Identifying Connectivity Patterns for Brain Diseases via Multi-side-view Guided Deep Architectures

Jingyuan Zhang* Bokai Cao* Sihong Xie* Chun-Ta Lu* Philip S. Yu*† Ann B. Ragin‡

Abstract

There is considerable interest in mining neuroimage data to discover clinically meaningful connectivity patterns to inform an understanding of neurological and neuropsychiatric disorders. Subgraph mining models have been used to discover connected subgraph patterns. However, it is difficult to capture the complicated interplay among patterns. As a result, classification performance based on these results may not be satisfactory.

To address this issue, we propose to learn non-linear representations of brain connectivity patterns from deep learning architectures. This is non-trivial, due to the limited subjects and the high costs of acquiring the data. Fortunately, auxiliary information from *multiple side views* such as clinical, serologic, immunologic, cognitive and other diagnostic testing also characterizes the states of subjects from different perspectives. In this paper, we present a novel **M**ulti-side-**V**iew guided **A**uto**E**ncoder (MVAE) that incorporates multiple side views into the process of deep learning to tackle the bias in the construction of connectivity patterns caused by the scarce clinical data. Extensive experiments show that MVAE not only captures discriminative connectivity patterns for classification, but also discovers meaningful information for clinical interpretation.

1 Introduction

In many neurological and neuropsychiatric disorders, ongoing neural changes may be clinically silent in the earliest stages. By the time clinical symptoms present, injury to the brain may be irreversible. Early detection has transformative potential for preservation of the brain cognitive function and for alleviating the burden of neural injury and associated disability. For brain interrogation and detection of early anomalies, the last several decades have witnessed rampant progress in non-invasive imaging technologies for acquiring quantitative brain data in vivo, including unprecedented capabilities for generating vast data concerning structural (Diffusion Tensor Imaging, abbreviated as DTI) and functional (resting state fMRI) brain connectivities.

The potential of mining the vast image data for the detection of alterations in neurological disorders has been demonstrated in many studies [3, 8, 15, 16, 19, 24]. However, connectivities of brain regions are not investigated in these studies. Thus, the results are usually hard to interpret and they cannot help practicing clinicians to better understand the underlying disease mechanisms. Therefore, discovering clinically meaningful patterns associated with disrupted brain connectomes has recently emerged as a potentially powerful strategy for identifying early changes in neurological disorders [6, 9, 13].

Brain connectomes are often represented as graphs, where nodes represent brain regions and edges represent connections between pairs of regions. Subgraph pattern mining has been used to discover connected subgraph patterns [6, 9, 13]. However, connections between brain regions are characterized by massive complexity. It is difficult for traditional subgraph mining models to capture the complicated interplay among patterns. When the selected subgraph patterns are employed for disease diagnosis, the classification performance is often not satisfactory (e.g., around 60% of accuracy for HIV disease [13]). Hence, existing approaches are unable to find both discriminative and clinically useful connectivity patterns for the study of brain disorders.

To address this issue, we propose to learn non-linear representations of brain connectivity patterns from deep learning architectures. This is non-trivial because imaging studies of neurologic disorders often involve small samples due to the limited number of available patients of interest and the high costs of acquiring the data. Nevertheless, available patients may be very well-characterized with many other sources of auxiliary data from clinical, serologic, immunologic, cognitive and other diagnostic testing. This auxiliary information characterizes the state of subjects from different perspectives that may also be meaningful for brain status, providing *multiple side views*.

In this paper, we devise a deep architecture named **M**ulti-side-**V**iew guided **A**uto**E**ncoder (MVAE), which can effectively discover useful patterns by exploiting

*University of Illinois at Chicago. {jzhan8, caobokai, sxie6, clu29, psyu}@uic.edu

†Institute for Data Science, Tsinghua University.

‡Northwestern University. ann-ragin@northwestern.edu

the auxiliary information associated with brain connectomes. We first apply a stacked autoencoder in MVAE to learn non-linear representations of connectivity patterns from labeled data. Because the autoencoder alone can overfit the scarce training data, we further exploit the multiple side views to guide the representation learning process.

Note that a simple concatenation of the side views and the brain connectome data is undesirable, since the non-linear features learned in the autoencoder would be mixtures of both brain region connectivities and other auxiliary features, making the learned features more difficult to interpret by clinical researchers. On the contrary, we propose a principled method to infer the pairwise similarities between instances from auxiliary data, such that the stacked autoencoder can learn complex connectivity patterns w.r.t the geometry of the data on a low-dimensional manifold [20]. By properly incorporating side views into the autoencoder method, MVAE can discover both discriminative and clinically useful connectivity patterns even with a limited number of brain connectomes.

Our contributions can be summarized as follows:

- We formulate the problem of learning deep connectivity patterns of brain connectomes with the help of multiple side views (Section 4). To the best of our knowledge, this is the first work that mines discriminative and clinically significant brain connectome patterns using non-linear deep architectures.
- We derive the pairwise relationships between instances from multiple side views and verify the importance of such relationships in real-world datasets (Section 3). Based on the derived geometric information, we propose MVAE and incorporate side views into the process of deep learning to tackle the bias in the construction of connectivity patterns caused by the scarce clinical data.
- Through extensive empirical studies on real brain connectomes, we demonstrate that MVAE outperforms all the baseline methods with an improvement of 28% and 22% of accuracy on average for fMRI and DTI, respectively (Section 5.2).
- We visualize the discriminative patterns learned from MVAE (Section 5.6). The rich information embedded within these patterns can be used to discover clinically meaningful patterns associated with disrupted brain connectivities.

2 Preliminary

In this section, we first describe the neuroimage datasets and the auxiliary information used in this study. We then present the notations and the problem formulation of learning deep connectivity patterns of brain connectomes with the help of multiple side views. Table 1 lists the main notations we use through this paper.

