}, s_{21}, s_{22}}$ **and** $\mathbf{s_{23}}$) so it is chosen as the best classification hyper-plane [35].

4.2. Principles of Linear SVMs

Consider a pattern classifier, which uses a hyper plane to separate two classes of patterns based on given examples$\{x_n, y_n\}$ for n= 1,…,l, where xn is a vector in the input space $S = R^p$ and y_n denotes the class index taking a value +1 or -1

Given a p-dimensional N-size training dataset of the form

$$\{(x_n, y_n) | x_n \in R^p, y_n \in \{-1, +1\}\} \quad n= 1,\ldots, N \tag{4}$$

International Journal of Latest Research in Engineering and Technology (IJLRET)
ISSN: 2454-5031
www.ijlret.com || Volume 03 - Issue 10 || October 2017 || PP. 60-69

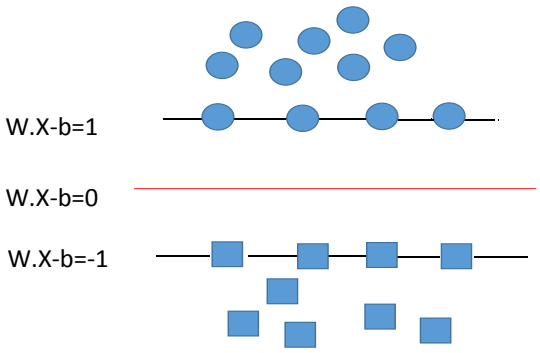

Figure 7. The concept of parallel hyper-planes (w denotes the weight, and b denotes the bias).

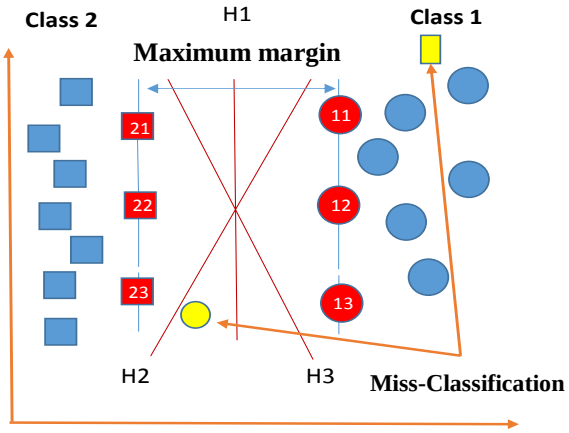

Figure 6. linear SVMs (H denotes for the hyper-plane, S denotes for the support vector).

Here$y_n$is either -1 or 1 corresponds to the class 1 or 2. Each xis a p-dimensional vector. The maximum-margin hyper-plane which divides class 1 from class 2 is the support vector machine we want. And we should know that any hyper-plane can be written in the form of

$$\boldsymbol{w}.\boldsymbol{x} - \boldsymbol{b} = \mathbf{0} \tag{5}$$

Here, W is the normal vector of hyper plane, $w.x$ is the dot protuct of vectors x and w,we want to choose the W and b to maximize the margin between the two parallel (as shown in Fig. 7) hyper-planes as large aspossible while still separating the data. by minimizing || w || we can have the maximize margin between the two hyper plans, So we define the two parallel hyper-planes by the equations as,

$$\boldsymbol{w}.\boldsymbol{x} - \boldsymbol{b} = \pm\mathbf{1} \tag{6}$$

Therefore, the task can be transformed to an optimization problem, i.e., we want to maximize the distance between the two parallel hyper-planes, subject to prevent data falling into the margin. Using simple mathematical knowledge, the problem can be formulated. as

$$\begin{aligned} &\min_{w,b} ||\, w \,|| \\ &s.t.\, y_n(w.x_n - b) \geq 1 \quad \text{n= 1,…, N} \end{aligned} \tag{7}$$

In practical situations the $||w||$ is usually be replace by

$$\begin{aligned} &\min_{w,b} 1/2||w||^2 \\ &s.t.\, y_n(w.x_n - b) \geq 1 \quad \text{n= 1,…, N} \end{aligned} \tag{8}$$

There as on leans upon the fact that$||w||$is involved in a square root calculation. After it is superseded with formula (8), the solution will not change, but the problem is altered into a quadratic programming optimization that is easy to solve by using Lagrange multipliers [36] and standard quadratic programming techniques and programs [37,38].

