

構造化深層生成モデルによる fMRI 画像を用いた精神疾患診断

Mental Disorder Diagnosis using fMRI Images by Structured Deep Generative Model

田代哲生 *¹ 松原崇 *¹ 上原邦昭 *¹
Tashiro Tetsuo Matsubara Takashi Uehara Kuniaki

*¹神戸大学 大学院システム情報学研究科 計算科学専攻
Graduate School of System Informatics, Kobe University

In mental disorder diagnosis based on fMRI images, conventional studies perform preprocessing such as feature-extraction using correlation analysis since the fMRI dataset is composed of a small number of high-dimensional samples. However, this preprocessing could miss features necessary for diagnosis. On the other hand, deep generative models achieved good accuracy even with a small dataset with limited preprocessing. In this paper, we model fMRI brain images using a deep generative model with a subject-wise variable. The proposed model explicitly separates individual differences from the mental disorder and noise in fMRI images. The proposed model achieves accuracy higher than conventional methods.

1. はじめに

機能的磁気共鳴画像 (fMRI) のような神経画像データセットを使った機械学習技術に関する多くの研究が行われている。これらの技術は統合失調症などの精神的障害が持つと考えられる特定のバイオマーカーを発見することが期待されている。加えて、適切な治療の機会を提供し、潜在的な治療の有効性を評価することもできる。しかしこれらの神経画像データセットは、そのほとんどが他の機械学習向けのデータセットと比較して、限られた数のデータによって構成されている。したがって、機械学習技術を適用するためには、相関解析や特徴抽出、次元削減を行うことが一般的である。これは、fMRI 脳画像内に存在する撮影時の雑音や、被験者ごとの個人差 (脳の大きさや形の違い) を抑制することが目的でもある。

こうして、特徴抽出で得られた特徴を使って、機械学習を行う手法は多様なものが提案されている。具体的には、特徴の時系列に生成モデルを適用する研究などがある。Suk らは、深層学習による特徴抽出の後に、隠れマルコフモデル (HMM) によって時間的変化をモデル化し、精神疾患の診断を行い精度の向上を達成した [Suk 16]。Yahata らは相関解析による次元削減の後に、Sparse Logistic 回帰 (SLR) を使って精神疾患の診断を行った [Yahata 16]。しかしながら、これらの手法においても、前処理では取り除ききれない被験者ごとの個人差や、雑音などがある程度残ってしまう。このような精神疾患の診断に関連しない情報を強く学習してしまう事が、診断精度の向上の妨げとなっていると考えられる。

一方で、近年、Deep Neural Network (DNN) が柔軟な機械学習フレームワークとして注目されている [Schmidhuber 15]。DNN は任意の関数を近似することで、データセットが持つ高度な特徴を自動的に学習することができる。DNN は教師あり分類器 (Multi-layer Perceptron; MLP) [Castro 16] と、教師なし特徴抽出器 (Auto-encoder; AE) として用いられている [Castro 16, Suk 16]。これらに限らず、深層生成モデル (DGM) と呼ばれる Neural Network は、ネットワーク構造上に複数の要因の関係性を記述する生成モデルを実装している [Tashiro 17]。深層生成モデルの研究では、fMRI 画像と、

精神疾患であるかどうかのクラスラベル、各時間のスキャンごとに起こる変動の 3 つの因子関係をモデル化し、分離する事で、他の比較手法よりも良い精度を達成した。これは、データの要因の関係性を正しく記述できていれば、生成モデルが小規模データセットにおいて識別モデルよりも良い精度が得られることによる [Raina 03]。よって、fMRI データセットの事前知識に基づいて、深層生成モデルを構築することで、診断精度が改善する可能性がある。以上を踏まえて、被験者の個人差の特徴を要因として明示的にモデル化する Structured subject-wise DGM (SDGM) を提案する。SDGM はそれぞれの被験者が、その fMRI 画像全てに共有する個々の変動 (すなわち、被験者特有の特徴) を要因として加える。この推論によって、データセットに対して、より妥当なモデル化を行う。本研究では、統合失調症と双極性障害の resting-state fMRI データセットを用いて実験を行い、SDGM を評価した。この実験では提案手法が、従来手法として行った SVM (Support Vector Machine) [Pereira 09], LSTM (Long Short-Term Memory) [Dvornek 17], DGN [Tashiro 17], AE+HMM [Suk 16], fMRI 脳画像解析でよく用いられるピアソン相関行列 (PCC) 用いて抽出された機能的接続性に基づいた手法 [Shen 10, Yahata 16] よりも良い精度の診断を達成できること示す。

