

# Assisted Magnetic Resonance Imaging Diagnosis for Alzheimer's Disease Based on Kernel Principal Component Analysis and Supervised Classification Schemes

Yu Wang\*, Wen Zhou\*, Chongchong Yu\*, and Weijun Su\*

## Abstract

Alzheimer's disease (AD) is an insidious and degenerative neurological disease. It is a new topic for AD patients to use magnetic resonance imaging (MRI) and computer technology and is gradually explored at present. Preprocessing and correlation analysis on MRI data are firstly made in this paper. Then kernel principal component analysis (KPCA) is used to extract features of brain gray matter images. Finally supervised classification schemes such as AdaBoost algorithm and support vector machine algorithm are used to classify the above features. Experimental results by means of AD program Alzheimer's Disease Neuroimaging Initiative (ADNI) database which contains brain structural MRI (sMRI) of 116 AD patients, 116 patients with mild cognitive impairment, and 117 normal controls show that the proposed method can effectively assist the diagnosis and analysis of AD. Compared with principal component analysis (PCA) method, all classification results on KPCA are improved by 2%–6% among which the best result can reach 84%. It indicates that KPCA algorithm for feature extraction is more abundant and complete than PCA.

## Keywords

Alzheimer's Disease, Feature Extraction, KPCA, Machine Learning, Structural Magnetic Resonance Imaging

## 1. Introduction

Alzheimer's disease (AD) is an insidious and degenerative neurological disease accompanying many clinical manifestations such as diminished memory, persistent cognitive decline and dyskinesia [1]. By the end of 2006, the number of AD patients worldwide reached 26.6 million, and by 2050, one in 85 is estimated as an AD patient worldwide. With the coming of global aging, the negative influence on AD to society and families will become more obvious [2]. Because the clinical manifestation of AD patients is not obvious, it also brings some difficulties for the early diagnosis of doctors. There is no accurate results and effective treatment for AD at present. Most researchers hope to find out some clue when patients are in the mild cognitive impairment (MCI) phase. Then some effective measures are taken to prevent disease from further deterioration [3]. Therefore, it is the key for researchers to accurately determine the stage of AD patients. At present, the methods on the early detection of AD mainly focus

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Manuscript received December 14, 2018; accepted October 14, 2019.

Corresponding Author: Yu Wang (wangyu@btbu.edu.cn)

\* School of Artificial Intelligence, Beijing Technology and Business University, Beijing, China (wangyu@btbu.edu.cn, wenzhoumail@163.com, ycc@163.com, suweijun@btbu.edu.cn)

on neuropsychological test, neuroimaging, electroencephalogram analysis, cerebrospinal fluid detection, and so on [4-7].

In recent years, with the development of computer and imaging technologies using machine learning methods to analyze AD magnetic resonance images and to assist doctors for the diagnosis and analysis has become a mainstream trend [8].

A large number of studies on structural magnetic resonance imaging (sMRI) have been found that the abnormal gray matter of the brain regions is the main manifestation of AD patients, among which the hippocampus, the parahippocampal gyrus, and the medial temporal lobe are obvious. Several approaches have been recently proposed in the literatures aiming at providing a tool which may guide the clinician in the AD diagnosis process. Dhikav et al. [9] analyzed the brain structure of AD/MCI with depression, and found atrophy phenomenon. Lotjonen et al. [10] used machine learning to analyze changes in the hippocampus of AD patients. Cuingnet et al. [11] detected the atrophy of the hippocampus to differentiate between Alzheimer's patients and normal subjects. Falahati et al. [12] used machine learning and multivariate data methods to analyze the sMRI images of AD. Riise et al. [13] studied the abnormal signal pathway of medial temporal lobe of AD patients. Salvatore et al. [14] used MRI brain images and support vector machine (SVM) to make early diagnosis of AD. Takahashi et al. [15] used the apparent diffusion coefficient atlas to analyze the morphology of whole brain voxels, and to achieve the auxiliary diagnosis of AD. Therefore, with the help of computer-aided diagnosis techniques such as machine learning, artificial intelligence etc., the features of MRI images can be extracted and classified so that the subjects can be diagnosed more accurately and better scientific references are provided for doctors.

In order to better assist doctors to early diagnose AD, inspired by previous research, in this paper a general method for aided diagnosis of AD is proposed in which kernel principal component analysis (KPCA) and supervised schemes are combined to classify sMRI data of brain. Preprocessing and correlation analysis on sMRI data are firstly made. Then KPCA is used to extract the features of brain gray matter images. Finally AdaBoost algorithm and SVM algorithm are used to classify.

The rest of this paper is organized as follows. Firstly, feature extraction and classification framework of brain gray images is detailedly described in Section 2. Then extensive experiments and comparisons of the proposed method and the other methods on the MRI dataset are presented in Section 3. Finally, the conclusion is drawn in Section 4.

