Received November 20, 2018, accepted December 19, 2018, date of publication December 24, 2018, date of current version January 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2889327

An Effective Dictionary Learning Algorithm Based on fMRI Data for Mobile Medical Disease Analysis

YUTING XU[1], (Member, IEEE), HAIMA YANG[1], (Member, IEEE), JUN LI[1], (Member, IEEE), JIN LIU[2], (Member, IEEE), AND NAIXUE XIONG[3,4], (Senior Member, IEEE)

[1]School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China
[2]School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China
[3]College of Intelligence and Computing, Tianjin University, Tianjin 300350, China
[4]Department of Mathematics and Computer Science, Northeastern State University, Tahlequah, OK 74464, USA

Corresponding authors: Haima Yang (snowyhm@sina.com.cn) and Naixue Xiong (dnxiong@126.com)

This work was supported in part by the Natural Science Foundation of Shanghai under Grant 17ZR1443500, in part by the Shanghai Aerospace Science and Technology Innovation Fund under Grant SAST2017-062, in part by the Joint Funds of the National Natural Science Foundation of China under Grant U1831133, and in part by the National Natural Science Foundation of China under Grant 61701296.

ABSTRACT With the continuous development and progress of the healthcare monitoring system, medical diagnosis for human health plays a, particularly, critical role, which can help doctors make correct choices and effective treatment plans. However, effective feature extraction is very important for the analysis of functional magnetic resonance imaging data; the traditional feature-based dictionary learning algorithm ignores the relationship between atoms and the input samples, and the small sample data is prone to over-fitting. In this paper, we propose a new weighting mechanism, which effectively considers the relationship between the atom and the input sample; meanwhile, the cross-validation method performed well on obtaining additional validation sets but proved to be over-fitting on small datasets in the traditional dictionary learning algorithm. Therefore, ℓ2-norm and $\not{f}$-norm regularization constraint is adopted to avoid over-fitting, achieve the limitations of the model space, and improve the generalization ability of the model. In order to extract features better, this paper uses the cosine similarity method to select good feature subsets, which effectively improves further the generalization ability and enhances the feature extraction accuracy. The results show that the improved dictionary classification algorithm has better performance in terms of accuracy, sensitivity, and specificity, and it also demonstrates that the proposed algorithm has an effective classification about mobile multimedia medical diseases, which can provide a better guidance for the diagnosis of later diseases, so as to promote the rapid development of medical feature extraction.

INDEX TERMS Healthcare monitoring, feature extraction, regularization, mobile multimedia, dictionary learning, classification.

I. INTRODUCTION

With the development of the healthcare monitoring system, data analysis is playing an important role in different industry, the combination of big data [1] analysis and healthcare are also closely related. Multimedia resources are growing at an amazing rate with the increase in users' demand. Mobile multimedia is a portable mobile device, which is a combination of computer and video technology. It is important for the mobile multimedia system to improve healthcare by collecting resource and managing information. Meanwhile, data analysis is powerful for healthcare monitoring, which can make good decisions for patients' diagnose. The term big data was put forward in 2008, whose main idea is to extract useful information from large amounts of data [2], so it has certain requirements for information processing capacity. As time goes by, the era of big data will meet the world that belongs to it, the United States invested much money for development in 2012 and incorporated it into the national strategy. As a result, the development of big data is out of control. As the four major features of big data, Volume, Velocity, Variety, and Value [3] have illustrated the potential capabilities of big data. Deep learning [4] and image recognition [5], as two important branches of large data, have been studied by enthusiasts. The underlying reason is that it can convert potential data values into useful information for human use. And it has been at the forefront of the era to provide a certain channel for future generations to study it and to help the rapid development of science and technology.

There are many kinds of big data, such as images big data [6], medical big data [7] and so on. Recently, the feature extraction and classification of data have become a hot spot for human research on medical big data, the reason is that the improvement of social science and technology have stimulated people's consideration of their own health, and better medical services are helpful to the rapid cure of diseases for meeting the disease needs of more people. Among them, the feature extraction [8] and data analysis of medical data are particularly important, which undoubtedly provides better guidance to doctors and experts to make treatment strategies [9] quickly. In general, if the patient's head is heavily impacted, the doctor will first gives him a function Magnetic Resonance Imaging (fMRI), which shows tiny changes in magnetic resonance signals from the oxygenation state of the veins in various regions of the brain [10], for example, depression and Attention Deficit Hyperactivity Disorder (ADHD), which can effectively study brain activity patterns, and which show that it does exist. At the same time, the diagnosis of mental disorders depends more on clinical symptoms, which may pose a threat to misdiagnosis [11]. Therefore, the classification and extraction of functional magnetic resonance data are beneficial to the diagnosis and treatment of mental diseases, which alleviates the working efficiency of medical experts to some extent and improves the reliability of medical diagnosis.

Dictionary classification algorithm, as one of the developments of medical big data, is widely applied in many aspects, such as pattern recognition [12], image processing [13] and so on. The advantage of the dictionary classification algorithm is to classify the features according to certain features and data, and analyze the relevant data effectively, and draw the corresponding conclusions, which can provide accurate information for medical diagnosis to prevent misdiagnosis. Among them, the accuracy, error rate and specificity of the dictionary classification algorithm are the factors that we must consider in the research algorithm. How to make the dictionary algorithm with high accuracy, low error rate, and specificity is the focus of our current study.

To address the problem that the traditional feature-based dictionary learning algorithm ignores the relationship between atoms and the input samples,which caused small sample data over-fitting, we design an improved mechanism, that is: an effective dictionary learning algorithm based on fMRI data. Specifically, this paper makes the following contributions:

1. It proposes a new weighting mechanism, which effectively considers the relationship between the atom and the input sample;

2. The traditional dictionary learning algorithm uses cross-validation method to obtain additional data validation set well, but there is an over-fitting phenomenon in the small training set. Regularization constraint on coding coefficients by $\ell\mathbf{2}$-norm and $\not{f}$-norm is adopted to avoid over-fitting, achieve the limitations of the model space, and improve the generalization ability of the model.

