IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 3, 2020, accepted January 19, 2020, date of publication January 22, 2020, date of current version January 31, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2968634

Deep Principal Correlated Auto-Encoders With Application to Imaging and Genomics Data Integration

GANG LI[1], CHAO WANG[1], DE-PENG HAN[1], YI-PU ZHANG[1], PENG PENG[1], VINCE D. CALHOUN[2], AND YU-PING WANG[3]

[1]School of Electronic and Control Engineering, Chang'an University, Xi'an 710064, China
[2]Mind Research Network, Albuquerque, NM 87106, USA
[3]Department of Biomedical Engineering, Tulane University, New Orleans, LA 70118, USA

Corresponding authors: Gang Li (15229296166@chd.edu.cn) and Chao Wang (iswang_ch@qq.com)

This work was supported in part by the Natural Science Foundation of Shaanxi under Grant 2019JM-536, in part by the China Scholarship Council under Grant 201806565009, and in part by the Fundamental Research Funds for the Central Universities, Chang'an University (CHD), under Grant 300102329102.

ABSTRACT In terms of complex diseases like schizophrenia, more and more studies are beginning to treat genetic variants and brain imaging phenotypes as an important factor. In this paper, a competent optimization model is exploited to overcome the weakness of deep canonical correlation analysis (DCCA). The model consists of principal component analysis (PCA) on the multi-modality linear features learning and multi-layer belief networks on multi-modality nonlinear features learning. In order to complete a better result of correlation analysis and classification, the output nodes of multi-layer belief network are used for back propagation (BP) network training. Previous works on solving canonical correlation analysis (CCA) had proposed several models based on deep neural network or regularization, typically involving either some form of norm or auto-encoders with a reconstruction objective. Many existing advanced models had been developed to find the maximal correlation in multi-modality data. However, these multi-modality data tend to have the number of feature dimensions which more than that of samples. Differ from these advanced models, our proposed model is applied to analyze the real set of multi-modality data and test several previous models, then comparing them experimentally on fMRI imaging and SNPs genomics. In experiments, the results show that our model, deep principal correlated auto-encoders (DPCAE), learns features with effectively higher correlation and better performance of classification than those previous models. In terms of classification accuracy, the classification accuracy of the datasets exceeds 90%, but that of the CCA-based model are about 65%, and that of the DNN-based model are about 80%, the classification accuracy of the DPCAE is significantly improved obviously. In the experiment of clustering performance evaluation, the DPCAE further verified its superior classification performance with an average normalized mutual information index of 93.75% and an average classification error rate index of 3.8%. In terms of maximal correlation analysis, the model outperforms other advanced models with a maximal correlation of 0.926, showing excellent performance in high-dimensional data analysis.

INDEX TERMS Classification, data fusion, dimensionality reduction, belief network, optimization algorithm, principal component analysis.

I. INTRODUCTION

For the past few years, many researchers have worked on exploiting different variants of canonical correlation analysis (CCA) models and apply them to analyze genome-wide association problems [1]. For examples, these problems include identifying correlation between single nucleotide polymorphisms (SNPs) and genes [2], as well as correlation between DNA copy number change and functional magnetic resonance imaging (fMRI) [3], [4], and so on.

The associate editor coordinating the review of this manuscript and approving it for publication was Farid Boussaid.

In this paper, two-modality data were generated, the datasets use fMRI data of 116 brain regions, which has a dataset of 183 examples, and every example includes more than forty thousand features, and use SNPs dataset of specific genes [5], [6]. For the purposes of correlation analysis validation between fMRI and SNPs data, various modified CCA algorithms and different deep networks were simultaneously applied on the datasets. Previous works [2], [7]–[12] have developed many variants of CCA model which to address more particular challenges in some large-scale studies.

Some of the statistical learning models mentioned above need to be reviewed. For example, canonical correlation analysis (CCA) [7] was not only used as a standard technique for searching linear projections of two groups of synthetic variables that are maximally correlated, but also as a design method of decreasing the dimension of the feature vector, sparse canonical correlation analysis (sparse CCA) [2], [8]. Joint sparse canonical correlation analysis [9] are the extended forms of CCA with a particular regularized constraint, which makes penalty model easier to interpret. Deep CCA (DCCA) was introduced by Andrew [10] to find out complex nonlinear relationships, general CCA has a limitation in that it can only detect simple linear relationships and does not have the significant advantages as DCCA. Deep canonically correlated auto-encoders (DCCAE) was introduced by Wang et al., which as a non-linear model was used to optimize the combination of nonlinear correlations between reconstruction error of general auto-encoders and the learned representations of neural network [11]. Deep collaborative learning (DCL) is an effective method, which was proposed by Hu et.al. It first uses a deep network to represent original data and then seeks their correlations, while also linking the data representation with phenotypical information [12].

It is greatly known that RBM is a popular tool for representing dependency structure between stochastic variables, and the weights of the general RBM are optimized by minimizing the logarithmic loss function [13]. Differ from the multi-layer belief network based on stack RBMs, the DPCAE model updates weights by minimizing the objective function on each modality and maximizing the canonical correlation among the fourth hidden layer on each modality. As canonical correlation analysis (CCA) has been used for describing the maximal correlation between two stochastic variants, it could be utilized to characterize the correlation among the hidden layers. In the DPCAE model, activation functions in visible or hidden units is a smooth monotonic non-linearity function, such as non-saturating sigmoid nonlinearity function. Additionally, as the multi-layer belief network shows the better representational learning ability and obtains a unified representation, then the model based on multi-layer belief network is updated to learn multi-modality data representations.

In this paper, we propose a novel model, deep principal correlated auto-encoders (DPCAE), which uses two back propagation neural networks and two multi-layer belief networks to integrate the maximal correlation of the model simultaneously. It's necessary to note that two multi-layer belief networks are applied to extract features of hidden units of two modal data respectively. These extracted features are used for correlation analysis, and then the maximal correlation is selected to extract features with maximal correlation coefficients. Two back propagation neural networks are added to receive the output eigenvectors of the multi-layer belief networks as its input eigenvectors, each layer of the belief networks can only ensure that the weight within its own layer achieve the optimal eigenvector mapping for the layer, but not for the entire multi-layer belief networks. The back propagation neural networks propagate the parameter information from the top down to each layer of the multi-layer belief networks, and fine-tuning the whole multi-layer belief networks in each modality respectively. Comparing to the model of CCA-based and that of DNN-based, our model joint multi-layer belief network is based on several RBMs and linear dimension reduction method, which is not only to speed up machine learning, but also to discard dimensions that carry less information. Moreover, different from nonlinear deep canonically correlated auto-encoders, the DPCAE model adjusts the network's in-layer parameters by using multi-layer belief network and fine-tunes all network parameters from top to bottom layer by way of BP network, resulting in the better acquisition of nonlinear correlation in multi-modality data, while applying PCA and penalty term for better data fitting. It overcomes the disadvantage of falling into local optimum and long training time due to random initialization of weight parameters [13].

The rest of the paper is organized as follows. Section II introduces the several existing state-of-the-art methods and explain the reasons and advantages of our model. The collection and application of image gene data can be found in Section III. Detailed discussions and analysis of the results were in Section IV. The conclusions of the work and future direction of the work were in Section V.

II. TECHNICAL METHODS

A. REVIEW OF TRADITIONAL CANONICAL CORRELATION ANALYSIS

Linear canonical correlation analysis (CCA) was proposed by Hotelling et al., which was a model generally used for determining linear correlations between two sets of data [7]. It provided a thought to learn synthetic problems using multi-modality data by studying covariance. The CCA have a disadvantage is that it can only seeks the linear projections that are maximally correlated with two-modality data.

Specifically, the data sets of two modalities were denoted by $X \in R^{n\times p}$ and $Y \in R^{n\times q}$. General linear CCA tried to find two linear projection vectors $u \in R^{p\times 1}$ and $v \in R^{q\times 1}$ that in order to maximize the CCA criterion, as shown in (1). The corresponding schematic illustration is given in Fig. 1

$$\begin{aligned} &\max_{u,v} u'X'Yv \\ &\text{s.t. } u'X'Xu = v'Y'Yv = I \end{aligned} \tag{1}$$

Traditional CCA algorithm can only determine the maximal correlation in a series of canonical independent vectors,

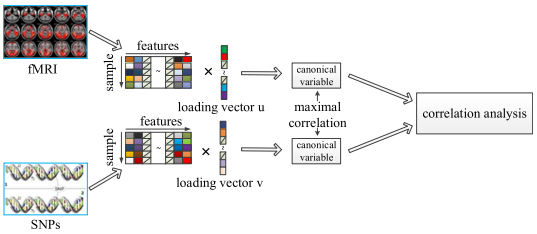


FIGURE 1. Illustration of the canonical correlation analysis model. In this model, two sets of data include the same number of samples and different numbers of high-dimensional features. Using loading vectors u, v to get canonical variable, and perform the correlation analysis to obtain the maximal correlation.

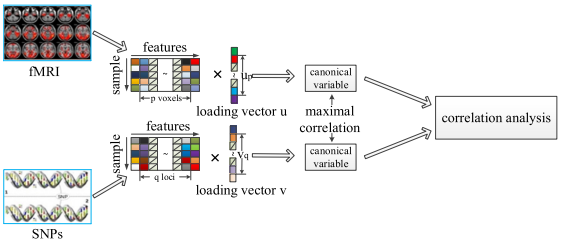


FIGURE 2. A figure showing the correlation analysis of combining fMRI data and SNPs data by sparse canonical correlation analysis algorithm. In this frame, two sets of data include samples size n and different numbers of features p, q by using the penalty to reduce the dimension. Then, loading vectors u, v are used to get the canonical variable, and perform correlation analysis to obtain the maximal correlation.

but a series of canonical vectors are pairwise independent, (1) is further improved to (2), which can guarantee the maximal total relationship. Then given two projection matrices $U \in R^{p\times k}$ and $V \in R^{q\times k}$, where $k = \min(rank(X), rank(Y))$, the objective function is shown in (2)

$$\begin{aligned} &max_{U,V}\, U'X'YV \\ &\text{s.t. } U'X'XU = V'Y'YV = I \end{aligned} \tag{2}$$

where data matrix X and Y have been standardized one as the standard deviation and mean zero, and consider $XU(Xu)$ and $YV(Yv)$ as the canonical variables in this paper.