Table 1: Important Notations

Symbols	Definitions
$\mathcal{X}$	feature representations of brain connectomes
$\mathbf{y}$	class labels
$\mathcal{Z}$	multiple side views for $\mathcal{X}$
$\mathcal{A}$	learned features for $\mathcal{X}$
N	number of data instances
M	number of features for $\mathcal{X}$
D_k	dimension of the k-th view
s^k	pairwise similarity under the k-th view
K	number of side views associated with $\mathcal{X}$
T	number of learned features
$\mathbf{W}$	weight parameters for autoencoder
$\mathbf{b}$	bias parameters for autoencoder
L	number of layers for autoencoder
h_l	number of hidden features for layer l

2.1 Data Description In this work, we focus on datasets collected from the Chicago Early HIV Infection Study at Northwestern University [17]. This study consists of records of 56 patients with HIV (positive) and 21 normal controls (negative). Both fMRI and DTI images are included. In addition to neuroimages, hundreds of other variables are available for each subject in this study, including *clinical, immunologic, serologic, cognitive* measures. While many brain image datasets[1] are publicly available, usually rich side view information is not included. Therefore, it is not possible to evaluate the effectiveness of incorporating multiple side views into deep learning on these public datasets.

We derive brain connectomes from fMRI and DTI images following the construction steps in [5]. In addition, we study seven groups of measurements associated with brain connectomes. They are *neuropsychological tests, flow cytometry, plasma luminex, freesurfer, overall brain microstructure, localized brain microstructure* and *brain volumetry*. Each group can be considered as a distinct view that partially reflects the status of a subject. Different groups of measurements provide complementary information from different perspectives. Moreover, we normalize both the brain connectome features and the side view features to the range of $[0, 1]$ before employing the proposed model.

2.2 Problem Description Let $\mathcal{X} = \{\mathbf{X}_1, ..., \mathbf{X}_N\}$ denotes the brain connectome dataset of N subjects. $\mathbf{X}_p = \{x_1^p, ..., x_M^p\}$ is the p-th set of features in $\mathcal{X}$, which represents the brain connectome of the p-th subject. The features in $\mathbf{X}_p$ denote the connections

[1] http://en.wikipedia.org/wiki/List_of_neuroscience_databases

between regions of interest (ROIs) in a brain. The values of these features derived from fMRI describe the functional correlations between ROIs, while those from DTI describe the structural connectivities (in terms of the number of connecting fibers) between ROIs. $\mathbf{y} = \{y_1, ..., y_N\}$ denotes the class labels, where $y_p \in \{-1, +1\}$ is the binary class label of $\mathbf{X}_p$. $+1$ means a brain is abnormal with a neurological disorder while -1 means a normal brain connectome.

Suppose we have K side views associated with a brain connectome. Given the p-th brain connectome $\mathbf{X}_p \in \mathcal{X}$, we use $Z^p = \{z_1^p, ..., z_K^p\}$ to represent the side view information. The k-th side view z_k^p contains a set of features $z_k^p = \{z_{k1}^p, ..., z_{kD_k}^p\}$, where D_k is the dimension of this view. We denote the multiple side views for $\mathcal{X}$ as $\mathcal{Z} = \{Z^1, ..., Z^N\}$.

Given the brain connectomes $\mathcal{X}$, our goal is to learn deep latent representations $\mathcal{A} = \{\mathbf{A}_1, ..., \mathbf{A}_N\}$ by exploring multiple side views. $\mathbf{A}_p = \{a_{p1}, ..., a_{pT}\}$ is the set of the hidden features for the p-th subject, where each hidden feature can represent a non-linear combination of brain region connections and T is the number of hidden features. From $\mathcal{A}$, we can discover various deep connectivity patterns of brain connectome.

Although taking the auxiliary information into consideration is promising, how to use the tremendous amount of side views and whether they are useful are unclear. In the following section, we will introduce a way of exploiting the multiple side views and conduct a data analysis to verify whether valuable information can be extracted to guide the process of deep learning.

3 Analysis of Multi-side-view Guidance

Brain connectomes are represented in complex graph structures while the multiple side views exist in the form of vector representations. This introduces the problem of how to leverage the relationship between the graph structures and multiple side views, and how to facilitate deep learning by exploring the vector-based auxiliary views. A simple way to utilize the side information is to concatenate it with brain connectomes. However, such transformation would fail to leverage the underlying correlations between different views. Moreover, direct employment of deep learning methods on such concatenation would discover non-linear features that are mixtures of both brain connectivities and side view features, making the learned features less interpretable for clinical researchers.

In this paper, we take advantage of the pairwise relationships between instances derived from the multiple side views to guide the process of deep learning. Similarity is typically an important type of pairwise relationship that measures the relatedness between any two data instances. In this section, we investigate the consistency of pairwise similarities between auxiliary views and pre-specified label information. Our basic assumption is that, if two instances are similar under all side views, they should have higher probability to be with the same class label (both positive or negative). We form a two-sample one-tail t-test to validate the above hypothesis in both fMRI and DTI datasets.

Table 2: The Max P-values of Hypothesis Testing for Different View Combinations

		Datasets	
#views	#combinations	fMRI dataset	DTI dataset
1	7	1.2700e-02	1.0400e-02
2	21	1.3000e-03	4.9674e-05
3	35	1.5999e-04	1.7004e-06
4	35	9.8788e-06	2.6410e-08
5	21	2.0746e-09	1.2150e-12
6	7	5.0407e-14	1.6554e-19
7	1	4.1716e-28	3.1066e-33

Give two data instances $\mathbf{X}_p$ and $\mathbf{X}_q$ and $\mathcal{K}$ ($1 \leq \mathcal{K} \leq K$) side views, we denote s_{pq}^k as the pairwise similarity from the perspective of the k-th view. s_{pq}^k can be calculated according to z_k^p and z_k^q. For example, we can utilize the radial basis function (RBF) kernel to calculate the pairwise similarity:

$$s_{pq}^k = \exp(\gamma \| z_k^p - z_k^q \|^2),$$

where $\gamma = -\frac{1}{D_k}$. The similarities from multiple views can be simply multiplied to denote the overall pairwise similarity s_{pq} over $\mathcal{K}$ views: $s_{pq} = \prod_{k=1}^{\mathcal{K}} s_{pq}^k$. The resulting similarity will be high only if all of the base similarities are high.

