

Deep Transformation Method for Discriminant Analysis of Multi-Channel Resting State fMRI

Abhay M S Aradhya

School of Computer Engineering
NTU, Singapore
abhay003@e.ntu.edu.sg

Aditya Joglekar

Computer Science Department
UCLA, USA
adivj123@gmail.com

Sundaram Suresh

Department Of Aerospace
IISC, India
sureshsundaram@iisc.ac.in

M. Pratama

School of Computer Engineering
NTU, Singapore
mpratama@ntu.edu.sg

Abstract

Analysis of resting state - functional Magnetic Resonance Imaging (rs-fMRI) data has been a challenging problem due to a high homogeneity, large intra-class variability, limited samples and difference in acquisition technologies/techniques. These issues are predominant in the case of Attention Deficit Hyperactivity Disorder (ADHD). In this paper, we propose a new Deep Transformation Method (DTM) that extracts the discriminant latent feature space from rs-fMRI and projects it in the subsequent layer for classification of rs-fMRI data. The hidden transformation layer in DTM projects the original rs-fMRI data into a new space using the learning policy and extracts the spatio-temporal correlations of the functional activities as a latent feature space. The subsequent convolution and decision layers transform the latent feature space into high-level features and provide accurate classification. The performance of DTM has been evaluated using the ADHD200 rs-fMRI benchmark data with cross-validation. The results show that the proposed DTM achieves a mean classification accuracy of 70.36% and an improvement of 8.25% on the state of the art methodologies was observed. The improvement is due to concurrent analysis of the spatio-temporal correlations between the different regions of the brain and can be easily extended to study other cognitive disorders using rs-fMRI. Further, brain network analysis has been studied to identify the difference in functional activities and the corresponding regions behind cognitive symptoms in ADHD.

Introduction

Blood Oxygen Level Dependent (BOLD) fMRI is widely used to map the brain functional activities (Song and Lu 2017). The correlations between the regional functional activities during a cognitive task are studied using the multi-channel time series fMRI data to evaluate the interactions

between the various regions of the brain (Yousefnezhad and Zhang 2017). fMRI data have been used to understand the etiopathogenesis of neurological disorders such as Attention Deficit Hyperactivity Disorder (ADHD) (Mahanand, Savitha, and Suresh 2013), autism (Subbaraju et al. 2017) and epilepsy (Gill, Mirsattari, and Leung 2017). Studies on the automatic diagnosis of neurological disorders from fMRI data have focused on analyzing the correlations between the regions to find discriminant activities. Among them, Spatial Filtering Method (SFM) (Subbaraju et al. 2017) proposed a spatial transformation method to obtain discriminatory features by projecting the fMRI data into a new dimension, such that the two classes are highly separable. Their results represent the state of the art diagnostic performance on the Autism dataset. Recently, (Aradhya et al. 2018) utilized a regularized SFM based technique to improve the accuracy in the automatic detection of accuracy of ADHD from rs-fMRI data. SFM and its derivative techniques have adopted the 'Fukunaga-Koontz' transform (Fukunaga 2013) to mathematically derive the spatial transformation filter which is greatly dependent on the mean distribution of the training data. Further, the mean co-variance based spatial transformation methodologies such as SFM overlooks the temporal information of the data, resulting in plausible loss of discriminative information.

Deep learning algorithms have recently gained importance due to their ability to extract ingenious discriminatory information from highly homogeneous data. With rapid advancements in computational power and parallelization techniques along with the increasing availability of large-scale dataset, there has been an increased effort to better understand the human body using deep learning methodologies. (Ravi et al. 2017), provided a brief overview of recent works utilizing deep learning approaches in health informatics. Deep learning methodologies have especially been advantageous in investigating neurological and psycholog-

ical disorders (Vieira, Pinaya, and Mechelli 2017). The results indicate that deep learning methodologies achieve better classification in comparison to the traditional methods. However, direct application of deep learning methodologies on neurological datasets is a challenge due to - the small size of the dataset, imbalanced class distribution, high variability in the data and lack of discriminatory information between the classes. Deep learning approaches have been explored in the diagnosis of ADHD, (Kuang and He 2014) used Restricted Boltzmann Machines (RBM) for classification of ADHD. They converted the fMRI time series into a 1D spectral feature vector and then trained a Deep Belief Network (DBN) for the classification task and reported a classification accuracy below 50% on the ADHD200 benchmark dataset.