3. The extracted features are measured by cosine similarity, so that the appropriate feature subset is selected, which effectively improves further the generalization ability and enhances the feature extraction accuracy.

The remainder of this paper is organized as follows. In Section 2, we introduce the Related Works and the data and method are described in detail in Section 3. Then, the improvement of the algorithm is also introduced in Section 3. Next, in Section 4, experiments of our method and the compared methods on the three parameters demonstrate the effectiveness and improved performance of the improved method. After that, we interpret our results in this Section. Finally, we conclude this paper and discuss the future works in Section 5.

II. RELATED WORKS

Dictionary learning, as a sparse representation of a number of branches, is widely used in pattern recognition and image processing. As a tool of image classification, sparse representation is widely used in face recognition [14]. In [15], a Supervised intra-class Discrimination Dictionary Learning (SCDDL) algorithm for face recognition is proposed. The main idea is to directly restrict the similar coefficient of expression in the same category and to supervise the dictionary learning scheme, which includes linear classification errors to derive a more differentiated dictionary for facial recognition. In this way, the proportion of fishermen expressing the coefficient and the classifier recognition rate are improved, but the complexity of the improved algorithm of this scheme is obviously higher than other algorithms, and the application range is too narrow. Based on supervised intra class discriminating dictionary learning (SCDDL) algorithm, Multi-feature Kernel SCDDL (MKSCDDL) is proposed in [16], The main idea is to introduce the multi-kernel learning technology into the dictionary learning scheme, so as to improve its recognition rate and effectively classify the features. Based on the strategy of multi-kernel learning, an iterative algorithm for the coincidence learning of dictionary matrices and multi-core functions is proposed in this paper [17]. The sparse coding, dictionary learning, and multiple kernel learning are used to complete the minimization of sparsely encoded data reconstruction, which makes the classification settings more appropriate. In order to achieve high precision and good classification effect, a multi-stage classification algorithm using spatio-temporal feature extraction and supervised dictionary learning is proposed in the literature [18], which includes classifying and marking information, differentiating and sparse representation of each view. In the literature [19], a novel discrimination dictionary is developed by formulating a learning framework consisting of a shared sub-dictionary and a set of category-specific dictionaries, which makes the differences between the sparse code groups better expressed, so that the classification of data features can be better carried out and the effect of classification is achieved best. In [20], a novel model based on the combination of linear coefficient classifiers and dictionary pairs for joint learning is proposed, which is a model of Discriminative Analysis-Synthesis Dictionary Learning (DASDL), through the same optimization procedure, and considering both the performance of the classifier and the representation ability of the dictionary pair, it is superior to other dictionary algorithms. For sparse coding coefficients, it has been studied in many articles, in the literature [21], the $\ell\mathbf{1}$ -norm and the ℓ–norm are used to restrict the sparse learning of a particular dictionary, so as to achieve the minimum difference between the intra-class and inter-class, which makes the classification more accurate and the classification's effect better. In order to avoid the phenomenon of over-fitting for small data sets, $\ell\mathbf{2}$-norm constraint on coding coefficients is proposed in the paper [22], and a weighting mechanism based on the relationship between samples and dictionary atoms is proposed. In this way, the effect of classification is fast and good.

III. DATA AND METHODS

A. DEPRESSION DATA SET

This paper uses 15 patients with depression, including eight women, seven men, and 15 sex-, age- and education-matched healthy controls, including eight men, seven women, which are studied in this research. All of them are right-handed native Chinese came from all over China, and their age is randomly selected at different stages of distribution, the data obtained from the Institute of Automation, Chinese Academy of sciences. In addition, in order to verify the effectiveness of our proposed algorithm, we will use ADHD data [23] to verify our conclusions.

B. DATA SET PROCESSING

Since the data we obtained are too complex, some useless data may have some negative effect on the experimental results. In order to maintain data fidelity, we will use Resting-State Fmri Data Analysis Toolkit (REST, http://restfmri.net/forum/index.php), Statistical Parametric Mapping (SPM8, http://www.fil.ion.ucl.ac.uk/spm/software/spm8/) and Data Processing Assistant for Resting-State fMRI(DPARST, http://www.restfmri.net/forum/taxonomy/term/36) software to process data, so as to obtain an efficient data set, thus ensuring the effectiveness of the data.

C. IMPROVED DICTIONARY CLASSIFICATION ALGORITHM

1) OBJECTIVE FUNCTION

First, we represent the data information feature as a matrix, denoted by $P^t \in R^{d\times e}$, where d represents the number of brain regions and e represents the number of time series. By collecting the data for each class, we divide it into two parts, Some represent matrix of the health class, denoted as MHC, the matrix is expressed as $P \in R^{d\times en}$ $(r > n,\ r \ll en)$, The other part represents the matrix of the patient class, denoted as MPC, and its matrix is expressed as $P' \in R^{d\times ek}(r > n, r \ll ek)$, where n and k represent sample sizes corresponding to each class. For dictionary learning, D and D' are the dictionary matrices for the MHC and MPC, respectively, both matrics are composed of a matrix R of d rows and $\mathbf{r}$ columns, that is $R^{d\times r}$, and r represents the number of atoms. In order to simplify the complexity, we will discuss the model of the MHC, the MPC model is similar. Therefore, for the idea of improving the algorithm, the objective function is as follows:

$$\begin{aligned} F = \mathop{\arg\min}_{D,M,M'} &\Big(\frac{1}{C}\sum_{i=1}^{C}\frac{W_i}{1+W_i}\|p_i - Dm_i\|_2^2 \\ &- \frac{\rho}{C'}\sum_{j=1}^{C}{}'\frac{W_j}{1+W_i}\left\|p_j' - Dm_j'\right\|_2^2\Big) \\ \text{s.t. } &\|d_c\|_2^2 = 1, \quad \|M\|_2 < \varepsilon_1, \\ &\left\|M'\right\|_2 < \varepsilon_2, \quad c = 1, 2, \ldots r \end{aligned} \tag{1}$$

where C is the number of columns of the MHC, which is equal to ek, p_i denotes the i-th column of the MHC, m_i and $W_i/1+W_i$ denotes denote the weight and coding coefficient of MHC. Similarly, C' represents the number of columns of the patient matrix MPC, which is equal to en, m_j' and $W_j/1+W_i$ is the weight and coding coefficient of MPC. ρ represents the regularization parameter, c is the index value of the dictionary atom, and the first expression requires that the intra-class differences be small, meanwhile, the second expression emphasize the inter-class difference.