The pairwise similarity set over $\mathcal{K}$ views can be denoted as $\mathcal{S}$. In order to conduct a two-sample one-tail t-test, we split $\mathcal{S}$ into two sets $\mathcal{S}_{sim} = \{s_{pq} | y_p y_q = 1\}$ and $\mathcal{S}_{dis} = \{s_{pq} | y_p y_q = -1\}$. We sample an equal number of elements from each set, denoted as $\mathbf{S}_{sim}$ and $\mathbf{S}_{dis}$, and test whether there is sufficient evidence to support the hypothesis, i.e., the pairwise similarity value in $\mathbf{S}_{sim}$ is larger than that in $\mathbf{S}_{dis}$. The null hypothesis is $H_0 : \mu_{sim} - \mu_{dis} \leq 0$, and the alternative hypothesis is $H_1 : \mu_{sim} - \mu_{dis} > 0$. Here μ_{sim} and μ_{dis} are the sample means of $\mathbf{S}_{sim}$ and $\mathbf{S}_{dis}$, respectively.

We conduct a t-test on every combination of $\mathcal{K}$ side views and present the maximum p-values in Table 2. The results validate the hypothesis that the pairwise similarity under multiple side views is clearly correlated with that in the label space, with significance level $\alpha = 0.05$. Further, we observe that, as the number of side views increases, the likelihood of having the same label becomes increasingly stronger. This preliminary study verifies the usefulness of multiple side views and also paves the way for the design of the MAVE architecture.

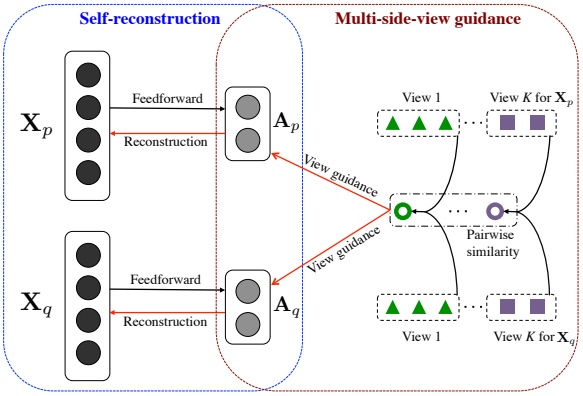


Figure 1: The procedure of MVAE. $\mathbf{X}_p$ and $\mathbf{X}_q$ represent two brain connectomes. $\mathbf{A}_p$ and $\mathbf{A}_q$ are deep hidden features learned by MVAE.

4 Methodology

In this section, we first introduce the basic model of the proposed MVAE with a shallow architecture (one hidden layer) and the optimization algorithm. After that, we describe how to stack MVAE to form a deep architecture, where higher-level features can be learned.

4.1 MVAE Framework Before proceeding, we first illustrate an example of the procedure of MVAE on two brain connectomes ($\mathbf{X}_p$ and $\mathbf{X}_q$) associated with their multiple side views in Figure 1. The pairwise similarities of the two brain connectomes are first extracted from multiple side views and then encapsulated into MVAE to help learn the hidden features $\mathbf{A}_p$ and $\mathbf{A}_q$. Each hidden feature in $\mathbf{A}_p$ (or $\mathbf{A}_q$) represents a non-linear combination of brain region connections. Hence, MVAE can discover various components of disconnected subgraph patterns for the brain connectome from $\mathbf{A}_p$ (or $\mathbf{A}_q$), while traditional subgraph pattern mining approaches cannot. Such patterns can help clinical neuroscientists understand disrupted connectivities.

In order to learn non-linear representations of connectivity patterns, we adopt the classic autoencoder (AE) using the self-reconstruction criterion [1]. Autoencoder is an unsupervised feedforward neural network that applies backpropagation by fitting the input using the reconstructed output. It is often used to reduce high-dimensional features and pre-train deep learning models.

The basic AE has three layers: an input layer, a hidden layer and an output layer. For simplicity, we use $\mathbf{A}^{(l)}$ and h_l to denote features (neurons) and the number of features in layer l, respectively. Suppose we have L layers in total. Here $L = 3$. Given the p-th data instance $\mathbf{X}_p = \{x_1^p, ..., x_M^p\}$, $\mathbf{A}_p^{(1)} = \mathbf{X}_p$ is the input layer (i.e., layer 1) and $h_1 = M$. We can learn a hidden representation $\mathbf{A}_p^{(2)} = \{a_{p1}^{(2)}, ..., a_{ph_2}^{(2)}\}$ by a feedforward propagation, where $h_2 = T$ is the number of features to be learned. Each feature $a_{pi}^{(2)}$ in $\mathbf{A}_p^{(2)}$ represents a non-linear combination of brain region connections for $\mathbf{X}_p$, from which we can get a connectivity pattern for clinical neurology research. After the feedforward pass, $\mathbf{A}_p^{(2)}$ is mapped back to a reconstruction $\hat{\mathbf{X}}_p = \{\hat{x}_1^p, ..., \hat{x}_M^p\}$. We denote the output as $\mathbf{A}_p^{(L)} = \hat{\mathbf{X}}_p$, where $h_L = h_1$. In general, $\mathbf{A}_p^{(l+1)}$ is computed as: $\mathbf{A}_p^{(l+1)} = f(\mathbf{W}^{(l)}\mathbf{A}_p^{(l)} + \mathbf{b}^{(l)})$, where $\mathbf{W}^{(l)} \in \mathbb{R}^{h_{(l+1)} \times h_l}$ is a weight matrix, $\mathbf{b}^{(l)} \in \mathbb{R}^{h_{(l+1)}}$ is a hidden bias vector, and $f(\cdot)$ is called the activation function, e.g., the sigmoid function $f(\tau) = \frac{1}{1+\exp(-\tau)}$. Each element $w_{ij}^{(l)} \in \mathbf{W}^{(l)}$ is a parameter associated with the connection between $a_{pj}^{(l)}$ and $a_{pi}^{(l+1)}$. Each element $b_i^{(l)} \in \mathbf{b}^{(l)}$ is the bias term associated with $a_{pi}^{(l+1)}$. Here each hidden neuron $a_{pi}^{(2)}$ corresponds to a non-linear combination of brain connectivities for $\mathbf{X}_p$ and each weight $w_{ij}^{(1)}$ shows the strength (contribution) of the connectivity $x_{pj} \in \mathbf{X}_p$ to the combination.

