Investigating Causality Between Genotype And Clinical Phenotype In Neurological Disorders Using Structural Causal Model and Normalizing Flow

Fanyang Yu
Department of Bioengineering
University of Pennsylvania
yfy@seas.upenn.edu

Rongguang Wang
Department of Electrical and Systems Engineering
University of Pennsylvania
rgw@seas.upenn.edu

Pratik Chaudhari
Department of Electrical and Systems Engineering
University of Pennsylvania
pratikac@seas.upenn.edu

Christos Davatzikos
Department of Radiology
University of Pennsylvania
Christos.Davatzikos@pennmedicine.upenn.edu

Abstract

Understanding the causal relationship between genotype and clinical phenotype is crucial for disease treatment and prognosis. Despite the existing literature on exploring associations of genetics with clinical phenotypes such as imaging patterns and survival in various diseases, there are few to none work address the causation of these correlated genotypes. This paper leverages recent advances in causal deep learning to formulate the phenotypical outcome given the change in genotype as a causal inference problem. We build upon structural causal model (SCM) with normalizing flows parameterized by deep networks to perform the counterfactual query to investigate the causal relationship between genotype and clinical phenotype in two types of neurological disorders. Specifically, we focus on the causal effect of (1) APOE4 allele on brain volumetric measures in Alzheimer's disease; (2) key driver gene mutations on overall survival (OS) in glioblastoma. Experimental results show that APOE4 noncarriers causally lead to greater gray matter atrophy in the frontal lobe, and survival-correlated genes do not exhibit causal effect on OS in glioblastoma.

1 Introduction

Uncovering the causality between genotype and clinical phenotype is of great interest to scientific research to further the understanding of disease mechanism as well as clinical practice to develop targeted therapy. While genetics could potentially play a significant role with respect to neurological disorders as shown from association studies Shen and Thompson [2019], the causal relationship between genotype and the phenotypical characteristics of these diseases is not known.

Causal deep learning (DL) models integrate causality into DL research which enable the answer to interventional and counterfactual questions. This paper builds upon the deep structural causal

1st Workshop on Deep Generative Models for Health at NeurIPS 2023.

model Pawlowski et al. [2020], Wang et al. [2021] to formulate the causes (APOE4 allele or gene mutation status, sex, age, imaging phenotype) and effects (brain volumetric measures or survival days after surgery) via a causal graph and learn the structural assignments by using normalizing flows parameterized by deep networks. To the best of our knowledge, this work should be among the first to apply causal DL models for studying the causality between genotype and clinical phenotype, and can be further extended to identifying causal relationship in other applications of biomedical domain.

Our work mainly focuses on two neurological disorders, Alzheimer's disease (AD) and glioblastoma (GBM), to demonstrate the utility of causal deep learning with generative models for healthcare. Alzheimer's disease is the most prevalent neurodegenerative disease and affects millions of people globally Association et al. [2016]. Apolipoprotein E (APOE4) is known as the major genetic risk factor for AD, but few studies have shown the phenotypical differences between APOE4 carrier with non-carrier Emrani et al. [2020]. We formulate the effect of APOE4 as a causal inference problem as the counterfactual question of "how the brain volumes will change if the APOE4 allele changes" is asked. Glioblastoma is the most common and aggressive malignant primary adult brain tumor with a median overall survival (OS) of 14.6 to 16.7 months Stupp et al. [2017]. Maximal safe surgical resection combined with radiotherapy and temozolomide (TMZ) chemotherapy has been the standard treatment pipeline for GBM in clinical practice Akbari et al. [2016]. There have been studies on using multi-omics data, i.e., clinical information, radiomics and genomics data, to predict the OS and derive radiomic signature or survival prediction index Fathi Kazerooni et al. [2022]. As a result, these studies often present the correlation between the risk factors and the OS without addressing the causal aspect. We also treat this similarly as a causal inference problem where the causal relationships between 5 key driver genes (EGFR, NF1, PTEN, RB1, TP53) and OS are studied.

