Deep Structural Causal Modeling for Magnetic Resonance Neuroimaging

July 12, 2023

Student: Lars Chen
Matriculation Number: 0465415
Program: MSc Computational Neuroscience, TU-Berlin
Supervisors: Prof. Dr. rer. nat. Kerstin Ritter / Moritz Seiler

Abstract

Scientific questions that are causal in nature can be answered through causal inference using tools such as structural causal models (SCMs). This project attempts to develop an SCM using deep learning that can models the relationship between demographics, disease covariates and the magnetic resonance (MR) images of subjects with multiple sclerosis (MS). Here, the framework developed by Pawlowski et al., which looks at healthy subjects, is extended and applied to a custom dataset containing brain scans of patients with varying severities of Clinically Isolated Syndrome (CIS), one of the four possible disease courses of MS. Causal inference using deep SCMs is applied here in hopes of better understanding the progression of MS, investigating confounding factors that contribute to MR scans and being used for data augmentation. The model obtained in this project was only able to answer answer queries at low levels of Pearl's Hierarchy of causality with some success and failed to realistically generate answers at high levels.

1 Introduction

The application of machine learning (ML) to medical imaging aims to improve the lives of many patients and decrease the burden of physicians. ML has shown interesting results to better understand disease progression (Ravi et al., 2019). However, ML models still perform poorly in many clinical applications because they are prone to learning spurious relationships that exist in the data they are trained on. The integration of causal information into DL methods is a novel and promising direction in that hopes to tackle this short-coming (Pawlowski, 2020).

This lab rotaion aims to use DL methods to model the causal relationships between hidden variables related to the disease course of Multiple Sclerosis (MS) and the appearance of brain images. MS is an autoimmune disease with a typical onset in early adulthood that affects millions of people world-wide. The severity of symptoms are rated using Expanded Disability Severity Score (EDSS), which describes the extent of ambulatory disabilities of patients with MS from 0-10 Dobson and Giovannoni (2019). In structural magnetic resonance (MR) images of the brain, MS presents as lesions where neurons have demylinated. The loss of myelination in the brain shows up as white spots on MR images which is the primary tool for diagnosing MS (Reinhold et al., 2021).

1.1 Causality and Machine Learning

Recently, deep networks have been used to generate images of how the brain of an Alzheimer's patient will look after a certain time duration (Ravi et al., 2019). However such methods are inherently limited since they rely on inference via association which is prone to learning biases in data (Varoquaux and Cheplygina, 2022). In other words, subjects in a study may not represent the true distribution of demographic information or disease covariates in a real world setting - i.e. a distribution shift can occur and can cause poor performance in clinical applications (Schölkopf, 2022).

Simply by knowing that two random variables are correlated with one another, the direction of causality may still be unclear. For example, if observing brain lesions in an MR scan reduces the amount of surprise that a person has MS and vice versa by the same amount (mutual information) then the causal direction is not known. If a physician then prescribes a symptomatic treatment that can reduce brain lesions, it will not be observed that the frequency of MS declines. This provides information that MS causes lesions and not the other way around, moving the framework of inference from statistical to causal (Neuberg, 2003).

Many such scientific inquiries can be categorized into Pearl's 3-level hierarchy of causation. The first level is called *association* in which data is collected and statistical relationships are learned. Here one can ask questions such as "is observing brain lesions related to observing MS?". This can be expressed as the conditional probability $P(\text{MS} \mid \text{lesions})$. The second is called *intervention* and is

higher because it involves enacting a change on the variables. In this level, a typical question would be "what happens when we apply a treatment to patients with MS?" This is expressed with 'do' statements that signify reassignment of a variable: $P(\text{MS} \mid \text{do(treatment)})$. At the highest level are *counterfactual* queries because they subsume the previous two, meaning if counterfactuals can be answered, then interventions and associations are known as well. *Counterfactuals* pose hypothetical questions, which for example can apply *interventions* retrospectively: "If a the patient displayed MS symptoms for 10 years instead of 5, how would the MR image look differently?" (Neuberg, 2003). This can be expressed: $P(\mathbf{x}'_{d=10} \mid \text{do}(d' = 10),\ d = 5,\ \mathbf{x}_{d=5})$.