Given N data instances $\mathcal{X} = \{\mathbf{X}_1, ..., \mathbf{X}_N\}$, the overall cost function of the classic AE is:

$$\mathcal{J}_{AE}(\mathbf{W}, \mathbf{b}) = \frac{1}{N}\sum_{p=1}^{N}\frac{1}{2}\|\hat{\mathbf{X}}_p - \mathbf{X}_p\|^2 + \frac{\lambda}{2}\sum_{l=1}^{L-1}\|\mathbf{W}^{(l)}\|^2,$$

where the first term is the average of reconstruction error on dataset $\mathcal{X}$ using sum-of-squares. The second term is a regularization term (also called a weight decay term) to prevent over-fitting. λ is the weight decay parameter.

In order to reduce noise, we add a third term $\beta\sum_{j=1}^{h_2} KL(\rho\|\hat{\rho}_j)$ as a sparsity term on the hidden layer. Here $KL(\rho\|\hat{\rho}_j)$ is the Kullback-Leibler (KL) divergence between two Bernoulli random variables with mean ρ and $\hat{\rho}_j$, respectively. It aims to control the sparsity of the weight and bias parameters $\mathbf{W}$ and $\mathbf{b}$. ρ is the sparsity parameter that specifies the level of sparsity. β is the weight of the sparsity term in the overall cost function. We use $KL(\rho\|\hat{\rho}_j) = \rho\log\frac{\rho}{\hat{\rho}_j} + (1-\rho)\log\frac{1-\rho}{1-\hat{\rho}_j}$ to penalize $\hat{\rho}_j$ deviating significantly from the sparsity parameter ρ. $\hat{\rho}_j = \frac{1}{N}\sum_{p=1}^{N} a_{pj}^{(2)}$ is the average activation of the j-th hidden unit.

Since there are limited subjects due to the high cost of clinical experiments, it is nontrivial to directly apply AE to brain connectomes. Therefore, we take advantage of the auxiliary information associated with brain connectomes (i.e., multiple side views) to address the issue of data scarcity. Our aim is to encapsulate multiple side views into the overall cost function of AE so that the pairwise relationship of the learned features is consistent with that in the space of side

views. Specifically, if two instances $\mathbf{X}_p$ and $\mathbf{X}_q$ are similar under multiple side views (i.e., s_{pq} is high), we hope the hidden representations $\mathbf{A}_p^{(l)}$ and $\mathbf{A}_q^{(l)}$ in layer l are also close to each other. On the other hand, if two instances are dissimilar under side views, $\mathbf{A}_p^{(l)}$ and $\mathbf{A}_q^{(l)}$ should be far away from each other. Since the basic AE has only one hidden layer, we use $\mathbf{A}_p$ to denote $\mathbf{A}_p^{(2)}$ for simplicity.

Given the cost function of AE with a sparsity constraint, we propose to add the fourth term as a multi-side-view guidance term and the goal of the proposed MVAE framework is to minimize the following objective function:

$$\mathcal{J}(\mathbf{W},\mathbf{b}) = \frac{1}{N}\sum_{p=1}^{N}\frac{1}{2}\|\hat{\mathbf{X}}_p - \mathbf{X}_p\|^2 + \frac{\lambda}{2}\sum_{l=1}^{L-1}\|\mathbf{W}^{(l)}\|^2 + \beta\sum_{j=1}^{h_2} KL(\rho\|\hat{\rho}_j) + \frac{\mu}{2}\sum_{p=1}^{N}\sum_{q=p+1}^{N} s_{pq}\|\mathbf{A}_p - \mathbf{A}_q\|^2.$$

The multi-side-view guidance term is designed to help the learned hidden features $\mathcal{A}$ follow the geometric of the side view information as in [20]. Here s_{pq} is employed to control the pairwise distance of the learned hidden features. A higher s_{pq} indicates that two instances are similar under multiple side views. By minimizing the multi-side-view guidance term, $\mathbf{A}_p$ and $\mathbf{A}_q$ would become closer to each other. Conversely, a lower s_{pq} will help $\mathbf{A}_p$ and $\mathbf{A}_q$ be far away from each other if we minimize the multi-side-view guidance term. In $\mathcal{J}(\mathbf{W},\mathbf{b})$, μ is a guiding parameter to control the influence of multiple side views. The objective of the proposed MVAE framework is to minimize $\mathcal{J}(\mathbf{W},\mathbf{b})$ with respect to the parameters $\phi = \{\mathbf{W},\mathbf{b}\}$.

4.2 Optimization Algorithm In order to solve the objective function of MVAE, we apply the batch gradient descent algorithm to learn $\mathbf{W}$ and $\mathbf{b}$. During each iteration, the parameters are updated as follows: $w_{ij}^{(l)} = w_{ij}^{(l)} - \alpha\frac{\partial}{\partial w_{ij}^{(l)}}\mathcal{J}(\mathbf{W},\mathbf{b})$ and $b_i^{(l)} = b_i^{(l)} - \alpha\frac{\partial}{\partial b_i^{(l)}}\mathcal{J}(\mathbf{W},\mathbf{b})$, where α is the learning rate.

