

# 深層生成モデルによる fMRI データの患者個人特徴量抽出と それを用いた精神疾患診断

## Modeling a Subject-Wise Variability in fMRI data and Diagnosis of Psychiatric Disorder by Deep Generative Model

松原崇<sup>1\*</sup>                      田代哲生<sup>1</sup>                      上原邦昭<sup>1</sup>  
Takashi Matsubara<sup>1</sup>              Tetsuo Tashiro<sup>1</sup>              Kuniaki Uehara<sup>1</sup>

<sup>1</sup> 神戸大学 大学院システム情報学研究科

<sup>1</sup> Graduate School of System Informatics, Kobe University

**Abstract:** Accurate and early diagnosis of psychiatric disorders is critical for patients' quality of life. Since neuroimaging datasets are still composed of a limited number of high-dimensional samples, they require sophisticated procedures for machine learning-based analyses. Recent studies reported a certain success by applying generative models to fMRI data: Generative models based on appropriate assumptions can build a model of small datasets accurately. Leveraging the flexibility of deep neural networks, we propose a structured deep generative model, which takes into account fMRI images, disorder, and individual variability. The proposed model estimates the subjects' conditions more accurately than existing diagnostic procedures, general discriminative models, and recently-proposed generative models.

## 1 Introduction

脳機能画像を始めとする多様な医用データの蓄積に伴い [1], 機械学習を用いて精神疾患などに関わるバイオマーカーを発見することが期待されている [2]. このようなバイオマーカーの発見は精神疾患の機序を明らかにすることに貢献するだけでなく, 正確な早期診断や定量的な評価を可能にし, 患者のクオリティ・オブ・ライフにも貢献するとともに, 治療の有効性を評価することも可能になる. しかし, 如何にデータの蓄積が進んだとは言えデータセットの大きさは一般画像認識などに比べて遥かに小さいため, 解析には特徴選択や次元削減のような複雑な前処理と手続きが必要になる [3-5]. 一方で, 近年のいくつかの研究は, 脳機能画像データの解析に生成モデルの手法を用いることで一定の成果を上げている [6-8]. Sukら [7] は隠れマルコフモデル (hidden Markov models; HMMs) を用いて, 脳機能画像データの背後にある脳の大規模神経系のダイナミクスのモデル化を試みた. Chenら [6] は刺激への応答と線形の生成モデルでモデル化することで, 脳機能画像データから刺激を推定することに成功した. しかし, これは全被験者に共通のモデルと, 各被験者の個人差を表現するモデルの組み合わせである. その

ため, 新しい被験者に適応するためには個人差の部分を新たに調整する必要がある, 精神疾患の診断に用いることは難しい [9].

近年, 深層学習と呼ばれる深い構造を持った深層ニューラルネットワーク (deep neural networks; DNNs) を用いる手法が注目を浴びている (詳細はレビュー論文 [10] を参考のこと). 深層学習は与えられたデータから, 目的に必要な高次の特徴を自動的に学習することで, 様々なタスクで高い性能を示している. 深層学習には主に教師あり分類を行う多層パーセプトロン (multilayer perceptron; MLP) [4, 11] と, 教師なし次元削減が可能な自己符号化器 (autoencoder; AE) [4, 7, 11] が存在に, すでに脳機能画像データ解析に広く用いられている. また深層学習には深層生成モデル (deep neural generative models; DGMs) と呼ばれる生成モデルを実装できる構造が存在する [12, 13]. 田代ら [8] は深層生成モデルを用い, 脳機能画像データに加え被験者の状態 (疾患の有無) をモデル化し, 高い精度で被験者の疾患を診断することに成功した.

分類に一般的に用いられる識別モデルに比べ, 生成モデルはその構造が一種の拘束条件として働くため, 想定した構造が真の構造に近ければ, 少ないサンプル数のデータセットに対しても, 高い精度で分類が可能であることが示されている [14, 15]. そのため, 脳機能画像データの背後にある構造を深層生成モデル上で実現

\*連絡先: 神戸大学 大学院システム情報学研究科  
兵庫県神戸市灘区六甲台町 1-1  
E-mail: matsubara@phoenix.kobe-u.ac.jp

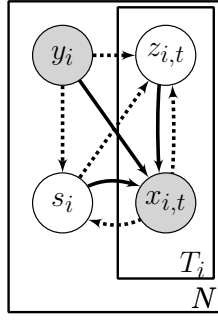


図 1: Our proposed generative model composed of fMRI images  $x_{i,t}$ , a diagnosis  $y_i$ , a subject-wise feature  $s_i$ , and scan-wise variabilities  $z_{i,t}$ .