4.3. Kernel SVM

Much of the flexibility and classification power of SVM's resides in the choice of kernel. These kernel functions have a few main disadvantages, they can 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 [39]. The resulting algorithm is formally similar, except that every dot product is replaced by a nonlinear kernel function. Let us consider a reproducing kernel function$K(x_i, x_j)$and let its eigenfunctions $\varphi_{(\alpha)}(x)$ be, so The kernel is related to the transform, by the equation$K(x_i, x_j) = \varphi_{(\alpha)}(xi)\varphi_{(\alpha)}(xj)$, $\alpha = 1,2, \ldots$ Then,The value w is also in the transformed space, and Dot products with w for classification can be computed by w.$\varphi_{(\alpha)}(x)$=$\sum_\alpha$ $\lambda\alpha$ $\varphi_{(\alpha)}K(x_i, x_j)$. And also we have $K(x_i, x_j) = \sum_\alpha \varphi_{(\alpha)}$ (xi) $\varphi_{(\alpha)}(xj)$=$\emptyset(xi).\emptyset(xj)$ *w*here $\emptyset(xi) = \emptyset_\alpha$ (xj).In another point of view, the KSVMs allow to fit the maximum-margin hyper-plane in a transformed feature space. The transformation may be nonlinear and the transformed space higher dimensional; thus though the classifieris a hyper-plane in the higher-dimensional feature space, it may be nonlinear in the original input space. For kernel, there need be at least one adjusting parameter so as to make the kernel flexible and tailor itself to practicaldata.RBF kernel was chosen due to its superior performances to other kernels reported in a mass of open literatures [43,44]. RBF is defined with the form Of:

$$K(x_i, x_j) = \exp(-\frac{||x_i - x_j||}{2\sigma^2}) \quad (9)$$

Where σ represents the scaling factor, and k the kernel function.

5. K-Fold Stratified Cross Validation

The K-fold cross validation is applied due to its properties as simple, easy, and using all data for training and validation. Since the classifier is trained by a given dataset, so it may achieve high classification accuracy only for this training dataset not yet other independent datasets. To avoid this over-fitting, we need to integrate cross validation into our work. Cross validation *set* does not increase the final classification accuracy, but it will make the classifier reliable and can be generalized to other independent datasets.

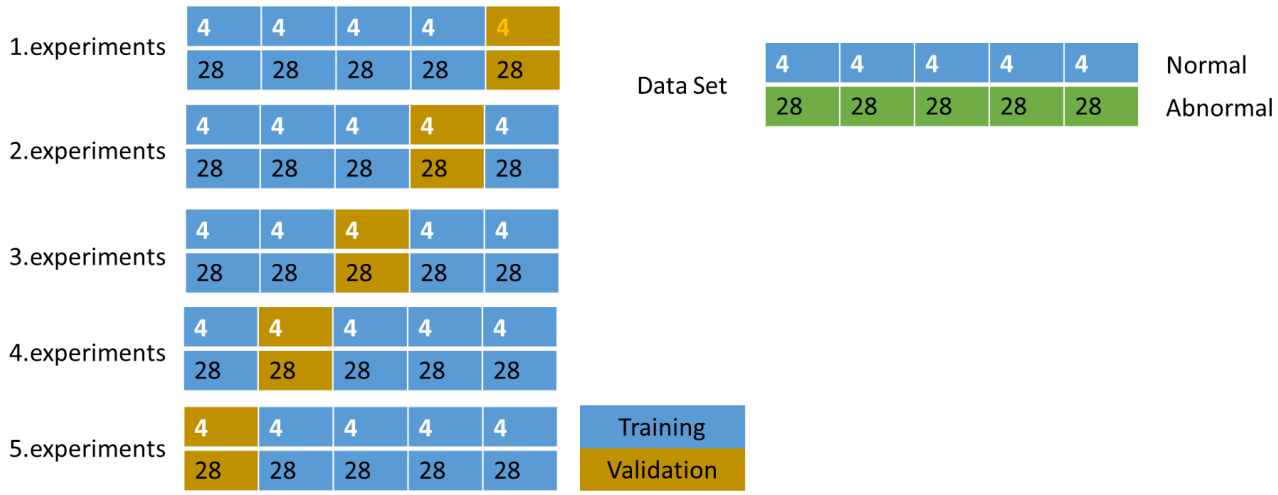

Figure 8. A 5-fold cross validation, Setting of training and validation images (5-fold stratified, Training (128) Validation (32)).

Some statistical learning methods have computationally intensive fitting procedures and data sets can have an extremely large number of observations. This makes leave one out cross validation less feasible. So, K-fold Cross Validation is a nice compromise between single data set validation and leave one out cross validation. In addition, K-fold Cross Validation often provides more accurate estimates of the test error rate, than does leave one out cross validation.