2. Structured subject-wise DGM

Structured subject-wise Deep Generative Model (SDGM) は、Auxiliary Deep Generative Models [Maaløe 15] の導出過程を応用し、DGM を fMRI 解析のために拡張したものである。まず、SDGM を構築するための各要因のモデル化を行う。 N を被験者数、 i を被験者の番号を示すとし、fMRI 画像を \mathbf{x}_i 、クラスラベルを y_i とすると、データセットは $\mathcal{D} = \{\mathbf{x}_i, y_i\}_{i=1}^N$ となる。各被験者 i は健常者 ($y_i = 0$) か、精神疾患患者 ($y_i = 1$) であり、それぞれ一人につき、 T_i 枚の fMRI スキャン画像 $\mathbf{x}_i = \{\mathbf{x}_{i,t}\}_{t=1}^{T_i}$ を持つ。また本手法では被験者 i は事前分布 $p(s)$ に従う特徴量 s_i (subject-wise feature) を持つものとする。この s_i は前処理によって取り除ききれなかった特徴を表し、各被験者が持つ、全スキャン画像に共通する個人の特徴 (脳の形など) である。ここで、fMRI 画像 $\mathbf{x}_{i,t}$ は、クラスラベル y_i 、被験者個人の特徴 s_i 、時間的変化を含む潜在変数 $z_{i,t}$ に関連しているとする。なお、この潜在変数 $z_{i,t}$ は事前分布 $p(z)$

連絡先: 連絡先: 田代哲生, 神戸大学大学院システム情報学研究科計算科学専攻, tashiro@ai.cs.kobe-u.ac.jp

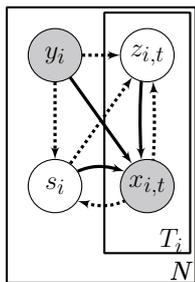


Fig. 1: Graphical model of our proposed generative model.

に従うものとする。\$z_i\$ は時間変化によって起こる脳活動の変化や頭の物理的な動きなどを表す。ここまでの仮定を用いて、fMRI 画像 \$x_i\$ を生成する、\$\theta\$ をパラメータとして持つモデル \$p_\theta\$ を構成する。この生成モデルを図で表したものを図. 1 に、式で表したものを以下に示す。

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

この式を変分法 [Kingma 13] によって変形することで、モデルエビデンス \$\log p_\theta(\mathbf{x}_i|y_i)\$ は、\$\phi\$ をパラメータとする推論モデル \$q_\phi\$ を用いて次のように表される。

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

それぞれ、\$D_{KL}(\cdot||\cdot)\$ はカルバック・ライブラー情報量を、\$\mathcal{L}_g(\mathbf{x}_i; y_i)\$ はモデルエビデンスの下界を表す。この \$\mathcal{L}_g(\mathbf{x}_i; y_i)\$ が、条件付き生成モデル \$p_\theta\$ と条件付き生成モデル \$q_\phi\$ のパラメータである \$\theta\$ と \$\phi\$ の最適化のための目的関数となる。

すなわち、\$\mathcal{L}_g(\mathbf{x}_i; y_i)\$ をモデルエビデンス \$\log p_\theta(\mathbf{x}_i|y_i)\$ の近似として使用する。加えて、本研究では生成モデルを識別に応用するために、識別的な目的関数を定義する。クラスラベル \$y_i\$ の事後分布 \$p(y|\mathbf{x}_i)\$ をベイズの公式を用いて計算すると以下のように変形できる。