Level	Activity	Typical Questions
1. Association: $P(y \mid x)$	Observing	How likely is Y if observed X?
2. Intervention: $P(y \mid do(x))$	Doing	What if I do X now?
3. Counterfactual: $P(y_x \mid x', y')$	Imagining	What if X happened instead of Y?

Table 1. Pearl's 3-Rung Hierarchy of Causation.

Using a plausible model of the relationships between covariates, causal questions at any level of Pearl's Hierarchy may be answered with observational data (Pawlowski, 2020).

1.2 Structural Causal Models

Causal relationships are modelled via Structural Causal Models (SCMs) which can answer causal questions as mentioned previously.

An SCM is a graph, $\mathcal{G} = (\{x_i\}, P(\epsilon))$, defined as a tuple of a set of structural assignments (eq1) and a joint distribution over mutually independent exogenous noise (eq2)), i.e. - unaccounted sources of variation.

$$x_i := f_i(\epsilon_i; \mathbf{pa_i}), \quad i = 1, ..., N \tag{1}$$

$$P(\epsilon) = \prod_{i}^{N} (\epsilon_i) \tag{2}$$

SCMs can be described with directed acyclic graph (DAG) such that observable variables, x_i, are the nodes and the functions, f_i, are edges that connect to parent nodes. These structural assignments are described then as functions of parents $\mathbf{pa_i}$ and exogenous noise ϵ_i (Schölkopf, 2022).

As opposed to Bayesian networks, which describe conditional probabilities of child nodes given their parents and describe to correlations between the two, SCMs have causal meaning and can provide predictions for interventional and counterfactual questions. Every SCM entails a unique joint observational distribution $P(\mathcal{G})$ which is Markovian: each random variable is independent of all non-parents, given its parents. It therefore factorizes as:

$$P(\mathcal{G}) = \prod_{i}^{N} P(x_i|\mathbf{pa}_i) \tag{3}$$

Given a realistic SCM, such a model fulfills all three levels of Pearl's hierarchy of causal inference - i.e. can answer associational, interventional and counterfactual questions. Counterfactual queries are a three-step procedure:

- **Abduction**: Predict the the exogenous noise that is compatible with the observations, $\mathbf{x}$, i.e. infer $P_{\mathcal{G}}(\epsilon|\mathbf{x})$.
- **Action**: Perform an intervention (e.g. $\text{do}(x_i := x_i')$) corresponding to the desired manipulation, resulting in a modified SCM $\mathcal{G}' = \mathcal{G}_{x;\ \text{do}(x')} = (S', P_{\mathcal{G}'}(\epsilon|\mathbf{x}))$.
- **Prediction**: Compute the quantity of interest based on the distribution entailed by the counterfactual SCM, $P_{\mathcal{G}'}(\mathbf{x})$.

1.3 Related Work and Context

Inference in SCMs using data such as medical images is too high dimensional and intractable to compute. The computational cost of variational inference can be spread across the samples of a data set instead of needing to infer a proposed distribution all at once. Pawlowski et al. (2020) use deep networks to amortize inference across data samples using encoder-decoder structured deep networks.

They implemented a deep SCM that was trained on MR images of healthy brains from UK Biobank dataset. Brain MRI scans were modeled using biological sex, person's age, brain volume and ventricle volume as nodes of an SCM. The model was shown to generate convincing counterfactual images which suggest how an MR scan of a specific subject could look given different biological or demographic information. For example, how an MR scan would look if the subject were 60 years old instead of 40 or if they were biologically male instead of female.

The proposed framework by Pawloski et al. could be promising in identifying anatomical changes in brain structure caused by various neurological diseases. This lab rotation aims to extend the work by Pawlowski to include clinical data relevant to the severity of MS. Here, existing code is adapted to work on custom dataset containing structural MR images of patients with varying severities of MS.

2 Methods

2.1 Data

A private dataset consisting of 684 MRI scans with 161 unique subjects was roughly split 60%, 20%, 20% into train, test, and validation. Each subject appears multiple times in the dataset with different slice numbers. Some subjects were excluded from the original dataset if they had missing values in the labels corresponding to variables that were modelled.

Randomly cropped brain slices from their original size of 218 px × 182 px to 182 px × 182 px during training and center crops during validation and testing. The cropped images are downsampled by a factor of 2 to a size of 81 px × 81 px. **Figure 1** shows samples from this dataset.