5.1The K-fold Mechanism

The mechanism is to create a K-fold partition of the whole dataset, repeat K times to use K- 1 folds for training and a left fold for validation, and finally average the error rates of K experiments. The schematic diagram of 5-fold cross validation is shown in Fig. 8.The K folds can be purely randomly partitioned, however, some folds may have a quite different distributions from other folds. Therefore, stratified K-fold cross validation was employed, where every fold has nearly the same class distributions [41]. Another challenge is to determine the number of folds.we need to specify the number of folds. If K is set too large, the bias of the true error rate estimator will be small, but the variance of the estimator will be large and the computation will be time-consuming. Alternatively, if K is set too small, the computation time will decrease, the variance of the estimator will be small, but the bias of the estimator will be large [42]In this article, we empirically determined K as 5 through the trial-and-error method, which means that we assume thatparameter Kvaring from 3 to 10 with increasing step as 1, and then we train the SVM by each value. Finally we select the optimal K value corresponding to the highest classification accuracy.

6. Case study

In this section, the proposed techniques have been implemented on a real human brain MRI dataset. All the input dataset (total images is 160: 140 images are abnormal and 20 normal) used for classification consists of T2-weighted MR brain images in axial plane and 256×256 in-plane resolution, which were downloaded from the website of Harvard Medical School, URL: http://med.harvard.edu/AANLIB/),[40].

Note that all diseases are treated as abnormal brains, and our task is a binary classification problem, i.e., to detect abnormal brains from normal brains. We randomly selected 5 images for each type of brain. Since in the dataset there are 4 normal brains and 28 diseases of abnormal brains, Figure 8**,** $5 * (4 + 28) = 160$ images was selected to construct the brain dataset, consisting of 20 normal and 140 abnormal brain images in total. In this study, we downloaded the open SVM toolbox, extended it to Kernel SVM, and applied it to the MR brain images classification. The programs can be run or tested on any computer platforms where Matlab is available.

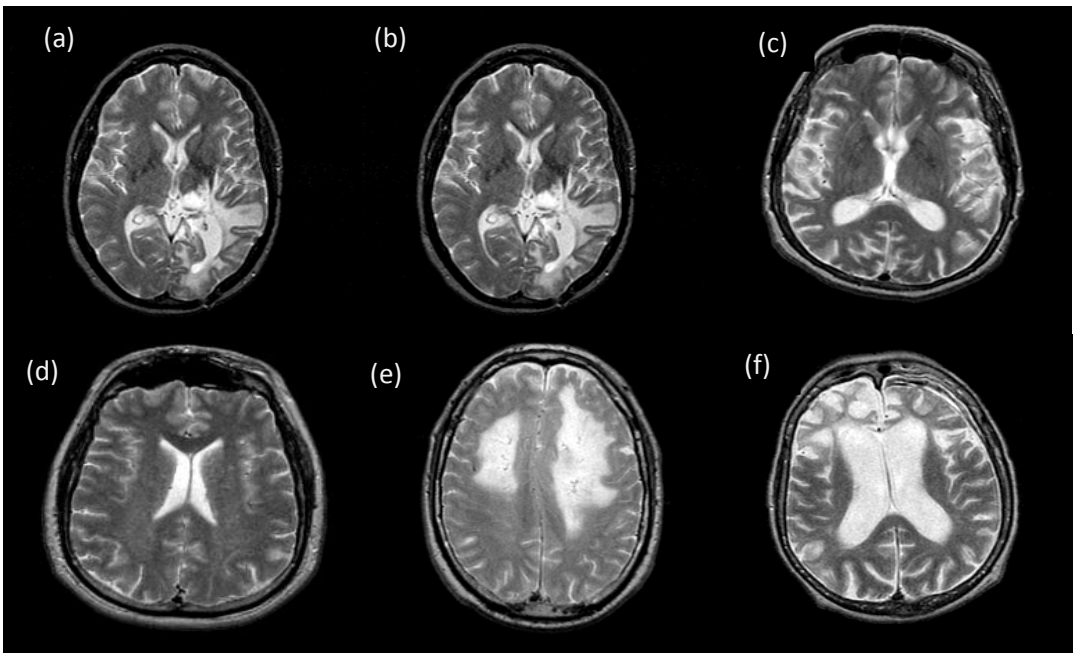

Figure 9. Sample of brain MRIs: (a) glioma; (b)Alzheimer's disease; (c) sarcoma;(d) Normal;(e) meningioma; (f) Pick's disease

7. Conclusions

In this study, we have developed a novel NLPCA+DWT+PCA+KSVM method to distinguish between normal and abnormal MRIs of the brain The medical decision making system designed by the Nonlocal Pca, wavelet transform, principal component analysis (PCA), and supervised learning method (SVM) that we have built gave very promising results in classifying the healthy and pathological brain. This technique of brain MRI classification based on PCA and KSVM is a potentially valuable tool to be used in computer assisted clinical diagnosis.

