

Original Research

Using Regularized Multi-Task Learning for Schizophrenia MRI Data Classification

Yu Wang^{1,*}, Jiantong Shi¹, Hongbing Xiao¹

¹Beijing Key Laboratory of Big Data Technology for Food Safety, School of Artificial Intelligence, Beijing Technology and Business University, 100048 Beijing, China

*Correspondence: wangyu@btbu.edu.cn (Yu Wang)

Academic Editor: Rafael Franco

Submitted: 14 December 2021 Revised: 20 February 2022 Accepted: 23 February 2022 Published: 24 June 2022

Abstract

Background: Machine learning techniques and magnetic resonance imaging methods have been widely used in computer-aided diagnosis and prognosis of severe brain diseases such as schizophrenia, Alzheimer, etc. **Methods:** In this paper, a regularized multi-task learning method for schizophrenia classification is proposed, and three MRI datasets of schizophrenia, collected from different data centers, are investigated. Firstly, slice extraction is used in image preprocessing. Then texture features of gray-level co-occurrence matrices are extracted from the above processed images. Finally, a p -norm regularized multi-task learning method is proposed to simultaneously learn the site-specific and site-shared features of the multi-site data, which can effectively discriminate schizophrenia patients from normal controls. **Results:** The classification error rate on 10 datasets can be reduced from 10% to 30%. **Conclusions:** The proposed method obtains excellent results and provides objective evidence for clinical diagnosis and treatment of schizophrenia.

Keywords: schizophrenia; magnetic resonance imaging; feature extraction; regularized multi-task learning

1. Introduction

According to the quantitative evaluation of the world health organization, brain diseases such as Alzheimer, Parkinson, and schizophrenia, etc. account for about 28% of all kinds of diseases in the world [1], which seriously threatens human health. Among them, schizophrenia is the most common psychosis. Its clinical manifestation is a syndrome with different symptoms involving many obstacles such as perception, thinking, emotion, behavior, as well as the disharmony of mental activities [2]. The diagnosis of schizophrenia in traditional medicine is mostly based on American DSM-IV, international ICD-10, and domestic classification and diagnostic criteria of mental disorders [3]. With the development of science and technology, various types of high-end medical imaging devices are developing rapidly. And medical images play an increasingly important role for assisting doctors to diagnose diseases. However, a large number of medical images have obviously increased the burden of doctors. At present, as a research hotspot in the field of medical science, image classification task is widely completed with the help of computer-aided means.

Among many medical images, magnetic resonance imaging (MRI) has been widely used in the clinical diagnosis of brain diseases due to its advantages of non-radiation and high resolution [4,5]. A great deal of studies on sMRI show that abnormal gray matter located in multiple parts of the brain such as temporal lobe, parietal lobe and frontal lobe is the main manifestation of schizophrenia patients [6,7]. In many papers, brain abnormalities in schizophre-

nia striatum [8,9] and hypothalamus [10] have been identified. In paper [11] the gray matter texture analysis of magnetic resonance images is used, and it is determined that there is heterogeneity in the cerebral gray matter structure of schizophrenic patients. Therefore, abnormal sMRI images can be used to diagnose schizophrenia disease according to biological characteristics. In paper [12], a method called volume local binary patterns (VLBP) was used to calculate texture features to classify fMRI images of schizophrenic patients. In paper [13], the gray level co-occurrence matrix texture features of sMRI images combined with XGBoost were used to classify schizophrenia patients, which effectively verified the role of computer-aided diagnosis.

At present, based on MRI images of some brain diseases, researchers usually study image segmentation, recognition and classification in single area. However, in the era of internet information explosion, it is possible to obtain MRI images of multiple regions of homologous brain diseases through multiple channels. In the literature [14] multi-site data with 900 subjects was used, and about 200 subjects from 2 sites were included in the paper [15]. Papers [16,17] show that, compared with a small number of samples in a single area, MRI data of the same kind of brain diseases in multiple sites can provide more sufficient statistical information, so as to better explore the functional mode of the brain structure of a patient. By studying the papers [18–20], it can be found that compared with the patients in a single area, the population distribution of the same disease in different regions is diverse. For example, there are certain differences in the structure and function of different



people's brains. The severity of the disease and the clinical symptoms in the population of multi-site are different. The type of patients is more extensive, etc. Obviously, studying the medical images of the same patients in many areas can not only get more comprehensive image information or a consistent pattern of abnormal pathological characteristics, but also analyze the characteristics of medical images in a single area. The experimental results are also more convincing. With the wide cooperation between international medical institutions and medical workers, it is an inevitable trend to study the pathological mechanism of the disease by using medical image data from multi-site of the sick people.

It can be seen from the above that, on the basis of multi-site MRI image data of schizophrenia brain disease, computer-aided diagnosis technology is used to distinguish normal and abnormal MRI images, and finally to correctly classify patients and normal people. The advantages are obvious. However, when any classifier is trained under limited sample conditions, it is difficult to replace the infinite sample pattern with a limited sample mode and to achieve a high degree of conformity with the actual pattern, especially under the conditions of the small number of MRI image sample and less diversity. Aiming at this problem, the advantage of multi-task learning method [21] is gradually presented. Multi-task learning is an optimal learning method by mining shared information among tasks while training multiple related tasks. It can significantly improve the learning effect of the algorithm, and has been applied to many fields such as spam filtering [22], natural image classification [23], various disease modeling, classification and prediction [24,25] and so on. In the articles [26] and [27], using mutual inductive bias, multi-task learning can obtain bias information to supplement the lack of samples, which can simultaneously learn single task's unique feature information and feature information shared by multiple tasks, and effectively improve the generalization ability of the model. Evgeniou *et al.* [28] proposed a regularized multi-task learning (rMTL) method based on support vector machine (SVM) model. This method added regularized penalty term which constrains the correlation parameters of different task model, and improves the generalization ability of the model. In the paper [29], a regularized multi-task learning method based on SVM and hybrid norm of and was proposed for MRI image classification of depressive disorders patients, and excellent results were obtained. The classification error rate on 10 datasets can be reduced from 10% to 30%.