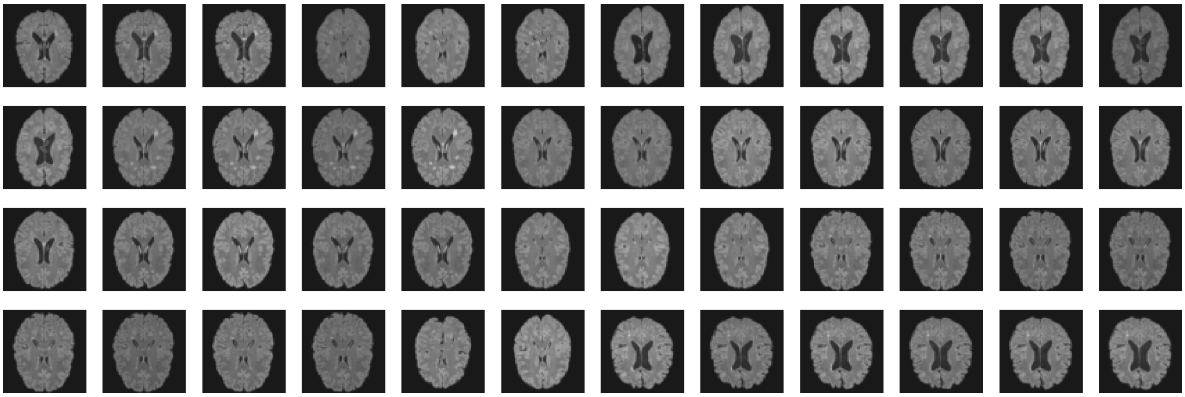

Figure 1: Samples from the CIS Dataset. Many of the samples are different slices of the same brain scan. White spots can be seen that indicate characteristic lesions caused by MS.

2.2 Experimental Setup

The SCM that was used in this experiment is shown in **Figure 2** with its endogenous variables. The SCM from Pawlowksi et al. was used as a starting point. Ventricle volume was removed as a node, since it was not labeled in the dataset, and duration of symptoms, lesion volume and EDSS scores were added similiarly to in Reinhold et al. 2021.

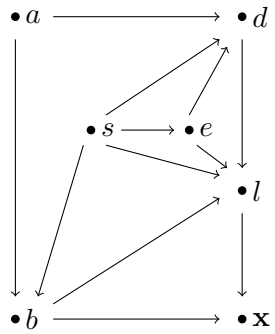

Figure 2: The structural causal model used: a is age, d is the duration of MS symptoms, l is the lesion volume of the subject, $\mathbf{x}$ is the image, b is the brain volume, s is biological sex, e is the EDSS number. The exogenous noise variables are not shown in the graph.

The same deep architectures and training procedures were used as by Pawlowski et al. to construct

the structural assignments of the SCM. The images were modelled with conditional variational auto-encoders (CVAEs), which are stochastic encoder-decoder architectures, were used with 5 kernel sizes with each size repeating the block (LeakyReLU(0.1)BNθConvθ) three times. LeakyReLU(0.1) is a leaky ReLU with a negative slope parameter, BN is a batch normalization layer, and Conv is a convolutional layer. A fully connected linear layer at the end transforms the network into the latent dimension of 100. The sex variable is learned by sampling from a Bernoulli distribution. Normalizing flows were used for all other structural assignments f_i Dinh et al. (2017). Normal distributions were used as base distributions for all exogenous noise distributions. PyTorch and Pyro probabilistic programming language (PPL) were used for all transformations (Bingham et al., 2019). The normalizing flows use the following invertible transforms to model the age a, brain volume b, lesion volume l, duration d, and EDSS score e as:

$$a := f_A(\epsilon_A) = (\exp \circ \text{AffineNormalization} \circ \text{Spline}_\theta)(\epsilon_A) \tag{4}$$

$$b := f_B(\epsilon_B\ ; s, a) = (\exp \circ \text{AffineNormalization} \circ \text{ConditionalAffine}_\theta([s, \hat{a}]))(\epsilon_B) \tag{5}$$

$$d := f_B(\epsilon_D\ ; s, a) = (\exp \circ \text{AffineNormalization} \circ \text{ConditionalAffine}_\theta([s, \hat{a}]))(\epsilon_D) \tag{6}$$

$$e := f_B(\epsilon_E\ ; s, d) = (\exp \circ \text{AffineNormalization} \circ \text{ConditionalAffine}_\theta([s, \hat{d}]))(\epsilon_E) \tag{7}$$

$$l := f_l(\epsilon_L\ ; a, b, d, e) = (\exp \circ \text{AffineNormalization} \circ \text{ConditionalAffine}_\theta([\hat{a}, \hat{b}, \hat{d}, \hat{e}]))(\epsilon_L) \tag{8}$$

These transforms are are parameterized as a fully-connected network with 8 and 16 hidden units, and a LeakyReLU(0.1) nonlinearity.