Considering that the distance between the sample and the dictionary atom is too small, the weight ratio increases, resulting in the representation error being small enough to satisfy the convergence condition of the objective function. In short, the purpose of increasing the weighting mechanism is to make the dictionary better test the similarity between the sample and the real information in order to improve its classification performance. To this end, we set the weight to be inversely proportional according to the distance between the sample and the mean of dictionary atoms, the weight is defined as follows:

$$W_i = \frac{e^{-\|p_i - \mathrm{avg}(d)\|_F^2}}{b * (\max(d) - \min(d))} \tag{2}$$

where b is the normalization constant, avg(d) is the mean vector of the dictionary atoms in D, and the weight W_j is also similar. The distance between the mean of dictionary atoms and the sample can measure the similarity between the sample and the dictionary atoms. Since the distance between dictionary atoms and sample is so small, the weight would be very large,which made the representation error small enough to satisfy the convergence condition. In other words, the aim of adding the weighting scheme can better represent the samples similar in learning the dictionary. The better representation can further enhance the classification performance. By this weight set, the optimization of the convergence condition can be achieved, so that the dictionary can better achieve its classification effect. Figures. 1 and Figures. 2 show the mean

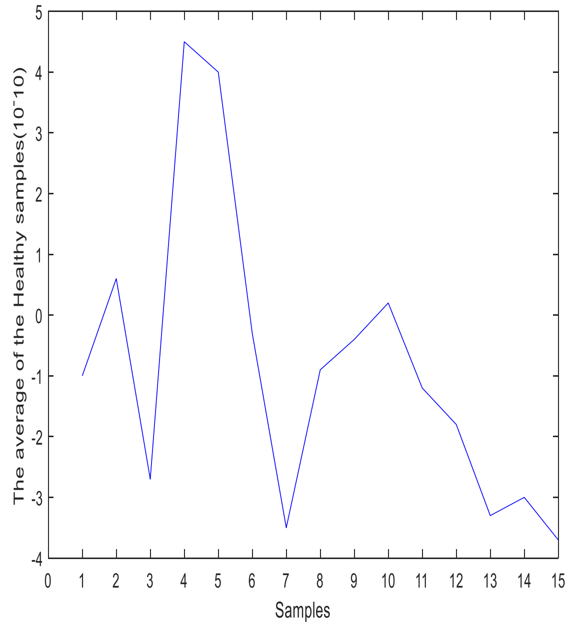

FIGURE 1. **The average of healthy samples.**

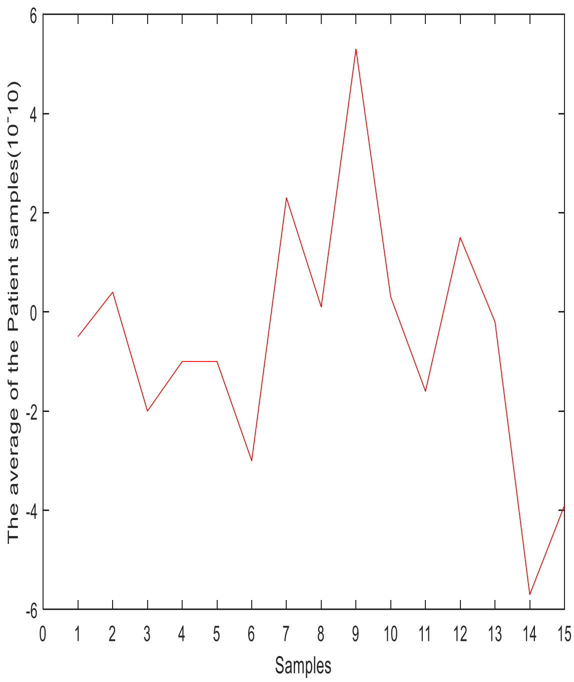

FIGURE 2. **The average of patient samples.**

values of the healthy sample and the patient sample data set, respectively.

2) ALGORITHM OPTIMIZATION PROCESS

We divide the optimization process into two parts: one is to keep the dictionary matrix D and D'' constant when updating the encoding coefficients m and m', and the other is to keep the encoding coefficients m and m' constant when updating the dictionary matrices D and D'. Assuming that the dictionary matrices D and D' are kept constant, the objective function will be solved by the solver, since the samples of the two classes share the same dictionary, so they can be solved in the same way:

$$m_i^* = arg\min_{m_i}\left(\frac{W_i}{1+W_j}||p_i - Dm_i||_2^2 + \lambda||m_i||_2^2\right) \tag{3}$$

$$m_j^{*\prime} = arg\min_{m_j}\left(\frac{W_j}{1+W_i}||p_j' - Dm_j'||_2^2 + \lambda||m_j'||_2^2\right) \tag{4}$$

where λ denotes normalization parameter, because the weight ratio is different, so the expression difference is limited to the weight ratio.