2 Dataset

Alzheimer's Disease We select 983 participants with mild cognitive impairment (MCI) from The Alzheimer's Disease Neuroimaging Initiative (ADNI) with only baseline included. 1.5T and 3T MRI were acquired for all subjects upon which a fully automated pipeline was applied for processing T1 structural MRI data. Intensity inhomogeneities were corrected for the scans Sled et al. [1998] followed by extra-cranial material removal via a multi-atlas skull stripping algorithm Doshi et al. [2013]. A multi-atlas label fusion method Doshi et al. [2016] is then used for identifying 145 anatomical regions of interest (ROIs) in gray matter (GM, 119 ROIs), white matter (WM, 20 ROIs) and ventricles (6 ROIs). As for APOE4, we denote homozygote as 2, heterozygote as 1 and non-carrier as 0, and in the deep SCM formulation we map all carrier to 1 for the binary experimental setting. We investigate the causal relationship between APOE4 and 145 brain ROI measures while age and sex are the confounders.

Glioblastoma Multi-parametric MRI (mpMRI) scans (T1, T1-Gd, T2, T2-FLAIR, DSC, DTI) of patients diagnosed with glioblastoma were retrospectively collected. Radiomics features, including histograms, morphological and textural descriptors, were derived using Cancer Imaging Phenomics Toolkit (CaPTk) Davatzikos et al. [2018]. We applied non-negative matrix factorization (NMF) to reduce the dimension of imaging features and used the coefficients as covariates. Genetic markers were obtained through a targeted next generation sequencing (NGS) panel on the resected tumor samples from the patients. A total number of 27 genes are included in the NGS panel from which 5 most frequent genes were selected for analysis. Clinical information include sex, age at the scan and survival days from the surgery. The final cohort consists of $n = 181$ IDH-wildtype patients with all radiomics, genomics, and clinical measures available. Specifically, we investigate the causal relationship of the survival days with respect to the 5 gene mutation status (EGFR, NF1, PTEN, RB1, TP53), while age, sex, 3 NMF coefficients are the other confounders.

3 Methodology

3.1 Structural Causal Model

A structural causal model (SCM) denoted as $M = (\mathbf{S}, P(\epsilon))$ consists of a collection $\mathbf{S} = (f_1, \ldots, f_K)$ of structural assignments $x_k = f_k(\epsilon_k; \mathbf{pa}_k)$, where $\mathbf{pa}_k$ represents the set of parents (direct causes) of x_k and ϵ_k are noise variables of the unknown sources of variation for x_k. A SCM satisfies the causal Markov condition that each variable x_k is independent of its non-effects given the direct causes. There-

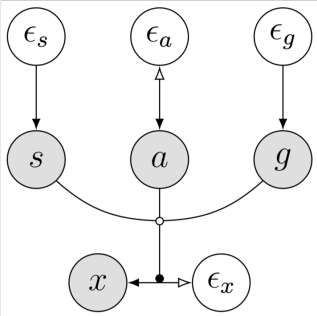

Figure 1: Causal graph of SCM for Alzheimer's disease where the causal effect of APOE4 on brain ROI measures is studied. The nodes consist of sex (s), age (a), gene (g) and brain ROIs (x), with corresponding exogenous noise terms ($\epsilon_s, \epsilon_a, \epsilon_g, \epsilon_x$). Invertible normalizing flow models are denoted by bidirectional arrows. The black dot indicates a conditional flow where x is conditioned on the parent nodes s, a and g. The causal graph for the glioblastoma case follows the same logic where more nodes are added for incorporating the NMF coefficients and multiple key driver genes besides sex and age, while the effect being the survival days after surgery.

fore the joint observational distribution of a SCM can be factorised as $P_M(\mathbf{x}) = \prod_{k=1}^{K} P_M(x_k|\mathbf{pa}_k)$, where each conditional distribution $P_M(x_k|\mathbf{pa}_k)$ is characterized by the corresponding structural assignment and noise distribution. The joint distribution of exogenous noise variables can be expressed as $P(\epsilon) = \prod_{k=1}^{K} P(\epsilon_k)$.