Inspired by previous research, in this paper the classification problem of schizophrenia MRI images in multi-site data centers is regarded as a multi-task learning problem. A regularized multi-task learning classification model with SVM and p -norm is constructed, and the gradient descent method is used to optimize this model. Finally, this model is used to classify the MRI images of schizophrenic patients and normal people.

The rest of this paper is organized as follows. In section 2, the detail of the data is given, and the data preprocessing and statistical analysis is specified. In section 3, detailedly presents the proposed classification method is introduced. The experimental results and analyses are provided to demonstrate the feasibility and effectiveness of our method in section 4. In section 5, a conclusion is drawn.

2. Data

2.1 Database

MRI data were collected in the United States of America, Brazil, and China (referring as site A, site B, and site C). 132 normal controls (NC) and 137 schizophrenia patients (SCZ) were recruited in site A, 94 NC and 62 SCZ in site B, 181 NC and 144 SCZ in site C. All patients met DSM-IV [30] criteria and were diagnosed as schizophrenia by psychiatrists. All the sMRI image scans were acquired on a GE 3-T Signa scanner (GE Medical Systems, Milwaukee WI, USA) with the following protocol: slice thickness = 1 mm, TE = 3.2 ms, TR = 8.2 ms, flip angle = 12°, acquisition matrix = 256 × 256, FOV = 25.6 cm. All participants were remained quiet, without moving, eyes closed, no sleeping, and no system thinking activities during functional MRI scanning. None of them has any history of other neurological diseases or serious drug diseases. Written informed consent was obtained from all subjects before MRI scanning. In order to get a balanced subject numbers of controls and patients, 60 NC and 60 SCZ were randomly chosen in each site.

The acquired MRI images were preprocessed using the statistical parametric mapping software package (SPM, Wellcome Trust Centre for Neuroimaging, Institute of Neurology, London, UK, <http://www.fil.ion.ucl.ac.uk/spm>) in which the following steps such as skull stripping, bias correction, tissue segmentation (four types of tissue including gray matter, white matter, cerebrospinal fluid, and lateral ventricles), spatial registration to a Montreal neurological institute (MNI) template, generation of the regional analysis of volumes examined in normalized space maps called RAVENS [31,32] of gray matter, white matter, cerebrospinal fluid by the deformable registration package, named DRAMMS, which is publicly available [33], and the smoothing of RAVENS maps using a 6-mm full width at half maximum (FWHM) Gaussian filter are included.

2.2 Statistical Analysis

To better illustrate the demographic and clinical characteristics of the study groups, the Student's t -test of the age means and Pearson Chi-square test for gender differences were calculated. The statistical analysis of 200 subjects is analyzed in the Table 1.

It can be seen from the Table 1 that these 200 subjects are matched in age and gender in dataset. And no statistically significant characteristic occurs when the difference p is smaller than 0.05.

Table 1. Statistic analysis of the participants' characteristics in this study.

Region	Class	Sample size	Gender (male/female)	Average age/years	Age range/years
A	SCZ	60	37/23	34.68	18~60
	NC	60	28/32	31.78	13~65
	<i>p</i> Value	–	0.09 ^a	0.22 ^b	–
B	SCZ	60	44/16	27.53	18~50
	NC	60	37/23	29.73	18~50
	<i>p</i> Value	–	0.17 ^a	0.13 ^b	–
C	SCZ	60	23/37	30.85	16~54
	NC	60	28/32	33.93	20~57
	<i>p</i> Value	–	0.36 ^a	0.11 ^b	–
A + B + C	SCZ	180	105/75	31.39	16~60
	NC	180	89/91	30.90	15~65
	<i>p</i> Value	–	0.09 ^a	0.65 ^b	–

^aPearson Chi-square test. ^bTwo-sample *t*-test. A, United States of America; B, Brazil; C, China; SCZ, Schizophrenia patients; NC, Normal Controls.

3. Methods

3.1 Image Preprocessing

Brain MRI data is typically stored in the form of three-dimension. In our study, we investigate the gray matter image of structural magnetic resonance imaging (sMRI) which has a size of $96 \times 113 \times 94$ voxel. Because dimension disaster problem will happen, or the model performance will be cut down by the large number of irrelevant and redundant feature information if features are directly extracted in the light of each voxel, a preprocessing method which includes slicing and calculating weighted sum of average gray images is proposed in this paper. The detailed steps are as follows:

(1) Original images are sliced. For each subject, the size of gray matter image is $96 \times 113 \times 94$ voxel. The volume image is sliced in the Z-axis direction. Therefore, 94 slices can be gotten.

(2) Sliced images are selected and converted to gray images. By removing 10 slices (the head most 5 slices and the backmost 5 slices) which don't include feature information in the sliced gray matter images, and converting the remaining slices into gray images, the sequentially numbered slices are obtained and denoted as ($i = 0, 1, 2 \dots 83$). The part of sliced and grayed image slices of the first subject numbered NC001 are shown in Fig. 1.