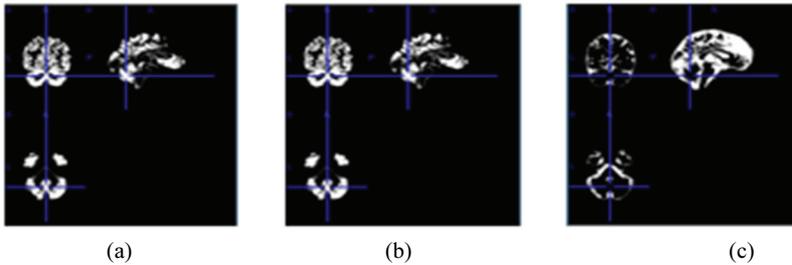
## 2. Method

### 2.1 Image Preprocessing

The image format downloaded from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database is DICOM (Digital Imaging and Communications in Medicine). MRIcro software is used to convert the picture format to NIFTI (Neuroimaging Informatics Technology Initiative). Then the image was preprocessed with Statistical Parametric Mapping (SPM8) software (The Wellcome Centre for Human Neuroimaging, UCL Queen Square Institute of Neurology, London, UK) including slice timing, realignment, co-register, segmentation, normalization, etc. Segmentation can divide the image into three parts which are gray matter, white matter, and cerebrospinal fluid, respectively. Fig. 1 is an example of image segmentation. Because the brain abnormalities of AD patients are mainly in the gray matter, main studies are involved gray matter in this paper.

## 2.2 Feature Extraction

Feature extraction is significant in medical image processing steps. Its main purpose is to measure the intrinsic, essential and important characteristic attributes of the research object, and to digitize the results, or to decompose and symbolize the object, so that a feature vector, symbol string, or a relational graph is obtained. The characteristics of medical images are mainly from the three aspects including color, texture and shape.



**Fig. 1.** An example of brain gray matter image segmentation: (a) gray matter, (b) white matter, (c) cerebrospinal fluid.

According to the MRICro software, the size of sMRI images of each subject is  $121 \times 145 \times 121$  voxel in this paper. A total of 121 slices were obtained along the z-axis, and 20 slices were selected to cover the hippocampus, temporal gyrus and other parts in the activated body concentration. In the case of AD-1 of the subject, the serial number of slices is set as AD-1-i, and  $i = 0, 1, 2, \dots, 19$ . The slice images are grayed and the features are extracted.

### 2.2.1 Principal component analysis

Principal component analysis (PCA) is used to analyze multivariate statistical distribution by feature quantity, and to generally identify multivariate and multivariate statistical analysis [16]. It not only reduces the dimensionality of high-dimensional data, more importantly, but also finds the main characteristic patterns in the data.

Because the adjacent pixels in the image are highly correlated, the input data is redundant. The image of the subject in the experiments is  $121 \times 145$  gray image which is recorded as a vector  $x$  with 17545 dimension.  $x \in R^{[17545]}$  and  $R$  is the real range. Each of these values  $x_j$  corresponds to the brightness value of each pixel. Due to the correlation between adjacent pixels, PCA algorithm can transform input vector into a low dimensional approximation vector, and the error is very small. The algorithm flow is as follows.

#### (1) Data standardization

A subject is taken as an example, and its data is represented as the high matrix  $X$  of 17545 rows and 20 columns.

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} = (x_1 \dots x_p) \quad (1)$$

where  $p=20$ ,  $n=17545$ .

(2) Matrix dimension reduction.

(3) The covariance matrix of the sample characteristics is calculated by

$$\text{cov}(x, x) = \frac{1}{t} x^T x \tag{2}$$

where  $\text{cov}(x, x)$  is the mean covariance of the sample  $x$ , and is a  $n \times n$  matrix.

After the covariance matrix is obtained, its eigenvalues and eigenvectors are calculated. There are many methods for calculating the eigenvalues and eigenvectors of covariance matrix such as Jacobi method, singular value decomposition, etc. which are solved by singular value decomposition method. Assuming that  $X$  is a  $t \times n$  matrix. It can be broken down into

$$X = U \Sigma V^T \tag{3}$$

where  $U$  is a  $t \times t$  square matrix which is called the left singular vector.  $\Sigma$  is a  $t \times n$  matrix. Other elements are all 0 except the diagonal elements, and the elements on the diagonal of the matrix are called singular values.  $V^T$  ( $V$  transpose) is a  $n \times n$  matrix which is called the right singular vector. By multiplying the  $X$ 's transpose with  $X$  and calculating the eigenvalues of  $XX^T$ , the following form is obtained.

$$(X^T X)V = \lambda V \tag{4}$$

The eigenvalues of the covariance matrix are  $\lambda = (\lambda_1, \lambda_2 \dots \lambda_p)$ . The corresponding normalized unit eigenvectors are  $\alpha_i = (\alpha_{i1}, \alpha_{i2} \dots \alpha_{ik} \dots \alpha_{in})$  ( $i = 1, 2, 3, \dots, n$ ) and all eigenvectors are orthogonal each other. At the same time,  $\sum_{i=1}^p \alpha_i^2 = 1$ .