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

この式変形では、クラスラベル \$y\$ の事前分布 \$p(y)\$ が \$p(y=0) = p(y=1) = 0.5\$ としている。したがって、\$\mathcal{L}_g(\mathbf{x}_i, y=1)\$ は大きな値を持つ場合、被験者 \$i\$ はより精神疾患である可能性が高くなる。加えて、クラスラベルの対数尤度である \$\mathcal{L}_d(\mathbf{x}_i, y_i)\$

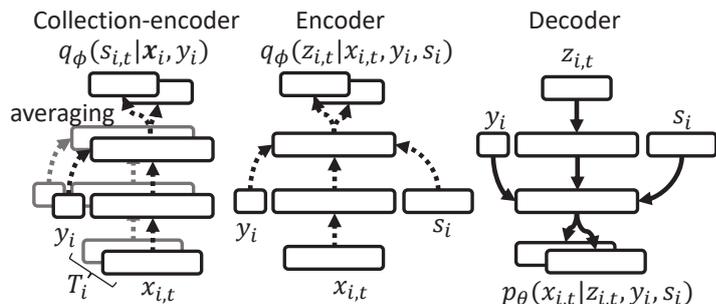


Fig. 2: Our proposed generative model on the deep neural networks.

は、これを目的関数とすることで、識別的な学習を可能とする。本研究ではこの二つの目的関数の調整を行うハイパーパラメータ \$\omega \in [0, 1]\$ を用いて、以下の式を目的関数とする。

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

ここまでで得られた生成モデル \$p_\theta\$ と推論モデル \$q_\phi\$ を、Neural Network を用いて構築する。確率モデルを出力する Neural Network の構築には、Kingma らと同様の手法によって行う [Kingma 13]。前処理済みの fMRI 信号 \$x_{i,t}\$、被験者個人の特徴 \$s_i\$、スキャンごとの変動 \$z_{i,t}\$ をそれぞれ次元が \$n_x, n_s, n_z\$ のベクトルとする。推論モデル \$q_\phi(z_{i,t}|\mathbf{x}_{i,t}, y_i, s_i)\$ と生成モデル \$p_\theta(x_{i,t}|y_i, s_i, z_{i,t})\$ を多変量ガウス分布とする。それぞれのモデルのパラメータを出力する Neural Network を構築する。推論モデルの実装では、\$x_{i,t}, y_i, s_i\$ を入力として、\$q_\phi(z_{i,t}|\mathbf{x}_{i,t}, y_i, s_i)\$ のパラメータを出力する encoder を構成する。生成モデルの実装には \$z_{i,t}, y_i, s_i\$ を入力として、\$p_\theta(x_{i,t}|y_i, s_i, z_{i,t})\$ のパラメータを出力する decoder を構成する。ここで、データとして存在する \$x_{i,t}, y_i\$ と異なり、\$s_i\$ は事前に得られない。そこで \$x_i, y_i\$ を使って、\$s_i\$ を推論するモデル \$q_\phi(s_i|\mathbf{x}_i, y_i)\$ を Neural Network を用いて構成する必要がある。ここで、\$s_i\$ が、各被験者に対して、全ての \$t\$ において共通する特徴であることを考え、各被験者 \$i\$ の全ての \$t\$ における \$x_{i,t}\$ と、\$y_i\$ を入力とし、\$q_\phi(s_i|\mathbf{x}_i, y_i)\$ のパラメータを出力とする encoder を構成する。この \$s_i\$ を推論するモデルは、\$T_i\$ 枚存在する \$x_{i,t}\$ と、\$y_i\$ をそれぞれ別々の Neural Network に入力する。その各 Neural Network の出力の平均値を入力とする Neural Network によってこれを構成する。以上のようにして、Neural Network で構成した提案手法のネットワーク構造を図. 2 に示す。

このモデルを用いて診断を行う場合には、\$y_i\$ が未知のデータに対して、精神疾患の尤度の近似として学習している \$\mathcal{L}_d(\mathbf{x}_i, y_i = 0)\$ と \$\mathcal{L}_d(\mathbf{x}_i, y_i = 1)\$ の値を比較することで、精神疾患であるか、健常者であるかを診断する。