B. EXTENSION: SPARSE CCA, AND SPARSE mCCA

Traditional CCA was widely used for linear correlation, but the number of high-dimensional features are considerably higher than that of real samples in the real datasets. For this reason, a lot of segmentation forms of penalized CCA algorithm [2], [3], [14], [15] were introduced to solve the confronting problem, when applying the datasets with high-dimensional features but small-size samples. They performed different sparse penalty functions on the canonical vectors u and v, in order to acquire some meaningful sparse vectors. The form of sparse CCA is proposed by Witten et al. (2009), which is given by (3), the corresponding schematic illustration is provided in Fig.2.

$$\begin{aligned} &\max_{u,v}\, u'X'Yv \\ &\text{s.t. } u'u \le 1, \quad v'v \le 1,\ P(u) \le c_1,\ P(v) \le c_2 \end{aligned} \tag{3}$$

where $P(*)$ is convex penalty function (e.g., L_1 norm), and c_1, c_2 are trade-off parameters which decide sparsity level of u and v respectively. The optimization challenge of CCA can be overcome by $(X'X)^{-1}$ and $(Y'Y)^{-1}$, but they may be singular for high dimensional data matrix so that the inverse matrix of $X'X$ and $Y'Y$ may not exist. Assume that the features are uncorrelated within both data matrix respectively, constraints $u'X'Xu \le 1$ and $v'Y'Yv \le 1$ in (1) become $u'u \le 1$ and $v'v \le 1$ in (3). Because of these constraints in (1), the penalized criterion in (3) tries to maximize the covariance $u'X'Yv$ instead of maximizing the Pearson correlation $u'X'Yv/XuYv$, this way is available given that it can reduce computational cost without computing $(X'X)^{-1}$ and $(Y'Y)^{-1}$.

Sparse multiple CCA (sparse mCCA) was introduced by Witten et al. (2009), which is an extended form of sCCA with two-modality data when three or more modality data are applied to analyze the maximal correlation [4]. Let the n data matrices be defined as $X_1, X_2, \ldots, X_n$ respectively, and assume that sample features of all data matrices are independent. After that, sparse mCCA finds canonical weight coefficients $w_1, w_2, \ldots, w_n$ that maximize the sum of all pairwise linear correlation, which is shown in the following objective function

$$\begin{aligned} &\max_{w_1,w_2,\ldots,w_n} \sum_{i<j} w_i' X_i' X_j w_j \\ &\text{s.t. } w_i' w_i \le 1, P_i(w_i) \le c_i, \quad \forall i \end{aligned} \tag{4}$$

where P_i is sparse penalty function, and w_i can be sparse when $P_i(*)$ is a proper penalty term and c_i is a suitable value. However, it is unjust to combine different coupled covariance, because these coupled covariance varies significantly in the dimensions of values. Some works of sparse multiple CCA [16], [17] could moderate the trouble of 'unjust combination of coupled covariance' by adding appropriate weighting coefficients to individual coupled covariance. For weight selection, cross-validation algorithm as limit functions [18] is used to evaluate the capacity of each appropriate weight.

C. INNOVATION AND BREAKTHROUGH: DEEP CCA

The employment of deep neural network (DNN) has increased explosively due to its breakthrough applications in cancer detection, image processing and so on. DNN can represent high dimensional nonlinear function, which has two or more layers. With the accumulation of various data resources and the improvement of computing power, the use of the deep network architecture is promoted, and the performance of this architecture is greatly optimized when the size and complexity of the data increases. The traditional CCA algorithm had encountered a problem that it can only detect the linear correlation between two-modalities, but cannot capture the nonlinear correlation between them. Sparse mCCA also faces a restriction in that it is unjust to combine different coupled covariance when the coupled covariance varies greatly in the scale of values.

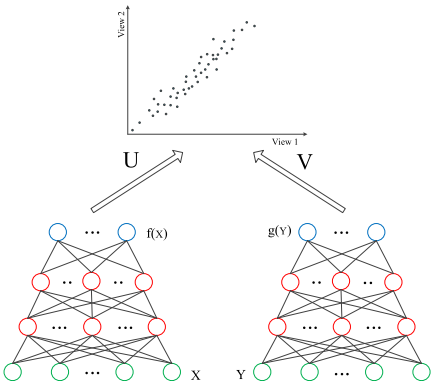


FIGURE 3. A work-flow of deep CCA for analyzing the maximal canonical correlation between two kinds of imaging data. The data of first (visual) layers are X, Y (the actual total numbers of nodes of two modalities are determined by the high-dimensional features of real data), the data of output layer are $f(X)$, $f(Y)$, deep networks are f, g and both networks include two hidden layers which mainly used to deal with the trouble of linear inseparable respectively.

Faced with these issues, deep CCA with a deep network capable of preprocessing data was introduced by Andrew et al. to capture nonlinear mappings [10].

They supposed that each hidden layer in the deep network of the two modalities has i units, and the final layer has d units. In this paper, the weighting coefficient matrix of first layer is defined as $W^1 \in R^{i\times n}$, then define the vector of biases as $b^1 \in R^i$, at last, the output of first hidden layer is $h^1 = s(W^1X + b^1) \in R^i$. The second hidden output is defined by applied output of first hidden layer as $h^2 = s(W^2h^1 + b^2) \in R^i$, until the final layer of deep network is defined by $f(X) = s(W^mh^m + b^m) \in R^d$, where m is the total number of layers in the deep network. The formulation of deep CCA is given in (5) and corresponding work-flow diagram is given in Fig. 3

$$\begin{aligned} &\max_{W_f,W_g,U,V} \frac{1}{N} Trace\left(U' f\,(X)'\, g\,(Y)\, V\right) \\ &\text{s.t. } U'\left(\frac{1}{N} f\,(X)' f\,(X)\right) U = I, \\ &\quad V'\left(\frac{1}{N} g\,(Y)'\, g\,(Y)\right) U = I \end{aligned} \tag{5}$$

The paired observations from two modalities, denoted $(x_1, y_1), \ldots, (x_N, y_N)$, where N is the sample size, $x_i \in R^{D_x}$ and $y_i \in R^{D_y}$ for $i = 1, \ldots, N$. Denoting the data matrices for each view by $X = [x_1, \ldots, x_N]$ and $Y = [y_1, \ldots, y_N]$. Where f and g are two deep neural networks which improve the ability to capture both linear and complex nonlinear correlations. U, V are projection matrices, which project outputs of the final layer in both deep neural networks. The dimensionality of $f(X)$ and $g(Y)$ are denoted s and r, $U'f(X)$ and $V'g(Y)$ are the final projection mapping. The DCCA objective function joins together all preprocessed sample data through the whitening constraints, so stochastic gradient descent algorithm is not suitable in a normal way. The algorithm was introduced by Wang et al., but it's only suitable for large-scale small batch optimization with stochastic gradient [19].

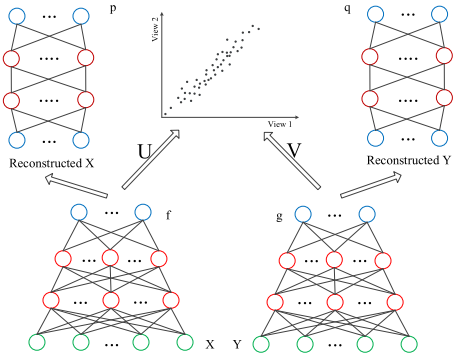


FIGURE 4. A work-flow of deep canonically correlated auto-encoders, in order to obtain the maximal correlation inspired by deep CCA and the minimum of reconstruction feature error. The model includes two feature extraction networks f, g and two reconstruction networks p, q.

D. DEEP CANONICALLY CORRELATED AUTO-ENCODERS

Wang et al. proposed an improved form of DCCA model termed deep canonically correlated auto-encoders (DCCAE; see Fig. 4) [11]. Their model was introduced in International Conference on Machine Learning that consists of a couple of auto-encoders, in addition, the model also optimized the combination of nonlinear correlations between reconstruction error of general auto-encoders and the learned representations of neural network. The objective function of DCCAE is

$$\begin{aligned} &\max_{W_f,W_g,W_p,W_q,U,V} \frac{1}{N} Trace\left(U^T f\,(X)^T\, g\,(Y)\, V\right) \\ &\quad -\frac{\gamma}{N}\sum_{i=1}^{N}\left(\|x_i - p\,(f\,(x_i))\|^2 + \|y_i - q\,(g\,(y_i))\|^2\right) \end{aligned} \tag{6}$$

s.t. The corresponding constraint is consistent with (5) where γ is trade-off parameter. As shown in (6), DCCAE finds two regularized deep neural network representation $f(X)$, $g(Y)$ and two reconstruction networks representations $p(f(x_i))q(g(y_i))$. Compared to CCA and DCCA, DCCAE further extracts features for better data fitting. This model considered the best consequence of multi-modality data correlation in a practical manner and stochastic gradient was applied to optimize the objective function.

E. DEEP PRINCIPAL CORRELATED AUTO-ENCODERS (DPCAE)

Inspired by nonlinear deep network models (DCCA, DCCAE) and linear sparse models (CCA, sCCA, smCCA), we propose an innovative model, deep principal correlated auto-encoders (DPCAE; see Fig.5). The model consists of two back propagation neural networks and two multi-layer belief networks, the multi-layer belief network is made up of Restricted Boltzmann Machines (RBM), while integrate the maximal correlation of the model simultaneously. It not only acquire multi-modality linear features by referencing the benefits of adding penalty items to dimensionality reduction of high-dimensional data, but also learn multi-modality nonlinear feature representation by developing deep network