In order to reduce the sample dimension to  $k$  dimension,  $k$  value can be selected.

(4) Determination of dimension  $k$  (Contribution rate)

The value of  $k$  which is the number of eigenvectors in the reduced dimension matrix  $\alpha_i$  need be chosen. The larger the  $k$  is, the more the eigenvectors are, the smaller the dimension-reduction error is, and the more the characteristics of the original features are retained. To preserve the authenticity of the sample as much as possible, the uncertainty is greater than 95% (It is usually between 0.95 and 0.99). The following formula can be used to determine the  $k$  value.

$$\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^n \lambda_i} \geq 0.95 \quad (\lambda_i \text{ is the eigenvalue of covariance}) \tag{5}$$

### 2.2.2 Kernel principal component analysis

KPCA algorithm is a nonlinear extension based on the PCA algorithm. Because PCA is mainly for linear data, it usually is deficient for analyzing nonlinear data. There must be a nonlinear relationship between brain images of different people. KPCA is more advantageous for extracting the main components and reducing the dimension, mainly because KPCA can mine the nonlinear information contained in the dataset.

In KPCA method the most important factor is the choice of nonlinear mapping function  $\Phi$ . The input vector  $X$  is mapped to a high dimensional linear feature space  $\Phi$ . Then in the space  $\Phi$  PCA method is used to calculate principal components.

(1) Determination of the nonlinear mapping function  $\Phi$ .

Training samples are  $x = (x_1, x_2 \dots x_n)$ . The training sample  $x$  is mapped to a high-dimensional space represented as  $\Phi$ . The feature space  $\Phi$  meets the following requirements

$$\sum_{i=1}^n \Phi(x_i) = 0, (i = 1, 2 \dots n) . \tag{6}$$

(2) Calculation of the covariance matrix  $\bar{c}$ .

The covariance matrix is described by

$$\bar{c} = \frac{1}{N} \phi(X_i) \phi(X_i)^T . \tag{7}$$

Because of the mapping in high-dimensional space, it is very difficult to solve  $\phi(X_i) \phi(X_i)^T$  directly. So the kernel function is usually used to solve the covariance. The usual kernel functions include radial basis kernel function (RBF)  $k(x_i, x_j) = (b \cdot s(x_i, x_j) + c) \cdot d$ , polynomial kernel function  $k(x_i, x_j) = (b \cdot s(x_i, x_j) + c)^d$ , and sigmoid kernel function  $k(x_i, x_j) = \tan \cdot h(e \cdot s(x_i, x_j) + f)$ , etc. RBF is used in this paper. To define a  $n \times n$  matrix  $K$ ,  $K_{uv} = (\phi(x_u) \cdot \phi(x_v))$  ( $u, v = 1, 2, \dots, n$ ) can be calculated. The covariance is calculated by the RBF,  $\bar{c} = \frac{1}{n} K$ .

(3) Centralization of kernel function matrix.

Before  $K = [K_{mn}]_{n \times n} = (x_i, x)$  is not established, the centralized kernel function matrix  $K_c$  need be confirmed, i.e.  $K_c = K - I_n K - K I_n + I_n K I_n \cdot I_n$  is a  $n \times n$  matrix. And each element is  $1/n$ .

(4) Calculation of eigenvalues and eigenvectors.

Eigenvalues  $\lambda = (\lambda_1, \lambda_2 \dots \lambda_p)$  and eigenvectors  $\alpha = (\alpha_1, \alpha_2, \alpha_r \dots \alpha_n)$  of the matrix  $K_c$  can be calculated. Then according to Schmidt's orthogonalization and unitization, a new feature vector  $\beta = (\beta_1, \beta_2, \beta_r \dots \beta_n)$  is obtained. Finally, by calculating the cumulative contribution rate, the main component eigenvector after features reduction calculation is determined as  $(\beta_1, \beta_2 \dots \beta_r)$ .

The main purpose of KPCA algorithm is to retain the main characteristic information of the sample as much as possible, and to simplify the representation of the data. At the same time, the cumulative contribution rate is used to select the useful eigenvectors so as to reduce the dimension of the feature matrix and to improve the classification accuracy. Compared with the traditional PCA, KPCA has two innovations:

- (i) a function for mapping the data from an original low-dimensional space to a high-dimensional space is introduced.
- (ii) A theorem, i.e., any vector in space (even a basis vector) can be represented linearly by means of all samples in the space, is introduced.

## 2.3 Classification Algorithm

Using machine learning algorithm for the classification of images is one of the main development trends in recent years. In this paper classic machine learning algorithm like AdaBoost algorithm and SVM algorithm are considered as classifiers, and contrast experiments are carried out.

### 2.3.1 AdaBoost

The AdaBoost method was proposed by Yoav Freund and Robert Schapire in 1997 [17]. It is an iterative algorithm. Repeated learning for the same data can get a series of weak classifiers. Then these weak classifiers are combined to get a strong classifier.