3. 実験

本研究では統合失調症患者、双極性障害患者、健常者の resting state fMRI (rs-fMRI) 脳画像のデータセットを使用した。データセットは OpenfMRI データベースの ds000030 (<https://openfmri.org/dataset/ds000030/>) より得られる。まず、このデータに SPM12 (<http://www.fil.ion.ucl.ac.uk/>) を使用して以下の前処理を施した。最初に fMRI 撮影開始時の磁気の乱れを取り除くために、各被験者の撮影した fMRI 画像の最初の 10 スキャンを除去した。その後、撮影時の時間的ズレを補正するために Slice Timing 処理を施した。続いて、撮

Table 1: Diagnostic Accuracies.

Model	Schizophrenia			Bipolar		
	BACC	SPEC	SEN	BACC	SPEC	SEN
PCC+Kendall+PCA+c-means[Shen 10]	0.640	0.635	0.645	0.602	0.565	0.640
PCC+SCCA+SLR[Yahata 16]	0.639	0.779	0.500	0.607	0.735	0.480
LSVM[Pereira 09]	0.505	0.788	0.223	0.512	0.855	0.169
LSTM[Dvornek 17]	0.661	0.854	0.467	0.571	0.802	0.340
DGM[Tashiro 17]	0.722	0.920	0.524	0.619	0.650	0.587
AE+HMM[Suk 16]	0.618	0.554	0.682	0.616	0.490	0.742
SDGM (proposed)	0.767	0.812	0.722	0.622	0.844	0.401

影中の被験者の頭や脳の揺れを除去するために、Realignement処理を施し、各被験者の画像の位置ズレを抑制した。次に脳画像の個人差や、MRI撮影時の環境差をある程度抑制するために、Normalization処理を行った。結果として、ボクセルサイズ $3 \times 3 \times 3\text{mm}$ の MNI 空間に正規化される。ここまでの処理で、SPM12によって正規化された画像を、AAL (Automated Anatomical Labeling) [Tzourio-Mazoyer 02] による 116 部位に分割した。この各分割した領域のボクセルの平均値を計算する。結果として、rs-fMRI 画像はベクトル化され、[116 部位 \times スキャン枚数] の形となる。ここで、rs-fMRI では、0.025 Hz から 0.06 Hz、または 0.07 Hz の間の周波数の帯域に、有用な特徴が存在することが知られている [Malinen 10]。そこで、本研究では、 $0.01 \leq f \leq 0.1$ Hz の周波数帯域が残るようにバンドパスフィルタを使用した。最後に、データのボクセル値が、時間方向において平均 0、分散 1 になるように、正規化を施した。なお、scrubbing として、スキャン全体での平均画像との、脳の位置のずれが 1.5mm を超えるスキャン画像と、任意の方向に 1.5 度を超える角度のズレが存在するスキャン画像を除去している。また、被験者ごとのスキャン枚数のバランスを保つために、scrubbing によって残ったスキャン画像が 100 枚以下となった被験者については学習するデータセットから取り除いた。よって、最終的に、健常者 113 人、統合失調症患者 44 人、双極性障害患者 45 人を使用した。

4. 結果

提案手法の効果を検証するために、統合失調症の診断と、双極性障害の診断の 2 つを行った。具体的には、患者の前処理済みデータと、健常者の前処理済みデータを学習し、識別を行った。なお、10 fold cross-validation を 5 回行った平均値を用いて精度を検証した。また精度の評価において、データに偏りが存在することを考慮して、Accuracy ではなく、Sensitivity と Specificity の平均値を使用する。これはデータの偏りに依らない精度となり、これを Balanced Accuracy として使用する。また学習がデータの偏りに影響されないよう、oversampling を行って学習した。SDGM のハイパーパラメータは探索範囲を狭めるために、encoder, decoder, s の推論器ともに、3 層のニューラルネットワークで構築し、入力層と隠れ層は同じ unit 数で、{50, 100, 200, 400} の範囲とした。また encoder, s の推論器の出力層は {5, 10, 20, 50, 100} の範囲から探索した。また、モデルの入力層には dropout を行った。ここでの dropout 率は {0.0, 0.5} の 2 つのパラメータで探索した。最後に学習には Adam アルゴリズム [Kingma 14b] を使用した。Adam は初期パラメータを使用し、学習率は 10^{-4} とした。