3 Results

3.1 Association

The learned deep SCM has with some success been able to model the observational distributions between some disease covariates. This can be seen in the noise vectors z_x which are plotted with varying degrees if brain and lesion volumes. It can be observed that the brain consistently becomes larger with larger labels of brain volume. However, no apparent changes occur when changing lesion volume.

Next, kernel density estimations (KDEs) were plotted comparing the distributions of the true data versus the learned model. The model was able to learn the relationship between covariates with relatively faithful representations of the true distributions. However some collapse of variance does occur.

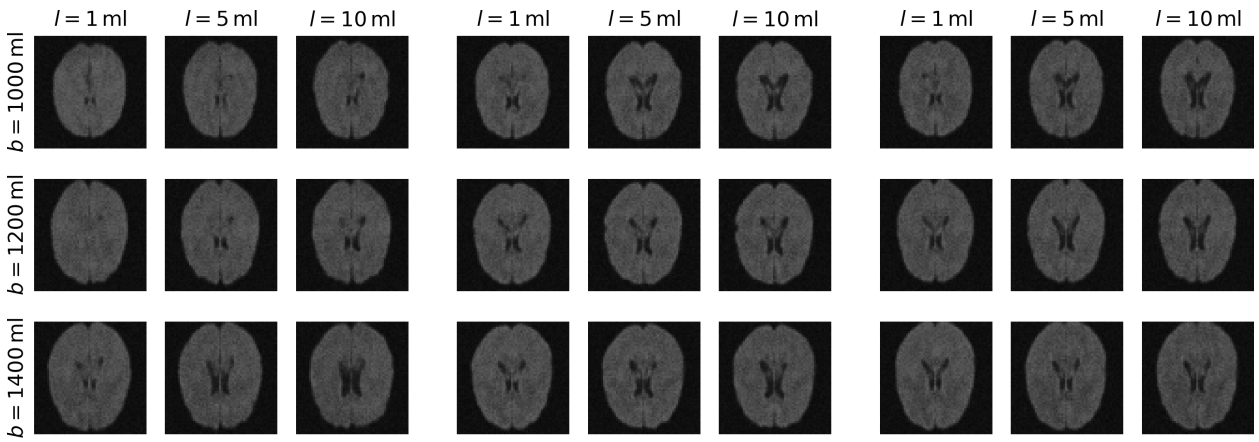

Figure 3: Conditional samples from the model trained on the CIS dataset. Images in each 3×3 block share the same the high-level noise vector, z_X . Each row consistently changes the brain size, however it is unclear if each column changes the lesion volume.