The sample category in this paper is three types including AD (116), mild cognitive impairment (MCI, 116) and normal controls (NC, 117).  $T$  stands for training data and  $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ ,  $x_i \in X \subseteq R^n$ ,  $y_i \in \{0, 1, 2\}$ , where  $N$  is the number of eigenvectors for each sample. The AdaBoost algorithm is mainly divided into three steps. First the weights distribution of training data is initialized, and each training data is set to the same weight which is  $1/N$ . Then each weak classifier  $h$  in weak class space  $H$  is trained. Finally each training-derived weak classifier group is synthesized into a strong classifier.

### 2.3.2 Support vector machine

SVM algorithm is a powerful mathematical tool for classification and regression of sample data [18] which have been successfully applied to many tasks such as speech recognition, face detection, image classification, etc. SVM algorithm is the most widely used machine learning method in recent years. Its main advantages are that it can effectively solve the high-dimensional problem of samples. At the same time SVM has good generalization ability, and can avoid the problem of neural network structure selection and local minima. However other traditional algorithms cannot avoid these problems.

Initially, SVM mainly solved the problem of two classifications, and could not be directly used for multi-classification. Now there are many algorithms to generalize SVM to multi-classification problems. These algorithms are called to multi-category support vector machines (M-SVMs). The usually used models for multi-classification are two ways including “one against all” and “one against one” [19]. In this paper, the “one against all” type of SVM algorithm is adopted.

## 2.4 The Algorithm Flow

A method for aided diagnosis of AD is proposed in this paper in which KPCA and supervised classifying schemes are used aiming at sMRI data of brain. A whole algorithm flow of the proposed method is described as follows.

- Step 1:** Preprocessing and correlation analysis on MRI images using SPM and two-sample T-tests are performed on brain gray matter.
- Step 2:** PCA and KPCA are used to extract features of gray matter images, and key information is determined by calculating eigenvalues and eigenvectors.
- Step 3:** AdaBoost algorithm is used for the classification of MRI images. Different classifiers are trained for the same training set. Then these weak classifiers are grouped to construct a stronger final classifier, and to achieve correct classification of AD, MCI, and NC. Or SVM algorithm is used for the classification of MRI images.

### 3. Experimental Results

In order to prove the validity of the proposed algorithm, a series of experiments were designed. The experimental hardware environment is personal PC, and the processor is Intel Core i7-1065G7, CPU@2.5 GHz, memory 4.00 GB. The compilation environment of the experiments is Matlab 2015a, Python 3.0. Image preprocessing software includes SPM8, xjView, and MRICro.

#### 3.1 Statistics Analysis on ADNI Database

The MRI data used in the experiments in this paper are from the ADNI database (<http://adni.loni.usc.edu/>). Therefore, staff of the ADNI research organization only participated in the establishment and maintenance of the ADNI database and provided data for this experiment, but did not participate in any analysis and suggestions in this paper. This data includes 116 AD, 116 MCI, and 117 NC. The basic statistics of the subjects are shown in Table 1.

In the Statistical Product and Service Solutions (SPSS), the Pearson chi-square test of “gender” showed that significant difference between these three groups does not exist. The ages of these three groups of subjects including AD and MCI, AD and NC, MCI and NC are performed two samples F-test and T-test, and the variance was obtained.

**Table 1.** Relevant feature statistics for subjects of this experimental data

Class <sup>a</sup>	Sample size <sup>b</sup>	Gender (male/female)	Average age (yr)	Age range (yr)
AD	116	64/52	75.36	58–90
NC	117	62/55	75.42	56–94
<i>p</i> -value	-	0.11 <sup>c</sup>	0.47 <sup>d</sup>	0.38 <sup>e</sup>
MCI	116	64/52	75.05	59–91
NC	117	62/55	75.42	56–94
<i>p</i> -value	-	0.11 <sup>c</sup>	0.10 <sup>d</sup>	0.47 <sup>e</sup>
AD	116	64/52	75.36	58–90
MCI	116	64/52	75.05	59–91
<i>p</i> -value	-	0.06 <sup>c</sup>	0.41 <sup>d</sup>	0.34 <sup>e</sup>

<sup>a</sup> Three different stages of Alzheimer's disease: AD, MCI, and NC.

<sup>b</sup> The number of images. The acquisition of medical image data is difficult, and this data is not opened in the hospital. So the data that can be used for research is very limited. In this paper, KPCA method which can perform feature analysis for limited data is used to extract features.