The least square method can be used to solve the objective function. In Formula (3) and Formula (4), we take the derivation of m_i and m_j' and obtain optimal solutions, respectively:

$$\frac{\partial m_i^*}{\partial m_i} = arg\min_{m_i}\left(\frac{2W_i}{1+W_j}D*(Dm_i - p_i) + 2\lambda m_i\right) \tag{5}$$

$$\frac{\partial m_j^{*\prime}}{\partial m_j'} = arg\min_{m_j}\left(\frac{2W_j}{1+W_i}D*\left(Dm_j' - p_j'\right) + 2\lambda m_j'\right) \tag{6}$$

Suppose that when the encoding coefficients m and m' are kept constant, the dictionary matrices D and D' can be updated through classes, and in Formula (1), the objective function can be reduced to:

$$\begin{aligned} F' = \arg\min_{D,M,M'}(&\frac{1}{C}\sum_{i=1}^{C}\frac{W_i}{1+W_i}\|p_i - Dm_i\|_2^2 \\ &- \frac{\rho}{C'}\sum_{j=1}^{C}{}'\frac{W_j}{1+W_i}||p_j' - Dm_j'||_2^2 \\ \text{s.t. } &\|d_c\|_2^2 = 1, \quad c = 1, 2, \ldots r \end{aligned} \tag{7}$$

The above objective function can also be written here as:

$$\begin{aligned} F' &= \arg\min_{D}\left(\frac{1}{C}\sum_{i=1}^{C}diag\left(\sqrt{\frac{W_i}{1+W_j}}\right)||p_i - Dm_i||_F^2\right. \\ &\quad \left.- \frac{\rho}{C'}\sum_{j=1}^{C}{}'diag\left(\sqrt{\frac{W_j}{1+W_i}}\right)||p_j' - Dm_j'||_F^2\right) \\ &= \arg\min_{D}\left(-2\text{tr}\left(\text{H}D^T\right) + \text{tr}(\text{DS}D^T)\right) \end{aligned} \tag{8}$$

where tr represents the sum of all the elements on the main diagonal of the matrix, we denote all matrices R with $||R||_F^2 = tr(RR^T)$, and derive further from Formula (8), the Formula denotes as follows:

$$H = \frac{diag\left(\sqrt{\frac{W_i}{1+W_j}}\right)}{C}PM^T - \frac{\rho diag(\sqrt{\frac{W_i}{1+W_j}})}{C'}P'M^{T'} \tag{9}$$

$$S = \frac{diag\left(\sqrt{\frac{W_i}{1+W_j}}\right)}{C}MM^T - \frac{\rho diag(\sqrt{\frac{W_i}{1+W_j}})}{C'}M'M'^T \tag{10}$$

In the objective function of Formula (8), if and only if the S matrix is semi-definite, the objective function can not be determined to be convex. In order to further determine that the matrix S is semi-definite, we use $S' = S - \lambda_{min}(S)I_k$ instead of S, where $\lambda_{min}(S)$ represents the minimum eigenvalue of S and I_k is the unit matrix. Since the eigenvalues of S' is non-negative, it is determined as semi-definite and the objective function is convex. Thus, the objective function is equivalent to:

$$F' = \arg\min_{D} \left(-2\mathrm{tr}\left(\mathrm{H}D^T\right) + \mathrm{tr}(\mathrm{D}S'D^T)\right)$$
$$\text{s.t. } ||d_c||_2^2 = 1, \quad c = 1, 2, \ldots r \tag{11}$$

In [24], update problem and optimization problems of dictionary atoms are very similar, so we rely on J.Mairal [24] to update d_c, the formula is as follows:

$$u_c \longleftarrow \frac{1}{S'_{cc}}\left(h_c - Ds'_c\right) + d_c \tag{12}$$
$$d_c \longleftarrow \frac{1}{||u_c||_2} u_c \tag{13}$$

where S'_{cc} represents the value of S' at position (c, c), h_c represents the c-th column of matrix H, and s'_c represents the c-th column value of S'.

The process of the improved dictionary algorithm is shown in Table 1:

TABLE 1. Algorithm process.

Algorithm : Improved Dictionary Learning Algorithm
1: Parameters Initialization.
2: Input: dictionary learning matrix P_i and P'_i, parameter ρ and λ, dictionary size r.
3: Choose dictionary D and initialize D by selecting r columns of P.
4: While algorithm isn't converged do
5: Update W_i and W_j by solving (2), so that weight ratio update.
6: Keep D constant, and update m and m' by solving (3) and (4)
7: Keep m and m' constant, update D by solving (13).
8: end While
9: Return D

3) COSINE SIMILARITY INTRODUCED

The similarity measure is calculating the degree of similarity between individuals, which is opposite to the distance measure, the smaller the value of the similarity measure is, the smaller the similarity between individuals is, the greater the difference is. Cosine similarity measure [25], that is, the two cosine values are calculated from the angles of the two vectors, and their similarity is compared by the following Formula:

$$\cos\theta = \frac{a \cdot b}{|a||b|} \tag{14}$$

where the value of $\cos\theta$ is in the range of -1 to 1, the closer the cosine is to 1, the closer the angle is to 0. In other words, the more similar the two vectors are. In order to avoid $\cos\theta$ less than 0, we take the absolute value of $\cos\theta$, so the modified cosine formula is as follows:

$$|cos\theta| = \frac{|a \cdot b|}{|a||b|} \tag{15}$$

Thus, the value of $\cos\theta$ is in the range of 0 to 1, and the similarity can be found by calculating the cosine similarity, and the error is reduced.

4) TEST SAMPLE CLASSIFICATION

In order to further verify the reliability of our proposed algorithm, we will classify the test samples. First, in the dictionary learning and training stage, the test samples P^t are denoted by m^T and $m^{T\prime}$, respectively. Using the representative error as the criterion, the test sample belongs to the class with smaller error, the representative of the error is expressed as:

$$E_1 = ||P^t - Dm^T||_F \tag{16}$$
$$E_2 = ||P^t - D'm^{T\prime}||_F \tag{17}$$

In this paper, the structure block diagram of the improved algorithm is divided into three stages, as shown in Figure. 3. The first stage is mainly data processing, the existing data is processed according to the relevant tools to obtain the experimental data needed in this paper. The second stage is