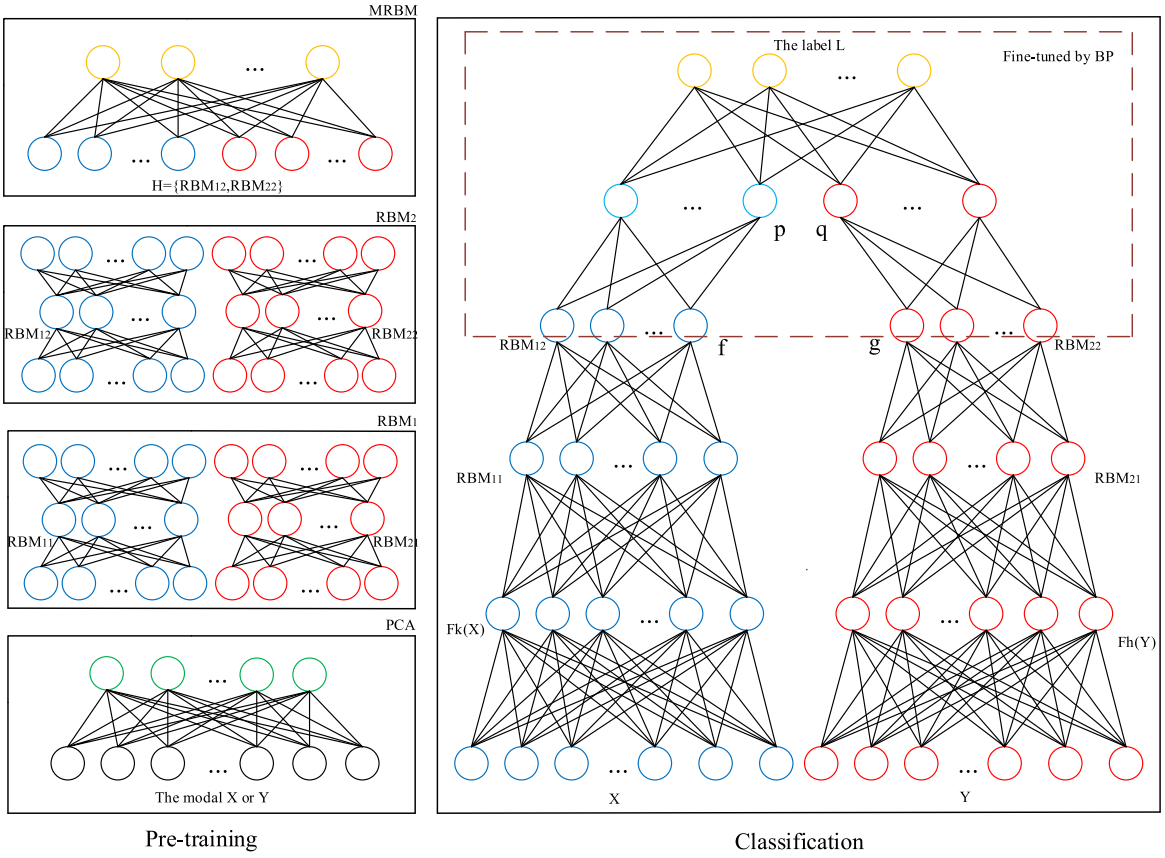


FIGURE 5. Schematic diagram of DPCAE, consisting of two back propagation neural networks and two multi-layer belief networks which is made up of RBM. The layer of each RBM are maximal correlation, then optimize the input layer of each BP network by using largest correlation coefficient, in order to get the best results of classification of multi-modal imaging genetic data.

structure. Here, it is worth noting that two multi-layer belief networks f and g are applied to extract consequential hidden unit feature representation while two back propagation neural networks p and q are used to extract top-level feature representations and fine-tune the entire multi-layer belief network in each modality respectively. In terms of multi-modality sample representations learning setting, paired observation is used from two modalities, then denote the data set of each modality by X and Y of dimension $n \times p_x$ and $n \times p_y$, where n denotes the total number of samples. p_x, p_y are the dimensionality of feature representations of X, Y respectively, and F_k, F_h are symbolic representations of PCA methods. Next, p_{x_1}, p_{y_1} are defined as the number of features after linear dimensionality reduction, p_{x_2}, p_{y_2} are the number of features after DBN networks training, $p_{x_{3i}}, p_{y_{3i}} (i = 1, 2, \ldots, m)$ are the number of features of hidden layers before reconstructing features, m denotes the number of stacked RBMs. The objective function of DPCAE is shown in (7)

$$\begin{aligned}
&\max_\theta \frac{1}{N} Trace\left(U' f\left(F_k\left(X\right)\right)' g\left(F_h\left(Y\right)\right) V\right) \\
&\quad -\frac{\gamma_1}{N} \sum_{i=1}^{m} \sum_{j=1}^{N} \begin{pmatrix} \left\| F_k\left(x_{ij}\right) - \theta_1 r_{1i}\left(F_k\left(x_{ij}\right)\right)\right\|^2 \\ + \left\| F_h\left(y_{ij}\right) - \theta_2 r_{2i}\left(F_h\left(y_{ij}\right)\right)\right\|^2 \end{pmatrix} \\
&\quad -\frac{\gamma_2}{N} \begin{pmatrix} \left\| L - W_p p\left(f\left(F_k\left(X\right)\right)\right)\right\|^2 \\ + \left\| L - W_q q\left(g\left(F_h\left(Y\right)\right)\right)\right\|^2 \end{pmatrix} - P\left(W^*\right) \\
&\text{s.t. } U'\left(\frac{1}{N} f\left(F_k\left(X\right)\right) f\left(F_k\left(X\right)\right)' + r_1 I\right) U = I, \\
&\quad V'\left(\frac{1}{N} g\left(F_h\left(Y\right)\right) g\left(F_h\left(Y\right)\right)' + r_2 I\right) V = I, \\
&\quad u_i' f\left(F_k\left(X\right)\right)' g\left(F_h\left(Y\right)\right) v_j = 0, \quad \text{for } i \neq j
\end{aligned} \tag{7}$$

where $P(W^*)$ is the combined function of convex penalty function $P(W_p)$ and $P(W_q)$. Note that convex penalty is mainly ridge penalty, $P(W^*)$ is a sum of squared L_2 penalty on W_p and W_q, namely $P(W^*) = \mu_p \left\| W_p \right\|_2^2 + \mu_q \left\| W_q \right\|_2^2$, ridge penalty can prevent overfitting of the model and help to ensure non-singularity. θ is a combination of $W_p, W_q, \theta_1, \theta_2, U$ and V,are weight-decay parameters which can be used to adjust the rationalization of the two sides in the function, and γ_1, γ_2 are trade-off parameters.

The model applies a 5-fold cross-validation pattern to select the weight-decay parameters. The total samples are grouped into 5 subgroups, and one subgroup is selected as testing sample and the remaining 4 subgroups are used as training sample in each stage. A calculated score is determined by the difference value between the correlation of

training sample and that of the test sample, which is applied to assess the property of selecting the weight-decay parameters. It is very vulnerable when utilizing cross-validation to pick out weight-decay parameters μ_p, μ_q, which is immediately applied as the threshold value when updating loading vectors. The vulnerability is due to the non-uniform distribution of loading vector values.

For (7), a reasonable explanation is given as follows: sparse CCA maximizes the correlational information between the two projected models to achieve specific and smooth distributions, while training an auto-encoder to minimize reconstruction error amounts to maximizing a lower bound on the correlational information between inputs and learned features. The DPCAE model gives a trade-off between the information captured in the (input, feature) mapping within each model on the one hand, and the information in the (feature, feature) relationship across modalities.

As $f(F_k(X))f(F_k(X))'$ and $g(F_h(Y))g(F_h(Y))'$ may be singular but they are essential when calculating the computation of $(f(F_k(X))f(F_k(X))')^{-1}$ and $(g(F_h(Y))g(F_h(Y))')^{-1}$, matrix regularization process is frequently accomplished on $f(F_k(X))f(F_k(X))'$ and $g(F_h(Y))g(F_h(Y))'$, as follow to ensure that they are positive define

$$\frac{1}{N}f\left(F_k\left(X\right)\right)f\left(F_k\left(X\right)\right)' \cong \frac{1}{N}f\left(F_k\left(X\right)\right)f\left(F_k\left(X\right)\right)' + \delta_1 \tag{8}$$

$$\frac{1}{N}g\left(F_h\left(Y\right)\right)g\left(F_h\left(Y\right)\right)' \cong \frac{1}{N}g\left(F_h\left(Y\right)\right)g\left(F_h\left(Y\right)\right)' + \delta_2 \tag{9}$$

where $\delta_1, \delta_2 > 0$ are regularization parameters in the multivariate estimation of variance [20].

The formulation of (7) can also be written as

$$\begin{aligned} &Z\left(R_1, R_2, P, Q\right) \\ &= \max_\theta \left\| \sum\nolimits_{11}^{-1/2} \sum\nolimits_{12} \sum\nolimits_{22}^{-1/2} \right\| \\ &\quad - \left\| F_k\left(X\right) - R_1\left(R_1'R_1\right)^{-1}R_1'F_k\left(X\right) \right\|_2^2 \\ &\quad - \left\| F_h\left(Y\right) - R_2\left(R_2'R_2\right)^{-1}R_2'F_h\left(Y\right) \right\|_2^2 \\ &\quad - \left\| L - P\left(P'P\right)^{-1}P'L \right\|_2^2 - \left\| L - Q\left(Q'Q\right)^{-1}Q'L \right\|_2^2 \end{aligned} \tag{10}$$

where $R_1 = r_1(F_k(X)) \in R^{n\times p_{x_3}}$, $R_2 = r_2(F_h(Y)) \in R^{n\times p_{y_3}}$, $P = p(f(F_k(X))) \in R^{n\times p_{x_2}}$, $Q = q(g(F_h(Y))) \in R^{n\times p_{y_2}}$ and $\sum_{22} = g(F_h(Y))g(F_h(Y))$ in (10) is equivalent to (7).

Linear CCA and linear sparse CCA models seem to have the maximal correlation with optimal projection vectors U, V, as given in (2) and (3). Nevertheless, the main reason for reviewing the models of Deep CCA and DCCAE in this paper, is to analysis limitations of traditional CCA which cannot obtain complex nonlinear relations. Compared to general model and structure of deep neural network, our model joint deep belief network based on several RBMs and linear dimension reduction method, the goal is not only to speed up machine learning, but also to discard dimensions that carry less information. Moreover, differ from nonlinear deep canonically correlated auto-encoders, DPCAE adjusts the network's in-layer parameters by using multi-layer belief network, and fine-tunes all network parameters from top to bottom layer by BP network. It acquire the best nonlinear correlation and data fitting in multi-modality data. In particular, the model of DPCAE further optimize significant feature information and parameter information, it can generate an excellent presentation of multi-modality data relationship in an available way.

In order to acquire more efficient computing efficiency, the experiment just uses the first derivative information and employs the conjugate gradient descent method to solve the optimization problem [21]. Multi-layer belief networks algorithm which is made up of RBM and back-propagation algorithm is used to pass the conjugate gradient backward to each layer of the network during each iteration process. In addition, the regularization technique is employed to avoid over-fitting. Here, we must compute the gradient of DPCAE's formulation (7) for using conjugate gradient descent method and BP.