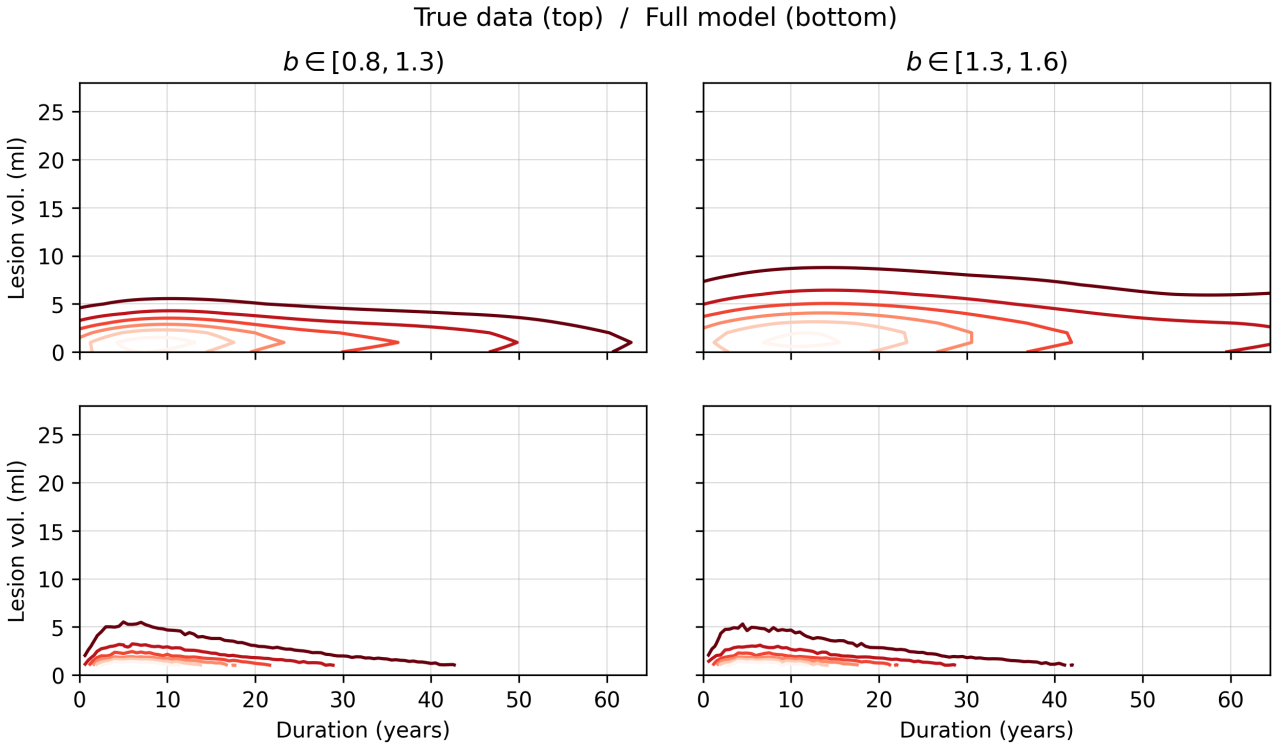

Figure 4: KDE showing the joint conditional distributions of duration of symptoms vs. lesion volume given a range of brain volumes: $p(a, v \mid b \in \cdot)$. The true data has no relationship between lesion volume and duration of symptoms.