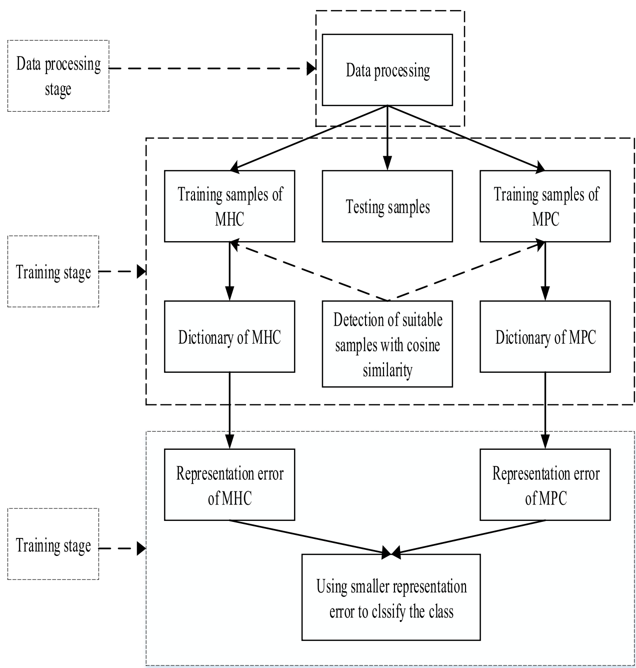

FIGURE 3. Block diagram of the improved algorithm.

the training data phase, which mainly includes training sample data phase, testing sample data phase, cosine similarity detection and so on, the third stage is the testing phase, mainly using the representative error to classify it.

IV. EXPERIMENT AND ANALYSIS

A. SIMULATION ENVIRONMENT

In this paper, we use simulation environment based on the windows7 system. By comparing the DFDL algorithm [21], this experiment will explain the effectiveness of the improved algorithm. The experimental parameters are set as follows: regularization parameter $\rho = 0.001, \lambda = 0.3$, the number of iterations is 20, the size of the dictionary is between 120 and 200.

B. EXPERIMENTAL RESULTS AND ANALYSIS

Before the experiment begins, we use the cosine similarity to calculate the similarity between the samples for selecting the part with the same cosine value, and then select the most suitable part according to the correlation between the samples. Considering the limited number of depression data sets, we will use cross-validation as a classifier. In 30 samples, we use half of the data set as a training sample and the other half as a test sample. In order to better evaluate the classification performance of the improved algorithm, we use the accuracy rate, sensitivity, specificity, and cosine similarity as the criterion of patient classification.

First we give a simple definition of accuracy, sensitivity and specificity:

$$Acc = \frac{TP + TN}{P + N} \tag{18}$$

$$Sen = \frac{TP}{TP + FN} = \frac{TP}{P} \tag{19}$$

$$Spe = \frac{TN}{TN + FP} = \frac{TN}{N} \tag{20}$$

where *Acc, Sen,* and *Spe* represent the accuracy, sensitivity, and specificity respectively, *TP* indicates the number of cases that are correctly divided into positive ones, *TN* indicates the number of cases that are correctly divided into negative ones, *FN* indicates the number of cases that are erroneously divided into negative ones, *FP* indicates the number of cases that are erroneously divided into positive ones, $P = TP + FN$ indicates the number of samples that are actually positive, and $N = TN + FP$ represents the number of samples that are actually negative.

First, we use the MHC data and the MPC data to construct two matrices Y1 and Y2, each of which is a $90 * 134$ matrix, the circular test was carried out on *i_s* from the 1th to the 30th group of sample data using leave one out cross-validation method, *i_s* samples are used as the validation samples, and other samples are used as training samples. One of the 30 sets of sample data is used as a test sample, and when the *i_s* is less than 15, the healthy class sample is selected as test samples, otherwise, the sample data of the MPC is selected as test samples.

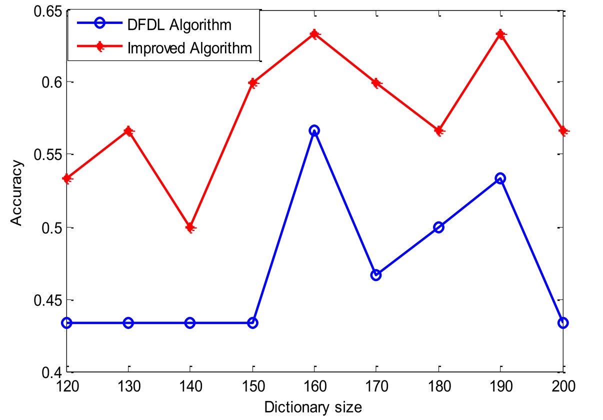

FIGURE 4. Accuracy of algorithms under different dictionary sizes.

Figure. 4 shows the comparison of two algorithms on accuracy. We can compare the accuracy by setting the step size is 10 and the range is 120 to 200. Obviously seen from the figure that the improved algorithm is more accurate than the DFDL algorithm. During this period, The average accuracy of the DFDL algorithm is 0.47, and the average accuracy of the improved algorithm is 0.58, and the average accuracy of the algorithm is improved by 11%, which indicates that the classification effect of the improved algorithm is more accurate than the DFDL, The classification performance of the algorithm is better than DFDL.

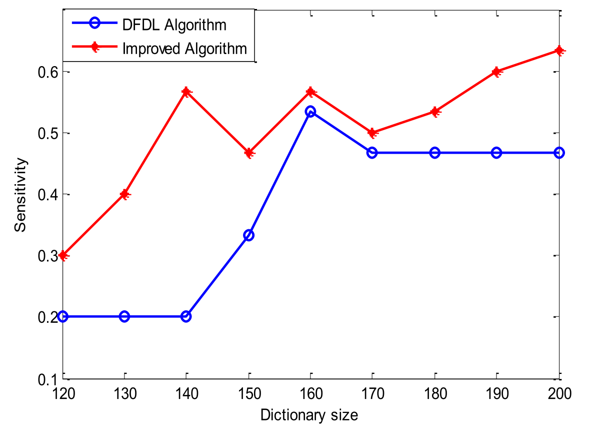

FIGURE 5. Sensitivity of algorithms under different dictionary sizes.