することで、診断精度の向上とさらなる解析が可能となる。本稿では、脳機能画像データ解析のために個人差をモデル化した深層生成モデルを提案する。個人差は単一の患者から得られたすべての脳機能画像が共有する潜在変数としてモデル化され、それらの脳機能画像から推論する。この推論によって、提案モデルは未知の被験者に対しても汎化することが可能であり、この点において Chen らの研究 [6] よりも優れている。

提案モデルを用いて統合失調症と双極性障害の被験者から得られた安静時脳機能画像をモデル化し、コントロール群と識別することを試みた。結果として、提案モデルはいくつかの既存手法よりも優れた診断精度を示した。既存手法とは脳領域間の機能的結合 (functional connectivity; FC) をピアソンの相関係数 (Pearson correlation coefficients; PCC) でモデル化したものを特徴量として用いる手法 [3, 5], 代表的な識別モデルであるサポートベクターマシン (support vector machine; SVM) [16] と人工ニューラルネットワークの時系列拡張である長期短期記憶 (long short-term memory; LSTM) [17], 生成モデルとして田代らが提案した深層生成モデル [8], Suk らの提案した自己符号化器と隠れマルコフモデルの組み合わせ [7] を用いた。

## 2 提案モデル

### 2.1 提案生成モデル

本節ではまず脳機能画像データに対し、構造化された生成モデルを提案する。深層生成モデルへの実装については次節に回す。脳機能画像データセット  $\mathcal{D} = \{\mathbf{x}_i, y_i\}_{i=1}^N$  を考える。ただし  $N$  は被験者数、 $i$  はその番号であり、 $\mathbf{x}_i$  は被験者  $i$  から得られた脳機能画像の集合、 $y_i$  は被験者  $i$  の診断結果 (つまりクラスラベル) である。 $y = 0$  はコントロール群、 $y = 1$  は疾患を持った患者を意味する。被験者  $i$  からは  $T_i$  枚の脳機能画像が得られていると考え、 $\mathbf{x}_i = \{x_{i,t}\}_{t=1}^{T_i}$  で表す。

各被験者  $i$  には疾患の有無とは別に個人差が存在すると仮定する。例えば脳の形状、抽象的な意味での脳の使い方や癖などが考えられる。脳の形状は前処理に置いて除去されることが期待される要素ではあるが、完全に除去することは実のところ困難である。この個人差は事前分布  $p(s)$  に従う特徴量  $s_i$  で表現され、脳機能画像  $\mathbf{x}_i$  を生成する潜在変数としてモデル化する。各脳機能画像  $x_{i,t}$  はクラスラベル  $y_i$ , 個人差  $s_i$ , そして潜在変数  $z_{i,t}$  に紐付けられる。潜在変数  $z_{i,t}$  は事前分布  $p(z)$  に従い、画像ごとのばらつきを表現する。例えば、データの採取中に被験者の考えていたこと、前処理で除去しきれなかった体動などが相当する [8]。これらを用いて、クラスラベル  $y_i$  を条件とした脳機能画像  $\mathbf{x}_i$  の生成モデル  $p_\theta$  を図 1 のように構築し、以下の式で表す。

$$p_\theta(\mathbf{x}_i|y_i) = \prod_{t=1}^{T_i} p_\theta(x_{i,t}|y_i) \\ = \prod_{t=1}^{T_i} \int_{s_i} \int_{z_{i,t}} p_\theta(x_{i,t}|y_i, z_{i,t}, s_i) p(z_{i,t}) p(s_i).$$

