Preprints.org

Causal Fuzzy Deep Learning Algorithm for the Differentiation of Gliosarcoma From Glioblastoma

Usef Faghihi [*] , Cyrus Kalantarpour , Ismaila Baldé [*] , Amir Saki

Article

Causal Fuzzy Deep Learning Algorithm for the Differentiation of Gliosarcoma from Glioblastoma

Usef Faghihi [1,*], Cyrus Kalantarpour [1], Ismaila Baldé [2,*] and Amir Saki [1]

[1] University of Quebec at Trois-Rivieres, QC, Canada; cyrus.kalantarpour@uqtr.ca (C.K.); amir.saki@uqtr.ca (A.S.)

[2] Université de Moncton, QC, Canada

* Correspondence: usef.faghihi@uqtr.ca (U.F.); ismaila.balde@umoncton.ca (I.B.)

Abstract. In this paper, we introduce an innovative approach to distinguish Gliosarcoma (GSM) from Glioblastoma (GBM). Our method combines causal fuzzy logic rules with the Big Bird architecture, a Transformer-based Deep Learning algorithm. Unlike prior research, which often relied on statistical models to reduce dataset dimensions before causal analysis, our approach harnesses the complete dataset in tandem with our causal fuzzy Big Bird architecture. Additionally, we benchmark our results not only against previous Gliosarcoma/Glioblastoma studies but also against GPT-2 for a comprehensive evaluation.

Keywords. fuzzy logic; causal inference; Causal Deep Learning algorithms

1. Introduction

Glioblastoma (GBM) is one of the most common brain cancers [1]. GBM generally happens in the glial cells [2] and has several variants. Gliosarcoma (GSM) is a cancer classified by the World Health Organization (WHO) as a variant of GBM [1,3]. It is particularly important to accurately distinguish GSM from GBM [3]. Radiomics is an emerging field for imaging data analysis. It has been successfully employed for the differentiation of the central nervous system tumors [4]. In a recent paper, Qian et al. [4] conducted a study to identify an optimal machine learning algorithm for the differentiation of GSM from GBM based on radiomics data analysis. Baldé and Ghosh [20] studied the causal effect of Edema on the differentiation of GSM from GBM by using machine learning algorithms in radiomics. To do so, the authors developed a two steps procedure: In the first step, they reduced the dataset dimension using the sure independence screening procedure [5,6]. In the second step, they employed the outcome adaptive lasso [7] or the generalized outcome adaptive lasso [8].

To find the causes of Gliosarcoma, most of the previous studies use dimension reduction and/or statistical machine learning algorithms.

In this paper, we will use the entire dataset without applying any dimension reduction algorithms. Specifically, we will use causal fuzzy logic rules from [9], combined with Transformer-based deep learning algorithms [10] to identify the possible causes of Gliosarcoma. To the best of our knowledge, we are the first to use Causal Fuzzy Deep Learning Algorithm to detect the causes of Gliosarcoma.

2. Causal reasoning

In causal reasoning, we are looking how or why something happened and the relationship between the causes and their effects. In causal reasoning, we examine the relationship between causes and effects, and we try to understand how or why something happened. To calculate the causes of an event, current causal models use Individual or Average Treatment Effect (I-ATE). For instance, Pearl [11] computes Average Causal Effect by subtracting the means of the treatment and control

groups. Pearl uses Directed Acyclic Graphs (DAGs) to visualize and compute the associations or causal relationships between a set of elements. He also uses do operators, which are interventions on the nodes of DAGs, as well as probability theory, the Markov assumptions, and other concepts/methods/tools [11].

However, in [12] the authors showed that ATE describes the linear relationship between variables and in some examples cannot reflect the causality. Also, Pearl's approach to causation does not allow reasoning in cases of degrees of uncertainty [9]. Since *do* operators cut the relation between two nodes, Pearl's approach cannot answer gradient questions such as: given that you smoke a little, what is the probability that you have cancer to a certain degree? To solve Pearl's do operator problem the authors in [9,13], used fuzzy logic rules which implement human language nuances such as "*small*" instead of mere zero or one. Furthermore, Janzing [14] showed that at a macro level, Pearl's causal model works well with situations that are rare, such as rare medical conditions but, at a micro level, fails with bidirectional nodes. Authors in [12] showed that Janzing's model [15] works well with bidirectional nodes, but fails with situations that are rare [12].

In [9] which uses fuzzy logic as the fundamental part of their causal model, there are two types of rules which we call association and causal rules. An association rule can be of type A → B. A causal rule can be ~A & B. That is, tell me what happens to B when A is missing. In other words, using fuzzy logic rules, we can estimate more than fourteen values in the presence and the absence of each element (in the dataset context this becomes columns or variables).