Given the above formulation, we perform a counterfactual query following the three-step process: abduction, action and prediction Pearl et al. [2000]. The abduction step is the prediction of the exogenous noise ϵ given the observations $\mathbf{x}$ which is to infer the posterior $P_M(\epsilon|\mathbf{x})$. Next comes the action step where an intervention, denoted by $\text{do}(x_k := \tilde{x_k})$, is conducted according to the desired manipulation. A modified SCM $\tilde{M} = M_{\text{do}(\tilde{x})} = (\tilde{\mathbf{S}}, P_M(\epsilon|\mathbf{x}))$ is obtained as a result. Finally, in the prediction step, the counterfactuals are predicted through sampling from the distribution $P_{\tilde{M}(x)}$ entailed by the modified SCM.

3.2 Normalizing Flow

Normalizing flow can model a complex probability density by applying transformation to simpler base distribution with same dimensionality. Given an observed variable x, an invertible and differentiable transformation $f(\cdot)$, and base variable $\epsilon \sim P(\epsilon)$ such that $x = f(\epsilon)$, the output density is computed as $p(x) = p(\epsilon)|\text{det}\nabla f(\epsilon)|^{-1}$ where $\epsilon = f^{-1}(x)$ and $\nabla f(\epsilon)$ as the Jacobian of the flow $f : \epsilon \mapsto x$. Denote a dataset with sample size of n as $D = \{x^i \sim p(x)\}_{i=1}^{n}$, a normalizing flow can be fit by maximizing the likelihood to obtain the parameters:

$$\theta^* = \text{argmax}\frac{1}{n}\sum_{i=1}^{n} \log p(\epsilon^i) - \log|\text{det}\nabla f_\theta(\epsilon^i)|$$

where we use a deep neural network to parameterize the normalizing flow. The above approach can be used for the conditional densities $P_M(x_k|\mathbf{pa}_k)$ in SCM. The flows f_{θ_k} will be fit to map exogenous noise ϵ to effect x_k (e.g., brain volumes or survival days) given parents $\mathbf{pa}_k$ (e.g., APOE4, sex, age, NMF coefficents of radiomics features, gene mutation). The abduction step involving the flow is essentially $\epsilon^i = f_\theta^{-1}(x^i)$. We can manipulate the value of a particular variable for the flow as the intervention $\text{do}(\tilde{x_k})$, as in our case we change the APOE4 allele or gene mutation status from 0 (non-carrier / non-mutant) to 1 (carrier / mutant) and vice versa. Then the modified flow $\tilde{f}_\theta$ can be conducted to sample from $P_{\tilde{M}(x)}$ to get counterfactuals (brain volumes w.r.t. change in APOE4 / survival days w.r.t. change in gene mutation status). Statistical testing will be applied to examine whether the counterfactual and original variable of interest has significant differences to imply if the causality can be established or not.