Recently, there has been increasing interest in the application of Convolution Neural Networks (CNNs) in the detection and identification of biomarkers of diseases from fMRI, X-ray images, computed tomography scans and positron emission tomography scans (Ravi et al. 2017). (Riaz et al. 2017) adopted a CNN based feature extraction method to identify the discriminant functional activities using rs-fMRI and phenotypic information. They obtained an average accuracy of 62%, with a Support Vector Machine (SVM) classifier. (Zou et al. 2017) introduced a multi-modal 3D CNN approach to study ADHD. Here, the encoded structural-MRI and fMRI are jointly extracted as features and achieved a classification accuracy of 69.15%. In (Vieira, Pinaya, and Mechelli 2017), the functional connectivity in the brain between the different regions was examined and the results concluded that the functional connectivity between the various regions of the brain encodes critical information in understanding ADHD. However, directly applying existing CNN algorithms to diagnose ADHD is not feasible due to the small size of the publicly available ADHD dataset and the lack of separability between the two classes. Motivated by the findings in the above studies and to address the challenges in the automatic diagnosis using rs-fMRI data, in this paper, we develop a new deep learning method to effectively capture the spatio-temporal functional correlations between the different regions of the brain.

In this paper, we propose a Deep Transformation Method (DTM), which project the rs-fMRI time series data into latent feature space using hidden transformation layer and the subsequent convolution/decision layers help in classification. The cross-correlations across the different regions are used by hidden transformation layer to capture the spatio-temporal correlations between the regions of the brain. Further, the convolution layer transforms the latent space to high-level features. A softmax activation function in the decision layer estimates the class conditional probabilities to classify the rs-fMRI data. DTM shows that the functional correlations and convolution filters are complementary, and not only help in extracting the spatio-temporal correlations in the multi-channel data but also boost the classification performance.

The three significant contributions of this paper are:

1. The hidden transformation layer in DTM handle the homogeneity and provide discriminant analysis by preserv-

ing the spatial-temporal information. Further, DTM is suitable to handle transformation from a small sample size, and uncertainties in time-series data due to differences in data acquisition methods.

2. We find that the functional activity disruptions in ADHD are spatio-temporally correlated and the proposed methodology provides support to the brain network dysfunction hypotheses proposed to explain dysfunctional functional brain activities in ADHD.
3. The proposed methodology achieves a state of the art classification accuracy of 70.36% on a hold-out testing dataset from the ADHD200 consortium dataset, demonstrating the significance of concurrently analyzing the spatio-temporal information in fMRI for the diagnosis of neurological disorders.

The rest of the paper is organized as follows, first, an overview of the problem, its implications and a brief literature review is presented. Further, a detailed methodology of the proposed Deep Transformation Method (DTM) is described. The results and discussion section presents the inferences and classification performance of DTM using the ADHD200 dataset. In the next section, a brief overview of the experimental setup and dataset is presented. The performance of the DTM is evaluated using the state of the art methodologies and the significance of the concurrent analysis of the spatio-temporal correlations are presented. Finally, the conclusions from the study are summarized and future scope of work is defined.

Problem Formulation

Diagnosis of ADHD from rs-fMRI is challenging due to a highly homogeneous dataset with large intra-class variability. ADHD is one of the most prevalent developmental neurological disorders in the world with (3-10% worldwide) (Burd et al. 2003). Traditional diagnostic methods for ADHD are based on cognitive tests, interviews and observations by the doctors and associates. However, these are highly prone to variations due to human errors like inaccurate observations and inter-observer variability. The intelligence quotient, culture and language of the subject is also known to have a major influence on the diagnostic procedure. These discrepancies in the current diagnostic procedure have led to a high risk of misdiagnosis and administration of improper medications.

To overcome the shortcomings of the traditional diagnostic procedures and to have a better understanding of the etiopathogenesis of ADHD, researchers have looked at the Blood Oxygen Level Dependent (BOLD) functional Magnetic Resonance Imaging (fMRI) to find discriminant brain activities in children with ADHD. Task based fMRI studies have shown variations in functional activity in various brain networks such as the default mode network (Konrad and Eickhoff 2010), cingulo-fronto-parietal network (Bush 2011), central executive network and the salience network (SN) (Menon 2011). However, a clear consensus on the pathophysiological cause of ADHD was not achieved due to variations in the tasks and the resulting activation regions between the studies. Therefore, researchers have looked at