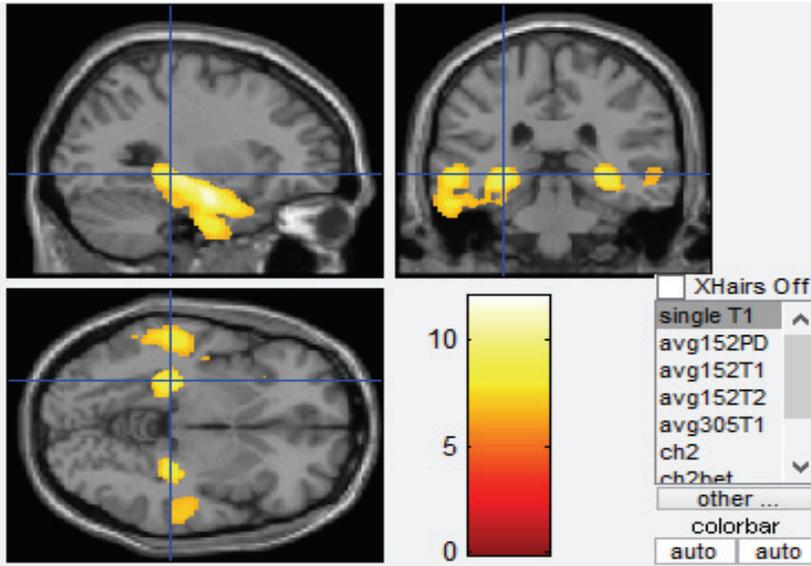
<sup>c</sup> Pearson chi-square test.

<sup>d</sup> Two-sample T-test.

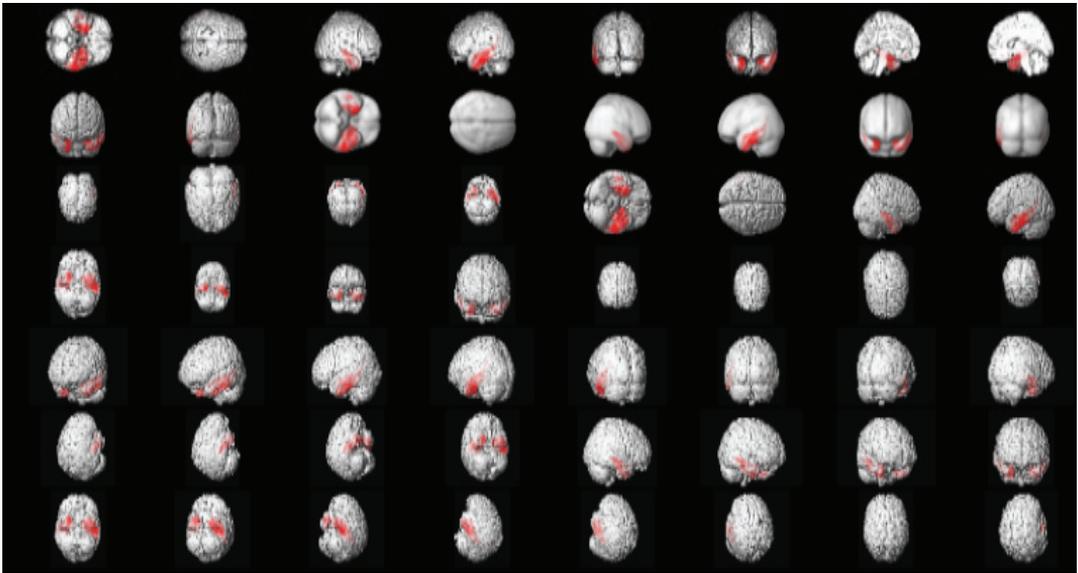
<sup>e</sup> Two-sample F-test.

Using SPM, two-sample T-tests were performed on brain gray matter of 116 cases of AD patients and 116 cases of NC to observe the significance level of tissue voxel values difference. So Alzheimer's lesions area can be gotten, and the disease reasons can be explored.

By means of analysis of false discovery rate (FDR), the lesion area of the AD patients can be obtained. The lesion area is overlaid on the standard template of brain T1 MRI, and different statistical thresholds and voxel sets are set. Then xjView software was used to display the differential brain regions of the AD patients and NC. The generated pseudocolor maps were shown in Fig. 2, and the whole brain rendering was shown in Fig. 3.