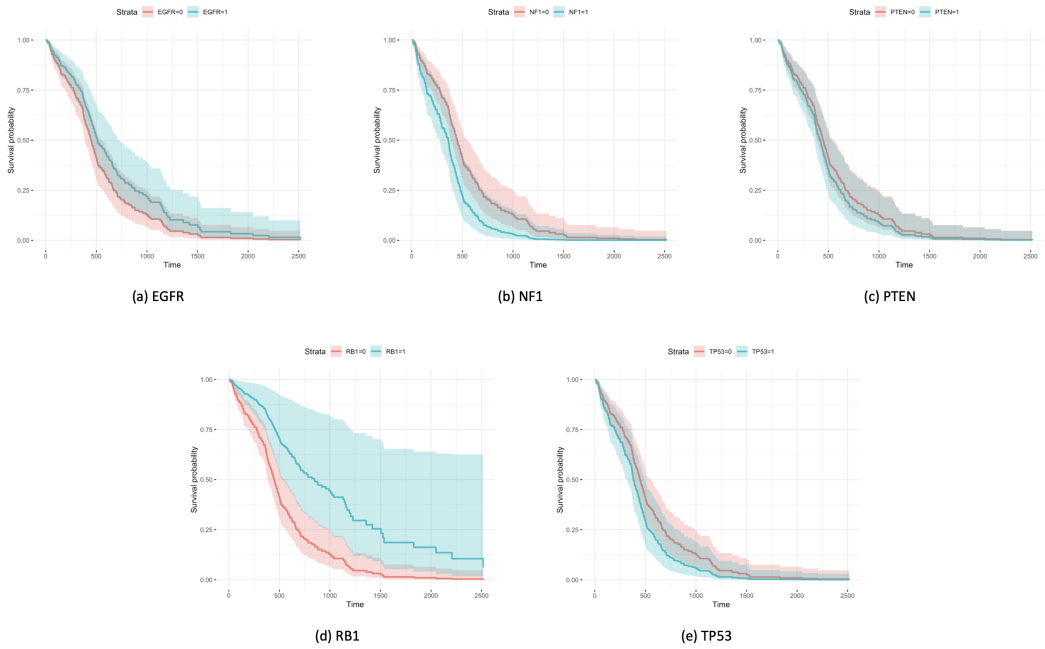

Figure 2: Survival curves show the association of gene mutations with survival after surgery in GBM patients. NF1 and RB1 are significantly associated with survival among the 5 most frequent genes.

4 Results

The flow-based SCM is implemented with quadratic autoregressive spline Durkan et al. [2019] by using PyTorch and Pyro Bingham et al. [2019]. For training details, we use Adam as the optimizer, set learning rate as 3×10^{-4}, weight decay as 10^{-4} and the number of epochs as 100 for both experiments.

Alzheimer's Disease As for the deep SCM implementation details, categorical distribution is used for sex and APOE4, and we use real-valued normalizing flow for other structural assignments. Specifically, the structural assignment for age is linear flow, while that for brain ROI volumes is a conditional flow conditioned on the activations of a fully-connected network with age, sex and APOE4 as input. The counterfactual outcomes of 145 ROI volumes are generated by changing either APOE4 status from non-carrier to carrier or carrier to non-carrier. Then we apply t-test adjusted for multiple testing (Bonferroni correction) comparing the original volume with the counterfactual outcome of each ROI, and find the region of right medial orbital gyrus has decreasing volume for APOE4 non-carriers. This observation is consistent with previous literature where the study shows non-carriers had greater frontoparietal atrophy Wolk et al. [2010].

Glioblastoma The multivariate Cox-PH model and statistical analysis is performed in R. Figure 2 shows that NF1 (p-value: $0.0133 < 0.05$) and RB1 (p-value: $0.0072 < 0.05$) were significantly associated with survival days after surgery. Categorical distribution is used for sex and gene mutations, with real-valued normalizing flow for others. The structural assignments for age and NMF coefficients are linear flows, while that for survival days is a conditional flow conditioned on the activations of a fully-connected network with age, sex, NMF coefficients and gene mutations as input. Figure 3 (b-c) show the distributions of factual and counterfactual survival days under the intervention of specific gene mutation status. No statistical significance was observed from the t-test. The genetic effects on survival days from the counterfactuals are in accordance with literature, e.g., mutation in RB1 may prolong the survival Dono et al. [2021]. These results indicate that correlation may not imply causation and it is desired to uncover the causal factors underlying the disease prognosis.