(3) The gray images are weighted and averaged. According to the structural integrity of the cerebral gray matter in each slice, the 84 slices were divided into three groups in sequence among which slices from 1st to 28th are a group. Similarly, slices from 29th to 56th and slices from 57th to 84th are respectively a group. Because the slices closer to the middle reflect more complete structure on the gray matter of the brain, they contain more feature information. And then the greater weight are given when calculating the average grayscale image. The equations for these three groups

of images are described by

$$\text{Img1} = \frac{m_0 \times 1 + m_1 \times 2 + m_2 \times 3 + \dots + m_{27} \times 28}{1 + 2 + 3 + \dots + 28} \quad (1)$$

$$\text{Img2} = \frac{m_{28} \times 1 + m_{29} \times 1 + m_{30} \times 1 + \dots + m_{55} \times 1}{28} \quad (2)$$

$$\text{Img3} = \frac{m_{56} \times 28 + m_{57} \times 27 + m_{58} \times 26 + \dots + m_{83} \times 1}{28 + 27 + 26 + \dots + 2 + 1} \quad (3)$$

Img1, *Img2*, and *Img3* are calculated using sMRI data of each subject after sMRI data is preprocessed, which puts a good way for the subsequent feature extraction.

3.2 Feature Extraction

Image feature extraction is a fundamental and critical step in medical image processing whose purpose is to show the characteristics or attributes of the samples in the form of numerical values, symbols and feature vectors. The results of feature extraction directly affect the classification accuracy. Because the texture information in the image is not sensitive to noise, light and color, the texture feature is chosen to use in this paper.

3.2.1 Texture Features Based on Gray-level Co-Occurrence Matrix

There is no universally mathematical model for texture feature extraction. Because the gray-level co-occurrence matrix (GLCM) model method is not restricted by the analysis object, it can well reflect the spatial gray distribution of the image and the texture features of the image, and has been widely used [34]. GLCM which describes the grayscale of adjacent pixels (or within a certain distance) is a statistical matrix, and reflects the comprehensive information which consists of the image gray change in the

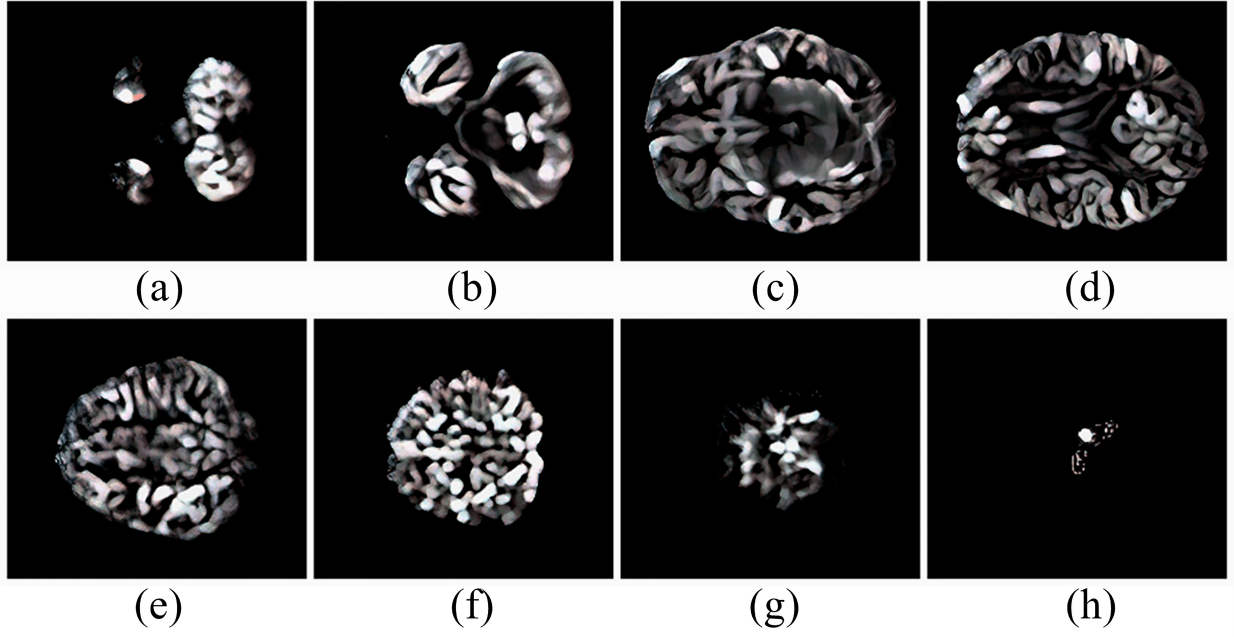


Fig. 1. Some examples of sliced gray matter images. From (a) to (h) they are respectively the 13th, 20th, 34th, 48th, 62nd, 69th, 76th, 83rd slices.

direction, interval, and amplitude. Assume the gray level of a digital image is N , $p(i, j)$ represents the possibility (or frequency) of the appearance of grayscale j under the condition that the starting grayscale is i , where it is assumed that j is along the direction θ of i and the space distance is d . GLCM shows statistical information, and can be calculated using the following equations.

(1) Mean:

$$\text{Mean} = \bar{x} = \sum_{i=0}^N \sum_{j=0}^N p(i, j) \times i \quad (4)$$

The mean reflects the regularity of texture. The smaller the mean is, the more disorganized the texture is.