**Fig. 2.** Difference brain maps  $p=0.05$ , cluster=20.



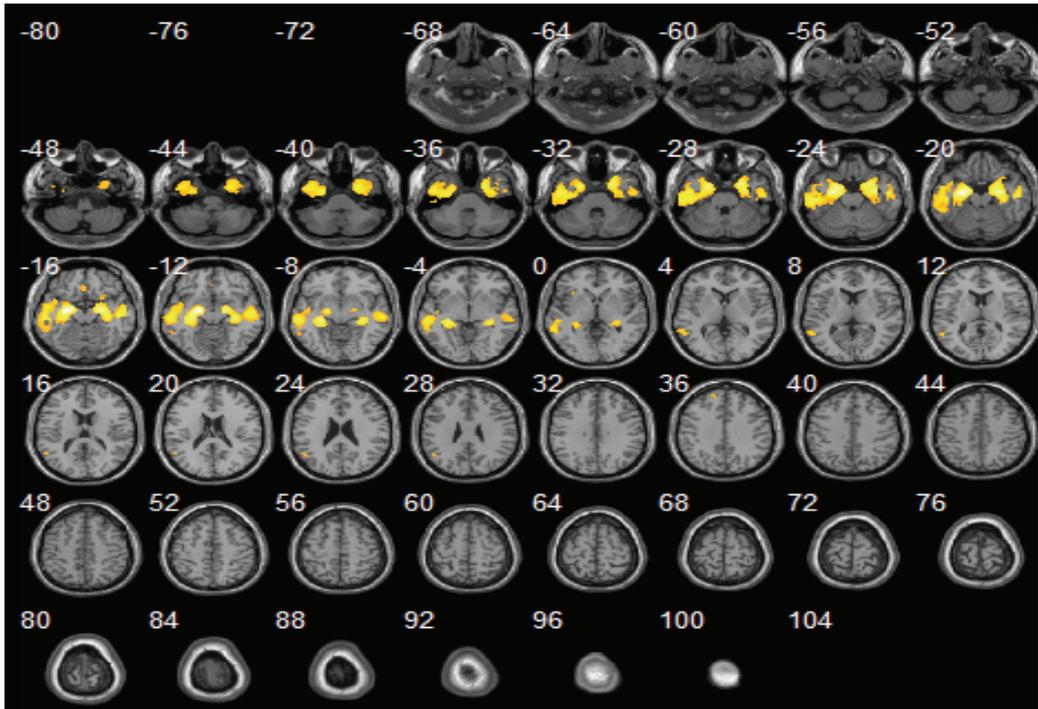
**Fig. 3.** The whole brain rendering images.

By observing the contrast of Figs. 2 and 3, it can be found that around the hippocampus of the brain obvious activation is found in AD patients. Parahippocampal gyrus and thalamus regions have color change of different degrees. According to the degree of activation of different brain regions and the color change, the lesion regions can be obtained which are shown with the yellow color in the map.

In order to display the lesion areas more intuitively, all images were activated and the lesion areas were stratified. The original image was divided differently cross-sectional layers, with a total of 172 layers. The cross-section through the eyebrows was the 0<sup>th</sup> layer. The positive layer was upward, and the negative

layer was downward. The whole brain activation slices are displayed every 4 slices in Fig. 4.

As can be seen from the Fig. 4, we find that from the negative 48<sup>th</sup> layer to the fourth layer, the number of activated voxels is more, and the number of activated voxels of the negative 20<sup>th</sup> layer is the most.



**Fig. 4.** The whole brain activation slices.

### 3.2 Experimental Results and Analysis

In this section, the purpose of the experiments in this paper is to correctly classify AD, NC, and MCI. 116 groups of AD, 116 MCI, and 117 NC were selected as experimental data. The results of the proposed method and other compared methods including gray level co-occurrence matrix (GLCM) and SVM method [20] and kernel support vector machine decision tree (KSVM-DT) method [21] are shown. In the proposed method, PCA and KPCA were used to analyze features of the preprocessed data. 80% of the random samples are selected as the training data, and the remained 20% of the samples are used as testing data. Then the extracted feature vectors are inputted to AdaBoost and SVM classifiers for training and testing results. In SVM classifier three types of kernel functions are used like RBF, polynomial kernel function and sigmoid kernel function. Classification accuracy and Kappa coefficient are used to evaluate the classification results.

Kappa coefficient is a measure of classification accuracy index. The estimation of the Kappa is called KHAT statistics, and the Kappa coefficient represents the proportion of the error reduction between the evaluated classification and the complete random classification. The relation between the statistical value and the classification accuracy is shown in Table 2. The formula of Kappa coefficient is  $Kappa = \frac{p_0 - p_e}{1 - p_e}$  ( $0 < Kappa < 1$ ), where  $p_0$  is the observed proportion of compliance and  $p_e$  is the proportion of

matches due to randomness. Table 2 shows the classification accuracy associated to a Kappa statistics value [22]. The results of the experiments are shown in Table 3.

**Table 2.** Classification accuracy associated to Kappa statistics value

	Kappa statistics				
	0.00–0.20	0.21–0.40	0.41–0.60	0.61–0.80	0.81–1.00
Classification accuracy	Slight	Fair	Moderate	Substantial	Almost perfect

**Table 3.** Comparison of experimental results

Criteria	GLCM-SVM	KSVM-DT	The proposed method							
			PCA			KPCA				
			Adaboost	SVM			Adaboost	SVM		
				RBF	Pol	Sig		RBF	Pol	Sig
Acc(%)	77.78	80.00	77.40	<b>80.10</b>	73.32	74.56	<b>82.62</b>	<b>84.34</b>	79.47	76.20
Kappa	-	-	0.753	0.782	0.724	0.737	0.806	0.823	0.775	0.753

## 4. Discussion

According to the experimental results in Table 3, compared with the PCA method all classification results on KPCA are improved by 2%–6% among which the best result can reach 84%. It proves that KPCA method can extract nonlinear features which are useful for lesion detection as more accurate judgment. When using SVM classifier, the classification performance of common sigmoid function, RBF kernel function and polynomial kernel function are compared. The results show that the classification accuracy of RBF kernel function is higher. And the results of SVM are better than those of AdaBoost, which indicates that SVM is more sensitive to the extracted features. And the above conclusion can be reflected by calculating the Kappa coefficient. Comparing with the experimental results of GLCM-SVM method and KSVM-DT method it can also be seen that the method used in this paper has more advantages.