To compute the partial derivatives, we apply the backpropagation algorithm. The basic idea is to compute an error term for each neuron after the feedforward step. The error term measures how much a neuron is responsible for any errors in the output. Given a data instance $\mathbf{X}_p = \{x_1^p, ..., x_M^p\}$, we denote the error term $\Delta_i^{(l)}(p)$ for the i-th neuron in the l-th layer. For the output layer (i.e., $l = L$), the error terms can be directly measured by the reconstruction error: $\Delta_i^{(L)}(p) = \frac{\partial}{\partial g_{pi}^{(L)}}\left(\frac{1}{2}\parallel x_i^p - \hat{x}_i^p \parallel^2\right) = (\hat{x}_i^p - x_i^p)\cdot\hat{x}_i^p\cdot(1-\hat{x}_i^p)$, where

Algorithm 1 The MVAE algorithm

Input: A set of brain connectomes $\mathcal{X} = \{\mathbf{X}_1, ..., \mathbf{X}_N\}$ and multiple side views $\mathcal{Z} = \{Z^1, ..., Z^N\}$ associated with $\mathcal{X}$, learning rate α, weight decaying parameter λ, sparsity parameter ρ, weight of the sparsity term β, and multi-side-view guidance parameter μ.

Output: An optimized parameter set $\phi = \{\mathbf{W},\mathbf{b}\}$.

```
Calculate pairwise similarities for X under each side view
Multiply the similarities under different side views together
Initialize φ = {W, b} randomly
while NOT converged do
    Perform a feedforward pass on X
    Perform the backpropagation on X
    Update φ
end while
```

$g_{pi}^{(l)}$ is the i-th element in $\mathbf{g}_p^{(l)} = \mathbf{W}^{(l-1)}\mathbf{A}_p^{(l-1)} + \mathbf{b}^{(l-1)}$. Specifically, $g_{pi}^{(l)} = \sum_{j=1}^{h_{(l-1)}} w_{ij}^{(l-1)} a_{pj}^{(l-1)} + b_i^{(l-1)}$.

For the error term $\Delta_i^{(l)}(p)$ of the hidden neuron i in the l-th layer, we can compute it according to the weighted average of the error terms of neurons that use $a_{pi}^{(l)}$ as an input:

$$\Delta_i^{(l)}(p) = \left[\sum_{j=1}^{h_{(l+1)}} w_{ji}^{(l)}\Delta_j^{(l+1)}(p) + \beta\left(-\frac{\rho}{\hat{\rho}_i} + \frac{1-\rho}{1-\hat{\rho}_i}\right) + \mu N\sum_{q=1,q\neq p}^{N} s_{pq}(a_{pi}^{(l)} - a_{qi}^{(l)})\right]\cdot a_{pi}^{(l)}\cdot(1-a_{pi}^{(l)}).$$

With the error terms, we can rewrite the iterative update of parameters: $w_{ij}^{(l)} = w_{ij}^{(l)} - \alpha\left[\left(\frac{1}{N}\sum_{p=1}^{N} a_{pj}^{(l)}\Delta_i^{(l+1)}(p)\right) + \lambda w_{ij}^{(l)}\right]$ and $b_i^{(l)} = b_i^{(l)} - \alpha\left(\frac{1}{N}\sum_{p=1}^{N}\Delta_i^{(l+1)}(p)\right)$.

The overall procedure of our proposed MVAE is summarized in Algorithm 1.

4.3 Building Deep Architectures In this section, we stack the above shallow MVAE to build a deep architecture in a similar way of stacking autoencoders in [2]. The stacked MVAE applies a two-step procedure: an unsupervised pre-training and a supervised fine-tuning, to achieve effective performances. In the unsupervised pre-training step, a greedy layer-wise strategy is employed to train each layer as a MVAE as shown in Figure 2. We first learn the parameters for the first hidden layer according to Algorithm 1, and then take the output of the first hidden layer as the input of the second hidden layer. Therefore, the parameters of the second hidden layer can be optimized in a similar way by Algorithm 1. We repeat such training for subsequent hidden layers in our proposed stacked MVAE model. This greedy layer-wise method has been proven to be very useful in pre-

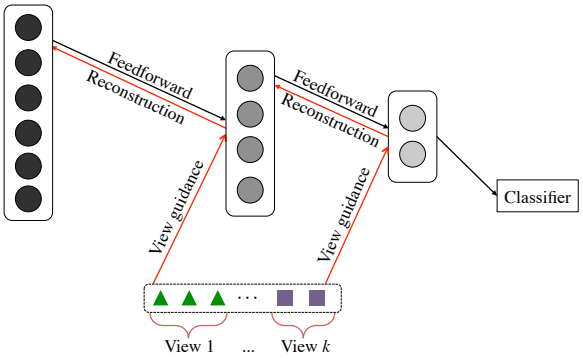


Figure 2: An example of MVAE-2.

training [7]. Unlike the pre-training of stacked autoencoder, in our model, we take advantage of multiple side views during the training of each layer. In other words, the information extracted from multiple side views helps guide the next layer as well as the current layer to learn accurate pairwise relationships between data instances. In this paper, we denote the stacked MVAE (or AE) with n hidden layers as MVAE-n (or AE-n).

Once the stacked MVAE is built, hidden representations in the top-level layer can be used as input to any supervised learning algorithm as shown in Figure 2. In our proposed stacked MVAE, we add a logistic regression model on top of the last layer, thereby yielding an unsupervised deep model to supervised learning. User-provided labels are considered as supervised information in this supervised fine-tuning step and a supervised backpropagation technique is applied to fine tune all the parameters of the whole system.

5 Experiments

In this section, we conduct extensive experiments to evaluate the proposed MVAE on real-world fMRI and DTI datasets as summarized in Section 2.1.

5.1 Compared Methods In order to demonstrate the usefulness of the connectivity patterns identified by the proposed MVAE, we compare our method with several state-of-the-art models that aim at selecting discriminative subgraph patterns. We summarize the compared methods as follows:

• gSSC: It is a semi-supervised subgraph feature selection method [12] based upon both labeled and unlabeled graphs.

• Frequent Subgraphs (Freq): In this approach, the top-k frequent subgraph features are selected. It is totally unsupervised.

• Discriminative Subgraphs (Conf, Ratio, Gtest, HSIC): They are supervised subgraph selection methods [13] based upon confidence, frequency ratio, G-test score and HSIC, respectively. The top-k discriminative subgraph features are selected in terms of different discrimination criteria.