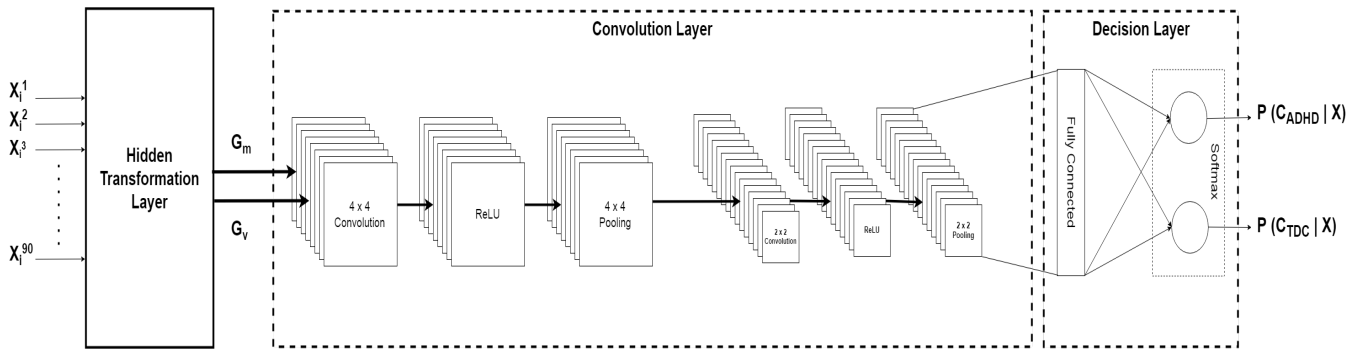


Figure 1: Architecture of the proposed DTM for the diagnosis of ADHD from rs-fMRI data.

resting state fMRI (rs-fMRI) which measures the intrinsic brain activities to underline the causal factors of ADHD. rs-fMRI although independent from task-directed cognitive processes, are extremely homogeneous and presents a challenge in identifying the discriminant functional activities.

(Mahanand, Savitha, and Suresh 2013) studied the regional anatomy of the brain using the amygdala and cerebellar vermis regions. They used a meta-cognitive learning classifier and achieved a classification accuracy of 65%. Analyzing rs-fMRI activity from specific regions of the brain leads to a potential loss of valuable discriminative information. (Anderson et al. 2014) used a decision tree based approach to achieve a maximum classification accuracy of 66.8% using both the phenotypic and rs-fMRI data. However, the classification accuracies of studies focusing of automated diagnosis of ADHD from rs-fMRI alone (without using phenotypic information or predetermined neural regions) have been substantially lower. (Ghiassian et al. 2013) used a histogram based feature reduction technique along with an Support Vector Machine (SVM) classifier and achieved a classification accuracy of 62.81%. (Guo et al. 2014) studied the regional connectivity between the regions of the brain and reported an accuracy of 63.75%. The existing approaches is not able to handle high homogeneity, large intra-class variability, limited samples and difference in acquisition technologies/techniques.

Deep Transformation Method (DTM)

DTM is a deep learning based classifier that projects the rs-fMRI data into a latent space and captures the spatio-temporal correlations to handle the above-mentioned issues and understand the discriminant functional activities of the brain. For this purpose, we propose a multi-layered Deep Transformation Method (DTM). The schematic representation of the DTM architecture is shown in Figure 1. DTM uses the Hidden Transformation Layer (HTL) to project the rs-fMRI data into a latent space where the classes are highly separable. The convolution layers extract high-level features from the grey scale correlation images and the decision layer uses a sigmoid activation function to classify the data.

Hidden Transformation Layer (HTL)

The hidden transformation layer transforms the time series rs-fMRI data $X_i = \{X_i^1, X_i^2, X_i^3, \dots, X_i^{90}\}$ (annotated in accordance to the automated anatomical labeling template) into an image data by preserving spatio-temporal information. These transformed images are highly discriminative for both the classes. Conventionally the transformation filters are determined using mathematical derivations such as the 'Fukunaga-Koontz' transform (Fukunaga 2013). Previous studies (Subbaraju et al. 2017) have used the derivative approaches and have achieved significant improvement in classification accuracy. However, the classification performance of such deterministic methodologies are greatly reliant on the accurate estimation of the spatial filter $W \in R^{90 \times 90}$, that transforms the BOLD rs-fMRI time series data $X \in R^{90 \times T}$ (from the 90 regions of the brain with T time points). The different classes of the rs-fMRI data are highly separable in the new space and the transformed rs-fMRI data $Y \in R^{90 \times T}$ calculated as

$$Y = WX \quad (1)$$

Although the eigen value decomposition based approaches provide a good approximation of the transformation filter, they are highly reliant on the mean distribution of the data. Also, such methods are susceptible to noise in problems with small dataset and high intra-class variability resulting in loss of discriminatory information. The discriminatory functional activities of the brain in the transformed rs-fMRI are represented as differences in functional connectivity between the different regions of the brain (Vieira, Pinaya, and Mechelli 2017). Therefore, it is beneficial to develop a non-deterministic approach to study the correlations between the rs-fMRI signals in order to validate the discriminant activities of the brain.