The gradient of DPCAE's formulation (7) is given as

$$\begin{aligned} &\frac{\partial Z\left(R_1, R_2, P, Q\right)}{\partial R_1} \\ &= -R_1 \sum_{11}^{-1/2} U_1 D_1 U_1' \sum_{11}^{-1/2} \\ &\quad + R_2 \sum\nolimits_{22}^{-1/2} V_1 U_1' \sum\nolimits_{11}^{-1/2} + 2F_k\left(X\right) F_k\left(X\right)' R_1 \left(R_1'R_1\right)^{-1} \\ &\quad - 2R_1\left(R_1'R_1\right)^{-1} R_1'F_k\left(X\right) F_k\left(X\right)' R_1 \left(R_1'R_1\right)^{-1} \end{aligned} \tag{11}$$

$$\begin{aligned} &\frac{\partial Z\left(R_1, R_2, P, Q\right)}{\partial R_2} \\ &= R_2 \sum\nolimits_{22}^{-1/2} V_1 D_1 V_1' \sum\nolimits_{22}^{-1/2} \\ &\quad + R_1 \sum\nolimits_{11}^{-1/2} U_1 V_1' \sum\nolimits_{22}^{-1/2} + 2F_h\left(Y\right) F_h\left(Y\right)' R_2 \left(R_2'R_2\right)^{-1} \\ &\quad - 2R_2\left(R_2'R_2\right)^{-1} R_2'F_h\left(Y\right) F_h\left(Y\right)' R_2 \left(R_2'R_2\right)^{-1} \end{aligned} \tag{12}$$

$$\begin{aligned} &\frac{\partial Z\left(R_1, R_2, P, Q\right)}{\partial P} \\ &= 2LL'P\left(P'P\right)^{-1} - 2P\left(P'P\right)^{-1}PLL'P\left(P'P\right)^{-1} \end{aligned} \tag{13}$$

$$\begin{aligned} &\frac{\partial Z\left(R_1, R_2, P, Q\right)}{\partial Q} \\ &= 2LL'Q\left(Q'Q\right)^{-1} - 2Q\left(Q'Q\right)^{-1}Q'LL'Q\left(Q'Q\right)^{-1} \end{aligned} \tag{14}$$

Optimization Many researchers use different forms of activation functions to obtain the output of network layers in a DPCAE model, however, after many experimental analyses and comparisons, the result indicated that the form of non-saturating sigmoid nonlinearity is a very effective activation function. If ρ is the function $\rho\left(y\right) = y^3 + y$, our nonlinearity form is $s\left(x\right) = \rho^{-1}\left(x\right)$. The traditional sigmoid function will quickly reach their asymptotic value which differential coefficient goes down to essentially zero, and s is not bounded, so its value of differential coefficient fall off with

Algorithm 1 Deep Principal Correlated Auto-Encoders

Input $X \in R^{n\times p_x}$, $Y \in R^{n\times p_y}$, phenotype $L \in R^{n\times 1}$, initial networks $f_1, g_1, p_1, q_1, r_{11}, r_{21}$

Output Optimal networks $\hat{f}_1, \hat{g}_1$ with $\hat{f}\ (F_k(X)) \in R^{n\times p_{x_1}}$, $\hat{g}\ (F_h(Y)) \in R^{n\times p_{y_1}}$; $\hat{p}_1, \hat{q}_1$ with $\hat{p}\left(\hat{f}\ (F_k(X))\right) \in R^{n\times p_{x_2}}$, $\hat{q}\left(\hat{g}\ (F_h(Y))\right) \in R^{n\times p_{y_2}}$; $\hat{r}_{11}, \hat{r}_{21}$ with $\hat{r}_1\left(\hat{f}_i\ (F_k(X))\right) \in R^{n\times p_{x_{3i}}}$, $\hat{r}_2\left(\hat{g}_i\ (F_h(Y))\right) \in R^{n\times p_{y_3\ i}}$, $\quad i = 1, 2, \ldots, m$

$\hat{f}\ (F_k(X)) \leftarrow \hat{f}_1\ (F_k(X)) \quad , \quad \hat{g}\ (F_h(Y)) \leftarrow \hat{g}_1\ (F_h(Y))$, $\hat{p}\left(\hat{f}\ (F_k(X))\right) \leftarrow \hat{p}_1\left(\hat{f}\ (F_k(X))\right)$, $\hat{q}\left(\hat{g}\ (F_h(Y))\right) \leftarrow \hat{q}_1\left(\hat{g}\ (F_h(Y))\right)$, $\hat{r}_1\left(\hat{f}_i\ (F_k(X))\right) \leftarrow \hat{r}_{11}\left(\hat{f}_i\ (F_k(X))\right)$, $\hat{r}_2\left(\hat{g}_i\ (F_h(Y))\right) \leftarrow \hat{r}_{21}\left(\hat{g}_i\ (F_h(Y))\right)$, $i = 1, 2, \ldots, m$

Iteration $k \leftarrow 0$

while $k <$ maxepoch **and** no convergence **do**

$\hat{p}_1 \leftarrow$ Eq. (11) and Eq. (12)

$\hat{f} \overset{\text{stack}}{\leftarrow} \hat{r}_{11} \leftarrow RBM_{1\ i}\left(\hat{r}_{11}, \nabla Z\ (R_1, R_2, P, Q)\ |_{R_{11}=\hat{r}_1, R_2=\hat{r}_{21}}\right)$

$\hat{g} \overset{\text{stack}}{\leftarrow} \hat{r}_{21} \leftarrow RBM_{2\ i}\left(\hat{r}_{21}, \nabla Z\ (R_{11}, R_2, P, Q)|_{R_1=\hat{r}_{11}, R_2} = \hat{r}_{21}\right)$

$\nabla Z\ (R_1, R_2, P, Q)|_{P=\hat{p}_1, Q=\hat{q}_1} \leftarrow$ Eq. (13) and Eq. (14)

$\hat{q}_1 \leftarrow$ BackProgation $\left(\hat{p}_1, \nabla Z\ (R_1, R_2, P, Q)\ |_{P=\hat{p}_1, Q=\hat{q}_1}\right)$

$\hat{q}_1 \leftarrow$ BackProgation $\left(\hat{q}_1, \nabla Z\ (R_1, R_2, P, Q)\ |_{P=\hat{p}_1, Q=\hat{q}_1)}\right)$

$\hat{p} \leftarrow$ ForwardProgation $\left(\hat{p}_1\left(\hat{f}\ (F_k(X))\right)\right)$

$\hat{q} \leftarrow$ ForwardProgation $\left(\hat{q}_1\left(\hat{g}\ (F_h(Y))\right)\right)$

$k \leftarrow k + 1$

return $\hat{f}, \hat{g}, \hat{p}_1, \hat{q}_1, \hat{p}, \hat{q}$

x gradually. These natures make s better-adapted for min-batch optimization with a special method which makes sure the model does not reach steady state early during optimization. In order to better conform to the datasets and prevent overfitting, on the one hand, L_2 penalty term is applied on W_p and W_q to increase the difference between each component of the weight vector, on the other hand, dropout is applied to the hidden layers in each training iteration, which is randomly omitted from the network with a certain probability, and in this way the hidden units do not rely on other hidden units to change their states.

By taking a DPCAE having three hidden layers for two modalities as an example, Fig. 5 shows the pre-training process of the DPCAE. The learning procedure of the DPCAE can be grouped into two processes, the pre-training process and the classification process. Moreover, the non-saturating sigmoid function is selected as the activation function of each layer in the DPCAE.

In the pre-training stage, PCA reduces dimension of the visible layer, then the RBM1 and RBM2 obtain weights connecting the layer after reducing dimension and the second hidden layer. At last, the DPCAE model makes use of the MRBM to train the weights $\{Wp, Wq, \mu p, \mu q\}$ connecting the second hidden layer and the third hidden layer and obtain a unified representation. The input of MRBM is $H = \{f(F_k(X)), g(F_h(Y))\}$ and the RBM updates the weights $\{Wp, Wq, \mu p, \mu q\}$ by maximizing (7).

In the classification stage, the DPCAE model utilizes the data with labels and the gradient descent method to fine-tune the weights for classification. As the first two hidden weights contain the correlation between two modalities, the model uses the gradient descent method to fine-tune the weights $\{Wp, Wq, \mu p, \mu q\}$ connecting the second hidden layer, the third hidden layer, and the weights $\{\gamma_1, \gamma_2\}$ connecting hidden units and label units. By taking a DPCAE model having three hidden layers for two modalities as an example, the specific flow of the DPCAE model is shown in Fig. 6. It's worth mentioning that the DPCAE model is not only suitable for two-class classification data but also for multi-class classification data.

III. APPLICATION OF IMAGING GENOMIC DATA

A. TYPES OF GRAPHICS DATA ACQUISITION AND ANALYSIS

Clinical Imaging Consortium (MCIC) data sets include many biology factors about schizophrenia patients like clinical characterization, functional magnetic resonance imaging (fMRI), single nucleotide polymorphisms (SNPs) and so on in the on-line data repository [22]. The data was dealt with cross-sectional learning to verify mensurable imaging genetic biomarkers of schizophrenia. Enlist patients early in the course of their disease is MCIC data's extraordinary preponderance, therefore, the MCIC data sets involve a remarkable ratio of many people with schizophrenia researched in early stages of the clinical sign as same as a comparatively equal sampling of disease persistence via standing and diagnosed sickness. The method satisfied the crucial medical purposes of the multi-modality schedule, research of the kernel cognitive skill deficits affiliated to schizophrenia and their correlation to the clinical manifestation of the disorder [23]–[28]. In this research, two modalities (fMRI, SNPs) were acquired from 184 subjects consisting of 81 patients with schizophrenia and 103 healthy controls [29]. The real samples include schizophrenia patients and healthy controls, which were supplied with an informed agreement. Healthy participants were free of any medical, neurological or psychiatric illnesses and had no history of substance abuse. By the clinical interview of patients for DSM IV-TR Disorders or the Comprehensive Assessment of Symptoms and History, patients met criteria for DSM-IV-TR schizophrenia.

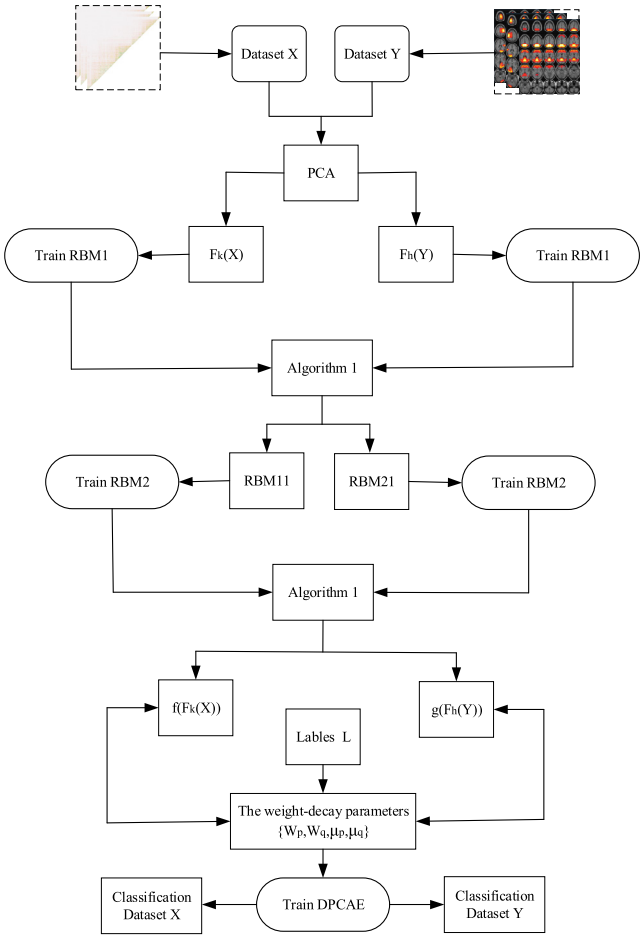


FIGURE 6. The flow chart of the DPCAE model.