変分推論を用い、モデルエビデンス  $\log p_\theta(\mathbf{x}_i|y_i)$  の下界を推論モデル  $q_\phi$  を用いて以下のように表す。

$$\log p_\theta(\mathbf{x}_i|y_i) \\ \geq \mathbb{E}_{q_\phi(s_i, \mathbf{z}_i|\mathbf{x}_i, y_i)} \left[ \log \frac{p_\theta(\mathbf{x}_i, \mathbf{z}_i, s_i|y_i)}{q_\phi(s_i, \mathbf{z}_i|\mathbf{x}_i, y_i)} \right] \\ = -D_{KL}(q_\phi(s_i|\mathbf{x}_i, y_i)||p(s_i)) \\ + \sum_{t=1}^{T_i} \mathbb{E}_{q_\phi(s_i|\mathbf{x}_i, y_i)} \left[ \begin{aligned} & -D_{KL}(q_\phi(z_{i,t}|\mathbf{x}_{i,t}, y_i, s_i)||p(z_{i,t})) \\ & + \mathbb{E}_{q_\phi(z_{i,t}|\mathbf{x}_{i,t}, y_i, s_i)} [\log p_\theta(x_{i,t}|y_i, z_{i,t}, s_i)] \end{aligned} \right] \\ =: \mathcal{L}_g(\mathbf{x}_i, y_i) \quad (1)$$

ただし、 $D_{KL}(\cdot||\cdot)$  はカルバック=ライブラー・ダイバージェンスであり、 $\mathcal{L}_g(\mathbf{x}_i; y_i)$  はモデルエビデンスの下界 (evidence lower bound; ELBO) である。ELBO は生成モデル  $p_\theta$  と推論モデル  $q_\phi$  の目的関数に当たる。

ELBO  $\mathcal{L}_g(\mathbf{x}_i; y)$  をモデルエビデンス  $\log p_\theta(\mathbf{x}_i|y)$  の近似値として使うことで、被験者  $i$  のクラスラベル  $y$  の事後確率  $p(y|\mathbf{x}_i)$  をベイズの定理を用いて以下のように表現できる。

$$p_\theta(y|\mathbf{x}_i) = \frac{p(y)p_\theta(\mathbf{x}_i|y)}{\sum_{y' \in \{0,1\}} p(y')p_\theta(\mathbf{x}_i|y')} \\ \approx \frac{p(y) \exp \mathcal{L}_g(\mathbf{x}_i, y)}{\sum_{y' \in \{0,1\}} p(y') \exp \mathcal{L}_g(\mathbf{x}_i, y')} \\ =: \exp \mathcal{L}_d(\mathbf{x}_i, y), \quad (2)$$

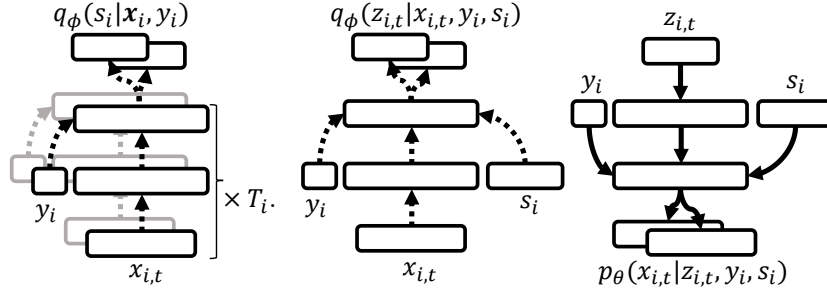


図 2: 提案生成モデルの深層ニューラルネットワーク実装.

ただしクラスラベル  $y$  の事前確率  $p(y)$  は  $p(y = 0) = p(y = 1) = 0.5$  と仮定した. つまり,  $\text{ELBO} \mathcal{L}_g(\mathbf{x}_i, y = 1)$  が大きいほど, 被験者  $i$  は対象となる疾患を持っている可能性が高いと言える.