DTM proposes a new deep learning based spatial transformation method, that projects the rs-fMRI data using a learn-able hidden transformation layer while preserving the spatio-temporal information using a cross-correlation based regional connectivity estimation. In DTM, the regional connectivity matrix is estimated as

$$S_{(i,j)} = \frac{Y_i[t]Y_j[t-d]'}{\text{trace}(Y_i[t]Y_j[t-d]')} \quad (2)$$

The above equation can be simplified as

$$\hat{S}_{(i,j)} = \frac{W_i[t]X_i[t]X_j[t-d]'W_j[t-d]'}{\text{trace}(W_i[t]X_i[t]X_j[t-d]'W_j[t-d]')} \quad (3)$$

where, $d = -t_n, \dots, t_n$ and $i, j = 1, 2, \dots, 90$

In Equation 3, t_n represents the total number of time points in the time series data X , d is the delay used in calculating the cross-correlations while i and j are indices representing a pair of regions. The regional covariance matrix $S \in R^{90 \times 90 \times t_n}$ is a non-symmetric three dimensional matrix where $0 \leq ||S|| \leq 1$. In order to determine the regional connectivity, it is beneficial to analyze the cross-correlation between the rs-fMRI time series signals as cross-correlation effectively captures the temporal correlations between the various regions of the brain, which is otherwise lost during simple linear correlation based estimation.

Further, as Equation 3 can be represented as a series of linear matrix operations and therefore is differentiable with respect to W . The transformation filter W is therefore updated using the backpropagation based learning strategy such that error in classification performance is minimized. Therefore, DTM uses a data-centric deep learning approach to derive the transformation filter W to optimize the classification performance. Moreover, this enables the formulation of HTL as an independent deep learning layer which can be used in a multi-layer deep learning network to transform the data into a latent feature space.

In this paper, we have applied a multi-layer convolutional network along with the HTL transformation layer in order to extract high-level features from the rs-fMRI data. Convolution neural networks use two-dimensional filters and therefore necessitate the conversion of the three-dimensional regional covariance matrix into a two-dimensional feature space. Hence, the mean (\hat{S}_m) and variance (\hat{S}_v) across the time dimension was calculated as

$$\hat{S}_m = \frac{1}{(2t_n - 1)} \sum_{t=1}^{(2t_n-1)} \hat{S}(t) \quad (4)$$

$$\hat{S}_v = \text{var}(\hat{S}(t)) \quad (5)$$

where, \hat{S}_m and \hat{S}_v are non-symmetric matrices of dimension 90×90 which encapsulate the spatio-temporal functional activity between the 90 regions of the brain.

Convolutional Layer

The functional relationship between the different regions of the brain is represented by the correlations between the neighbouring (spatially correlated) elements in the regional covariance matrices (Vieira, Pinaya, and Mechelli 2017).

Convolutional neural networks is a popular deep learning method that is adept at capturing spatial information and has been effective in the classification of natural images (LeCun, Bengio, and Hinton 2015). DTM exploits the convolutional layers to extract low-level spatial features and transform them into high-level features. The regional connectivity matrices \hat{S}_m and \hat{S}_v are converted into normalized mean greyscale image G_m and the variance greyscale image G_v are given as

$$G_m = 0.5 (\hat{S}_m + 1) \quad (6)$$

$$G_v = 0.5 (\hat{S}_v + 1) \quad (7)$$

G_m and G_v are used as inputs to train the two convolution layers. Convolution operations are generally data intensive, wherein the accuracy is directly proportional to the quantity of data used to train the network and the number of layers (depth) used.

However, neurological datasets pose a challenging problem as the number of data samples available are generally limited. Using a large number of convolution layers would thus lead to over-fitting, resulting in a bad generalization of the data. Therefore, DTM uses a shallow LeNet (LeCun and others 2015) based architecture with two layers of Conv-RELU-Maxpool followed by a decision layer.

The first convolutional layer consists of eight 4×4 filters with a depth two (corresponding to the 2 input channels). A stride of one was adopted in both horizontal and vertical directions to ensure that none of the discriminatory information is neglected. In order to maintain the dimension of the inputs, they are padded with zeros ('SAME' padding). Further, a Leaky Rectified Linear Units (ReLU) layer with $\alpha = 0.2$ was used to add non-linearity to the network. A max-pooling layer with a 8×8 filter and 8×8 stride was introduced which although leads to some information loss, prevents having too many learnable parameters compared to the size of the dataset to avoid overfitting. The second convolution layer had more depth along with narrower filters as per the standard practice. It consisted of sixteen 2×2 filters with 8 input channels with a unit stride, followed by a leaky ReLU ($\alpha = 0.2$) non-linearity and a max-pooling layer with 4×4 filter and 4×4 stride.

Decision Layer

The decision layer consists of a fully connected layer and a softmax layer. The output from the last Conv-RELU-Maxpool layer is connected to a fully connected layer with about 1000 neurons, which are then connected to the softmax layer with two output neurons. The softmax layer with a sigmoid activation function was used to estimate the conditional probability $P(C_x|X)$; that is, the confidence of DTM in predicting the class of the current input X correctly as C_x . The class with the maximum conditional probability was given as the predicted class for the input X .