• AE-1: It is **A**uto**E**ncoder (AE) with one hidden layer.

All the above baseline methods ignore the usefulness of side views in learning discriminative patterns from brain connectomes. The following two approaches use multiple side views to facilitate learning connectivity patterns.

• gMSV: It is a subgraph selection method using multiple side views [5].

• MVAE-1: It is the proposed MVAE method with one hidden layer.

In order to test the usefulness of the connectivity patterns, we feed the learned features to a logistic regression classifier for each compared method. For the subgraph mining models, we select the top 100 features as the input of the classifier. For a fair comparison, we also set 100 as the number of hidden neurons (features) for AE-1 and MVAE-1. The guidance parameter μ for MVAE-1 is set as 10^{-2} and 10^{-3} for fMRI and DTI datasets, respectively. Due to the limited number of subjects in our experiments, we perform 3-fold cross validations as in [4] on balanced datasets and report the average results.

5.2 Performance for Detection of Brain Disorders In this subsection, we study the effectiveness of the learned connectivity patterns by feeding them to a binary classifier. Tables 3 presents the average classification performances with respect to four evaluation metrics: accuracy, precision, recall and F1-measure. It can be observed that MVAE-1 consistently outperforms other baseline methods on every metric for both fMRI and DTI datasets.

In particular, among the baselines without side information, AE-1 can achieve similar performances with the subgraph mining approaches. Therefore, the connectivity patterns discovered by deep learning approaches are helpful in detecting brain disorders. Our proposed MVAE-1 outperforms AE-1 because effective guidance can be extracted from multiple side views to help discover better connectivity patterns. Furthermore, compared with the baseline method gMSV that considers side views, the proposed MVAE-1 can again have better performances on both datasets. Though the side view information is leveraged in the subgraph mining approach gMSV, its search space of candidate patterns is limited to connected subgraphs. In contrast, MVAE-1 is able to identify more complex connectivity patterns. It supports our premise that more discriminative connectivity patterns can be identified by applying

Table 3: Performances of the compared methods. The results are reported as "average performance + (rank)". "↑" indicates the larger the value the better the performance.

(a) Results on the fMRI dataset.

Groups	Methods	Criteria			
		Accuracy ↑	Precision ↑	Recall ↑	F1 ↑
No Side-view Guidance	gSSC	0.600 (3)	0.655 (2)	0.605 (3)	0.625 (3)
	Freq	0.543 (6)	0.591 (7)	0.579 (5)	0.582 (8)
	Conf	0.586 (5)	0.609 (6)	0.684 (2)	0.642 (2)
	Ratio	0.543 (6)	0.569 (8)	0.684 (2)	0.620 (4)
	Gtest	0.500 (7)	0.543 (9)	0.524 (7)	0.525 (9)
	HSIC	0.587 (4)	0.637 (4)	0.598 (4)	0.595 (7)
	AE-1	0.629 (2)	0.645 (3)	0.576 (6)	0.604 (6)
Side-view Guidance	gMSV	0.587 (4)	0.627 (5)	0.605 (3)	0.614 (5)
	MVAE-1	0.729 (1)	0.740 (1)	0.715 (1)	0.726 (1)

(b) Results on the DTI dataset.

Groups	Methods	Criteria			
		Accuracy ↑	Precision ↑	Recall ↑	F1 ↑
No Side-view Guidance	gSSC	0.592 (4)	0.644 (3)	0.560 (4)	0.596 (3)
	Freq	0.646 (2)	0.703 (2)	0.588 (2)	0.639 (2)
	Conf	0.524 (7)	0.506 (9)	0.430 (7)	0.461 (8)
	Ratio	0.593 (3)	0.638 (4)	0.535 (6)	0.579 (6)
	Gtest	0.593 (3)	0.629 (5)	0.559 (5)	0.585 (4)
	HSIC	0.579 (5)	0.617 (6)	0.560 (4)	0.583 (5)
	AE-1	0.578 (6)	0.573 (8)	0.575 (3)	0.568 (7)
Side-view Guidance	gMSV	0.513 (8)	0.576 (7)	0.412 (8)	0.460 (9)
	MVAE-1	0.698 (1)	0.729 (1)	0.628 (1)	0.670 (1)

deep learning approaches.

In summary, with multi-side-view guidance, the proposed MVAE-1 outperforms the baseline methods with an average improvement of 28% and 22% on accuracy for fMRI and DTI, respectively.

5.3 Performance with Deep Architectures The hidden features learned from the shallow version of MVAE (i.e., MVAE-1) have potential clinical significance for detection of disrupted connectivities. In this section, we study the usefulness of the hidden features learned from deeper architectures, i.e., MVAE-2 and MVAE-3 in our experiments. After learning the hidden features in the top-level layer for MVAE-1, MVAE-2 and MVAE-3, we feed the learned hidden features to a logistic regression classifier as shown in Figure 2. In the experiments, we constrain the number of neurons between 50 and 500 for each layer. Table 4 presents the best classification performances of MVAE-1, MVAE-2 and MVAE-3 for fMRI and DTI datasets. It shows that MVAE-2 performs best on both datasets. When the number of hidden layers increases from 1 to 2, the classification results also improve. This indicates that a deep architecture is more effective than a shallow one. However, with 3 hidden layers, MVAE-3 does not perform as well, which may be due to overfitting, because there are many parameters in MVAE-3 but available training data is limited. In addition, we constrained the number of neurons between 50 and 500 for each hidden layer in MVAE-3. Searching in a larger parameter space may achieve better results for MVAE-3. However, there are associated time costs. Therefore, in the following, we only show experimental results for MVAE-2.

5.4 Performance with Each Side View In this section, we investigate the guidance contribution of each side view. Figure 3 shows the accuracy performance for fMRI and DTI datasets by considering a single side view each time in MVAE-2. Specifically, the side views *brain volumetry* and *plasma luminex* provide the most valuable side information for learning latent features from

Table 4: Results of the stacked MVAE with different numbers of hidden layers.