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, andC. 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, Phasedarray 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 alterscircadian organization, spatial memory, DNA structure in thebrain 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 WuNo. 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 SignalProcessing 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, \An 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]. Yaroslavsky L.P., Digital Picture Processing - An Introduction, Springer Verlag,1985.
[24]. Tomasi C. and Manduchi R. Bilateral Filtering for Gray and Color Images.IEEE International Conference on Computer Vision, India. 1998
[25]. 25.MRI noise estimation and denoising using non-local PCA. *José V. Manjón* [1]Medical Image Analysis, Elsevier, 2015, 22 (1), pp.35-47.

[26]. Joseph salmon, Zachary, Charles-alban deledalle "Poision noise Reduction with non local PCA"Acoustics, Speech And Signal Processing (ICASSP),2012 IEEE International Conference On,30 March 2012.
[27]. P.S. Bradley, O.L. Mangasarian, Feature selection via concave minimization and support vector machines, in: Proceedings of the 15th International Conference on Machine Learning, 1998, Madison, WI, USA, pp. 82–90
[28]. Zhang, Y. and L. Wu, Crop classification by forward neuralProgress In Electromagnetics Research, Vol. 130, 2012 387network with adaptive chaotic particle swarm optimization,"Sensors, Vol. 11, No. 5, 4721{4743, 2011.
[29]. 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.
[30]. 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.
[31]. Iqbal, A. and V. Jeoti, \A novel wavelet-Galerkin method for modeling radio wave propagation in tropospheric ducts," Progress In Electromagnetics Research B, Vol. 36, 35,52, 2012.
[32]. Messina, A., Refinements of damage detection methods based on wavelet analysis of dynamical shapes," International Journal of Solids and Structures, Vol. 45, Nos. 14,15, 4068{4097, 2008.
[33]. Martiskainen, P., et al., Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines,"Applied Animal Behaviour Science, Vol. 119, Nos. 1,2, 32,38,2009.
[34]. Bishop, C. M., Pattern Recognition and Machine Learning (Information Science and Statistics), Springer-Verlag New York,Inc., 2006.
[35]. Vapnik, V., The Nature of Statistical Learning Theory, Springer Verlag New York, Inc., 1995.
[36]. Jeyakumar, V., J. H. Wang, and G. Li, \Lagrange multiplier characterizations of robust best approximations under constraint data uncertainty," Journal of Mathematical Analysis and Applications,Vol. 393, No. 1, 285{297, 2012.
[37]. Cucker, F. and S. Smale, On the mathematical foundations of learning," Bulletin of the American Mathematical Society, Vol. 39,1,49, 2002.
[38]. Poggio, T. and S. Smale, The mathematics of learning: Dealing with data," Notices of the American Mathematical Society (AMS),388 Zhang and WuVol. 50, No. 5, 537{544, 2003.
[39]. Acevedo-Rodriguez, J., et al., "Computational load reductionin decision functions using support vector machines," SignalProcessing, Vol. 89, No. 10, 2066{2071, 2009.
[40]. http://adni.loni.ucla.edu/
[41]. May, R. J., H. R. Maier, and G. C. Dandy, Data splitting forartificial neural networks using SOM-based stratified sampling,"Neural Networks, Vol. 23, No. 2, 283{294, 2010.
[42]. Armand, S., et al., Linking clinical measurements and kinematic gait patterns of toe-walking using fuzzy decision trees," Gait &Posture, Vol. 25, No. 3, 475{484, 2007.
[43]. P.J.G. Nieto, E. Garcia-Gonzalo, F.S. Lasheras and F.J.D. Juez, Hybrid PSO-SVM-based method for forecasting of the remaining useful life for aircraft engines and evaluation of its reliability. Reliability Engineering & System, Safety 138, 219 (2015).
[44]. S. Amari*S.Wu, Improving support vector machine classifiers by modifying kernel functions, Elsevier Science, Neural Networks 12 (1999) 783–789