In conclusion, by statistical analysis, there is no difference on age and gender of the subjects. By preprocessing the sample data, the lesion areas of AD patients are found by using the statistical method of the two-sample T-test. The hippocampus area is used as the reference area, and important brain sections are selected to provide basis for follow-up experiments. It is found through experiments that KPCA implicitly maps data to high-dimensional linear separable spaces. The kernel function is used to process images, because it can extract the nonlinear characteristics of the data. Thus, KPCA is more conducive to classification than PCA method, and better assists doctors for the diagnosis of AD.

## 5. Conclusion

Machine learning methods can be used to assist diagnosis and to solve medical problems, which is one of the powerful evidences of its continuous development. For using machine learning to classify AD, in traditional methods texture features and morphological features are often extracted, and classifiers are

used. But because of the variety of medical data and the large amount of storage, medical data often suffers from “dimension disaster.” KPCA is used to obtain the feature vectors and to reduce dimensionality of the data in this paper. Compared with PCA, better classification performance can be gotten. Analysis of the compared experimental results shows that the proposed algorithm can effectively classify the sMRI data of the subjects. So doctors can be assisted to make correct diagnosis and analysis of AD patients. Our future work will collect more modalities and rich data. At the same time, we will try our best to analyze more types of diseases, and will build a set of data analysis models based on medical images. In addition, we will cooperate with medical institutions to prove the validity of the models.

## Acknowledgement

This work is supported by Joint Project of Beijing Natural Science Foundation and Beijing Municipal Education Commission (No. KZ202110011015), and National Natural Science Foundation of China (No. 61671028).

## References

- [1] K. Hahn, N. Myers, S. Prigarin, K. Rodenacker, A. Kurz, H. Forstl, C. Zimmer, A. M. Wohlschlagler, and C. Sorg, “Selectively and progressively disrupted structural connectivity of functional brain networks in Alzheimer's disease: revealed by a novel framework to analyze edge distributions of networks detecting disruptions with strong statistical evidence,” *Neuroimage*, vol. 81, pp. 96-109, 2013.
- [2] R. Brookmeyer, E. Johnson, K. Ziegler-Graham, and H. M. Arrighi, “Forecasting the global burden of Alzheimer's disease,” *Alzheimer's & Dementia*, vol. 3, no. 3, pp. 186-191, 2007.
- [3] X. Zhu, H. I. Suk, and D. Shen, “A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis,” *NeuroImage*, vol. 100, pp. 91-105, 2014.
- [4] J. Harrison, S. L. Minassian, L. Jenkins, R. S. Black, M. Koller, and M. Grundman, “A neuropsychological test battery for use in Alzheimer disease clinical trials,” *Archives of Neurology*, vol. 64, no. 9, pp. 1323-1329, 2007.
- [5] S. G. Mueller, M. W. Weiner, L. J. Thal, R. C. Petersen, C. Jack, W. Jagust, J. Q. Trojanowski, A. W. Toga, and L. Beckett, “The Alzheimer's disease neuroimaging initiative,” *Neuroimaging Clinics*, vol. 15, no. 4, pp. 869-877, 2005.
- [6] H. Ramshini, A. Ebrahim-Habibi, S. Aryanejad, and A. Rad, “Effect of cinnamomum verum extract on the amyloid formation of hen egg-white lysozyme and study of its possible role in Alzheimer's disease,” *Basic and Clinical Neuroscience*, vol. 6, no. 1, pp. 29-37, 2015.
- [7] H. Zetterberg, T. Skillback, N. Mattsson, J. Q. Trojanowski, E. Portelius, L. M. Shaw, et al., “Association of cerebrospinal fluid neurofilament light concentration with Alzheimer disease progression,” *JAMA Neurology*, vol. 73, no. 1, pp. 60-67, 2016.
- [8] E. Moradi, A. Pepe, C. Gaser, H. Huttunen, and J. Tohka, “Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects,” *Neuroimage*, vol. 104, pp. 398-412, 2015.
- [9] V. Dhikav, M. Sethi, and K. S. Anand, “Medial temporal lobe atrophy in Alzheimer's disease/mild cognitive impairment with depression,” *The British Journal of Radiology*, vol. 87, no. 1042, article no. 20140150, 2014. <https://doi.org/10.1259/bjr.20140150>

