# Discovering the manifold of Psychiatric disorders using deep generative models

**Rajat Mani Thomas** * AMC/UvA
**Paul Zhutovsky** AMC/UvA
**Guido van Wingen** AMC/UvA
**Max Welling** UvA

## Abstract

Psychiatric disorders are amongst the most difficult to accurately diagnose and design a treatment plan for. Imaging the structural and functional properties of an individual's brain is the key to solving this challenge. Current machine learning approaches fail to utilize the information in these brain scans. Broadly, we propose finding a manifold of structural and/or functional brain scans in an embedding space that clusters mental disease states into clinically recognized classes. Specially, we propose a flavour of the adversarial autoencoder to accomplish this task. The goal is wide and ambitious enough to accommodate several other alternatives to solve this challenge.

## 1 Introduction

Improving the diagnosis and treatment strategies for individuals with neuropsychiatric disorders has widespread impact on the individual and our society. Increase in quality of life of patients and their relatives, lowering of the economic burden on the healthcare system and a better understanding of these disorders are only few of the benefits.

Currently, diagnosis of psychiatric disorders is based on clinical interviews. This is known to lead to variability in diagnosis across clinicians. Moreover, personalized clinical prognosis and the prediction of treatment outcome is currently impossible, and treatment is therefore performed on a trial-and-error basis.

In this interdisciplinary proposal, we aim to investigate the feasibility of using advanced machine learning techniques like deep convolutional neural networks [1] and adversarial autoencoders [2] on neuroimaging data, to discover the latent space of structural and/or function magnetic resonance images of psychiatric patients.

We are at a time in history where a combination of large datasets (>10000), advanced algorithms and cheap computational power has given us the opportunity to tackle a problem of this magnitude and importance. During the course of this project, we will assimilate a large dataset combining both publically available and clinical data from our hospital and collaborators to 1) develop a deep learning algorithm that will learn the clustering of brain images from psychiatric patients in a latent space , 2) discern healthy individuals from psychiatric patients, and 3) test whether or not the extensions of these algorithms can predict responders from nonresponders in a cohort of 300 severely depressed patients that underwent electroconvulsive therapy (ECT). The results from this proposal will enable the further development of personalized treatment of psychiatric disorders.

## 2 Related Works

Unsupervised machine learning techniques have been recognized as a way to allow the discovery of biologically relevant clusters in the data without requiring user defined target labels. These

*Correspondence: rajatthomas@gmail.com, or m.welling@uva.nl

clusters may correspond to 'new' data-driven phenotypic subtypes or demographic groupings and there is active research work in this area. For example Aljabar et al.[10] utilized an unsupervised manifold learning approach to characterize neonatal brain development. Most recently, Brodersen and Colleagues [11] reported an unsupervised machine-learning proof-of-concept study examining the feasibility of defining subgroups in psychiatric spectrum disorders using generative embedding techniques. Fair and Colleagues [12] utilized a graph theory approach to identify unique data-driven neuropsychological subgroups in children with attention deficit hyperactivity disorder (ADHD).

The main limitation of these methods were that although MRI images were used, they had to be preprocessed and the relevant features extracted. For example, [10] used an atlas based approach in which the brain is normalized into a common space and then split into > 100 regions, following which the grey matter values in these regions are averaged.

The proposed approach will be totally data-driven and there would be no need to use subjective preprocessing or feature-extraction techniques.

## 3 Adversarial Autoencoder

The goal is to cluster psychiatric disorders in the latent space of brain images. Figure 1 below illustrates the idea:

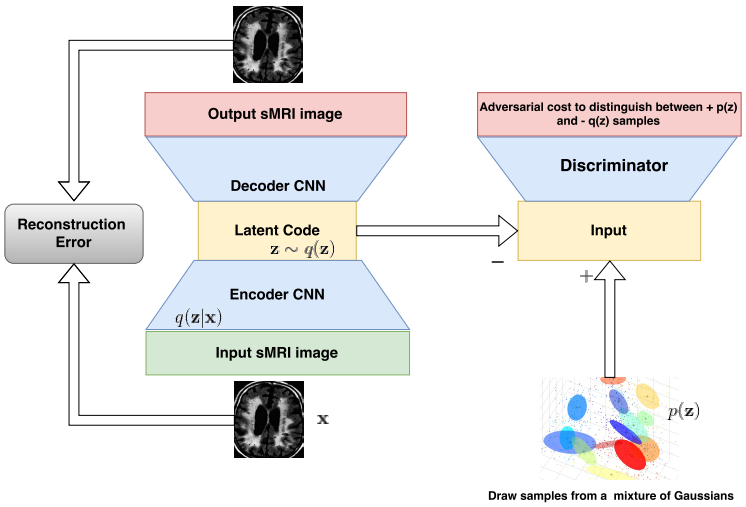


Figure 1: Sample figure caption.

A variational autoencoder [3], not only minimizes the reconstruction error but also tries to impose a probability structure in the latent representational space. An extension of this idea, potentially to arbitrary structures of the probability distribution in latent space, is the adversarial autoencoder.

The key idea is to use an adversarial autoencoder to impose a structure in the latent space which is a mixture of Gaussians. Each Gaussian representing a particular psychiatric diagnostic category.