B. INTRODUCTION OF fMRI DATA

The fMRI dataset which used in the experiments was collected by fMRI scanning, all people must early take part in a ''mock scanner'' task for adapting to the sensor motor setting. In the meantime, there are four fMRI tasks need all subjects to complete until they adapt themselves to the session and the sensor motor environment, for instance, a sensory-motor task (SM), a breath-hold task, an auditory oddball (AOD) and Sternberg item recognition paradigm (SIRP) [30]. For the four fMRI tasks, every part used a specific functional input equipment with equal terms for button presses to collect subjects' reaction, then all people were requested to respond to a suitable signal by pressing a button, at last, researchers connect the input equipment and specialized E-Prime software to provide visual stimuli for subjects and collect subjects' reaction. It is worth mention-ing that researchers collected four runs of SIRP task of fMRI tasks each contained 177 time frames for five minutes and fifty-four seconds per run, four runs of the AOD task of four tasks each contained 96 time frames for three minutes and twelve seconds per run, one run of breath-hold task and two runs of SM task each contained 120 time frames for four minutes per run.

The clinical data were pretreated using SPM5 (the Statistical Parametric Mapping package, SPM), then created brain imaging data of $63\times46\times53$ voxels [31]. Finally, 41,236 voxels were picked up, which were extracted from 116 ROIs for this experiment, it could well adapt the research of brain development and was used for biology identification, correlation analysis and classification [32].

C. INTRODUCTION OF SNPs DATA

All subjects were commanded to collect blood samples and extract candidate genes and genotyping. At the fewest 200 ng of candidate genes were used to type each schizophrenia patient and healthy control in the light of the manufacturer's protocol. After professional clinical techno-logy handling, the particular candidate genes were fluorescently labeled and detected making use of an appropriative scanner. Then some nonspecific hybridized fragments were separated by washing while these residual specific hybridized candidate genes were further processed [22]. To permit the pursual of subjects, capability control was used by statistical analysis, the genome-wide data management system was used to import raw genotypic data. At the end of the process, experimenters created 1,140,419 SNPs loci dataset. After capability control procedures, the dataset result in 777,365 SNPs loci left for our experiment finally. Lots of SNPs with linkage disequilibrium (LD) were spotted to reflect the genetic variant at different loci in learning high-dimensional features of samples. A subset of these simulated SNPs data showed the representative LD framework reflected the genetic variant, where special genetic markers that approximate one another on the candidate genes were in stronger LD, issuing in a representative block-like framework.

IV. EXPERIMENTS

Firstly, comparing the proposed algorithm and the other algorithms, and then demonstrate related work on deep principal correlated auto-encoders (Sections A–D), sections A describes that classification difference of two data types between different subjects, sections B gives a description of the performance measures about clustering and classification respectively, sections C describes receiver operating characteristic of multi-modality data, and the last section describes that the maximal correlation of various advanced methods on the different number of dimensions.

The DPCAE's experiment on genetic variants and brain imaging phenotypes focused on two works: classification (to classify different subjects using MCIC dataset), and correlation analysis (to explore the correlation between SNPs loci and brain imaging data). The classification verifies the classification capability of the DPCAE algorithm, while correlation analysis verifies the performance of the DPCAE algorithm in terms of finding the maximal correlation. All hyper-parameters, including energy function, activation function type, regularization parameters, learning rate, trade-off parameters, maximum epochs, types of deep networks, the number of layers, and the total number of nodes in different network layers, were selected using conjugate gradient descent [12]. The model is proposed by us and other advanced models can be divided into two categories

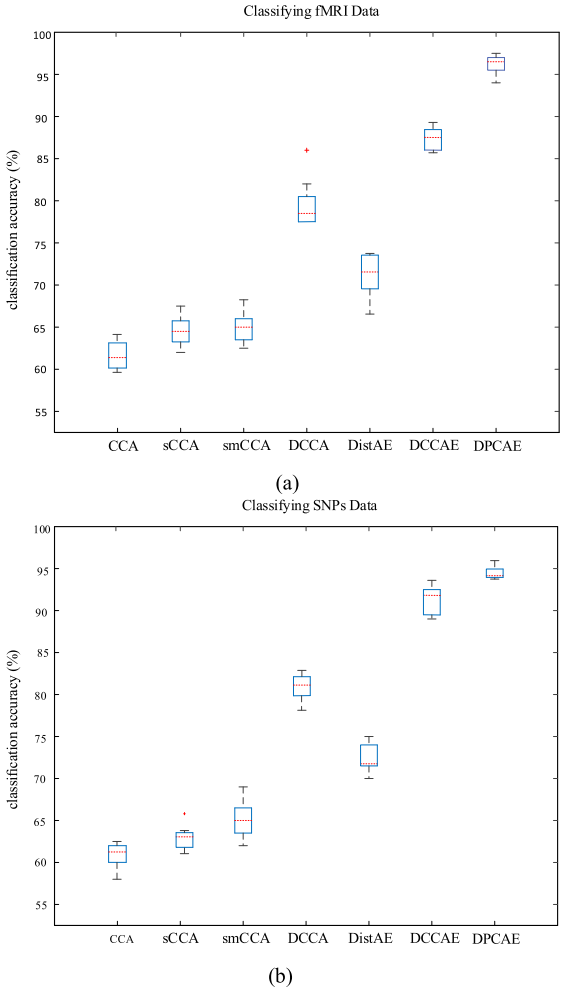


FIGURE 7. Maps showing the comparison of the power of each model on classifying two modalities (sub-figs a-b) (schizophrenia patients vs. healthy controls). (a) describes the power of classification on different subjects using fMRI data, (b) gives a description of the power of classification on different subjects using SNPs data, X-axis represents several different algorithms applied on MCIC data, the full names of the algorithms are traditional CCA, sparse CCA, sparse mCCA (smCCA), DCCA, distAE, DCCAE, and DPCAE. Y-axis represents classification accuracy (%).

generally, one kind is DNN-based models, which can detect non-linear relationship, including DCCA, DCCAE, DistAE and DPCAE, another CCA-based models, corresponding to DNN-based models with only a linear network with no hidden layers, including CCA, sparse CCA, sparse mCCA. We developed these programs using Tensorflow as the deep learning framework. These programs were run under the Tulane University High Performance Computing system, code named "Cypress" with an operating system of Linux. It is a 124-node cluster, with each node providing dual 10-core 2.8 GHz Intel Xeon E5-2680 v2 CPUs, 64 GB of RAM, and dual Xeon Phi 7120P coprocessors.

A. CLASSIFICATION DIFFERENCE OF MULTI-MODALITY DATA

Comparing the power of the DPCAE algorithm to that of other representative algorithms, including CCA, sparse CCA, sparse mCCA (smCCA), deep CCA (DCCA), minimum-distance auto-encoders (distAE), deep canonically correlated auto-encoders (DCCAE), and deep principal correlated auto-encoders (DPCAE) in terms of both fMRI classification and the classification of SNPs data. DPCAE's network uses MCIC dataset to train and then the trained network was used to classify. Some meaningful preprocessing methods, including mainly data augmentation, data standardization, and so on, were implemented on the multi-modality data [12]. The DPCAE model applied two modal data to study brain development of schizophrenia. It needs a large sample size in order to train deep networks. However, collecting fMRI data is very expensive, and therefore the sample sizes of existing fMRI cohorts are limited. A reasonable method to generate more valid data is data augmentation. Data augmentation is a widely used method in deep learning fields, notably when dealing with images. For application of deep learning in image classification, data augmentation methods, e.g., image rotation, image reflection, scaling, are continually applied to generate added 'real' images.

Like CCA-based models need at lowest two modality data as input, here data-pair combination of fMRI and SNPs data are used. For this data combination, we experimented with the property of each model, and the results of the experiment were given in Fig. 7. From this figure, the proposed method, deep principal correlated auto-encoders, acquired better classification accuracies than both CCA-based models and other DNN-based models for classifying different subjects using MCIC dataset separately. The classification accuracy is evaluated by comparing the trained datasets with the real labels. This classification method can observe the accuracy of datasets through the box diagram, and the calculation formula for the classification accuracy is shown in (15).

$$Pre = TP/(TP + FP) \tag{15}$$

where TP is the number of datasets correctly classified as positive samples after training of each modality, while FP is the number of modal data wrongly classified as positive samples after training of each modality.

Learning based deep network, DCCA and DCCAE performed better than CCA-based models and minimum-distance auto-encoders but worse than DPCAE in the aspect of classification, which may be due to the incorporation of DNN outputs and regularization parameters for assessment of example variance. While minimum-distance auto-encoders performed worse than DCCA and DCCAE, which may result from weakness of the average discrepancy term between projected example pairs.

The excellent classification accuracy (over 90% for SNPs data, over 95% for fMRI data) indicates the subjects (e.g., schizophrenia patients and healthy controls). The most important factor must be pointed out here is that it could be seen from Fig. 7 that the classification accuracy of fMRI data is higher than the classification accuracy of SNPs data which might be due to the fact that fMRI data is an imaging data which were pre-processed with SPM5, while SNPs

data may be deletions of gene fragments due to human or environmental factors with linkage disequilibrium which is not as accurate and consistent as fMRI data. There is great interest in studying how different original features of imaging genomics impact the classification accuracy of SNPs data and fMRI data, however, the data representation of DPCAE is created by different deep networks in which nonlinear activation functions are used to each hidden layer. As a consequence, it is not easy to analyze how each original variable is displayed in the deep network representation and hence it is challenging to analyze the identification performance of each original biomarker variable.

B. PERFORMANCE MEASURES OF PROJECTED MULTI-MODALITY DATA

The class distance is estimated in the extracted feature spaces by clustering around the projected SNPs data and fMRI data inputs into two clusters, and assessing the performance of the clusters in line with sample labels respectively. Spectral clustering [33] is applied in this task, in order to explain potentially non-convex cluster shapes. Firstly, it needs to be further pointed out that a k-nearest-neighbor diagram could be constructed on the projected the samples of two-modality data with a binary weighting manner, then these projected samples are embedded in R^2 employing the eigenvectors of the normalized diagram Euclidean, eventually run K-means procedure in the embedding to find an available segmentation of the samples [34]. The final step of the process needs to be explained that K-means procedure is run ten times with optimal initialization of maximal correlation and the best K-means value is used to run spectral clustering. The normalized mutual information (NMI) is used as a clustering evaluation index [35], what's more, tuning value '2' is used as the size of neighbor diagram k. The calculation formula of normalized mutual information is given

$$\begin{aligned} &Min\,(X_{tr}; Y_{tr}) = 2I\,(X_{tr}; Y_{tr})\,/H\,(X_{tr}) + H\,(Y_{tr}) \\ &\quad I\,(X_{tr}; Y_{tr}) = \sum_{x_{tr}} \sum_{y_{tr}} p\,(x_{tr}, y_{tr}) * \\ &\text{s.t.}\quad \log\{p\,(x_{tr}, y_{tr})\,/[p\,(x_{tr})\,p\,(y_{tr})\} \\ &\quad H\,(X_{tr}) = -\sum_{i} p\,(x_{tri}) \log p\,(x_{tri}) \\ &\quad H\,(Y_{tr}) = -\sum_{i} p\,(y_{tri}) \log p\,(y_{tri}) \end{aligned} \tag{16}$$

where $p(x_{tr}, y_{tr})$ is the joint distribution of two groups of data X_{tr}, Y_{tr} after training, $p(x_{tr})$ and $p(y_{tr})$ are probability distribution of x_{tr} and y_{tr} respectively, $H(X_{tr})$ and $H(Y_{tr})$ are information entropy of X_{tr} and Y_{tr} respectively, while $I(X_{tr}, Y_{tr})$ is relative entropy of $p(x_{tr}, y_{tr})$ and $p(x_{tr})p(y_{tr})$.