またクラスラベル  $y_i$  の事前確率の対数  $\log p(y_i | \mathbf{x}_i)$  の近似値  $\mathcal{L}_d(\mathbf{x}_i, y_i)$  も目的関数になり得る. これにより, ふたつのクラスラベルを明確に識別するようにモデルを学習させることができる [14, 18]. 一方で, 単なる識別モデルと化すことを防ぐため, 重みパラメータ  $\omega \in [0, 1]$  を用いてふたつの目的関数を以下のように調整し,  $\mathcal{L}(\mathbf{x}_i, y_i)$  を最終的な目的関数とする.

$$\mathcal{L}(\mathbf{x}_i, y_i) = \omega \mathcal{L}_g(\mathbf{x}_i, y_i) + (1 - \omega) \mathcal{L}_d(\mathbf{x}_i, y_i). \quad (3)$$

## 2.2 深層学習を用いた実装

本節では深層ニューラルネットワークを用い生成モデル  $p_\theta$  と推論モデル  $q_\phi$  を実装することで, 深層生成モデルを構築する方法について述べる. 脳機能画像を前処理して得られた信号  $x_{i,t}$ , 個人差  $s_i$ , 潜在変数  $z_{i,t}$  をそれぞれ  $n_x$  次元,  $n_s$  次元,  $n_z$  次元のベクトルとする.

生成モデル  $q$  は復号化器 (decoder) と呼ばれる単一の深層ニューラルネットワークで構成され, 脳機能画像  $x_{i,t}$  の事後分布  $q_\theta(x_{i,t} | y_i, z_{i,t}, s_i)$  を, 対角行列を分散共分散行列にもつ  $n_z$  次の多変量ガウス分布  $\mathcal{N}(\mu_{x_{i,t}}, \text{diag}(\sigma_{x_{i,t}}))$  で表現する. これには再パラメータ化の手法が用いられる [8, 12, 13, 18]. すなわち推論モデル  $q$  は  $2 \times n_x$  個の出力ユニットを持ち, そのうち  $n_x$  個のユニットは恒等関数を活性化関数として持ち, 事後分布の平均ベクトル  $\mu_{x_{i,t}}$  を表現する. 残る  $n_x$  個のユニットは指数関数を活性化関数として持ち, 事後分布の分散ベクトル  $\sigma_{x_{i,t}}$  を表現する (図 2 右参照). 識別モデル  $q_\phi$  のうち, 潜在変数  $z_{i,t}$  を推論する部分に関しては, 同じく符号化器 (encoder) 単一の深層ニューラルネットワークによって実装され, 再パラメータ化によって事後分布  $q_\phi(z_{i,t} | x_{i,t}, y_i, s_i)$  を表現する (図 2 中央参照). この実装法は先行研究の構造に従っている.

一方, 推論モデル  $q_\phi$  のうち個人差  $s_i$  を推論する事後分布  $q_\phi(s_i | \mathbf{x}_i, y_i)$  については工夫が必要である. 定

義からして, このモデルは被験者  $i$  ごとに, 長さの一定ではない脳機能画像の集合  $\mathbf{x}_i = \{x_{i,t}\}_{t=1}^{T_i}$  を入力として受け取る必要がある. そこで本稿では集団符号化器 (collection-encoder) を提案する. これは図 2 左のようにふたつの人工ニューラルネットワークを重ねたものである. 下のネットワークは 1 枚の脳機能画像  $x_{i,t}$  とクラスラベル  $y_i$  を受け取り, 活性値  $h_{i,t}$  を出力する. 上のネットワークは脳機能画像の集合  $\mathbf{x}_i = \{x_{i,t}\}_{t=1}^{T_i}$  から得られた活性値  $\mathbf{h}_i = \{h_{i,t}\}_{t=1}^{T_i}$  から求められた平均  $\bar{h}_i = \frac{1}{T_i} \sum_{t=1}^{T_i} [h_{i,t}]$  を入力とし, 再パラメータ化を用いて個人差  $s_i$  の事後分布  $q_\phi(s_i | \mathbf{x}_i, y_i)$  を出力する.

いくつかの既存研究においては, 異なる構造の深層生成モデルが提案されている. たとえば Maaløra [13] の Skip 深層生成モデルでは, ひとつのサンプルがそれぞれ固有の複数の潜在変数に依存しているモデルが提案されている. 本稿のモデルは複数のサンプルがひとつの潜在変数を共有しているというモデル化であり, その構造に置いて新規性がある. またこのような構造は脳機能画像が持つ構造を模しており, 適切な拘束条件として働くことが期待できる.