- [10] J. Lotjonen, R. Wolz, J. Koikkalainen, V. Julkunen, L. Thurfjell, R. Lundqvist, et al., “Fast and robust extraction of hippocampus from MR images for diagnostics of Alzheimer’s disease,” *Neuroimage*, vol. 56, no. 1, pp. 185-196, 2011.
- [11] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehericy, M. O. Habert, et al., “Automatic classification of patients with Alzheimer’s disease from structural MRI: a comparison of ten methods using the ADNI database,” *Neuroimage*, vol. 56, no. 2, pp. 766-781, 2011.
- [12] F. Falahati, E. Westman, and A. Simmons, “Multivariate data analysis and machine learning in Alzheimer’s disease with a focus on structural magnetic resonance imaging,” *Journal of Alzheimer’s Disease*, vol. 41, no. 3, pp. 685-708, 2014.
- [13] J. Riise, N. Plath, B. Pakkenberg, and A. Parachikova, “Aberrant Wnt signaling pathway in medial temporal lobe structures of Alzheimer’s disease,” *Journal of Neural Transmission*, vol. 122, no. 9, pp. 1303-1318, 2015.
- [14] C. Salvatore, P. Battista, and I. Castiglioni, “Frontiers for the early diagnosis of AD by means of MRI brain imaging and support vector machines,” *Current Alzheimer Research*, vol. 13, no. 5, pp. 509-533, 2016.
- [15] H. Takahashi, K. Ishii, N. Kashiwagi, Y. Watanabe, H. Tanaka, T. Murakami, and N. Tomiyama, “Clinical application of apparent diffusion coefficient mapping in voxel-based morphometry in the diagnosis of Alzheimer’s disease,” *Clinical Radiology*, vol. 72, no. 2, pp. 108-115, 2017.
- [16] K. Pearson, “LIII. On lines and planes of closest fit to systems of points in space,” *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559-572, 1901.
- [17] Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119-139, 1997.
- [18] V. Vapnik, *Estimation of Dependences Based on Empirical Data*. New York, NY: Springer, 1982.
- [19] Z. Qi and J. Zhang, “Design of fuzzy SVM multi-category classifier model and application in engine fault diagnosis,” in *Proceedings of 2012 5th International Congress on Image and Signal Processing*, Chongqing, China, 2012, pp. 1782-1785.
- [20] S. Pazhanirajan and P. Dhanalakshmi, “Alzheimer disease classification using SVM and multi-SVM,” *Advances in Natural and Applied Sciences*, vol. 9, no. 6(SE), pp. 29-35, 2015.
- [21] Y. D. Zhang, S. Wang, and Z. Dong, “Classification of Alzheimer disease based on structural magnetic resonance imaging by kernel support vector machine decision tree,” *Progress In Electromagnetics Research*, vol. 144, pp. 171-184, 2014.
- [22] J. R. Landis and G. G. Koch, “An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers,” *Biometrics*, vol. 33, no. 2, pp. 363-374, 1977.



**Yu Wang** <https://orcid.org/0000-0001-7284-9175>

She received the B.S. and M.S. degrees in mechatronic engineering from Northeastern University, Shenyang, China in 1999 and Yanshan University, Qinhuangdao, China in 2005 respectively, and the Ph.D. degree in controlling theory and controlling engineering from University of Science and Technology Beijing, Beijing, China, in 2009. She worked as a postdoctor in Tsinghua University from 2009 to 2011, and a Visiting Scholar in University of Pennsylvania, USA, from 2016 to 2017. Since 2011, she has been with the faculty of Beijing Technology and Business University, Beijing, China, and is currently a professor of the School of Artificial Intelligence, Beijing Technology and Business University. Her research interests include medical image processing, pattern recognition, and computer vision.



**Wen Zhou** <https://orcid.org/0000-0002-5287-3247>

He received B.S. degrees in School of information engineering from Tong Ji Zhe Jiang College, Jiaxing, China in 2015, and M.S. degree in the School of Artificial Intelligence, Beijing Technology and Business University, Beijing, China in 2019. He is currently an algorithm engineer in Shenzhen Weiaipu Information Technology Co., Ltd. His research research interests include image processing and pattern recognition.



**Chongchong Yu** <https://orcid.org/0000-0001-8685-3909>

She received her M.S. degree from China University of Mining and Technology, Beijing, China in 1996, and Ph.D. degree in University of Science and Technology Beijing, Beijing, China. She was a Visiting Scholar at Royal Melbourne Institute of Technology in Australia, and at University of Waterloo in Canada. Since 1996, she has been with the faculty of Beijing Technology and Business University, Beijing, China, and is currently a professor and dean of the School of Artificial Intelligence, Beijing Technology and Business University. Her research interests include intelligent information processing, machine learning and pattern recognition, and computer network.



**Weijun Su** <https://orcid.org/0000-0002-2836-5445>

He received his M.S. degree from Beijing Technology and Business University, Beijing, China. He is currently a professor of the School of Artificial Intelligence, Beijing Technology and Business University. His research interests include detection technologies and optimal control, the theory and application of the Internet of Things, etc.