Brain Functional Activity Maps

The activity maps of the brain are useful in analyzing the differences in functional activities in the regions of the brain

between subjects with neurological disorders and neurotypical subjects. The differences in neural activity are obtained using the inverse of the spatial filter W , derived such that the original time series can be regenerated from the inverse spatial filter (W^{-1}) and the projected time series data Y as

$$\hat{X} = W^{-1}Y \quad (8)$$

where \hat{X} is the regenerated BOLD rs-fMRI data. The columns of W^{-1} is referred to as spatial distributions. Each element in the column of W^{-1} is called as a spatial weight and is assigned to each of the corresponding of the 90 Automated Anatomical Labelling (AAL) regions. The spatial weights with significant variance from the mean of each column highlights the significant regions with differences in BOLD rs-fMRI signals between the classes. The polarity of the spatial weights do not hold any significance, and only the absolute value of the spatial weight needs to be examined. The mean spatial distribution and standard deviation of each of the spatial distribution were calculated and the regions with spatial weights whose magnitude is greater than two times the standard deviation away from the mean were considered significant in this paper. The inverse spatial filter from the trail with the highest classification accuracy was chosen and a representative spatial distribution is plotted and the distribution of the spatial weights are analyzed. The regions are further visualized using a brain functional map to depict the regions with discriminant functional activity between the classes.

Experimental Setup

The deep transformation method was implemented as a "TensorFlow" graph in "Python" scripting language and was trained over one hundred (100) iterations with a learning rate of 0.001 using a mini-batch gradient descent approach with a batch size of 32. The batch learning approach prevents convergence is a trade-off between avoiding local minima and the convergence time. Although the DTM took thirty minutes (30) of training time using an 'Intel(R) Xeon(R) CPU E5-1630 v4 @ 3.70GHz' central processing unit, it significantly reduced to seven (7) minutes with the use of 'Tesla P100-SXM2-16GB' graphical processing unit. Each fold of data consisted of non-intersecting training and testing data samples drawn in the ratio 371:54. Every consecutive fold of test data was ensured to have no overlap with the previous test data to obtain generalized results and the average performance over the five folds are reported as the mean classification performance measure.

ADHD200 Dataset

In order to validate the performance of the DTM, we have used the benchmark ADHD200 dataset from the International Neuroimaging Data-sharing Initiative (INDI). INDI has facilitated Deep learning research on ADHD by aggregating rs-fMRI data from 947 people between the age of 7 years to 22 years, aggregated from 8 different institutions under the ADHD200 consortium. However, for the study presented in this paper the data from Brown University were excluded as the diagnostic information of the rs-fMRI data

were not available at the time of compilation of this study. Of the 947 people in the ADHD200 consortium dataset, 563 people (59.45%) were found to be male right-handed. Previous studies (Skounti, Philalithis, and Galanakis 2007) have shown that the gender and handedness of a person severely influence the presentation of functional brain activities in rs-fMRI data. Hence, in order to preserve the homogeneity in the data, this study is confined to male right-handed people.

The rs-fMRI data from the ADHD200 consortium was pre-processed using the ATHENA pipeline (Bellec et al. 2017) in order to remove the variations due to the physiological noise, head motion and scanner drifts. The denoised rs-fMRI data without bandpass filtering provided by (Bellec et al. 2017) was anatomized using the AAL template into 116 regions. The 26 regions corresponding to the cerebellum were excluded to minimize the effects of involuntary activities such as breathing and cardiac activity. Finally, a manual quality control (QC) check was performed and samples with incomplete data and that failed the QC in (Bellec et al. 2017) were removed and rs-fMRI data from the 90 AAL regions were obtained from 465 people. Further details on the preprocessing and QC measures can be obtained from (Bellec et al. 2017).

To facilitate cross-validation and evaluate the performance of the algorithm, the dataset was partitioned into training and test data-sets of 371 and 94 subjects respectively. A 10-fold random partition was done ensuring the ratio of ADHD to TDC remained constant across the folds. The mean classification accuracy and the standard deviation across the partitions are considered as the key indicators of the performance of DTM.

Results and Discussion

In this section, the diagnostic performance of the DTM in classifying ADHD from BOLD rs-fMRI time series data has been presented. The performance is measured in terms of accuracy, sensitivity, specificity and F-score of the classification. First, we compare the performance of the proposed DTM with previous studies on the automatic diagnosis of ADHD in *Table 1*. Studies (Ghiassian et al. 2013; Guo et al. 2014) on the diagnosis of ADHD using BOLD fMRI data from the ADHD200 consortium dataset utilizing various feature extraction algorithms and a linear classifier like SVM have achieved a maximum accuracy of 63.75%. However, (Mahanand, Savitha, and Suresh 2013) used a metacognitive classifier to achieve the best performance of 65%. Using phenotype data along with BOLD fMRI data improved the classification accuracy. (Anderson et al. 2014) explored a decision tree based approach to achieve a classification accuracy of 66.8%, whereas (Riaz et al. 2017) adopted a CNN based process in order to extract the functional connectivity as features and then evaluated them using an SVM classifier to obtain an accuracy between 63.4% and 68.6%. Also, (Riaz et al. 2017) used data from individual acquisition sites in order to validate the results in order to avoid inter-site variability. (Zou et al. 2017) made use of the s-MRI and rs-fMRI data to extract six types of 3D features and introduced a 3D CNN based classifier. They achieved a maximum classification accuracy of 69.15%. These results