(2) Variance:

$$\text{Variance} = \sum_{i=0}^N \sum_{j=0}^N p(i, j) \times (i - \bar{x})^2 \quad (5)$$

The variance measures the deviation of the pixel value from the mean. The larger the variance is, the more the change of gray scale is.

(3) Entropy:

$$\text{Entropy} = - \sum_{i=0}^N \sum_{j=0}^N p(i, j) \times \ln p(i, j) \quad (6)$$

Entropy is the measurement of information contained in an image. The greater the value of entropy is, the more complex the texture is.

(4) Contrast:

$$\text{Contrast} = \sum_{i=0}^N \sum_{j=0}^N p(i, j) \times (i - j)^2 \quad (7)$$

Contrast reflects the total amount of local gray scale changes in an image. The greater the contrast of an image is, the clearer the visual effect of this image is.

(5) Correlation:

$$\text{Correlation} = \sum_{i=0}^N \sum_{j=0}^N \frac{(i - \text{Mean}) \times (j - \text{Mean}) \times p(i, j)^2}{\text{Variance}} \quad (8)$$

The correlation is a measure of the linear relationship of the gray scale. The longer the extension of the gray value in a certain direction is, the greater the correlation is.

(6) Homogeneity:

$$\text{Homogeneity} = \sum_{i=0}^N \sum_{j=0}^N p(i, j) \times \frac{1}{1 + (i - j)^2} \quad (9)$$

Homogeneity is used to measure the uniformity of the local gray level of an image. The more homogeneous the local gray scale is, the greater the value is.

(7) Energy:

$$\text{Energy} = \sum_{i=0}^N \sum_{j=0}^N p(i, j)^2 \quad (10)$$

Energy is the measurement of uniformity of gray distribution in the image.

Using the above 7 kinds of texture features, a texture vector representing an image can be obtained.

3.2.2 Normalization

The feature vectors need to be normalized so that any feature does not dominate among all of these features. All the feature vectors are normalized to contain zero mean and unit variance. The normalization is done by using the equation:

$$x_i = \frac{x_i - \mu}{\sigma_x} \quad (11)$$

where μ and σ_x are respectively the mean and standard deviation of all the features x_i .

In summary, for each of these three weighted and averaged gray images (*Img1*, *Img2*, *Img3*), the above 7 statistics are calculated and normalized. Thus every subject can be represented using a vector containing 21 features.

3.3 Classification Algorithm

3.3.1 Multi-Task Learning and Single Task Learning

Machine learning algorithm usually learns a task every time, and decomposes the complex problem into the theoretically independent sub-problems. Then it learns each sub-problem separately. Finally, it constructs the mathematical model of complex problems by combining the learning results of the sub-problems, namely single task learning [35,36]. Multi-task learning is a machine learning method relative to single task learning. It uses information shared by multiple tasks to learn multiple tasks simultaneously, and solves multiple problems simultaneously. The obtained results interact with each other. Sharing information between tasks is the prerequisite for multi-task learning. On this basis, the training of multiple tasks can improve the overall generalization performance of the model. The main difference between multi-task learning and single task learning is that the training process of the model is different [31,32]. In the training process of single task learning, each task is independent and does not affect each other. Its disadvantage is that it ignores the information contained in other tasks during the process of single task training. To some extent, the loss of relevant training information is caused, and this part of the lost information may be very useful for the training process. Nevertheless, the training of multi-task learning takes into account the correlation and useful information shared between tasks. At the same time it learns multiple tasks in parallel. The difference between the two training model is shown in the Fig. 2 below.

3.3.2 Support Vector Machine Classification Algorithm with Multi-Task p -norm

The most important issue of multi-task learning is how to build the model of relationship between tasks, and to

make the relevant tasks share information. Finally, the goal of using the correlation between different tasks to improve the learning performance of the algorithm is achieved. We expect that the model fits the training data as much as possible, and is not too complicated at the same time. Therefore, the support vector machine algorithm with regularized multi-task is adopted to solve MRI image classification problems of mental illness in multi-site data centers.

Assuming that there are t supervised learning tasks in a multi-task learning problem, for each task i , the learning function is assumed to be $f_t : R^d \rightarrow R$. Training set is $X_i = [x_1, x_2, \dots, x_n] \in R^{d \times n} (i = 1, 2, \dots, t)$, where n is the number of input sample, and d is the dimension of the sample feature vector. $Y_i = [y_1, y_2, \dots, y_n] \in R^n (i = 1, 2, \dots, t)$, and $y_i \in \{+1, -1\}$ is the label of each sample in the i -th task. The weight coefficient matrix of t -th supervised learning tasks is $W = [w_1, w_2, \dots, w_t] \in R^{d \times t}$. The goal of multi-task learning is to get t related tasks' regression or classification function $f_t(x)$ by learning the training data. In order to accurately find out the $f_t(x)$ function, the multi-task objective function should be determined first. Assuming that $f_t(w_i^T x_i, Y_i)$ is the loss function of the t -th task, in the classification problem, the classical loss function includes log-likelihood function, exponential function and hinge function. Support vector machine model with regularized multi-task learning and a least empirical error can be expressed by [28].

$$\min_W \frac{1}{n} \sum_{i=1}^n f(w_i^T X_i, Y_i) + \lambda \Omega(W) \quad (12)$$

where the first item is the empirical loss function on training data. $f(w_i^T X_i, Y_i)$ uses hinge loss function. The second one is the regularized term which can encode correlation between tasks. λ is the parameter of the regularized term, and $\lambda > 0$.