比較手法としては、まず脳画像解析によく用いられる手法である、脳の機能的接続性を考慮したピアソン相関係数 (PCC) を計算する手法を用いる。ここでは PCC の後にケンドールの順位相関係数 (Kendall) を使って特徴選択を行い、その後主成分分析による特徴抽出の後に、C-means によって識別を行う手法 (以後これを PCC+Kendall+PCA+c-means とする) [Shen 10] と、Sparse canonical correlation analysis (SCCA) によって特徴選択を行った後に Sparse Logistic 回帰 (SLR) によって識別を行う手法 (PCC+SCCA+SLR) [Yahata 16] を行う。fMRI の時間的変化を学習し、識別する手法として、LSTM (Long Short-Term Memory) [Dvornek 17] と Auto-encoder による特徴抽出の後に隠れマルコフモデルによって識別を行う手法 (AE+HMM) [Suk 16] を行った。他の手法としては、一般的な機械学習の手法として Linear Support Vector Machine (LSVM) [Pereira 09] と、同様に深層生成モデルを使った手法 (DGM) [Tashiro 17] による識別とも比較した。各種手法のハイパーパラメータは、Kendall, SCCA による特徴選択数は {50, 100, 200, 400, 600}, PCA による特徴抽出の次元数は {2, 5, 10, 20, 50}, LSVM の Cost パラメータは {0.1, 0.2, 0.5, 1, 2, 5, 10} の範囲から探索した。LSTM, Auto-encoder, DGM は、提案手法である SDGM と同様の条件で探索を行った。また HMM においては {2, 3, 4, 5, 6, 7} の範囲から探索した。

この実験結果を表 1 に示す。統合失調症、及び双極性障害両方の診断において、提案手法である SDGM が最も良い精度を達成した。特に統合失調症において大きな改善が見られた。SDGM と、PCC+Kendall+PCA+c-means, PCC+SCCA+SLR との比較によって、本手法が PCC のような相関分析を行わずとも、深層学習によって、一定の精度を得られることを示している。これは SDGM のネットワーク内で、各入力部位間での相関性や、機能的接続性を PCC と同様、あるいはそれ以上に、より良く学習していると考えられる。また、DGM と提案手法である SDGM との比較によって、被験者固有の特徴として推論した s を分離する事によって、精度の向上が行われたことが分かる。一方で、双極性障害の診断においては、精度の改善こそあったものの、大きな改善とはならなかった。ここで、AE+HMM の精度が統合失調症の場合に比べて、相対的に高い精度を示している事が分かる。このことから、双極性障害の診断に置いては、統合失調症の場合よりも、脳の時間的変化に診断に必要な情報が存在すると考えられる。よって、SDGM が DGM に対して大きな精度の向上を見込めなかったのも、今回の提案手法で行った、subject-wise feature を加えるというモデルの拡張が、時間的変化を捉えられるようになるものではなかったからであると考えられる。

Table 1: Top 5 Weights of AAL for Diagnosis.

Schizophrenia		Bipolar	
ROI	weight	ROI	weight
Cerebellum_6_L	0.0555	Cingulum_Ant_R	0.0132
Postcentral_L	0.0532	Frontal_Inf_Orb_L	0.0121
Cingulum_Mid_L	0.0531	Cerebellum_7b_R	0.0116
Lingual_R	0.0529	ParaHippocampal_L	0.0114
Lingual_L	0.0526	Temporoal_Mid_L	0.0106