本稿では符号化器と復号化器には三層の人工ニューラルネットワークを, 集団符号化器には二層と一層の人工ニューラルネットワークを積んだものを用いた. 各隠れ層は  $u_h$  個のユニットで構成され, layer normalization [19] による正規化を行い, 活性化関数として ReLU [20] を用いた. 目的関数には式 (1) 中の期待値の計算が必要だが, 訓練時には個人差  $s_i$  と潜在変数  $z_{i,t}$  を訓練の 1 イテレーションごとに 1 回サンプルすることで近似した. また, 検証時にはそれぞれの MAP 推定 (つまり事後分布を表現する平均ベクトル  $\mu_{z_{i,t}}$ ) を用いた. この簡略化は先行研究でも多く採用されており, 精度への大きな悪影響はないことが確認されている [12]. 脳機能画像を前処理して得られた信号  $x_{i,t}$  は確率  $p$  の dropout [21] によってデータ増強した. すべての人工ニューラルネットワークは Adam optimization algorithm [22] (パラメータ  $\alpha = 10^{-4}$ ,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ) を用いて訓練した. またハイパーパラメータは以下の範囲から探索した.  $p \in \{0.0, 0.5\}$ ,

$n_h \in \{50, 100, 200, 400\}$ ,  $n_z = n_s \in \{5, 10, 20, 50, 100\}$  (ただし  $n_h > n_z = n_s$ ),  $\omega \in \{0.0, 0.9, 0.99\}$ . またクラス間の被験者数のインバランスはオーバーサンプルによって調整した.

## 3 実験と結果

### 3.1 使用データ

OpenfMRI データベースより, 統合失調症と双極性障害のデータベースを入手した<sup>1</sup>. SPM12<sup>2</sup> を用い, 以下のように前処理を行った. すなわち, 磁場の安定を確保するために各被験者の最初の 10 枚の脳機能画像を破棄し, 時間方向の調整 (time-slice adjustment), 体動の補正 (realignment of brain positions), そして MNI 座標系を用いた空間正規化 (spatial normalization) を行った. Automated anatomical labeling (AAL) template [23] により, 116 の領域 (Regions-Of-Interest; ROI) に分割し, それぞれで信号強度を平均化し, 116 次元のベクトル  $x_{i,t}$  を得た. データの洗浄として, 1.5mm もしくは 1.5 度以上の体動が検出された瞬間から以降の脳機能画像をすべて破棄した. 破棄した結果, 脳機能画像が 100 枚未満しか残らなかった被験者は被験者のデータ全てを破棄した. また撮像位置のずれなどにより, MNI 座標系に変換できなかった被験者のデータも破棄した. 結果的にコントロール群 113 人, 統合失調症 44 人, 双極性障害 45 人のデータを得た.

### 3.2 比較手法

既存手法として, ROI 間のピアソン相関係数 (PCC) を機能的結合 (FC) として用いた手法を用いる [3, 5]. Shen ら [3] は FC からケンドールのタウ係数 (Kendall  $\tau$  coefficient) を用いて  $m$  個の特徴を選択し, 局所線形埋め込み (locally linear embedding; LLE) により近傍パラメータ  $k$  のもとで  $d$  次元の多様体に埋め込んだ. そして c-means 法でクラスタリングした. この手法は SVM や MLP を用いた手法よりも高性能であることが示されている. ハイパラメータはもとの論文に従って  $m \in \{50, 100, 200, 400, 600\}$ ,  $k \in \{5, 8, 10, 12, 15\}$ ,  $d \in \{2, 5, 10, 20, 50\}$  から選択した.

また, PCC を用いずに前処理済みの脳機能画像を用いる方法として, サポートベクターマシン (support vector machine; SVM) [16] と長期短期記憶 (long short-term memory; LSTM) [17] を用いた. SVM は脳機能画像  $x_{i,t}$  を一枚ずつ受け取り, それぞれに対し二値分

類の結果を返す. 被験者  $i$  の診断には  $T_i$  枚の脳機能画像に対する分類結果の多数決を用いた. ハイパラメータは  $C \in \{\dots, 0.1, 0.2, 0.5, 1, 2, 5, 10, \dots\}$  から選択した. LSTM は再帰構造を持ったニューラルネットワークであり, 一人の被験者から得られた脳機能画像  $\mathbf{x}_i = \{x_{i,t}\}_{t=1}^{T_i}$  を順番に受け取り, クラスの事後確率  $p(y|\mathbf{x}_i)$  をロジスティック関数を用いて出力する. その他の条件は提案モデルと同一である.