Figure. 5 shows the sensitivity of the two algorithms under different dictionary sizes. In the beginning, the sensitivity of the two algorithms is relatively low because of the low level of data learning and the inability to sensitively perceive data. With the deepening of learning, the degree of sensitivity is effectively improved. When the size of the dictionary is 160, the sensitivity of the DFDL maximized, however, the improved algorithm reached the same sensitivity with a dictionary size of 138. When the size of the dictionary is greater than 160, the sensitivity of the DFDL algorithm is almost constant, however, the sensitivity of the improved algorithm shows a steady trend of rising. It can be seen that the sensitivity of the improved algorithm is better than the DFDL algorithm, which means that the improved algorithm is better on data classification.

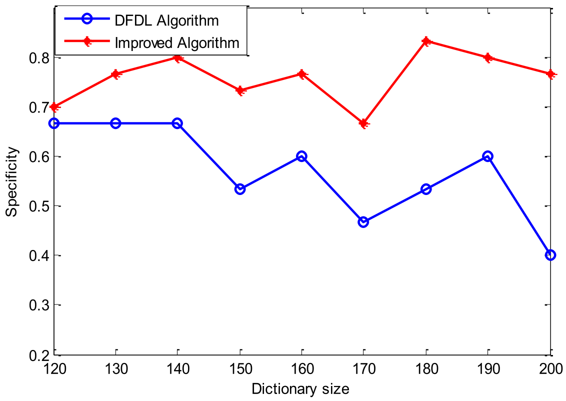

FIGURE 6. Specificity of the algorithm under different dictionary sizes.

Figure. 6 shows the comparison of the specificity of the algorithm under different dictionary sizes. First, it can be seen from the curve that the DFDL algorithm is decreasing and the fluctuation is more serious as the dictionary size changes, while the curve of the improved algorithm is smoother than DFDL. In practice, as the dictionary size continues to grow, the DFDL decreases from the initial maximum of 0.68 to the lowest value of 0.4, and the improved algorithm increases from the lowest 0.7 to the maximum 0.83. The average specificity of DFDL is 0.57, and the average specificity of the improved algorithm is 0.76, the average specificity of the algorithm is improved by 19%. It can be seen that the specificity of the improved algorithm is better than the DFDL algorithm, which further shows that the performance of the improved algorithm is superior to DFDL algorithm.

The performance of the classification algorithm is based on the representative error, and the representative error is closely related to each dictionary, and the small representative error is classified into the test sample. Figure. 7 and Figure. 8 show the comparison of presentation error of the

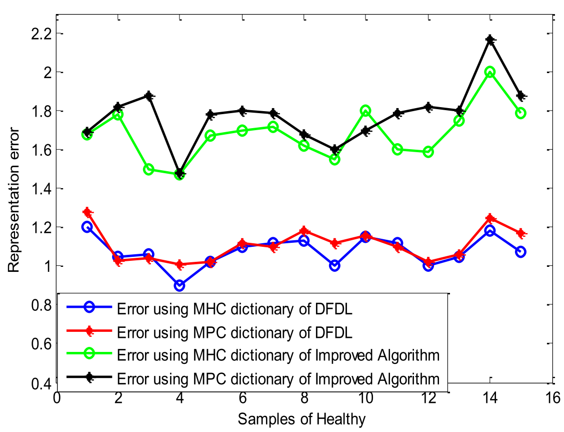

FIGURE 7. Representation error of the MHC based on the two algorithms.

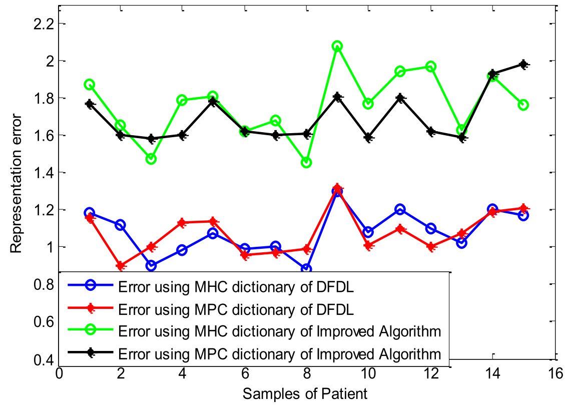

FIGURE 8. Representation error of the MPC based on the two algorithms.

two algorithms based on depression. As can be seen from the curve, the fluctuation of the two curves is different, the DFDL algorithm fluctuates smoothly, and the improved algorithm fluctuates greatly, which shows that the improved algorithm has enough discriminant to the sample data, so that it can be classified, if the representative errors are similar to each other, it is easy to classify a sampling error into another category. Therefore, the high representative error can effectively improve the efficiency of the algorithm classification, so that the classification effect achieve the best.

C. TIME COMPLEXITY ANALYSIS

Time complexity analysis describes the execution time of the algorithm, the quality of an algorithm depend on many aspects, and the algorithm execution time will improve the efficiency of the algorithm. Therefore, we will compare the time complexity of the improved algorithm and the DFDL algorithm, which will provide a foundation for future study. As we all know, the complexity of dictionary learning mainly depends on the sparse coding part. In [21], the time complexity for encoding coefficients is $O(cer(n+k)(2d+L^2))$, where c represents the class number (c=2), e represents the column number of the sample matrix (e=134), r represents the number of atoms (r=140), L represents sparse levels, n and k represent sample sizes (n=k=15), and d represents number of rows (d =90). In [26], the time complexity without weighting coefficients is O(2red), in this paper, the complexity of the algorithm is increased by the addition of the weighting factor to improve the classification performance of the algorithm. The time complexity is $O(cer\ (n+k)\left(2d+L^2\right)+O(2red)$, although the weight factor increases the time of the algorithm complexity, the execution time of the improved algorithm is still lower than the DFDL algorithm. Table 2 shows the execution time of the two algorithms, and it is clear that the improved algorithm is less time-consuming than the DFDL algorithm.

D. RESULTS FOR OTHER DATASETS

In order to further validate the effectiveness of the proposed improvement method, we will use the ADHD data set

TABLE 2. Time complexity analysis of DFDL and Improved alogrithm.