In [9], instead of cutting the relationships between the treatment and its confounding parents, the model using fuzzification method assigns fuzzy membership values, such as very low, low, medium, and high, to the Treatment. To automatically assign membership values, the authors used fuzzy c-mean algorithm [16]   which is widely used for clustering the datasets using fuzzy logic.

Once, the membership assignments step is done, the model then applies causal fuzzy **interventional** rules from [9] as following $max(1 - \mu(a), \mu(b))$. Doing so, it can calculates more than fourteen different possible membership degree based fuzzy **counterfactual** values for each variable using different causal fuzzy rules such as $max(1 - \mu(a), \mu(b))$, where $\mu(a)$ and $\mu(b)$ are the highest membership degrees of a and b among all considered fuzzy attributes for a and b, respectively [9]. We would like to put more emphasis that using the following causal rules, we are applying interventions and estimate fuzzy counterfactuals at the same time.

Here are some of our fuzzy interventional rules.

$max(1 - \mu(a), \mu(b))$,
$min(1 - \mu(a), \mu(b))$,
$max(\min(1 - \mu(a), 1 - \mu(b)), b)$,

We note that based on the model introduced in [9], each of the above rules has a meaning based on a subjective random selection. For instance, $min(1 - \mu(b), \mu(a))$ could be seen as the probability of subjectively selecting $A = a$ as F_A and $B = b$ as $F(\neg B)$, where F_A and F_B are the fuzzy attributes that $\mu(a)$ and $\mu(b)$ come from, respectively. The model then applies defuzzification to the dataset in order to obtain the outputs. The big changes in the fuzzy output values can indicate possible causes.

3. Model description

Causal models can reason but cannot learn, and Deep Learning algorithms can learn but have limited capacity for reasonings [17]. Causal fuzzy rules for the first time were integrated with DLs such as Variational Autoencoders [17] and Big Bird architecture [10]- the latter outperformed ChatGpt 2 in reasoning [10]. To handle long sequences, Big Bird is equipped with a sparse attention mechanism which reduces the computational complexity of self-attention from quadratic to linear.

Causal fuzzy rules, fuzzification and defuzzification steps were integrated with Big Bird as follows: the model first creates a matrix W with randomly generated values between zero and one

(Figure 1). W will represent the potential causal relationships between the columns of the dataset D [17]. To do this, the model first initialized matrix W with size (n*n) where n is the number of columns in D. It then multiplied W by the dataset. The result of this step was then fed into the Transformer's Embedding module. The output of the Embedding module was then fed into the Transformer's Encoder module which then partitioned using the Fuzzy c-means algorithm [16] (shown as FC-mean in Figure 1). This partitioning was done to automatically find fuzzy membership intervals without requiring an expert to define them. Using the output of the Fuzzy c-means algorithm, the architecture then used Fuzzification method (Figure 1) to fuzzify the dataset. In the next step, the Fuzzy causal rule set engine (Figure 1) applies the above causal fuzzy rules from [9] to each of the dataset D's columns. Finally, the architecture applies Defuzzification (Figure 1).
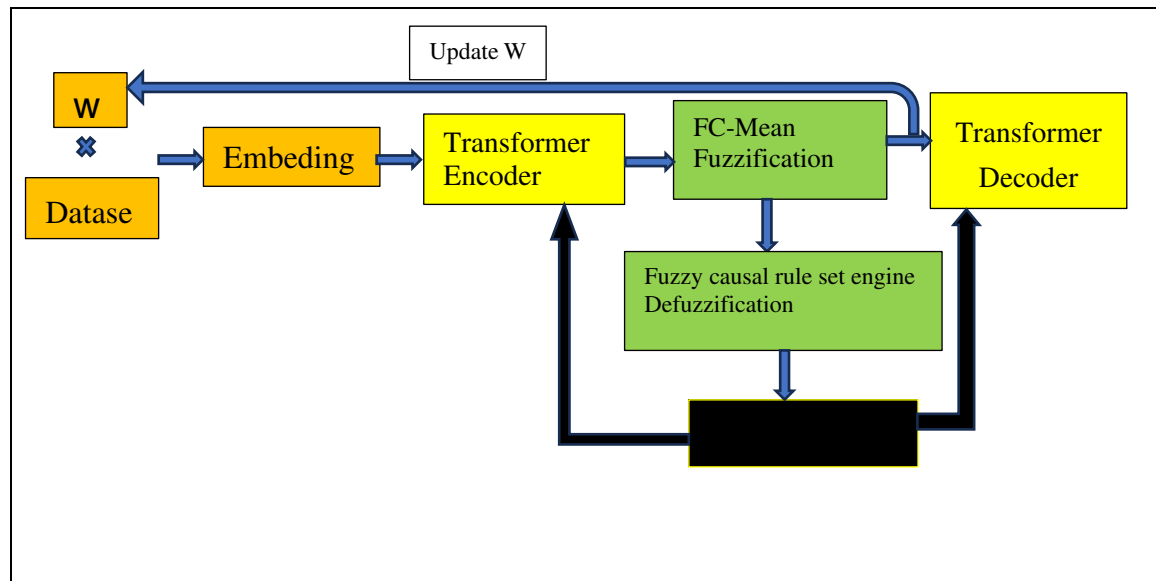

Figure 1. Causal Fuzzy Big Bird Architecture (CFT).