According to the training data of different models, the classification error rate of the datasets was calculated:

$$Err = (FP + FN)/(P + N) \tag{17}$$

where FP is the number of trained datasets that are actually negative samples but are divided into positive samples by the classifier; FN is the number of trained datasets that are actually positive samples but are divided into negative samples by the classifier; and $P + N$ is the total number of samples of trained datasets.

TABLE 1. Performance evaluation of each typical research methods on the dataset of SNPs and FMRI.

Method	S-NMI(%)	S-Error(%)	F-NMI(%)	F-Error(%)
CCA	47.8	23.3	50.3	21.1
sCCA	54.2	20.1	58.7	18.6
smCCA	57.7	18.7	63.3	15.7
DCCA	84.5	8.4	85.1	6.4
distAE	63.2	13.7	71.5	11.8
DCCAE	88.4	6.3	91.4	4.8
DCL	86.6.	4.9	89.5	3.7
DPCAE	92.1	4.5	93.8	3.2

The better NMI of clustering (92.1% for SNPs data, over 93.8% for fMRI data) indicates that different subjects (e.g., schizophrenia patients and healthy controls). While projections by distAE model perform somewhat better segmentation for some clusters, worse than other DNN-based models. For another, CCA-based models can approximate the same variable, but the part of separate classes is failed, probably because the original data are too complex to be better applied by linear mappings. Overall, DPCAE shows the best result of NMI of clustering and cleanest embedding between SNPs data and fMRI data using different methods. The model expects that an easy linear classifier can acquire better accuracy on DPCAE projections. Then, the linear SVM is trained on the projected dataset (now using two-modality sample labels), and applying the projected tuning set to select the SVM hyper-parameters [36]. Error rates (including error rates of SNPs data and fMRI data, and abbreviated form of them are S-Error and F-Error) on the optimal classification of each model are given in Table 1. When the model we used gets the maximal value of NMI, the value of classification error rate of the model is the largest. In generally, the smaller the classification error rate, the higher the classification accuracy, so these error rates consistent with the clustering results. The result of DCL is very close to DPCAE in classification accuracy and normalization of normalized mutual information, which may be due to the incorporation of phenotype information. But comprehensive indicators, deep principal correlated auto-encoders makes classification accuracy much higher than other methods, the result is sufficient to prove the correctness of the classification results of the two modal data. Take the place of employing a hard nonlinear classifier on the feature representations of high-dimension, a very ordinary linear classifier that can be educated effectively on low-dimensional projections already acquires the best result of accuracy.

C. RECEIVER OPERATING CHARACTERISTIC OF MULTI-MODALITY DATA

In some scenarios, varying feature types are favorable for different modalities. For instance, in scenario of video and

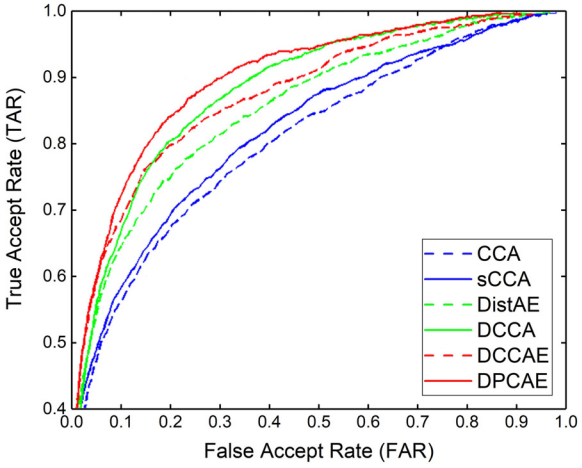


FIGURE 8. Data classification across feature type on MCIC dataset in terms of ROC.

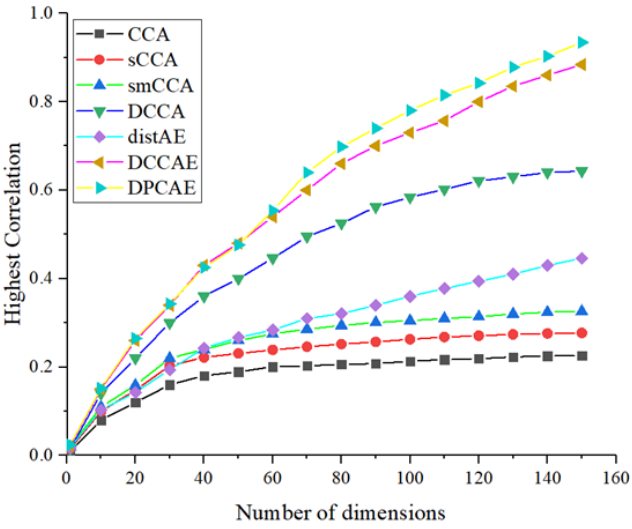


FIGURE 9. A comparison of maximal correlation of CCA-based models and DNN-based models on different number of dimensions.

images, intensity and covariance of intensity are preferred for representing the videos and images respectively, or varying lighting pre-processing are preferred for varying images. In these scenarios, classification is conducted across feature type. To analyze two-modality data across feature type, different models conduct the experiment on the MCIC dataset with two modalities.

Similar to the two experiments above, CCA, sCCA, DistAE, DCCA, DCCAE and DPCAE are tuned to report the best result. The evaluation consequences are given in Fig. 8. In the Fig. 8, the general linear models for evaluation include CCA and sCCA, denoted in blue lines. For non-linear deep models, the experiment tries to evaluate the DistAE, DCCA, DCCAE and our DPCAE which share similar objective as the general linear models. The task on this experiment is to do data classification between the SNPs and fMRI with different types of feature respectively. The property is measured in terms of ROC curve. It defines false accept rate (FAR) as the X-axis and true positive rate true accept rate (TAR) as the Y-axis. The calculation formulas of FAR and TAR are given

$$FAR = FP/(FP + TN) \tag{18}$$

where FP is the number of trained datasets that are actually negative samples but are divided into positive samples by the classifier, TN is the number of datasets correctly classified as negative samples after training of each modality.

$$TAR = TP/(TP + FN) \tag{19}$$

where TP is the number of datasets correctly classified as positive samples after training of each modality, FN is the number of trained datasets that are actually positive samples but are divided into negative samples by the classifier.

All algorithms are evaluated in Fig. 8. The unsupervised CCA and sCCA perform the worst followed by DistAE. Furthermore, DCCA and DCCAE perform much better. Finally, the proposed DPCAE performs the best, with effective improvement, benefited from the deep non-linear and penalty term. To all appearances, our DPCAE exceeds other algorithms, demonstrating the efficacy of our deep multi-modality scheme.

D. MAXIMAL CORRELATION OF VARIOUS ADVANCED METHODS ON THE DIFFERENT NUMBER OF DIMENSIONS

To compare the power of DPCAE with that of other advanced models, DPCAE is applied to integrate SNPs dataset and fMRI dataset, and use other advanced models to integrate SNPs dataset and fMRI dataset as an instance for comparison. After that, the experiment compared the power of maximal correlation of DPCAE and other advanced models for different number of dimensions. In the experiment, 5-fold cross-validations were used to test the data set. 80% of samples were randomly selected from all samples as the training set, and the remaining samples were used as the test set. In order to minimize the adverse effect on the results caused by the difference between the training set and the test set, we chose the result with the smallest correlation coefficient difference between the training set and the test set in the 5-fold cross-validations. Here both DPCAE and other advanced models were performed 50 times, and the maximal correlation was calculated, each using 50% of the SNPs data and fMRI data for learning projections, 30% for adjusting hyper-parameters (regularization parameters learning rate, the total number of neurons in all intermediate layers and so on), and 20% for eventual testing.

Fig. 9 shows different maximal correlations acquired using traditional CCA, sCCA with L_1 penalty, special sCCA with more than two data sets, Deep CCA with deep network including non-linear mapping, minimum-distance auto-encoders, deep canonically correlated auto-encoders and deep principal correlated auto-encoders. The DCCAE model is inclined to find slightly non-linear relationship in the first little margin, after which the DPCAE model exceed them

TABLE 2. The captured best correlation by different MODELs including our proposed DPCAE, which represent only what is shown in Fig.9.

Method	CCA	sCCA	smCCA	DCCA	distAE	DCCAE
HC-150	0.23	0.28	0.33	0.64	0.45	0.88

by a large margin. It's worth noting that DPCAE should particularly have a superiority when PCA technology is used to better fitting multi-modalities data, two deep networks are used to extract features of hidden units, two back propagation neural networks are used to extract top-level features representation and fine-tune the entire deep belief network in each modality respectively. The results indicate that DPCAE model can indeed detect more correlation than CCA-based models and other DNN-based models (the result is given in Fig. 9).

The maximal correlation coefficients of the top-level output nodes are calculated in different dimensions, and made maximal correlation analysis for these models:

$$\sigma = Corr(X_{train}u_{test}, Y_{train}v_{test}) \tag{20}$$

where X_{train} and Y_{train} are the outputs of trained datasets, u_{test} and v_{test} are the weight vectors for the final test. These models have be tested for several times, then select arg max σ as the maximal correlation coefficients.

To identify and clearly show the influence of different models on the performance of maximal correlation, we have combined the proposed model with other advanced models into a table. For DNN-based models, the experiment decreases the total number of intermediate units in all layers so as to the total sum of all parameters approximatively consistent with CCA-based models. Table 2 gives the maximal correlation, noting that the maximal correlation of both datasets increases monotonically with the number of parameters of all models.

V. DISCUSSION AND CONCLUSION

In the work, we propose the DPCAE model to effectively combine prediction and correlation using multi-layer belief networks. DPCAE seeks the optimal network representation that can maximize cross-data correlation while minimize the data fitting error. As we demonstrate, the DPCAE model overcomes the limitation of several existing models in that it can detect complex nonlinear relationship and acquire maximal correlation. As a result, the model can lead to better performance in both prediction and correlation detection. The superior power of DPCAE on both correlation detection and classification makes it a suitable model for genomic data integration, here we apply the model to analyze the correlations of functional networks and the difference of multi-modality features between how different subject groups.