加えて, 田代らの提案した深層生成モデル [8] と, Suk らの提案した自己符号化器 (autoencoder; AE) と隠れマルコフモデル (hidden Markov model; HMM) の組み合わせ [7] と比較する. 既存深層生成モデルは脳機能画像  $\mathbf{x}_i$ , クラスラベル  $t_i$ , そして画像ごとの差異  $z_{i,t}$  を, 符号化器  $q(z_{i,t}|\mathbf{x}_i, t_i)$  と復号化器  $p(\mathbf{x}_i|y_i, z_{i,t})$  を用いて実装している. しかし, 被験者の個人差  $s_i$  はモデル化していない [8]. Suk らの研究 [7] に従い, 脳機能画像を AE を用いて  $d$  次元に埋め込み, その上でふたつの HMM を学習させた. ひとつの HMM は患者用 ( $p_\theta(x_{i,t}|y=1)$ ), もうひとつはコントロール群用 ( $p_\theta(x_{i,t}|y=0)$ ) である. それぞれの HMM はガウス分布でモデル化され, 提案モデルと同様にベイズの法則によってクラス事後確率  $p(y|\mathbf{x}_i)$  を計算する. ハイパラメータの探索範囲は以下のとおりである. AE のボトルネック層のユニット数  $n_z \in \{2, 3\}$ , 混合分布のコンポーネント数  $n \in \{2, 3, 4, 5, 6, 7\}$ . AE のその他の条件は提案モデルと同一である.

### 3.3 診断精度

データセットがインバランスであるため, 感度 (sensitivity)  $SEN = TP/(TP + FN)$ , 特異度 (specificity)  $SPEC = TN/(TN + FP)$ , バランスした正解率 (balanced accuracy)  $BACC = 0.5 \times (SEN + SPEC)$  を用いた. ただし TP, TN, FP, FN はそれぞれ true positive, true negative, false positive, false negative を表す. 結果は 10-fold 交差検定 5 階の平均で採取し, 表 1 にまとめた. 提案モデルは統合失調症のデータセットに対して比較手法より明らかに優れた結果を, 双極性障害のデータセットに対してほぼ同等あるいは優れた結果を示した. 特に既存の深層生成モデルよりも高い精度が得られたという点は, 提案モデルに与えた個人差  $s_i$  という構造が適切な拘束条件となっていることを示唆している.

## 4 まとめと今後の課題

本稿では, 脳機能画像データ解析のために個人差をモデル化した深層生成モデルを提案し, 精神疾患診断のタスクにおいて高い精度を示した. 今後の課題とし

<sup>1</sup>We used datasets obtained from the OpenfMRI database. Its accession number is ds000030 (<https://openfMRI.org/dataset/ds000030/>).

<sup>2</sup><http://www.fil.ion.ucl.ac.uk/spm/software/spm12/>

表 1: 診断精度の比較.

Model	統合失調症			双極性障害		
	BACC	SPEC	SEN	BACC	SPEC	SEN
PCC+Kendall+PCA+c-means [3]	0.640	0.635	0.645	0.602	0.565	0.640
SVM [16]	0.505	0.788	0.223	0.512	0.855	0.169
LSTM [17]	0.661	0.854	0.467	0.571	0.802	0.340
DGM [8]	0.722	0.920	0.524	0.619	0.650	0.587
AE+HMM [7]	0.618	0.554	0.682	0.616	0.490	0.742
提案モデル	<b>0.744</b>	0.889	0.599	<b>0.623</b>	0.730	0.516

て, 個人特徴量が実際に個人を表しているのかの検証, 疾患に関連する部位や機能的結合の同定が挙げられる.

## 謝辞

本研究は科研費(16K12487)の支援を受けて行われた.