Algorithm	Complexity	Running Time
DFDL algorithm	$O(cer(n+k)(2d+L2))$	321.02s
Improved algorithm	$O\,(cer(n+k)(2d+L2)+2red)$	154.36s

TABLE 3. Comparison of the algorithms performance on the ADHD dataset.

Algorithm	Accuracy (%)	Sensitivity (%)	Specificity (%)
SRC[27]	72.73	66.67	75
SVM[28]	72.73	33.34	87.5
DFDL	63.64	33.34	75
Improved algorithm	82.53	65.37	83.65

to verify. First, the classifier is trained and tested separately in the training set and the test set, and we use cross-validation method to verify it. The experimental results show that the accuracy of the improved algorithm is 82.53%, the sensitivity is 65.37%, and the specificity is 83.65%. Table 3 shows the comparison of the improved algorithm with other algorithms.

V. CONCLUSIONS AND FUTURE WORK

An improved method is proposed in this paper, which is applied to classify medical disease in multimedia. With the help of mobile multimedia technology, we can exchange data information of patients to make the better decision for Disease Analysis. In this paper, we first propose a new weighted mechanism to consider the relationship between atoms and input samples effectively. Then, according to the traditional dictionary learning algorithm, the cross-validation method is used to have the phenomenon of over-fitting in the small training set. Regularization constraint on coding coefficients by $\ell\mathbf{2}$-norm and the $\not{f}$-norm is adopted to avoid over-fitting. Finally, the extracted features are measured by cosine similarity, so as to select the appropriate subsets of features. In this paper, we compare the MPC by setting up the MHC. The simulation results indicate that the improved algorithm is superior to the DFDL algorithm, the combination of healthcare and data analysis can play a great role in medical analysis, and can better improve efficiency at the end of mobile multimedia for healthcare.

Although this paper improves the performance of the algorithm, there are still some deficiencies. In future works, the research includes the following aspects:

(a) The study of the algorithm is limited to the simulation software, and its effect is unknown in practice. Therefore, to validate the superiority and possibility of the algorithm is imperative.

(b) The selection of the dictionary size is very important. If the dictionary is too small, there is not enough information for data representation; If the dictionary is too large, it can be very easy to generate over-fitting and reduces the accuracy, so it is necessary to select the dictionary size.

(c) The verification and validation of our algorithm are limited by the dataset. We hope to verify the feasibility of the algorithm on other datasets.

(d) The number of data sets is too limited, and the number of data sets must be expanded to improve the classification performance.

(e) The stability of the algorithm should be considered, and it is necessary to compare with more classification algorithms in the future.

ACKNOWLEDGMENT

The authors would like to thank all anonymous reviewers for their insightful comments and constructive suggestions to publish this paper in high quality.

AUTHOR CONTRIBUTIONS

Y.T. Xu and H.M. Yang conceived the idea, designed the experiments and analyzed the data; Y.T. Xu and H.M. Yang performed the experiments and conducted the analysis. J. Li and J. Liu collected and processed the data; N.X. Xiong interpreted the results and drew the conclusions; Y.T. Xu wrote the paper. All authors agree with the above contribution details.

CONFLICTS OF INTEREST

All authors declare no conflict of interest.

REFERENCES

[1] O. R. Team, *Big Data Now: Current Perspectives From O'Reilly Radar Sebastopol*. Sebastopol, CA, USA: O'Reilly Media, 2011.
[2] T. H. Mccormick *et al.*, "Big data, big results: Knowledge discovery in output from large-scale analytics," *Stat. Anal. Data Mining*, vol. 7, no. 5, pp. 404–412, 2015.
[3] S. Sagiroglu and D. Sinanc, "Big data: A review," in *Proc. IEEE Int. Conf. (CTS)*, May 2013, pp. 42–47.
[4] S. Hegelich, "Deep learning and punctuated equilibrium theory," *Cogn. Syst. Res.*, vol. 45, pp. 59–69, Oct. 2017.
[5] N. Zeng, Y. Xiong, and P. Gou, "Using kernel sparse representation to perform coarse-to-fine recognition of face images," *Optik*, vol. 140, pp. 528–535, Jul. 2017.
[6] C. Duan, B. Peng, Y. Min, X. Xianga, W. Zhana, and H. Zou, "Distributed in-memory vocabulary tree for real-time retrieval of big data images," *Ad Hoc Netw.*, vol. 35, pp. 137–148, Dec. 2015.
[7] X. Kong, M. Feng, and R. Wang, "The current status and challenges of establishment and utilization of medical big data in China," *Eur. Geriatric Med.*, vol. 6, no. 6, pp. 515–517, 2015.
[8] H. Zhang, B. Yang, and L. Huang, "Feature extraction of EEG signals using power spectral entropy," in *Proc. Int. Conf. BioMedical Eng. Inform.*, May 2008, pp. 435–439.
[9] B. Yang, "A new medical model by bundle of care in telemedicine in the big data area," *J. China Pract Diagn Ther.*, vol. 30, no. 8, pp. 729–731, 2016.
[10] M. E. Raichle, A. M. MacLeod, A. Z. Snyder, W. J. Powers, D. A. Gusnard, and G. L. Shulman, "A default mode of brain function," *Proc. Nat. Acad. Sci. USA*, vol 98, no. 2, pp. 676–682, Jan. 2001.
[11] A. J. Ferrari *et al.*, "Burden of depressive disorders by country, sex, age, and year: Findings from the global burden of disease study," *PLoS Med.*, vol. 10, no. 11, p. e1001547, Nov. 2013.
[12] N. Sengupta and N. Kasabov, "Spike-time encoding as a data compression technique for pattern recognition of temporal data," *Inf. Sci.*, vols. 406–407, pp. 133–145, Sep. 2017.
[13] L. Fan, Y. Bi, Y. He, M. Shipinga, S. Gaoa, and C. Lib, "Low-level structure feature extraction for image processing via stacked sparse denoising autoencoder," *Neurocomputing*, vol. 243, pp. 12–20, Jun. 2017.