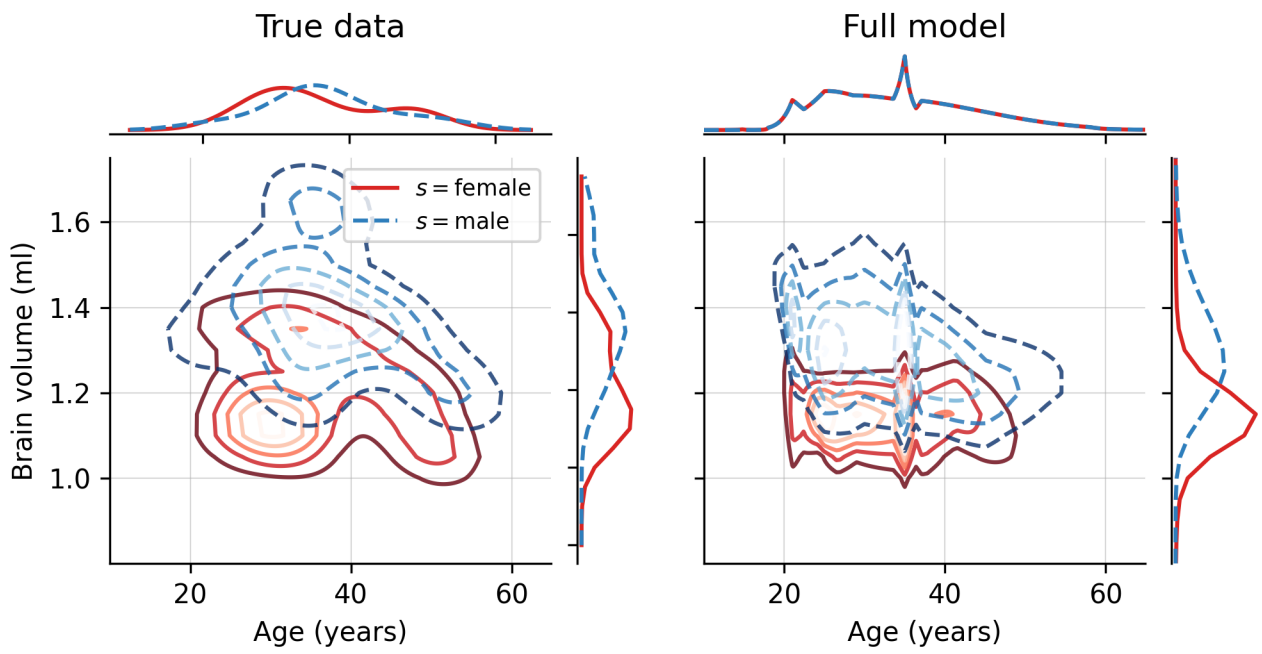

Figure 5: KDE showing the joint conditional distributions of Age vs. brain volume given biological sex: $p(a, b|s)$. Here we see differences in brain volumes, as well as a downward trend in brain volume in males as age progresses. This trend is less pronounced in females since the dataset includes relatively young subjects.

3.2 Counterfactuals

The learned SCM was not able generate convincing counterfactual images. The generated images do not preserve the unique shapes and cortical structures of the brain when changing the images. For example, in making the brain volume smaller, the edges of the brain are simply darkened, destroying the cortical structure instead of making the whole brain smaller.

Additionally, the model did not learn how to change the intensity of lesions in the images. The learned associations previously mentioned appear to have an affect overall. Increasing brain volume changing the biological sex from female to male and decreasing age all make the brain larger in the images, however not in a biologically realistic manner.