## 参考文献

- [1] T. J. Sejnowski, P. S. Churchland, and J. A. Movshon, “Putting big data to good use in neuroscience,” *Nature Neuroscience*, vol. 17, no. 11, pp. 1440–1441, 2014.
- [2] B. D. W. Group, “Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework,” *Clinical Pharmacology and Therapeutics*, vol. 69, no. 3, pp. 89–95, 2001.
- [3] H. Shen *et al.*, “Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fMRI,” *NeuroImage*, vol. 49, no. 4, pp. 3110–3121, 2010.
- [4] E. Castro *et al.*, “Deep Independence Network Analysis of Structural Brain Imaging: Application to Schizophrenia,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 7, pp. 1729–1740, 2016.
- [5] N. Yahata *et al.*, “A small number of abnormal brain connections predicts adult autism spectrum disorder,” *Nature Communications*, vol. 7, no. 7, p. 11254, 2016.
- [6] P.-h. Chen *et al.*, “A Reduced-Dimension fMRI Shared Response Model,” in *Advances In Neural Information Processing Systems (NIPS)*, 2015, pp. 460–468.
- [7] H.-I. Suk *et al.*, “State-space model with deep learning for functional dynamics estimation in resting-state fMRI,” *NeuroImage*, vol. 129, pp. 292–307, 2016.
- [8] T. Tashiro, T. Matsubara, and K. Uehara, “Deep Neural Generative Model for fMRI Image Based Diagnosis of Mental Disorder,” in *International Symposium on Nonlinear Theory and its Applications (NOLTA)*, 2017, p. accepted.
- [9] A. Abraham *et al.*, “Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example,” *NeuroImage*, vol. 147, no. 15 February 2017, pp. 736–745, 2017.
- [10] J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural Networks*, vol. 61, pp. 85–117, 2015.
- [11] S. Liu *et al.*, “Multimodal Neuroimaging Feature Learning for Multiclass Diagnosis of Alzheimer’s Disease,” *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 4, pp. 1132–1140, 2015.
- [12] D. P. Kingma, D. J. Rezende, and M. Welling, “Semi-supervised Learning with Deep Generative Models,” in *Advances In Neural Information Processing Systems (NIPS)*, 2014, pp. 3581–3589.
- [13] L. Maaløe *et al.*, “Auxiliary Deep Generative Models,” in *International Conference on Machine Learning (ICML)*, vol. 48, 2015, pp. 1445–1453.
- [14] J. Lasserre, C. Bishop, and T. Minka, “Principled Hybrids of Generative and Discriminative Models,” in *IEEE Computer Society Conference*

on *Computer Vision and Pattern Recognition (CVPR)*, vol. 1, no. 6. IEEE, 2006, pp. 87–94.

- [15] A. Prasad, A. Niculescu-Mizil, and P. K. Ravikumar, “On Separability of Loss Functions, and Revisiting Discriminative Vs Generative Models,” in *Advances in Neural Information Processing Systems (NIPS)*, 2017, pp. 7053–7062.
- [16] F. Pereira, T. Mitchell, and M. Botvinick, “Machine learning classifiers and fMRI: a tutorial overview.” *NeuroImage*, vol. 45, pp. S199–S209, 2009.
- [17] N. C. Dvornek *et al.*, “Identifying Autism from Resting-State fMRI Using Long Short-Term Memory Networks,” in *International Workshop on Machine Learning in Medical Imaging (MLMI)*, 2017, vol. 10541, pp. 362–370.
- [18] T. Matsubara, R. Akita, and K. Uehara, “Stock Price Prediction by Deep Neural Generative Model of News Articles,” *IEICE Transactions on Information and Systems*, p. accepted, 2018.
- [19] J. L. Ba, J. R. Kiros, and G. E. Hinton, “Layer Normalization,” *arXiv*, pp. 1–14, 2016.
- [20] V. Nair and G. E. Hinton, “Rectified Linear Units Improve Restricted Boltzmann Machines,” in *International Conference on Machine Learning (ICML)*, 2010, pp. 807–814.
- [21] N. Srivastava *et al.*, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.
- [22] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *International Conference on Learning Representations (ICLR)*, pp. 1–15, 2015.
- [23] N. Tzourio-Mazoyer *et al.*, “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, 2002.