The optimal solution for a single t task is equivalent to the global problem of solving the target function of the joint t task, and is described by

$$\min_W \sum_{t=1}^t \frac{1}{n} \sum_{i=1}^n f(w_i^T X_i, Y_i) + \lambda \sum_{t=1}^t \Omega(W) \quad (13)$$

The norm of the model parameter vector is usually used as regularized term in machine learning. The regularization order needs to be set in advance. l_0 , l_1 and l_2 norms are commonly used. In our experiments, we found that different regularization order can improve the classification accuracy of different data. So SVM classification algorithm with p -norm regularized multi-task learning is proposed in this paper. p -norm is not only effective during processing image data, but also easy to optimize. And it can reduce the computational complexity of the model. The formula is

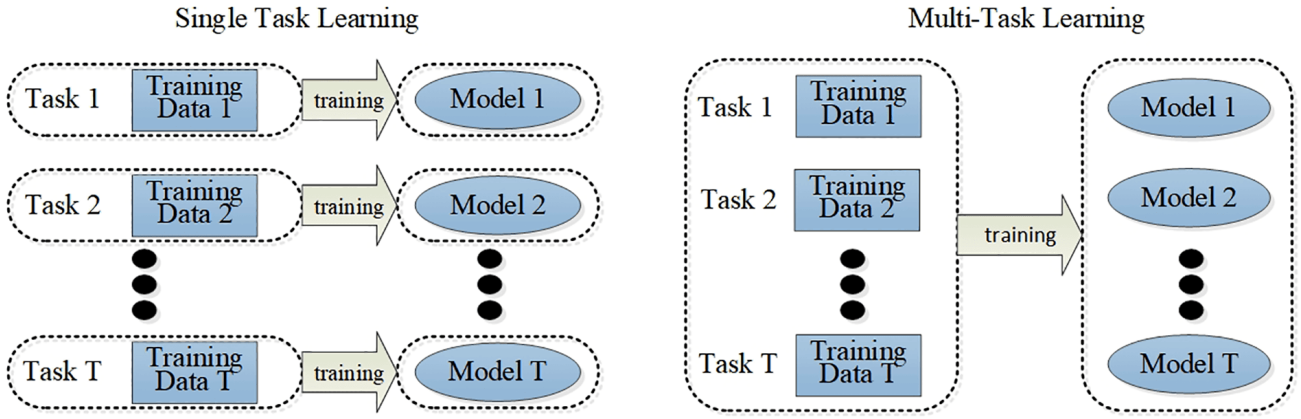


Fig. 2. The comparison of training process of single task learning model and multi-task learning model.

described by [29]

$$h(w) = \|w\|_p = \left(\sum_{i=1}^n |w_i|^p \right)^{\frac{1}{p}} \quad (14)$$

where $x = \{x_1, \dots, x_n\}$ is a vector. p -norm is a measure of the sparsity of the vector. The desirable range of the order p is $0 < p \leq 2$, and the choice of p depends on the related degree between the tasks. The more correlation and shared information between the tasks is, the larger the p value is. Let $k(w) = h^p(w)$, when $0 < p \leq 2$. Its derivative equation is shown by

$$\frac{\partial k(w)}{\partial w_i} = p |w_i|^{p-1} \times \text{sgn}(w_i) \quad (15)$$

where $\text{sgn}(w_i) = \frac{w_i}{|w_i|}$. So the Eqn. 15 can be written by

$$\frac{\partial k(w)}{\partial w_i} = p |w_i|^{p-2} \times w_i \quad (16)$$

Finally, the objective function of multi-task learning SVM with p -norm regularization is shown by

$$L = \min_w \sum_{t=1}^t \frac{1}{n} \sum_{i=1}^n \max(0, 1 - y_i w_i^T x_i) + \lambda \sum_{t=1}^t \|w_i\|_p \quad (17)$$

According to the different situation, the derivation of the Eqn. 17 is as follows.

If $1 - y_i w_{t,i}^T x_i < 0$, then

$$\frac{\partial L}{\partial w_{t,i}} = 0 \quad (18)$$

If $1 - y_i w_{t,i}^T x_i > 0$, then

$$\frac{\partial L}{\partial w_{t,i}} = -y_{t,i} x_{t,i} + p |w_{t,i}|^{p-2} \times w_{t,i} \quad (19)$$

The gradient descent method is used to update the weight coefficient matrix $W_{t,i} = \{\omega_{t,1}, \omega_{t,2}, \dots, \omega_{t,k}\}$ of t -th task, i.e., the following equation

$$W_{t,i} = W_{t-1,i} + r \cdot \nabla L(W_{t-1,i}) \quad (20)$$

4. Results

The experiments are elaborately designed and carried out using PC with Intel Core i5 (Intel Inc., CA, USA), CPU@2.40Ghz, speed 800 MHz, and 32G RAM. The compiling environments are Matlab2013a (American Math-Works company, MA, USA) and Python2.7 (Python Software Foundation, DE, USA).

4.1 Experimental Settings

In order to verify the effectiveness and robustness of the proposed method, comparative experiments are performed. (I) Single-site classification, i.e., that SVM classification algorithm was used to learn features of each single-site data separately for classification. (II) Pooling classification, i.e., that the three sites data were pooled together as a larger dataset regardless of the site differences. And SVM classifier was used to classify the remaining samples. (III) Multi-site classification, i.e., that SVM classification model with p -norm regularized multi-task learning was used to learn the site-specific and site-shared features simultaneously in the three data sites, and the two kinds of features were combined to classify the data corresponding to the data site.