Several experiments have been tested in both CCA-based and DNN-based representation learning. The results have discovered that on several experiments, all aspects of performance of DNN-based models consisting of DCCA, distAE, DCCAE and DPCAE outperform that of CCA-based models consisting of CCA, sCCA and smCCA. The best overall performer is a competent optimization algorithm for solving DCCA with PCA on multi-modality linear features learning and a multi-layer belief network based on RBM on multi-modality nonlinear features learning introduced here, deep principal correlated auto-encoders (DPCAE). It can be interpreted that DNN-based models can acquire preferable representations concerning the interrelated objective measured on multi-modalities. DPCAE not only provides a flexible nonlinear mapping but also provides a simple linear mapping. Another noteworthy preponderance of DPCAE is that, like the CCA method, it does not need an inner product. It is difficult to the pattern recognition problem faced by deep learning. There is lack of general conclusion about what conditions can be converged on the training set and what is the minimum upper bound of loss after many epochs. The main reason is that there are too many variables to describe the time complexity of deep learning. In the back-propagation algorithm, the gradient of parameters can only be obtained after a complete ''forward'' calculation and ''reverse'' calculation, and the parameters are updated. Therefore, all the intermediate gradients need to be saved, and the space complexity of utilizing deep networks is high. Therefore, the experiments on genetic variants and brain imaging phenotypes focused on two works: classification and correlation analysis. Meanwhile, in many learning tasks of DNN-based models, for instance, classification and visualization, obtaining the high performance of correlation is not the last aim and the maximal correlated representations are applied in the learning of other experiments [37], [38].

On the basis of the empirical researches, it is appealing to think about again the essential learning performances of different function types and corresponding penalty terms. Auto-encoder-based technology is grounded in the concept that the output variables should be in a position to exactly reconstruct the variables of the visual layer. For another, CCA method only focuses on how fine multi-modalities' learning feature forecasts the other feature while neglecting the power to reconstruct multi-modalities. The CCA method is anticipated to accomplish well when the two modalities are uncorrelated given the sample labels [39]. The penalty terms in the miscellaneous algorithms also have a crucial impression. Objectively speaking, the stronger DPCAE penalty term is not strong enough yet; an even slightly fine penalty term would be to need the learned feature representations to be mutually independent.

DPCAE could be further improved by using other neural network structures, e.g., self-organizing map network (SOM), which is a competitive unsupervised neural network, and also can map the high-dimensional feature representations to the low-dimensional space and maintain the topological formation of the input in the high-dimensional space [40]. However, directly using the SOM network to multi-modalities learning of SNPs-fMRI may face many challenges because brain imaging and single nucleotide polymorphisms data cannot have some output coming back as input. Despite the difficulties of using SOM networks, using recurrent neural

network structure might be an attractive way to exploit the complex non-linear information or simple linear information within SNPs-fMRI data [41], another prospective schedule is to compare DNN-based methods with methods grounded on deep Boltzmann machines [42], [43]. Multi-label learning arouses great interests in many applications, it can not only enhance the uncertainty but also improve the similarity measurement of multi-label data with labels information [44], [45]. But the lack of the labeled data and the complex structures of various data may make learn the uncertainty and representativeness accurately become hard, a multiple kernel active learning framework could be used to solve this problem [46]. We will consider applying multi-layer belief networks or SOM networks to represent original data and then seeks their correlations, while also linking the data representation with phenotypical information.

REFERENCES

[1] M. G. Naylor, X. Lin, S. T. Weiss, B. A. Raby, and C. Lange, "Using canonical correlation analysis to discover genetic regulatory variants," *PLoS ONE*, vol. 5, no. 5, May 2010, Art. no. e10395, doi: 10.1371/journal.pone.0010395.

[2] E. Parkhomenko, D. Tritchler, and J. Beyene, "Sparse canonical correlation analysis with application to genomic data integration," *Stat. Appl. Genet. Mol. Biol.*, vol. 8, no. 1, pp. 1–34, Jan. 2009, doi: 10.2202/1544-6115.1406.

[3] S. Waaijenborg, P. C. de Witt Hamer, and A. H. Zwinderman, "Quantifying the association between gene expressions and DNA-markers by penalized canonical correlation analysis," *J. Stat. Appl. Genet. Mol. Biol.*, vol. 7, no. 1, pp. 1–29, 2008, doi: 10.2202/1544-6115.1329.

[4] D. M. Witten and R. J. Tibshirani, "Extensions of sparse canonical correlation analysis with applications to genomic data," *Stat. Appl. Genet. Mol. Biol.*, vol. 8, no. 1, pp. 1–27, Jan. 2009, doi: 10.2202/1544-6115.1470.

[5] S. Cao, H. Qin, A. Gossmann, H.-W. Deng, and Y.-P. Wang, "Unified tests for fine-scale mapping and identifying sparse high-dimensional sequence associations," *Bioinformatics*, vol. 32, no. 3, pp. 330–337, Feb. 2016, doi: 10.1093/bioinformatics/btv586.

[6] S.-P. Deng, W. Hu, V. D. Calhoun, and Y.-P. Wang, "Integrating imaging genomic data in the quest for biomarkers of schizophrenia disease," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 15, no. 5, pp. 1480–1491, Sep. 2018, doi: 10.1109/tcbb.2017.2748944.

[7] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3/4, pp. 321–377, Dec. 1936, doi: 10.2307/2333955.

[8] D. M. Witten, R. Tibshirani, and T. Hastie, "A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis," *Biostatistics*, vol. 10, no. 3, pp. 515–534, Jul. 2009, doi: 10.1093/biostatistics/kxp008.

[9] Fang, J., D. Lin, C. Schulz, Z. Xu, V. D. Calhoun, and Y. P. Wang. "Joint sparse canonical correlation analysis for detecting differential imaging genetics modules," *J. Bioinf.* vol. 32, no. 22, pp. 3480–3488, 2011, doi: 10.1093/ bioinformatics/btw485.

[10] G. Andrew, R. Arora, J. Bilmes, and K. Livescu, "Deep canonical correlation analysis," in *Proc. Int. Conf. Mach. Learn.*, 2013, vol. 5, no. 6, pp. 1247–1255.

[11] W. Wang and R. K. J. Arora Livescu Bilmes, "On deep multi-view representation learning," in *Proc. Int. Conf. Mach. Learn.*, 2015, vol. 7, no. 2, pp. 1083–1092.

[12] W. Hu and B. V. Y.-P. Cai Calhoun Wang, "Multi-modality brain connectivity study using deep collaborative learning," *J. Springer Nature Amer.*, vol. 7, no. 10, pp. 1–9, 2018, doi: 10.1109/TBME.2019.29.04301.

[13] B. Du, W. Xiong, J. Wu, L. Zhang, L. Zhang, and D. Tao, "Stacked convolutional denoising auto-encoders for feature representation," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 1017–1027, Apr. 2017, doi: 10.1109/tcyb.2016.2536638.

[14] E. Parkhomenko and D. J. Tritchler Beyene, "Genome-wide sparse canonical correlation of gene expression with genotypes," *BMC Proc.*, vol. 1, no. 11, pp. 1–5, 2007, doi: 10.1186/1753-6561-1-s1-s119.

[15] K. A. L. Cao, P. G. Martin, C. Robert-Granié, and P. Besse, "Sparse canonical methods for biological data integration: Application to a cross-platform study," *BMC Bioinf.*, vol. 10, no. 34, pp. 1–17, 2009, doi: 10.11.86/1471-2105-10-34.

[16] P. Chalise, A. Batzler, R. Abo, L. Wang, and B. L. Fridley, "Simultaneous analysis of multiple data types in pharmacogenomic studies using weighted sparse canonical correlation analysis," *OMICS, A J. Integrative Biol.*, vol. 16, nos. 7–8, pp. 363–373, Jul. 2012, doi: 10.1089/omi.2011.0126.

[17] S. M. Gross and R. Tibshirani, "Collaborative regression," *Biostatistics*, vol. 16, no. 2, pp. 326–338, Apr. 2015, doi: 10.1093/biostatistics/kxu047.

[18] W. Hu, D. Lin, S. Cao, J. Liu, J. Chen, V. D. Calhoun, and Y.-P. Wang, "Adaptive sparse multiple canonical correlation analysis with application to imaging (EPI)Genomics study of schizophrenia," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 2, pp. 390–399, Feb. 2018.

[19] W. Wang, R. Arora, K. Livescu, and J. A. Bilmes, "Unsupervised learning of acoustic features via deep canonical correlation analysis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, vol. 12, no. 3, pp. 1–5, doi: 10.1109/ICASSP.2015.7178840.

[20] T. D. Bie and B. D. Moor, "On the regularization of canonical correlation analysis," *Int. Symp. Integr. Comput.-Aided Eng.*, vol. 76, no. 2, pp. 257–284, 2003.

[21] Y.-H. Dai, L.-Z. Liao, and D. Li, "On restart procedures for the conjugate gradient method," *Numer. Algorithms*, vol. 35, nos. 2–4, pp. 249–260, Apr. 2004, doi: 10.1023/b:numa.0000021761.10993.6e.

[22] R. L. Gollub, J. M. Shoemaker, M. D. King, T. White, S. Ehrlich, S. R. Sponheim, V. P. Clark, J. A. Turner, B. A. Mueller, V. Magnotta, D. O'Leary, B. C. Ho, S. Brauns, and D. S. Manoach, "The MCIC collection: A shared repository of multi-modal, multi-site brain image data from a clinical investigation of schizophrenia," *Neuroinform*, vol. 11, no. 3, pp. 367–388, Jul. 2013, doi: 10.1007/s12021-013-9184-3.

[23] C. C. Abbott, D. Il Kim, S. R. Sponheim, J. Bustillo, and V. D. Calhoun, "Decreased default mode neural modulation with age in schizophrenia," *Amer. J. Geriatric Psychiatry*, vol. 18, no. 10, pp. 897–907, Oct. 2010, doi: 10.1097/jgp.0b013e3181e9b9d9.

[24] O. Demirci, V. P. Clark, and V. D. Calhoun, "A projection pursuit algorithm to classify individuals using fMRI data: Application to schizophrenia," *NeuroImage*, vol. 39, no. 4, pp. 1774–1782, Feb. 2008, doi: 10.1016/j.neuroimage.2007.10.012.

[25] O. Demirci, V. P. Clark, V. A. Magnotta, N. C. Andreasen, J. Lauriello, K. A. Kiehl, G. D. Pearlson, and V. D. Calhoun, "A review of challenges in the use of fMRI for disease classification/characterization and a projection pursuit application from a multi-site fMRI schizophrenia study," *Brain Imag. Behav.*, vol. 2, no. 3, pp. 207–226, Sep. 2008, doi: 10.1007/s11682-008-9028-1.

[26] S. Ehrlich, S. Brauns, A. Yendiki, B.-C. Ho, V. Calhoun, S. C. Schulz, R. L. Gollub, and S. R. Sponheim, "Associations of cortical thickness and cognition in patients with schizophrenia and healthy controls," *Schizophrenia Bull.*, vol. 38, no. 5, pp. 1050–1062, Sep. 2012, doi: 10.1093/schbul/sbr018.