Table 1: Comparison of diagnostic performance between the proposed DTM and state of the art methodologies using ADHD200 consortium dataset

Reference	Data	Classifier	Testing accuracy
(Ghiassian et al. 2013)	BOLD rs-fMRI	SVM	62.81%
(Guo et al. 2014)	BOLD rs-fMRI	SVM	63.75%
(Mahanand, Savitha, and Suresh 2013)	BOLD rs-fMRI	Meta-cognitive Radial Basis Function	65%
(Anderson et al. 2014)	BOLD rs-fMRI and phenotype data	Decision tree	66.8%
(Riaz et al. 2017)	BOLD rs-fMRI and phenotype data	SVM	63.4% - 68.6%
(Zou et al. 2017)	BOLD rs-fMRI and s-MRI	3D CNN	69.15%
DTM	BOLD rs-fMRI	Softmax	70.36%

show the highest accuracy in diagnosis achieved by previous studies. Although from the above results it is observed that phenotype data and s-MRI data when used along with the rs-fMRI data improves the classification performance, it hinders the derivation significant conclusions on the discriminant functional activities in ADHD. Therefore, DTM uses only BOLD rs-fMRI as input and in comparison yields a better diagnostic performance on a hold-out test dataset. DTM achieves a mean classification accuracy of 70.36% with a low standard deviation of 0.02 which is a significant improvement of 8.25% over the existing methods. The confusion matrix of the classification performance is given in Table 2.

Table 2: Classification performance of DTM

Parameter	Train Percentage	Test Percentage
True Positive	49.8%	41.5%
True Negative	39.5%	28.9%
False Positive	10.5%	13.7%
False Negative	0.2%	15.9%

Evaluation of significance of concurrent analysis of spatio-temporal information

Table 3: Comparison of diagnostic accuracy of DTM with different region correlation methodologies

Region correlation measure	Training		Testing	
	Mean	S.D	Mean	S.D
Linear correlation	95.12%	0.12	55.27%	0.01
Cross-correlation	89.23%	0.007	70.36%	0.02

The advantages of substituting the linear correlation regional covariance matrix with the mean and variance of the region covariance matrix derived using the cross-correlation

approach is shown in Table 3. Adopting the linear correlation approach to derive the regional covariance matrix although leads to an increase in mean training accuracy, the mean testing accuracy sharply reduces to 55.27% due to overfitting. These results indicate that the time series signals of the different regions of the brain have higher correlations when shifted in time. These findings prove that the functional activity disruptions in ADHD are spatio-temporally correlated and supports the brain network dysfunction theory proposed by (Menon 2011).

Identification of distinguishable brain activities in ADHD right handed male subjects

The spatial weight distribution and brain activity maps of the regions of the brain highlight the discriminant regions of the brain with significant differences in functional activity. The spatial weights from the 90 regions of one of the prominent spatial distributions is plotted in Figure 2. The magnitude of the spatial weights in the figure represent the importance of the region for classification. As the signs of the spatial weights do not signify any information the absolute value of the spatial weights is considered. The magnitude of each of the spatial weights corresponds to the influence of the corresponding regions functional activity in the classification process.

From the Figure 2 it shows that the Rolandic Operculum (18), Medial Frontal Gyrus (23), Posterior Cingulate Gyrus (36), Hippocampus (38) and the Middle Temporal Gyrus (85) have significantly greater spatial weights. Therefore, it implies that these regions of the brain display discriminative functional activity among male right-handed subjects diagnosed with ADHD. The corresponding brain functional activity map is presented in Figure 3, where the regions of the brain are marked by their AAL region number. Prominent differences in functional activity are observed in the temporal lobe and the posterior cingulate cortex. The temporal lobe is primarily responsible for the long-term memory (Simons and Spiers 2003), while the posterior cingulate cortex is an integral part of the Default Mode Network (DMN) and acts as the functional link between the prefrontal and parietal regions of the brain (Fransson and Marrelec 2008). It is