In experiment (I), 72 cases were selected as training set from each site A, B, C, and the remaining samples serve as test sets. The feature vectors were input into the SVM classifier with the sigmoid kernel function. The main parameters include that penalty factor- c is 0.05, the fold of cross validation - v is selected 5 and 10, and the coefficient of kernel function - g is 0.05. The algorithm can be achieved through LIBSVM tools, and the performance of the classifier was evaluated by cross-validation. The classification accuracy was obtained at last. In experiment (II), 72 cases

Table 2. Comparison of experimental results.

Experimental method		A Site		B Site		C Site	
		ACC	AUC	ACC	AUC	ACC	AUC
(I)	GLCM + 5 folds	56.52%	0.58	58.97%	0.60	60.26%	0.56
	GLCM + 10 folds	57.02%	–	59.14%	–	60.22%	–
	GLCM + LBP + 5 folds	58.33%	0.59	60.42%	0.65	62.50%	0.58
	GLCM + LBP + 10 folds	57.69%	–	60.46%	–	61.84%	–
(II)	GLCM + 5 folds	60.00%	0.65	67.60%	0.70	69.00%	0.62
	GLCM + 10 folds	60.21%	–	67.64%	–	68.79%	–
	GLCM + LBP + 5 folds	60.83%	0.67	66.67%	0.71	69.17%	0.62
	GLCM + LBP + 10 folds	60.65%	–	68.17%	–	68.75%	–
(III)	GLCM + 5 folds	66.67%	0.73	75.00%	0.72	70.83%	0.67
	GLCM + 10 folds	68.70%	–	76.30%	–	72.83%	–
	GLCM + LBP + 5 folds	68.75%	0.76	77.08%	0.73	72.92%	0.70
	GLCM + LBP + 10 folds	68.55%	–	77.25%	–	72.67%	–

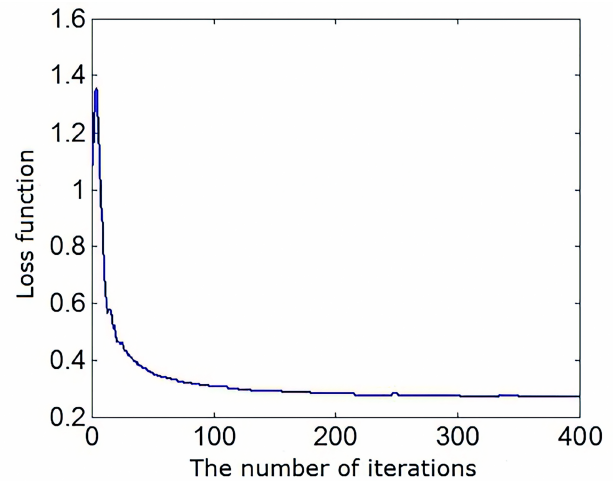
GLCM, Gray-level Co-occurrence Matrix; LBP, Local Binary Pattern; ACC, Accuracy; AUC, Area Under Curve; A, United States of America; B, Brazil; C, China.

from each site A, B, C were selected for fusion, so a total of 216 cases are used as training set. The remaining samples of A, B, and C data centers were classified after the model was trained. Other experimental conditions were set in accordance with the experiment (I). In experiment (III), 72 cases were selected as training set from each site A, B, C, and the remaining samples serve as test sets. The feature vectors were input into the proposed support vector machine classifier with p -norm multi-task. Hinge function was selected as loss function of the model, and gradient descent method was used to solve optimization of objective function (15). The optimal value of each parameter was based on the principle in which only one variable is changed. In the experiment, Gaussian kernel function [37] was selected as the SVM kernel function. In the SVM classifier the penalty coefficient c is 20 and rbf [37] kernel parameter g is 1.2. For verifying the role of multiple texture features, local binary pattern (LBP) [38] features are used to fuse in series GLCM ones because of the their advantages such as simplicity, validity, and spectrum form.

4.2 Experimental Results

The multi-task learning method was proposed to simultaneously learn the site-specific and site-shared features of the multi-site data. According to Eqns. 11,12,13,14,15,16,17 the gradient descent method is used to optimize the hinge loss function, and to verify the convergence of the proposed algorithm. As the number of iteration increases, the value of loss function shows a decreasing trend as shown in Fig. 3. It can be seen that the algorithm has good convergence property.

The best classification accuracy (ACC) and area under receiver operating characteristics curves (AUC) obtained by each experiment are shown in the Table 2.

**Fig. 3. The proof of algorithm convergence.**

5. Discussion

It can be seen from Table 2 that, under the premise of using 5-fold cross-validation, the classification accurate rates of the three data centers (A, B, C) are 56.52%, 58.97%, and 60.26% respectively in the experiment of single task learning algorithm. In joint classification experiment, the classification accurate rates of the three data centers (A, B, C) are 60.00%, 67.60%, and 69.00% respectively. However, in multi-task learning classification experiment, the classification accuracy rates of the three data centers (A, B, C) can reach 66.67%, 75.00%, and 70.83% respectively. The AUC of the three data centers (A, B, C) are 0.73, 0.72, and 0.67 respectively. The results of joint classification have increased to some extent compared to single-task learning classification. But multi-task learning is clearly superior to single task learning classification results. The classification performance of the multi-task learning algorithm is better than single-task learning system in this ex-

periment, because the multi-task learning process considers the association of multiple tasks. The model uses the shared information between tasks to enhance the inductive bias of the system, when training multiple tasks at the same time. Because the p -norm regularized term is added, the redundant features are effectively removed and the computational complexity of the model is reduced.