[27] S. Ehrlich, A. Yendiki, D. N. Greve, D. S. Manoach, B.-C. Ho, T. White, S. C. Schulz, D. C. Goff, R. L. Gollub, and D. J. Holt, "Striatal function in relation to negative symptoms in schizophrenia," *Psychol. Med.*, vol. 42, no. 2, pp. 267–282, Feb. 2012, doi: 10.1017/s003329171100119x.

[28] D. I. Kim, J. Sui, S. Rachakonda, T. White, D. S. Manoach, V. P. Clark, B.-C. Ho, S. C. Schulz, and V. D. Calhoun, "Identification of imaging biomarkers in schizophrenia: A coefficient-constrained independent component analysis of the mind multi-site schizophrenia study," *Neuroinform*, vol. 8, no. 4, pp. 213–229, Dec. 2010, doi: 10.1007/s12021-010-9077-7.

[29] D. Lin, V. D. Calhoun, and Y.-P. Wang, "Correspondence between fMRI and SNP data by group sparse canonical correlation analysis," *Med. Image Anal.*, vol. 18, no. 6, pp. 891–902, Aug. 2014, doi: 10.1016/j.media.2013.10.010.

[30] S. Sternberg, "High-speed scanning in human memory," *Science*, vol. 153, no. 3736, pp. 652–654, Aug. 1966, doi: 10.1126/science.153.3736.652.

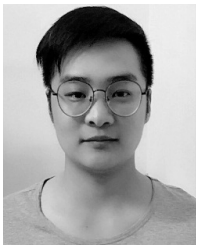
[31] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *NeuroImage*, vol. 15, no. 1, pp. 273–289, Jan. 2002, doi: 10.1006/nimg.2001.0978.

[32] B. Ng and R. Abugharbieh, "Generalized sparse regularization with application to fMRI brain decoding," in *Proc. Int. Conf. Inf. Process. Med. Imag.*, 2011, vol. 8, no. 6, pp. 612–623, doi: 10.1007/978-3-642-22092-0_50.

[33] A. Y. Ng and M. I. Y. Jordan Weiss, "On spectral clustering: Analysis and an algorithm," in *Proc. 14th Int. Conf. Neural Inf. Process. Syst.*, 2001, vol. 9, no. 13, pp. 1–8.
[34] L. Zhang, Q. Zhang, L. Zhang, D. Tao, X. Huang, and B. Du, "Ensemble manifold regularized sparse low-rank approximation for multiview feature embedding," *Pattern Recognit.*, vol. 48, no. 10, pp. 3102–3112, Oct. 2015, doi: 10.1016/j.patcog.2014.12.016.
[35] D. Cai, X. He, and J. Han, "Document clustering using locality preserving indexing," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 12, pp. 1624–1637, Dec. 2005, doi: 10.1109/tkde.2005.198.
[36] C.-C. Chang and C.-J. Lin, "A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–21, 2011, doi: 10.1145/19.61189.1961199.
[37] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodality deep learning," in *Proc. Int. Conf. Mach. Learn.*, 2011, vol. 9, no. 6, pp. 41–51.
[38] N. Srivastava and R. Salakhutdinov, "Multimodality learning with deep boltzmann machines," in *Proc. Int. Conf. Neural Inf.*, 2012, vol. 8, no. 1, pp. 1–8.
[39] S. M. Kakade and D. P. Foster, "Multi-view regression via canonical correlation analysis," in *Proc. Conf. Learn. Theory*, 2007, vol. 15, no. 3, pp. 1–15, doi: 10.10.07/978-3-540-72927-3_8.
[40] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biol. Cybern.*, vol. 43, no. 1, pp. 59–69, 1982, doi: 10.1007/bf00337288.
[41] J. L. Elman, "Finding structure in time," *Cognit. Sci.*, vol. 14, no. 2, pp. 179–211, Mar. 1990, doi: 10.1207/s15516709cog1402_1.
[42] K. Sohn and W. H. Shang Lee, "Improved multimodality deep learning with variation of information," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2014, vol. 7, no. 3, pp. 2141–2149.
[43] B. Mathias and R. Tapani, "Measuring the usefulness of hidden units in Boltzmann machines with mutual information," in *Neural Information Processing*, vol. 5, no. 9. Berlin, Germany: Springer, 2013, pp. 1–9, doi: 10.1007/978-3-642-42054-2_60.
[44] B. Du, Z. Wang, L. Zhang, L. Zhang, and D. Tao, "Robust and discriminative labeling for multi-label active learning based on maximum correntropy criterion," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1694–1707, Apr. 2017.
[45] B. Du, T. Xinyao, Z. Wang, L. Zhang, and D. Tao, "Robust graph-based semisupervised learning for noisy labeled data via maximum correntropy criterion," *IEEE Trans. Cybern.*, vol. 49, no. 4, pp. 1440–1453, Apr. 2019, doi: 10.1109/tcyb.2018.2804326.
[46] Z. Wang, B. Du, W. Tu, L. Zhang, and D. Tao, "Incorporating distribution matching into uncertainty for multiple kernel active learning," *IEEE Trans. Knowl. Data Eng.*, to be published, doi: 10.1109/tkde.2019.2923211.

GANG LI is currently an Associate Professor and a Postdoctoral Fellow with the School of Electronics and Control Engineering, Chang'an University. He has published more than 20 academic articles, among which many were published in *Automation in Construction*, *China Highway Journal*, and the *Journal of Transportation Engineering*. Apply for hosting and complete research projects have a Department of Shaanxi province natural science fund project the lower part of bridge structure fracture identification of multisensor fusion and localization, Xi'an technology bureau industry-university-institute cooperation to promote engineering based on image analysis of cracks in the bridge foundation over a long distance monitoring method research, Chang'an University, central university basic scientific research business expenses special fund project uneven illumination pavement cracks based on mutual information automatic identification algorithm research, and so on. In recent years, he has been engaged as the Project Applicant in the research work of bio-image genetic big data correlation analysis, and actively tracking the work of variable selection, dimension reduction and correlation analysis in the biological genetic big data based on machine learning, so as to make breakthroughs in methods and technologies.

CHAO WANG received the master's degree in electronic and control engineering from Chang'an University. He has published several academic articles, including the *Journal of Computer Science*. The scientific research projects, he have participated in and completed include the natural science foundation project of Shaanxi Science and Technology Department, identification and positioning of multisensor fusion bridge substructure cracks, and the industry-university-research cooperation promotion project of Xi'an Science and Technology Bureau, research on remote bridge substructure cracks monitoring method based on image analysis, and so on. In recent years, he have been engaged in the research on the correlation analysis of genetic big data in biological image, actively tracking the variable selection, dimension reduction and correlation analysis of biological genetic big data based on machine learning, and also engaged in the research on intelligent algorithm processing, and have made some progress.



DE-PENG HAN received the master's degree in electronic and control engineering from Chang'an University. He has published several academic articles. He participates in and completes the scientific research project with Chang'an University in central basic scientific research in colleges and universities were a special fund project based on mutual information of uneven illumination pavement crack automatic identification algorithm research, Xi'an Technology Bureau Industry-University-Institute cooperation to promote engineering based on image analysis of cracks in the bridge foundation over a long distance monitoring methods research, and so on. In recent years, he has been engaged in the research work of bio-image genetic big data correlation analysis, and the research of machine learning and data mining.

YI-PU ZHANG is currently an Associate Professor and a Postdoctoral Fellow with the School of Electronics and Control Engineering, Chang'an University. During Ph.D. study, he was involved in Ph.D. thesis entitled Algorithm Research on Identification of Transcription Factor Binding Sites, and participated in a number of scientific research projects as the Main Researcher. Among them, in terms of data analysis, as the main participants completed on national natural fund project multicore system under control graphs model and algorithm of pattern recognition research, despite the focus in the study of the project for the parallel algorithm design, but the algorithm model analysis of genetic data in pattern recognition under the hidden Markov model have a deeper understanding. In terms of big data processing, he participated in the project research on big data compression index and search algorithm of the National Natural Science Foundation of China, which consolidated the theoretical foundation of pattern recognition for large-scale data.

PENG PENG received the B.S. degree in communication engineering from Anhui Engineering University, Wuhu, Anhui, in 2013, and the M.S. degree in control engineering from the Shaanxi University of Science and Technology, Xi'an, Shaanxi, in 2016. He is currently pursuing the Ph.D. degree in traffic information engineering and control with the Department of Electronics and Control Engineering, Chang'an University, Xi'an. His research interests include machine learning, pattern recognition, data mining, and the correlation analysis of imaging genetic data.

VINCE D. CALHOUN is currently the President of the Mind Research Network and a Distinguished Professor with the Department of Electrical and Computer Engineering, University of New Mexico. He has authored more than 550 full journal articles and more than 650 technical reports, abstracts and conference proceedings. His works include the development of flexible methods to analyze functional magnetic resonance imaging such as independent component analysis (ICA), data fusion of multimodality imaging and genetics data, neuroinformatics, and the identification of biomarkers for disease. Among other things, he leads an NIH P20 COBRE center grant on multimodality imaging of schizophrenia, bipolar disorder, and major depression and an NSF EPSCoR grant focused on brain imaging and epigenetics of adolescent development. He is a Fellow of the Institute of Electrical and Electronic Engineers, The American Association for the Advancement of Science, The American Institute of Biomedical and Medical Engineers, The American College of Neuropsychopharmacology, and the International Society of Magnetic Resonance in Medicine.

YU-PING WANG received the B.S. degree in applied mathematics from Tianjin University, China, in 1990, and the M.S. degree in computational mathematics and the Ph.D. degree in communications and electronic systems from Xi'an Jiaotong University, China, in 1993 and 1996, respectively. After his graduation, he had visiting positions at the Center for Wavelets, Approximation and Information Processing, National University of Singapore, and the Washington University Medical School, St. Louis. From 2000 to 2003, he worked as a Senior Research Engineer with Perceptive Scientific Instruments, Inc., and then Advanced Digital Imaging Research, LLC, Houston, TX, USA. In the fall of 2003, he returned to academia as an Assistant Professor of computer science and electrical engineering with the University of Missouri–Kansas City. He is currently a Professor of biomedical engineering and biostatistics & bioinformatics with the School of Science and Engineering, Tulane University, and also with the School of Public Health and Tropical Medicine, Tulane University. He is also a member of the Tulane Center of Bioinformatics and Genomics, the Tulane Cancer Center, and the Tulane Neuroscience Program. His research interests have been computer vision, signal processing and machine learning with applications to biomedical imaging and bioinformatics, where he has about 200 peer reviewed publications. He has served on numerous program committees and NSF/NIH review panels, and served as editors for several journals such as *Neuroscience Methods*.