Thus, an adversarial autoencoder will be trained to minimize two objectives, (i) the reconstruction error of the MRI images $\{x_1, x_2, ..., x_S\}$ and (ii) minimize the adversarial cost incurred in distinguishing samples that are drawn from a mixture of Gaussians (+ samples) from that obtained from the latent space (- samples) .

Initially, we intend to keep $p(z)$ in Figure 1 fixed. We will assume about 10 clusters (to match the number of current major mental disorders) and distribute the means of the Gaussians to be equidistant from each other. So, for means $\{\mu_1, \mu_2, ..., \mu_{10}\}$, $\|\mu_i - \mu_j\| = \mathcal{C}$ and $\mu_i \in \mathcal{R}^N$, where $N$ is the dimension of the latent code.

# 4 Experiment

## 4.1 Hypothesis

- A Psychiatric disorder and its subtypes would belong to the same clusters.
- It should be possible to outperform other clustering techniques in native space.
- This representation could be used to classify new subjects into one of these categories or as being healthy.

## 4.2 Project Milestones

This project would involve the following tasks:

### 4.2.1 Data

The success of the project relies on gathering a substantial amount of heterogeneous dataset across varied populations. To this end we will aid in assembling data from [2];

- ENIGMA consortium [6]
- The UKBioBank data set [8] with 10K sMRI and fMRI images (publically available).
- ABIDE data set [7] with 3K images (publically available).
- GEMRIC [9]: a cohort of 300 patients that have undergone ECT. (a limited consortia data set we are part of).

Amongst the above datasets, we propose researchers to start with the ABIDE dataset (I and II). Together they have well over 1700 scans and provides a good platform to test the proposed techniques. And importantly, the ABIDE dataset is open to anyone to just download.

## 4.3 Implement the variational autoencoder

- Define the prior mixture of Gaussians with an appropriate set of mean and variance parameters
- Train the adversarial model using method presented in [2].
- Examine the latent space and generate novel MRI images for further training
- Use the autoencoder for labelling the test data set in a probabilistic fashion.

As mentioned in the abstract, the adversarial autoencoder is one of the proposed solutions. Researchers are encouraged to come up with alternative solutions, for example to learn the distribution $p(z)$ (instead of being fixed as we propose).

## 4.4 Evaluation

For evaluation, we will have a subset of held-out unlabeled brain scans and the goal would be to assign the appropriate cluster to each of these data points.

## 4.5 Resources

There are several implementations of the adversarial autoencoder to get started with; in TENSORFLOW [3], as shown here [4] and in PYTORCH [5] as shown here [6].

[2] restrictions apply for some of the datasets but researchers interested can contact the author for further details

[3] www.tensorflow.org

[4] https://github.com/takat0m0/AAE

[5] www.pytorch.org

[6] https://github.com/fducau/AAE_pytorch

## 5 Impact

- An objective protocol based on neuroimaging data to label psychiatric disorders will fundamentally change the way in which the healthcare system currently handles psychiatric patients. Insurances will evaluate covering neuroimaging expenses for long term benefits.
- The amount of time spent by a psychiatric patient to be diagnosed and appropriately treated will be drastically reduced.

## References


1. "ImageNet Classification with Deep Convolutional Neural Networks", Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, NIPS, 2012.
2. "Adversarial Autoencoder", Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, Ian Goodfellow, Brendan Frey, ICLR, 2016.
3. "Auto-Encoding Variational Bayes", Kingma and Welling, The International Conference on Learning Representations (ICLR), Banff, 2014.
4. "A survey of Deep Learning in medical imaging", Litjens et al., Medical Image Analysis, 2, pp. 60-88, 2017.
5. "Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications", Vieira et al., Neuroscience & Biobehavioral Reviews, 74A, pp. 58-75, 2017.
6. "The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data", Thompson et al., Brain Imaging and Behavior, 8, pp. 153-182, 2014.
7. "The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism", di Martino et al., Mol Psychiatry, 19, pp. 659-667, 2014.
8. "UK BIOBANK DATA: COME AND GET IT", Allen, M. E. et al., Science Translational Medicine, 6 (224), pp. 224, 2014.
9. "The Global ECT-MRI Research Collaboration (GEMRIC): Establishing a multi-site investigation of the neural mechanisms underlying response to electroconvulsive therapy", Oltedal L., et al., Neuroimage:clinical, 14, pp. 422-432, 2017.
10. "A combined manifold learning analysis of shape and appearance to characterize neonatal brain development", P. Aljabar, R. Wolz, L. Srinivasan, S.J. Counsell, M.A. Rutherford, A.D. Edwards, et al.,IEEE Trans Med Imaging, 30, pp. 2072-2086, 2011.
11. "Dissecting psychiatric spectrum disorders by generative embedding, K.H. Brodersen, L. Deserno, F. Schlagenhauf, Z. Lin, W.D. Penny, J.M. Buhmann, et al. NeuroImage Clin, 4 , pp. 98-111. 2014
12. "Distinct neuropsychological subgroups in typically developing youth inform heterogeneity in children with ADHD, D.A. Fair, D. Bathula, M.A. Nikolas, J.T. Nigg, Proc Natl Acad Sci, 109, pp. 6769-6774, 2012.