Furthermore, in order to verify the effectiveness of various features, we conduct experiments by merging LBP (Local Binary Pattern) and GLCM features in series. The experimental results show that effective fusion of multiple features such as LBP and GLCM can improve the classification accuracy to a certain extent.

Under the premise of using 10-fold cross-validation, the experimental results did not show a significant improvement on accuracy. Sometimes the mean accuracy even lower than the results of 5-fold cross validation. Actually, in cross-validation, the choice of k value can refer to the empirical formula which is $k \approx \ln(n)$ and $n/k > 3d$ [39], where n represents the amount of data and d represents the characteristic number. For the data settings of the experiments in this paper, the above theories explain why the accuracy drops sometimes.

6. Conclusions

In this paper, for discriminating schizophrenia patients from healthy controls, image processing and machine learning are introduced into the aided diagnosis and analysis of schizophrenia disease based on SMRI. Firstly, for achieving the effect of reducing dimension, the gray matter image is sliced, weighted and averaged preprocessing. Then GLCM texture features are extracted and normalized. Besides, the experimental samples are analyzed from the statistical point of view, excluding the influence of sex, age factors on the experimental results. At last, the main contribution of this work is that a support vector machine method with p -norm regularized multi-task learning is proposed and used to train and to establish the binary classification model. The experimental results show that multi-task learning approach has a superior performance compared with the single task learning method. It provides new ideas for studying multi-regional data and disease analysis. Furthermore, this experiment also provides guidance for computer-aided diagnosis and prognosis of mental illness. In the future work, more features will be considered to fuse, and the methods for mining deeper features of schizophrenia will be found, which can better improve the classification accuracy and assist doctors to diagnose schizophrenia disease.

Author Contributions

These should be presented as follows: YW and HX designed the research study. JS performed the research. YW analyzed the data. YW and JS wrote the manuscript. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

We thank the anonymous reviewers for their excellent criticism of the article.

Funding

This research was funded by Joint Project of Beijing Natural Science Foundation and Beijing Municipal Education Commission, grant number No. KZ202110011015.

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] Pu M, Xu B, Tan T. Brain science and brain inspired intelligence technology – an overview. *Bulletin of the Chinese Academy of Sciences*. 2016; 31: 725–736.
- [2] Wheeler AL, Voineskos AN. A review of structural neuroimaging in schizophrenia: from connectivity to connectomics. *Frontiers in Human Neuroscience*. 2014; 8: 653.
- [3] Tan Z, Luo L. Advances in magnetic resonance imaging of schizophrenia. *Diagnostic Imaging and Interventional Radiology*. 2015; 11: 507–511.
- [4] Bi W. A research of pattern recognition methodology for brain diseases & disorders classification with magnetic resonance imaging [M.S. thesis]. University of Electronic Science and Technology of China, Chengdu, China. 2016.
- [5] Seraj E, Yazdi M, Shahparian N. Instantaneous fMRI based cerebral parameters for automatic Alzheimer, mild cognitive impairment and healthy subject classification. *Journal of Integrative Neuroscience*. 2019; 18: 261–268.
- [6] Gupta CN, Calhoun VD, Rachakonda S, Chen J, Patel V, Liu J, *et al.* Patterns of Gray Matter Abnormalities in Schizophrenia Based on an International Mega-analysis. *Schizophrenia Bulletin*. 2015; 41: 1133–1142.
- [7] Radulescu E, Ganeshan B, Shergill SS, Medford N, Chatwin C, Young RC, *et al.* Grey-matter texture abnormalities and reduced hippocampal volume are distinguishing features of schizophrenia. *Psychiatry Research - Neuroimaging*. 2014; 223: 179–186.
- [8] Egloff L, Lenz C, Studerus E, Harrisberger F, Smieskova R, Schmidt A, *et al.* Sexually dimorphic subcortical brain volumes in emerging psychosis. *Schizophrenia Research*. 2018; 199: 257–265.
- [9] Fan FM, Xiang H, Wen Y, Zhao YL, Zhu XL, Wang YH, *et al.* Brain Abnormalities in Different Phases of Working Memory in Schizophrenia. *Journal of Nervous and Mental Disease*. 2019; 207: 760–767.
- [10] Tognin S, Rambaldelli G, Perlini C, Bellani M, Marinelli V, Zoccatelli G, *et al.* Enlarged hypothalamic volumes in schizophrenia. *Psychiatry Research*. 2012; 204: 75–81.
- [11] Eugenia R, Balaji G, Nicholas M, Sukhi S, Hugo C. Structural Brain Heterogeneity in Schizophrenia: A Gray Matter Texture Analysis on MR Images. *Schizophrenia Research*. 2012; 136: S109.
- [12] Pouyan AA, Shahamat H. A texture-based method for classification of schizophrenia using fMRI data. *Biocybernetics and Biomedical Engineering*. 2015; 35: 45–53.
- [13] Wang Y, Zhang N, Yan F, Gao Y. Magnetic resonance imaging study of gray matter in schizophrenia based on XGBoost. *Journal of Integrative Neuroscience*. 2018; 17: 331–336.