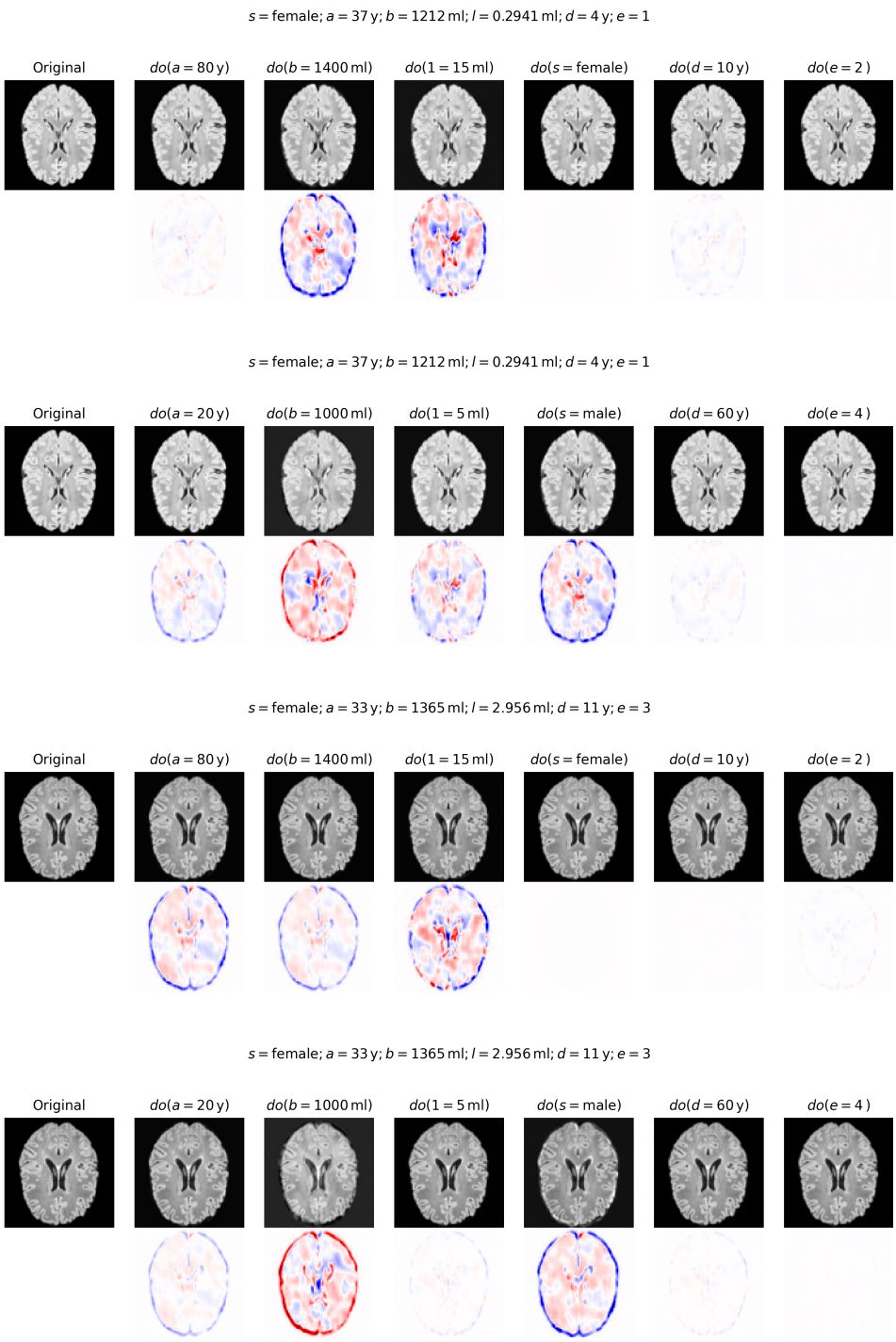

Figure 6: Counterfactuals showing different interventions on the same original brain. The first two and last two rows of plots depict the same subjects with two different sets of interventions to contrast the effects.

4 Discussion

In this work, a deep SCM is proposed to generate counterfactual images that could explain the effect of disease covariates on MS MR scans. This saw some success in learning associative relationships between covariates, meaning the model was able to learn the conditional distributions of variables reasonably well. However, realistic counterfactual images were not able to be produced. In applying counterfactuals that affect overall brain size, such as changing brain volume or sex, it can be observed that the unique cortical folding of each subject was not preserved. Rather, the brains are reduced in size simply by darkening the edges or increased in size by adding lighter pixels around the edges. Additionally, variables that should affect lesion size were observed to have no apparent affect on the appearance of lesions in the images.

Some of the lack of success of this can be attributed to the dataset used. Having access to more data could have improved the quality of the counterfactuals. The original work by Pawlowski was trained on 13,750 images from the UK Biobank and models anatomical data in the scope of healthy subjects. In contrast, the dataset used for this experiment contains only 684 scans from 161 unique subjects with MS. The model has no apparent effect on the lesions in the generated images. This could be related to the pre-processing of the data which measured lesion volumes for the whole brain volume and not per slice. Perhaps the relationship between the image and lesion volume was not learned because it is obscured by the method of pre-processing.

The proposed SCM also could be refined. Not all confounders were not included in the SCM. For example, the slice number was not included which could have led the model to over emphasize changes in the edges of the brain. In future work, other variables such as treatment type and ventricle volume should be included as well. Accounting for all relevant variables could only improve the plausibility of the SCM and its performance.

The limited success of this deep causal modelling experiment emphasizes the challenges and complexities involved in uncovering causal relationships using advanced techniques. The lessons learned from this experiment can guide future research into deep causal modelling, facilitating the development of more robust and interpretable models that accurately capture intricate causal mechanisms in psychiatric illnesses.

References

E. Bingham, J. P. Chen, M. Jankowiak, F. Obermeyer, N. Pradhan, T. Karaletsos, R. Singh, P. Szerlip, P. Horsfall, and N. D. Goodman. Pyro: Deep universal probabilistic programming. *The Journal of Machine Learning Research*, 20(1):973–978, 2019.

L. Dinh, J. Sohl-Dickstein, and S. Bengio. Density estimation using real nvp, 2017.

R. Dobson and G. Giovannoni. Multiple sclerosis–a review. *European journal of neurology*, 26(1): 27–40, 2019.

L. G. Neuberg. Causality: models, reasoning, and inference, by judea pearl, cambridge university press, 2000. *Econometric Theory*, 19(4):675–685, 2003.

G. Pawlowski, Coelho de Castro D. Deep structural causal models for tractable counterfactual inference. *NEURIPS*, 2020.

D. Ravi, D. C. Alexander, N. P. Oxtoby, and A. D. N. Initiative. Degenerative adversarial neuroimage nets: generating images that mimic disease progression. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22*, pages 164–172. Springer, 2019.

J. C. Reinhold, A. Carass, and J. L. Prince. A structural causal model for mr images of multiple sclerosis. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24*, pages 782–792. Springer, 2021.

B. Schölkopf. Causality for machine learning. In *Probabilistic and Causal Inference: The Works of Judea Pearl*, pages 765–804. 2022.

G. Varoquaux and V. Cheplygina. Machine learning for medical imaging: methodological failures and recommendations for the future. *NPJ digital medicine*, 5(1):48, 2022.