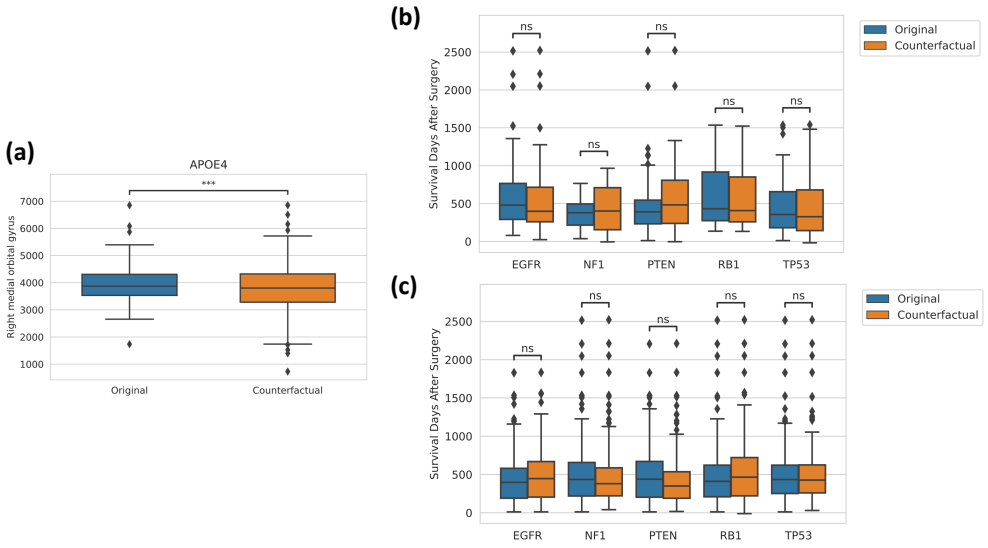

Figure 3: **(a) Results from Alzheimer's disease** Intervention on APOE4 to change its status of carrier (homozygote and heterozygote) to non-carrier where the counterfactual outcome of the right medial orbital gyrus shows significant decrease in the volume. **(b-c) Results from glioblastoma** Intervention on a specific gene to alter its status from mutant to non-mutant to observe how the counterfactual survival days would change comparing with the original survival days of subjects with mutant status of that gene. **(c)** follows the same logic as **(b)** while the alteration direction is from non-mutant to mutant.

5 Conclusion

The current study demonstrated the utility of deep generative models with causal framework in biomedical domain and drew the conclusion that APOE4 non-carrier can cause greater atrophy in frontal lobe in AD patients, and significantly associated genes may not be the underpinning causes for influencing the survival days after surgery in glioblastoma patients. The limitations of this study includes the model scalability for the input nodes and the lack of validation for the clinical findings. Future directions include defining causal graphs with intermediate nodes for secondary phenotypes, i.e, taking account of the causal relationship from gene to imaging features via causal mediation, adding more genes or other risk factors into the model, and also applying the causal inference method on multi-institutional data consortium Davatzikos et al. [2020] with larger sample size.

6 Broader Impact

Uncovering causality in complex real-world scenarios is challenging in nature, especially for biomedical domain where the clinically relevant conclusions and decision-making require rigorous validation Shen et al. [2020]. However, it is quite hard to validate the causal findings, i.e., as we refer to the literature on causal discovery where most of the methodological work rely on synthetic or benchmark dataset Yu et al. [2019], which can be less ideal than the realistic cases.

This work serves as a novel application of the deep SCM framework to provide a deeper understanding of causality between genotype and clinical phenotype in two real-world datasets of neurological disorders, thereby shed light upon the causality research for healthcare, i.e., causal inference and causal discovery from high-dimensional biomedical data. That said, the results from the current work are in very preliminary stages and demand further investigations as well as clinical validation. We do not foresee any negative social impact of this work since it is still far from being used for clinical decision-making. We hope our work can motivate applications of causal deep learning models in other diseases, such as different types of cancers or neurodegenerative diseases for conducting genotype-phenotype studies.

References

Li Shen and Paul M Thompson. Brain imaging genomics: integrated analysis and machine learning. *Proceedings of the IEEE*, 108(1):125–162, 2019.

Nick Pawlowski, Daniel Coelho de Castro, and Ben Glocker. Deep structural causal models for tractable counterfactual inference. *Advances in Neural Information Processing Systems*, 33: 857–869, 2020.

Rongguang Wang, Pratik Chaudhari, and Christos Davatzikos. Harmonization with flow-based causal inference. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24*, pages 181–190. Springer, 2021.

Alzheimer's Association et al. 2016 alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 12(4):459–509, 2016.

Sheina Emrani, Hirra A Arain, Cassandra DeMarshall, and Tal Nuriel. Apoe4 is associated with cognitive and pathological heterogeneity in patients with alzheimer's disease: a systematic review. *Alzheimer's Research & Therapy*, 12:1–19, 2020.

Roger Stupp, Sophie Taillibert, Andrew Kanner, William Read, David M Steinberg, Benoit Lhermitte, Steven Toms, Ahmed Idbaih, Manmeet S Ahluwalia, Karen Fink, et al. Effect of tumor-treating fields plus maintenance temozolomide vs maintenance temozolomide alone on survival in patients with glioblastoma: a randomized clinical trial. *Jama*, 318(23):2306–2316, 2017.