[14] Y. Xu, M. Li, B. Zhang, Y. Jiane, and Y. Jane, "Sample diversity, representation effectiveness and robust dictionary learning for face recognition," *Inf. Sci.*, vol. 375, pp. 171–182, Jan. 2017.

[15] L. Xu, X. Wu, W. Chen, and L. Yao, "Supervised within-class-similar discriminative dictionary learning for face recognition," *J. Vis. Commun. Image Represent.*, vol. 38, pp. 561–572, Jul. 2016.

[16] X. Wu, Q. Li, L. Xu, K. Chen, and L. Yao, "Multi-feature kernel discriminant dictionary learning for face recognition," *Pattern Recognit.*, vol. 66, pp. 404–411, Jun. 2017.

[17] T. Zare and M. T. Sadeghi, "A Novel multiple kernel-based dictionary learning for distributive and collective sparse representation based classifiers," *Neurocomputing*, vol. 234, pp. 164–173, Apr. 2017.

[18] H. Khamisa, G. Zurakhova, V. Azara, A. Raza, Z. Friedmanb, and D. Adam, "Automatic apical view classification of echocardiograms using a discriminative learning dictionary," *Med. Image Anal.*, vol. 36, pp. 15–21, Feb. 2017.

[19] T. Yao, Y. Wang, Z. Xie, J. Gao, and D. D. Feng, "Learning universal multiview dictionary for human action recognition," *Pattern Recognit.*, vol. 64, pp. 236–244, Apr. 2017.

[20] M. Yang, Y. Chang, and X. Luo, "Discriminative analysis-synthesis dictionary learning for image classification," *Neurocomputing*, vol. 219, pp. 404–411, Jan. 2017.

[21] T. H. Vu, H. S. Mousavi, V. Monga, G. Rao, and U. K. A. Rao, "Histopathological image classification using discriminative feature-oriented dictionary learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 3, pp. 738–751, Mar. 2016.

[22] X. Wang, S. Ren, H. Yang, W. Zhang, and N. N. Xiong, "A weighted discriminative dictionary learning method for depression disorder classification using fMRI data," in *Proc. IEEE Int. Conf. Big Data Cloud Comput.*, Oct. 2016, pp. 618–623.

[23] J. Ford *et al.*, "Patient classification of fMRI activation maps," in *Proc. Int. Conf., Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, Montréal, QC, Canada, Nov. 2003, pp. 58–65.

[24] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, Jan. 2010.

[25] F. S. Al-Anzi and D. Abuzeina, "Toward an enhanced arabic text classification using cosine similarity and latent semantic indexing," *J. King Saud Univ.- Comput. Inf. Sci.*, vol. 29, no. 2, pp. 189–195, Apr. 2017.

[26] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 471–478.

[27] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[28] K. Ota, N. Oishi, K. Ito, H. Fukuyama, SEAD-J Study Group, and Alzheimer's Disease Neuroimaging Initiative, "Effects of imaging modalities, brain atlases and feature selection on prediction of Alzheimer's disease," *J. Neurosci. Methods*, vol. 256, pp. 168–183, Dec. 2015.

YUTING XU was born in 1991. She is currently pursuing the degree with the University of Shanghai for Science and Technology, Shanghai, China. Her research interests include image processing, server programming, algorithm analysis, and medical big data.

HAIMA YANG was born in 1979. He received the Ph.D. degree in signal and information processing from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Science, Xi'an, China, in 2015. He is currently an Associate Professor with the Instrument Science and Technology, School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, China. His research interests include digital signal analysis and processing, SPR sensor mechanism and simulation, pattern recognition system development, and symbolic slider variable structure control.

JUN LI was born in 1975. She received the Ph.D. degree in measuring and testing techniques and instruments from the Beijing Institute of Technology, China, in 2002. She is currently an Associate Professor with the School of Information and Communication Engineering, School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, China. Her research interests include signal and information processing, error analysis, and photoelectric measurement.

JIN LIU was born in 1978. She is currently an Associate Professor with the School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai, China, where she is also an Associate Dean. Her research interests include intelligent detection and control technology, distributed sensor networks, and test information acquisition and processing.

NAIXUE XIONG received the Ph.D. degree in sensor system engineering from Wuhan University and the Ph.D. degree in dependable sensor networks from the Japan Advanced Institute of Science and Technology. He was with Northeastern State University, Georgia State University, the Wentworth Institute of Technology, and Colorado Technical University (Full Professor for about five years) for about 10 years. He is current a Professor with the College of Intelligence and Computing, Tianjin University, China. His research interests include cloud computing, security and dependability, parallel and distributed computing, networks, and optimization theory.

He has published over 300 international journal papers and over 100 international conference papers. Some of his works were published in the IEEE JSAC, the IEEE or ACM Transactions, the ACM Sigcomm Workshop, IEEE INFOCOM, ICDCS, and IPDPS. He has received the Best Paper Award from the 10th IEEE International Conference on High Performance Computing and Communications (2008) and the Best student Paper Award in the 28th North American Fuzzy Information Processing Society Annual Conference (2009). He has been a General Chair, Program Chair, Publicity Chair, PC Member, and OC Member of over 100 international conferences and a Reviewer of about 100 international journals, including the IEEE JSAC, the IEEE SMC (Park: A/B/C), the IEEE Transactions on Communications, the IEEE Transactions on Mobile Computing, and the IEEE Transsctions on Parallel and Distributed Systems. He is serving as the Editor-in-Chief, an Associate Editor, or an Editorial Member for over 10 international journals (including as an Associate Editor for the IEEE Transactions on Systems, Man & Cybernetics: Systems and the *Information Science* and the Editor-in-Chief for the *Journal of Internet Technology* and the *Journal of Parallel & Cloud Computing*) and a Guest Editor for over 10 international journals, including the *Sensor* Journal, WINET, and MONET.

Dr. Xiong is a Senior Member of the IEEE Computer Society. He is the Chair of Trusted Cloud Computing Task Force, IEEE Computational Intelligence Society, and the Industry System Applications Technical Committee.