Datasets	Methods	Criteria			
		Accuracy	Precision	Recall	F1
fMRI	MVAE-1	0.727	0.727	0.770	0.742
	MVAE-2	**0.772**	**0.761**	**0.831**	**0.786**
	MVAE-3	0.745	0.760	0.775	0.756
DTI	MVAE-1	0.710	0.729	0.684	0.704
	MVAE-2	**0.736**	**0.755**	**0.735**	**0.737**
	MVAE-3	0.670	0.644	0.761	0.697

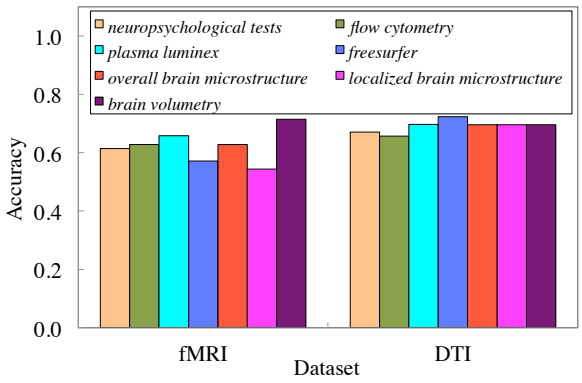


Figure 3: Accuracy performances for each single view.

deep architectures on the fMRI brain connectomes. The *brain volumetry* side view provides critical information concerning structural meningoencephalitic changes subsequent to brain viral invasion in HIV infection. Similarly, the *plasma luminex* side view, which reflects circulating levels of cytokines and chemokines, provides information concerning the inflammatory milieu in an individual subject and in this real world example, captures critical information concerning the immune perturbances and cytokine storm characteristic of early HIV infection. Inflammation and chronic immune activation have been implicated in brain injury in various neurological disorders, nevertheless available blood assay data is usually ignored in imaging studies due to the challenges of integrating the side view information.

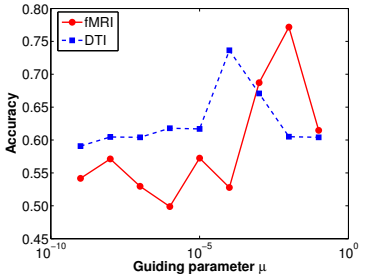
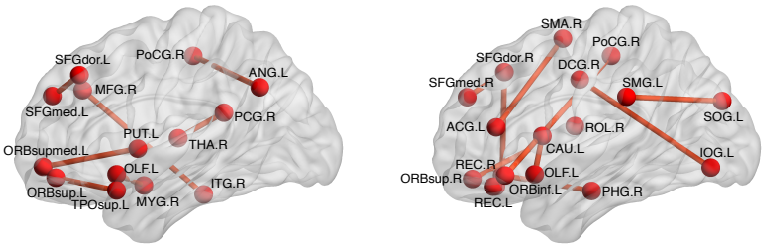


Figure 4: The guiding ability of MVAE-2.



(a) The first pattern (b) The second pattern

Figure 5: Two most discriminative connectivity patterns.

5.5 Effect of the Guiding Parameter μ In our proposed MVAE framework, the parameter μ controls the relative importance of the multi-side-view guidance. Here we assess the benefit of MVAE-2 with different values of parameter $\mu \in \{10^{-9}, 10^{-8}, ..., 10^{-1}\}$. Figure 4 shows the classification accuracies on fMRI and DTI datasets. It can be observed that for both datasets, the performance improves when increasing the guiding parameter μ. The best accuracy is achieved when $\mu = 10^{-2}$ for fMRI and $\mu = 10^{-4}$ for DTI. Similar results can be observed for the effect of other parameters. Due to space limits, we will not show the performances here.

5.6 Visualization of Connectivity Patterns Recall that MVAE can discover discriminative connectivity patterns to investigate brain dysfunction. Here we visualize discovered patterns in the fMRI dataset.

Since each hidden neuron $a_{pi}^{(2)}$ in the first hidden layer corresponds to a discriminative pattern of brain connectivities for $\mathbf{X}_p$ and each weight $w_{ij}^{(1)}$ shows the strength of the connectivity $x_{pj} \in \mathbf{X}_p$, MVAE can learn many different connectivity patterns from different hidden neurons. Each pattern demonstrates its significance in distinguishing normal connectivities from disrupted ones. Given a group of normal brain connectomes, we can get the average activation (hidden feature value) for each neuron in the first hidden layer. Similarly, we can calculate such values for abnormal brain connectomes. For each hidden neuron, the absolute difference between the within-group averaged activation values indicates its significance in distinguishing the two groups. So we select two representative neurons with the largest differences between two groups and display their corresponding connectivity patterns in Figure 5 (a) and (b). The most significant connectivity patterns identified from the fMRI brain connectomes are visualized with the BrainNet Viewer[2] [23].

We can observe that these two complex patterns are different, each providing information concerning connectivity changes occurring early in HIV infection. For example, the identified connection between Postcentral gyrus (PoCG) and Angular gyrus (ANG) in Figure 5 (a) is consistent with early parietal involvement with implications for memory function and spatial cognition. Further, these results suggest early changes in key brain hubs, such as putamen and thalamus that are involved in the default mode network. While the observed patterns are beyond the scope of this paper, these connectome results yield new clinical insights concerning brain alterations that occur with threats such as viral invasion, attendant meningoencephalitic changes, neuroinammation and immune activation. The rich information conveyed from connectivity patterns supports the promise of the MVAE framework.

6 Related Work

Our work is related to deep learning techniques, multi-view learning problems, and brain data analysis. We briefly discuss them in the followings.

Learning feature representations from deep learning models has rapidly developed over the past few years [14, 18, 26]. Deep learning has also been studied in the diagnosis of brain disease [3, 15, 16, 19]. However, all these existing methods focus on neuroimages and ignore rich information from multiple side views. A recent work by Zhang et al. [26] is worth mentioning in this context, although it does not perform deep learning on brain disorders directly. They extract pairwise constraints from labeled data to supervise the pre-training process of deep learning. Our study is different since we take advantage of side views and do not use any supervision information in our pre-training procedure.