Hamed Akbari, Luke Macyszyn, Xiao Da, Michel Bilello, Ronald L Wolf, Maria Martinez-Lage, George Biros, Michelle Alonso-Basanta, Donald M O'Rourke, and Christos Davatzikos. Imaging surrogates of infiltration obtained via multiparametric imaging pattern analysis predict subsequent location of recurrence of glioblastoma. *Neurosurgery*, 78(4):572–580, 2016.

Anahita Fathi Kazerooni, Sanjay Saxena, Erik Toorens, Danni Tu, Vishnu Bashyam, Hamed Akbari, Elizabeth Mamourian, Chiharu Sako, Costas Koumenis, Ioannis Verginadis, et al. Clinical measures, radiomics, and genomics offer synergistic value in ai-based prediction of overall survival in patients with glioblastoma. *Scientific Reports*, 12(1):8784, 2022.

John G Sled, Alex P Zijdenbos, and Alan C Evans. A nonparametric method for automatic correction of intensity nonuniformity in mri data. *IEEE transactions on medical imaging*, 17(1):87–97, 1998.

Jimit Doshi, Guray Erus, Yangming Ou, Bilwaj Gaonkar, and Christos Davatzikos. Multi-atlas skull-stripping. *Academic radiology*, 20(12):1566–1576, 2013.

Jimit Doshi, Guray Erus, Yangming Ou, Susan M Resnick, Ruben C Gur, Raquel E Gur, Theodore D Satterthwaite, Susan Furth, Christos Davatzikos, Alzheimer's Neuroimaging Initiative, et al. Muse: Multi-atlas region segmentation utilizing ensembles of registration algorithms and parameters, and locally optimal atlas selection. *Neuroimage*, 127:186–195, 2016.

Christos Davatzikos, Saima Rathore, Spyridon Bakas, Sarthak Pati, Mark Bergman, Ratheesh Kalarot, Patmaa Sridharan, Aimilia Gastounioti, Nariman Jahani, Eric Cohen, et al. Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. *Journal of medical imaging*, 5(1):011018–011018, 2018.

Judea Pearl et al. Models, reasoning and inference. *Cambridge, UK: CambridgeUniversityPress*, 19 (2):3, 2000.

Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Neural spline flows. *Advances in neural information processing systems*, 32, 2019.

Eli Bingham, Jonathan P Chen, Martin Jankowiak, Fritz Obermeyer, Neeraj Pradhan, Theofanis Karaletsos, Rohit Singh, Paul Szerlip, Paul Horsfall, and Noah D Goodman. Pyro: Deep universal probabilistic programming. *The Journal of Machine Learning Research*, 20(1):973–978, 2019.

David A Wolk, Bradford C Dickerson, Alzheimer's Disease Neuroimaging Initiative, Michael Weiner, Marilyn Aiello, Paul Aisen, Marilyn S Albert, Gene Alexander, Heather S Anderson, Karen Anderson, et al. Apolipoprotein e (apoe) genotype has dissociable effects on memory and attentional–executive network function in alzheimer's disease. *Proceedings of the National Academy of Sciences*, 107(22):10256–10261, 2010.

Antonio Dono, Arvind V Ramesh, Emily Wang, Mauli Shah, Nitin Tandon, Leomar Y Ballester, and Yoshua Esquenazi. The role of rb1 alteration and 4q12 amplification in idh-wt glioblastoma. *Neuro-Oncology Advances*, 3(1):vdab050, 2021.

Christos Davatzikos, Jill S Barnholtz-Sloan, Spyridon Bakas, Rivka Colen, Abhishek Mahajan, Carmen Balaña Quintero, Jaume Capellades Font, Josep Puig, Rajan Jain, Andrew E Sloan, et al. Ai-based prognostic imaging biomarkers for precision neuro-oncology: the respond consortium. *Neuro-oncology*, 22(6):886–888, 2020.

Xinpeng Shen, Sisi Ma, Prashanthi Vemuri, and Gyorgy Simon. Challenges and opportunities with causal discovery algorithms: application to alzheimer's pathophysiology. *Scientific reports*, 10(1): 2975, 2020.

Yue Yu, Jie Chen, Tian Gao, and Mo Yu. Dag-gnn: Dag structure learning with graph neural networks. In *International Conference on Machine Learning*, pages 7154–7163. PMLR, 2019.