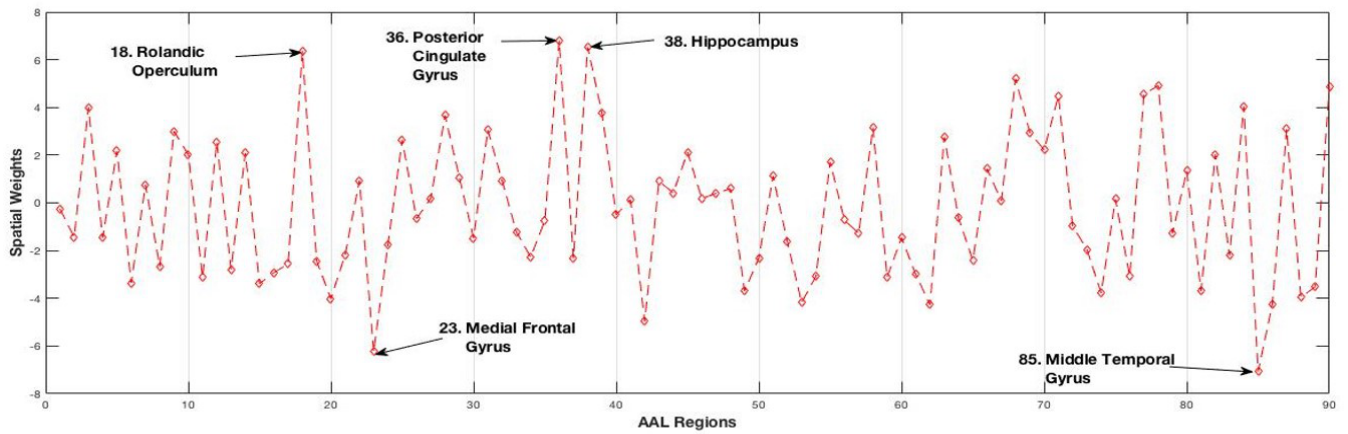


Figure 2: AAL regions exhibiting discriminant functional activity and the associated spatial weights obtained from DTM

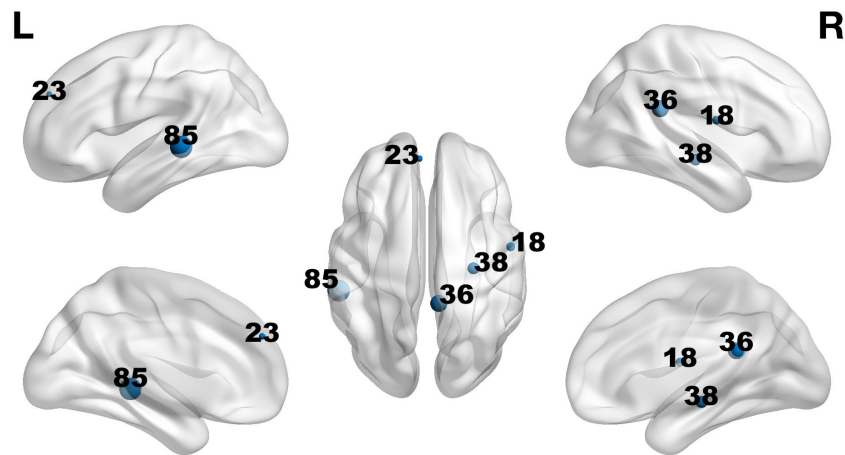


Figure 3: Brain functional activity map showing the lateral and dorsal view of the brain, presenting the regions of the brain with significant variances in rs-fMRI signal between ADHD patients and neurotypical males

Note: Rolandic Operculum (18), Medial Frontal Gyrus (23), Posterior Cingulate Gyrus (36), Hippocampus (38) and the Middle Temporal Gyrus (85)

observed that the regions identified with discriminant functional activity by DTM are related to the cognitive symptoms associated with ADHD such as hyperactivity, impulsivity and inattentiveness.

Conclusions

The proposed deep transformation method uses a hidden transformation layer to project the rs-fMRI into discriminative latent space and preserve the spatio-temporal information. Further, the convolution layers and decision layer

helps in accurate classification. The experimental study using ADHD200 right-handed male dataset clearly shows that the proposed DTM method are higher (approximately 8%) when compared to the existing results reported in the literature. DTM extracts the spatio-temporal correlations in functional activities of the brain instead of the simple spatially correlated functional activities in order to train the CNN based classifier. The results demonstrate that ADHD is characterized by differences in connectivity between the different regions of the brain and therefore should be treated

as a brain network dysfunctional neurological disorder. The brain network analysis reveals that the difference in functional activities is predominant in the temporal lobe and posterior cingulate cortex which implies inattentiveness and impulsivity in ADHD.