さらに定性的な評価として、提案手法である SDGM を用いて、診断に特に寄与した部位の特定を行った。これは入力、前処理において、AAL で部位ごとに分割して計算した平均値となっているために、116 次元の入力の中で、特に誤差が大きかった入力、識別に関連したと判断できる。具体的には、SDGM のクラスラベルを変化させた時に、もっとも誤差の変化 (weight) が大きい部位を特定する事によって行う。具体的な計算式は、DGM における方法 [Tashiro 17] と同様である。結果は表. 1 の通り。各部位において、ある程度疾患との関連性がある事を示している。統合失調症においては、小脳 (Cerebellum) の関連性 [Yeganeh-Doost 11] や、中心後回との関連性 [Ferro 14] などと言及した研究も存在する。一方で、双極性障害においては、提案手法自体の診断精度が特別に高いわけではないため、今回の実験で得られた部位が、特別に強く双極性障害に関連すると判断するのは難しい。しかしながら、それぞれ、帯状回 (Cingulum) との関連性 [Fei 08] や、海馬傍回 (ParaHippocampal) との関連性 [Almeida 09] などを示す研究も存在する。

5. 結論

本研究では、fMRI 脳画像と機械学習を用いた精神疾患の診断に有効な深層学習モデルである SDGM を提案した。SDGM では、fMRI 画像を生成モデルとしてモデル化した。加えて、機械学習における診断の障害になる考えられる、被験者ごとの個人差を明確にモデル化し、分離する事によって、診断精度の向上を試みた。この診断精度を検証するために、本研究では、統合失調症と双極性障害のデータセットを用いて実験を行った。結果として、SDGM は、両方の精神疾患の実験において、従来手法よりも良い精度を達成した。これにより、SDGM が小規模データの場合の fMRI 画像を用いた診断に有効である事示した。また、DGM と同様に、入力と出力の誤差から、脳の部位と精神疾患との関連性を確認する事ができた。これにより、精神疾患の持つバイオマーカーの発見に役立つ可能性を示している。

本研究は科研費 (16K12487) の支援、総務省 SCOPE(受付番号 172107101) の委託を受けて行われた。

参考文献

- [Schmidhuber 15] Schmidhuber, J. (2016) *Neur. Netw.*, vol.61 pp.85-117
- [Suk 16] Suk, H.I., et al. (2016), *NeuroImage*, vol.129, pp.292-307
- [Castro 16] Castro, E., et al. (2016) *IEEE Trans. Med. Imaging*, vol.35(7), pp.1729-1740
- [Raina 03] Raina et al. (2003), *Advances in Neural Information Processing Systems* 16, pp.545-552
- [Pereira 09] Pereira, F., et al. (2009) *NeuroImage*, vol.45 pp.S199-S209
- [Dvornek 17] Dvornek, N.C., et al. (2017) *In Proc. of Machine Learning for Multimodal Interaction*, pp.362-370
- [Tashiro 17] Tashiro, T., et al. (2017) *In Proc. of International Symposium on Nonlinear Theory and its Applications*.
- [Shen 10] Shen, H., et al. (2019), *NeuroImage*, vol.49(4), pp.3110-3121
- [Yahata 16] Yahata, N., et al. (2016), *Nature Communications*, vol.7(7), 11254
- [Maaløe 15] Maaløe, et al. (2015) *In Proc. of International Conference on Machine Learning* vol.48, pp.1445-1453
- [Yeganeh-Doost 11] Yeganeh-Doost et al. (2011), *Clinics*, vol.66(1), pp.71-77
- [Ferro 14] Ferro et al. (2014), *European Psychiatry*, vol.29(1), pp.1
- [Fei 08] Fei et al. (2008), *The British journal of psychiatry: the journal of mental science*, vol.193(2), pp.126-129
- [Almeida 09] Almeida et al. (2009), *Psychiatry research*, vol.174(3), pp.195-201
- [Kingma 13] Kingma et al. (2013), *In Proc. of International Conference on Learning Representations*
- [Tzourio-Mazoyer 02] Tzourio-Mazoyer et al. (2002), *NeuroImage*, vol.15, pp.273-289
- [Malinen 10] Malinen et al. (2010), *Proceedings of the National Academy of Sciences of the United States of America*, vol.107, pp.6493-6497
- [Kingma 14a] Kingma et al. (2014), *Advances in Neural Information Processing Systems* 27, pp.3581-3589
- [Kingma 14b] Kingma et al. (2014) *In Proc. of International Conference on Learning Representation*, p.1-15