There are a few research studies on multi-view features using deep learning methods [10, 11, 22]. In [10, 11], in order to learn individual and shared hidden spaces from multiple views of data, each hidden layer is composed of view-specific and shared neurons. The method in [22] learns representations of data in the

[2] http://www.nitrc.org/projects/bnv

setting that multiple views are known during training while only one view is available at test time. In our work, we consider the multiple side views as auxiliary information to guide the process of deep learning on the primary brain connectome view, which differs from the basic idea in previous studies.

The analysis of brain data is also related to our work. In [6, 8], the neuroimage data are directly mined and represented as tensors to handle extremely high dimensionality within the data. In addition, many efforts have focused on mining important brain regions and estimating region/neuron connections [9, 21, 25]. After the construction of brain connectomes, subgraph patterns are learned in [5, 13]. However, these works can only find connected subgraph patterns while our study can discover non-linear representations of brain connectivity patterns.

7 Conclusion

In this paper, we propose a deep architecture to learn clinically significant connectivity patterns of brain connectomes using multiple side views. This helps the diagnosis of neurological disorders. We show that by leveraging a plurality of side views that are available along with the brain connectomes, the proposed method MVAE can not only capture discriminative connectivity patterns for classification tasks, but also discover meaningful information for clinical interpretations.

There are several interesting directions for future work. Since brain connectomes and neuroimages can provide complementary information for brain diseases, one interesting direction of our future work is to explore both brain connectomes and neuroimages in deep learning methods. Another potential direction is to combine fMRI and DTI brain connectomes together, because the functional and structural connections together can provide rich information for learning deep feature representations.

Acknowledgement

This work is supported in part by NSF through grants III-1526499, CNS-1115234, and OISE-1129076, and Google Research Award.

References

[1] *Autoencoders and Sparsity.* `http://ufldl.stanford.edu/wiki/index.php/Autoencoders_and_Sparsity`.

[2] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle. Greedy layer-wise training of deep networks. *NIPS*, 19:153, 2007.

[3] T. Brosch and R. Tam. Manifold learning of brain mris by deep learning. In *MICCAI*. 2013.

[4] B. Cao, L. He, X. Kong, P. Yu, Z. Hao, and A. Ragin. Tensor-based multi-view feature selection with applications to brain diseases. In *ICDM*, 2014.

[5] B. Cao, X. Kong, J. Zhang, P. Yu, and A. Ragin. Mining brain networks using multiple side views for neurological disorder identification. In *ICDM*, 2015.

[6] I. Davidson, S. Gilpin, O. Carmichael, and P. Walker. Network discovery via constrained tensor analysis of fmri data. In *KDD*, 2013.

[7] D. Erhan, Y. Bengio, A. Courville, P. Manzagol, P. Vincent, and S. Bengio. Why does unsupervised pre-training help deep learning? *JMLR*, 11, 2010.

[8] L. He, X. Kong, P. Yu, A. Ragin, Z. Hao, and X. Yang. Dusk: A dual structure-preserving kernel for supervised tensor learning with applications to neuroimages. In *SDM*, 2014.

[9] S. Huang, J. Li, J. Ye, A. Fleisher, K. Chen, T. Wu, and E. Reiman. Brain effective connectivity modeling for alzheimer's disease by sparse gaussian bayesian network. In *KDD*, 2011.

[10] Y. Kang and S. Choi. Restricted deep belief networks for multi-view learning. In *ECMLPKDD*. 2011.

[11] Y. Kang, S. Kim, and S. Choi. Deep learning to hash with multiple representations. In *ICDM*, 2012.

[12] X. Kong and P. Yu. Semi-supervised feature selection for graph classification. In *KDD*, 2010.

[13] X. Kong, P. Yu, X. Wang, and A. Ragin. Discriminative feature selection for uncertain graph classification. In *SDM*, 2013.

[14] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012.

[15] R. Li, W. Zhang, H. Suk, L. Wang, J. Li, D. Shen, and S. Ji. Deep learning based imaging data completion for improved brain disease diagnosis. In *MICCAI*. 2014.

[16] S. Plis, D. Hjelm, R. Salakhutdinov, E. Allen, H. Bockholt, J. Long, H. Johnson, J. Paulsen, J. Turner, and V. Calhoun. Deep learning for neuroimaging: a validation study. *Frontiers in neuroscience*, 8, 2014.

[17] A. Ragin, H. Du, R. Ochs, Y. Wu, C. Sammet, A. Shoukry, and L. Epstein. Structural brain alterations can be detected early in hiv infection. *Neurology*, 79(24), 2012.

[18] R. Socher, J. Pennington, E. Huang, A. Ng, and C. Manning. Semi-supervised recursive autoencoders for predicting sentiment distributions. In *EMNLP*, 2011.

[19] H. Suk and D. Shen. Deep learning-based feature representation for ad/mci classification. In *MICCAI*. 2013.

[20] J. Tenenbaum, V. De Silva, and J. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500), 2000.

[21] V. Veeriah, R. Durvasula, and G. J. Qi. Deep learning architecture with dynamically programmed layers for brain connectome prediction. In *KDD*, 2015.

[22] W. Wang, R. Arora, K. Livescu, and J. Bilmes. On deep multi-view representation learning. In *ICML*, 2015.

[23] M. Xia, J. Wang, and Y. He. Brainnet viewer: a network visualization tool for human brain connectomics. *PloS one*, 8(7), 2013.

[24] S. Xiang, L. Yuan, W. Fan, Y. Wang, P. Thompson, and J. Ye. Multi-source learning with block-wise missing data for alzheimer's disease prediction. In *KDD*, 2013.

[25] S. Yang, Q. Sun, S. Ji, P. Wonka, I. Davidson, and J. Ye. Structural graphical lasso for learning mouse brain connectivity. In *KDD*, 2015.

[26] J. Zhang, G. Tian, Y. Mu, and W. Fan. Supervised deep learning with auxiliary networks. In *KDD*, 2014.