- [14] Oh J, Oh BL, Lee KU, Chae JH, Yun K. Identifying schizophrenia using structural MRI with a deep learning algorithm. *Frontiers in Psychiatry*. 2020; 11: 16.
- [15] Yamamoto M, Bagarinao E, Kushima I, Takahashi T, Sasabayashi D, Inada T, *et al*. Support vector machine-based classification of schizophrenia patients and healthy controls using structural magnetic resonance imaging from two independent sites. *PLoS ONE*. 2020; 15: e0239615.
- [16] Ma Q, Zhang T, Zanetti MV, Shen H, Satterthwaite TD, Wolf DH, *et al*. Classification of multi-site MR images in the presence of heterogeneity using multi-task learning. *NeuroImage: Clinical*. 2018; 19: 476–486.
- [17] Li C, Ning M, Fang P, Xu H. Sex differences in structural brain asymmetry of children with autism spectrum disorders. *Journal of Integrative Neuroscience*. 2021; 20: 331–340.
- [18] Wei P, Zou T, Lv Z, Fan Y. Human locomotion-control brain networks detected with independent component analysis. *Journal of Integrative Neuroscience*. 2021; 20: 695–701.
- [19] Al-Momani S, Dhou S. ‘Spinal functional Magnetic Resonance Imaging (fMRI) on Human Studies: a Literature Review’, 2019 Advances in Science and Engineering Technology International Conferences (ASET). Dubai, United Arab Emirates, Mar 26 - Apr 10, 2019. IEEE: New York, USA. 2019.
- [20] Hasan AM, Jalab HA, Meziane F, Kahtan H, Al-Ahmad AS. Combining Deep and Handcrafted Image Features for MRI Brain Scan Classification. *IEEE Access*. 2019; 7: 79959–79967.
- [21] Zhang Y, Yang Q. A Survey on Multi-Task Learning. *IEEE Transactions on Knowledge and Data Engineering*. 2021; 1–1.
- [22] Menikdiwela M, Nguyen C, Shaw M. ‘Deep Learning on Brain Cortical Thickness Data for Disease Classification’, 2018 Digital Image Computing: Techniques and Applications (DICTA). Canberra, Australia, Dec 10-13, 2018. IEEE: New York, USA. 2018.
- [23] Liu C, Peng Y. Research on natural image classification based on multi-task learning. *Application Research of Computers*. 2012; 29: 2773–2775.
- [24] Marquand AF, Brammer M, Williams SCR, Doyle OM. Bayesian multi-task learning for decoding multi-subject neuroimaging data. *NeuroImage*. 2014; 92: 298–311.
- [25] Nikhil R, Christopher C, Rob N, Timothy TR. Sparse overlapping sets Lasso for multitask learning and its application to fMRI analysis. *Computer Science*. 2013; 2202–2210.
- [26] Takanori W, Daniel K, Clayton S, Chandra S. ‘Multisite disease classification with functional connectomes via multitask structured sparse SVM’, *International Workshop on Sparsity Techniques in Medical Imaging*. Boston, USA, Sep 14-18, 2014. Elsevier: USA. 2014.
- [27] Wang X, Zhang T, Chaim TM, Zanetti MV, Davatzikos C. ‘Classification of MRI under the Presence of Disease Heterogeneity using Multi-Task Learning: Application to Bipolar Disorder’, *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich, Germany, Oct 05-09, 2015. Springer-Verlag: Berlin, Germany. 2015.
- [28] Evgeniou T, Pontil M. ‘Regularized multi-task learning’, the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Seattle, Washington, USA. Aug 22-25, 2004. ACM: New York, USA. 2004.
- [29] Liu J, Li S, Luo X. Classification algorithm of support vector machine via p -norm regularization. *Acta Automatica Sinica*. 2012; 38: 76–87.
- [30] Hu RJ. *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*. Encyclopedia of the Neurological Sciences. 2003; 25: 4–8.
- [31] Zhou Z. *Machine learning* (pp. 173–177). Tsinghua University Press: Beijing, China. 2016.
- [32] Pu J. *Research on multi-task learning algorithm* [Ph.D. dissertation]. Fudan University, Shanghai, China. 2013.
- [33] Yu B. *Multi-task learning and its application in spectral multivariate calibration* [Ph.D. dissertation]. University of Science and Technology of China, Hefei, China. 2015.
- [34] Traverso A, Wee L, Dekker A, Gillies R. Repeatability and Reproducibility of Radiomic Features: a Systematic Review. *International Journal of Radiation Oncology, Biology, Physics*. 2018; 102: 1143–1158.
- [35] Khagi B, Lee CG, Kwon G. ‘Alzheimer’s disease Classification from Brain MRI based on transfer learning from CNN’, *Biomedical Engineering International Conference (BMEiCON)*. Chiang Mai, Thailand, Nov 21-24, 2018. IEEE: New York, USA. 2018.
- [36] Li H. *Statistical learning Method* (pp. 10–20, 137–151). Tsinghua University Press: Beijing, China. 2012.
- [37] Steinwart I, Scovel C. Fast rates for support vector machines using Gaussian kernels. *The Annals of Statistics*. 2007; 35: 575–607.
- [38] Wang Y, Zhao Y, Chen Y. Texture classification using rotation invariant models on integrated local binary pattern and Zernike moments. *EURASIP Journal on Advances in Signal Processing*. 2014; 2014: 1–12.
- [39] Jung Y. Multiple predicting k -fold cross-validation for model selection. *Journal of Nonparametric Statistics*. 2018; 30: 197–215.