References

- Anderson, A.; Douglas, P. K.; Kerr, W. T.; Haynes, V. S.; Yuille, A. L.; Xie, J.; Wu, Y. N.; Brown, J. A.; and Cohen, M. S. 2014. Non-negative matrix factorization of multimodal mri, fmri and phenotypic data reveals differential changes in default mode subnetworks in adhd. *NeuroImage* 102:207–219.
- Aradhya, A. M. S.; Subbaraju, V.; Sundaram, S.; and Sundararajan, N. 2018. Regularized spatial filtering method (r-sfm) for detection of attention deficit hyperactivity disorder (adhd) from resting-state functional magnetic resonance imaging (rs-fmri). In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 5541–5544.
- Bellec, P.; Chu, C.; Chouinard-Decorte, F.; Benhajali, Y.; Margulies, D. S.; and Craddock, R. C. 2017. The neuro bureau adhd-200 preprocessed repository. *Neuroimage* 144:275–286.
- Burd, L.; Klug, M. G.; Coumbe, M. J.; and Kerbeshian, J. 2003. Children and adolescents with attention deficit-hyperactivity disorder: 1. prevalence and cost of care. *Journal of Child Neurology* 18(8):555–561.
- Bush, G. 2011. Cingulate, frontal, and parietal cortical dysfunction in attention-deficit/hyperactivity disorder. *Biological psychiatry* 69(12):1160–1167.
- Fransson, P., and Marrelec, G. 2008. The precuneus/posterior cingulate cortex plays a pivotal role in the default mode network: Evidence from a partial correlation network analysis. *Neuroimage* 42(3):1178–1184.
- Fukunaga, K. 2013. *Introduction to statistical pattern recognition*. Elsevier.
- Ghiassian, S.; Greiner, R.; Jin, P.; and Brown, M. 2013. Learning to classify psychiatric disorders based on fmri images: Autism vs healthy and adhd vs healthy. In *Proceedings of 3rd NIPS Workshop on Machine Learning and Interpretation in NeuroImaging*.
- Gill, R. S.; Mirsattari, S. M.; and Leung, L. S. 2017. Resting state functional network disruptions in a kainic acid model of temporal lobe epilepsy. *NeuroImage: Clinical* 13:70–81.
- Guo, X.; An, X.; Kuang, D.; Zhao, Y.; and He, L. 2014. Adhd-200 classification based on social network method. In *International Conference on Intelligent Computing*, 233–240. Springer.
- Konrad, K., and Eickhoff, S. B. 2010. Is the adhd brain wired differently? a review on structural and functional connectivity in attention deficit hyperactivity disorder. *Human brain mapping* 31(6):904–916.
- Kuang, D., and He, L. 2014. Classification on adhd with deep learning. In *Cloud Computing and Big Data (CCBD), 2014 International Conference on*, 27–32. IEEE.
- LeCun, Y., et al. 2015. Lenet-5, convolutional neural networks. URL: <http://yann.lecun.com/exdb/lenet> 20.
- LeCun, Y.; Bengio, Y.; and Hinton, G. 2015. Deep learning. *nature* 521(7553):436.
- Mahanand, B.; Savitha, R.; and Suresh, S. 2013. Computer aided diagnosis of adhd using brain magnetic resonance images. In *Australasian Joint Conference on Artificial Intelligence*, 386–395. Springer.
- Menon, V. 2011. Large-scale brain networks and psychopathology: a unifying triple network model. *Trends in cognitive sciences* 15(10):483–506.
- Ravi, D.; Wong, C.; Deligianni, F.; Berthelot, M.; Andreu-Perez, J.; Lo, B.; and Yang, G. Z. 2017. Deep learning for health informatics. *IEEE Journal of Biomedical and Health Informatics* 21(1):4–21.
- Riaz, A.; Asad, M.; Al-Arif, S. M. R.; Alonso, E.; Dima, D.; Corr, P.; and Slabaugh, G. 2017. Fcnet: a convolutional neural network for calculating functional connectivity from functional mri. In *International Workshop on Connectomics in Neuroimaging*, 70–78. Springer.
- Simons, J. S., and Spiers, H. J. 2003. Prefrontal and medial temporal lobe interactions in long-term memory. *Nature Reviews Neuroscience* 4(8):637.
- Skounti, M.; Philalithis, A.; and Galanakis, E. 2007. Variations in prevalence of attention deficit hyperactivity disorder worldwide. *European journal of pediatrics* 166(2):117–123.
- Song, X., and Lu, H. 2017. Multilinear regression for embedded feature selection with application to fmri analysis. In *AAAI*, 2562–2568.
- Subbaraju, V.; Suresh, M. B.; Sundaram, S.; and Narasimhan, S. 2017. Identifying differences in brain activities and an accurate detection of autism spectrum disorder using resting state functional-magnetic resonance imaging: A spatial filtering approach. *Medical image analysis* 35:375–389.
- Vieira, S.; Pinaya, W. H.; and Mechelli, A. 2017. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. *Neuroscience & Biobehavioral Reviews* 74:58 – 75.
- Yousefnezhad, M., and Zhang, D. 2017. Local discriminant hyperalignment for multi-subject fmri data alignment. In *AAAI*, 59–65.
- Zou, L.; Zheng, J.; Miao, C.; Mckeown, M. J.; and Wang, Z. J. 2017. 3d cnn based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural mri. *IEEE Access* 5:23626–23636.