The next step was to incorporate fuzzy rules into the Big Bird's loss function. According to [5], this was done by dividing the Big Bird Loss Function by the variations of fuzzy causal reasoning rules obtained from the previous defuzzification step.

The result of the previous step was then multiplied by the original dataset before being fed into the Transformer's Encoder module. These steps were repeated until the architecture converged. The elements of the matrix W, which resulted from the previous steps, are possible causes.

3. Dataeset Description

We used the radiomics data studied in Qian et al. [4]. The dataset contained a sample size n=183 patients including 100 with GBM and 83 with GSM with 1 303 radiomic features extracted from MRI images. In our study, we took "Edema" A (Yes: 1 / No: 0) as the exposure variable (treatment) and the outcome variable Y (Y = 1 if the patient had gliosarcoma and Y = 0 if the patient had glioblastoma. We used 1 303 radiomics features as potential confounders of the relationship between A and Y.

5. Results

In this section, we compare the Causal Fuzzy Big Bird Architecture (CFT) [10] (code on GitHub) and the GPT-2 architecture [18] in terms of features selection and accuracy. GPT-2 architecture is a powerful deep learning algorithm which is described in the following address[1]. We applied both CFT and GPT-2 to analyze the radiomics dataset. The objective was to identify the key radiomics confounders of the causal relationship of Edema and Gliosarcoma. As these confounding features

---

[1] https://github.com/huggingface/transformers

can inform the distinction of GSM from GBM. Based on the analysis, the radiomics features were grouped into four categories (Please see Appendix): Group A, Group B, Group C, and Singleton D, each representing varying probabilities of being an important potential confounder of the relationship between Edema and GSM.

We added a new embedding layer of size 20K to GPT-2, a language model, so that it could learn from numerical data. We used the same embedding size for CFT. We split the dataset into 70% for training and 30% for testing. Since the dataset had only 183 rows, we augmented it by 200 folds. We trained both models for 500 epochs on a Nvidia A100 with 24 GB RAM.

Since, Big Bird and GTP-2 are both language models, we used perplexity formula [19] to to measure the accuracy of both models. As we can see, GPT-2's perplexity score is higher than CFT. However, this cannot be considered as a good measure for causal inference.

Table 1. Train/Test accuracy of the Architectures that we have used.

Architecture	Train Accuracy	Test Accuracy
GPT-2	96.5%	93%
CFT	95%	92%

One way to support this claim is to compare how GPT-2 and CFT performance on detecting the 'original_shape_Sphericity', which is the most important radiomics feature. GPT-2 fails to detect it, while CFT succeeds, even though CFT has a lower test accuracy than GPT-2.

In the following we provided the result of the CFT and GPT-2 and their comparison. Our results agree with the findings in [20], who identified the radiomics feature original_shape_Sphericity as one of the key factors informing the distinction of GSM from GBM. That is this feature is a strong potential confounder of the relationship of Edema and GSM.

Once more, we wish to underscore the methodology distinction. In prior research, including the work referenced in [20], the standard practice involved the initial application of dimension reduction techniques followed by the utilization of statistical and machine learning algorithms, such as Lasso [8], to derive their findings. In contrast, our approach for this study involved the direct utilization of the entire dataset with our architecture, without any alterations or preprocessing steps.

The Appendix shows the groups A, B, and C of variables that are most likely to contribute to the distinction of GSM from GBM by CFT and GPT-2. However, GPT-2 detected _fewer_ variables than CFT in each group.

Conclusion:

Unlike previous studies on gliosarcoma detection that relied on statistical models to reduce the size of the dataset and then used machine learning, we used the entire dataset for causal inference. To do so, we used causal fuzzy logic rules from [9] integrated with Big Bird architecture which is a Transformer based Deep Learning algorithm [10]. Our finding is in accordance with the findings in [8]. We also compared causal fuzzy Big Bird results with GPT-2 which is one of the well-known DLs. Causal fuzzy Big Bird outperformed GPT-2 by detecting "original_shape_Sphericity," as a key potential radiomics confounder of the relationship of Edema and GSM. It is worth mentioning that DLs need lots of data to learn. However, our causal fuzzy Big Bird architecture used very small amount of data to discover the causal variables. This is because our model is equipped with casual fuzzy logic rules from [9].

Appendix

In the following list, we have enumerated the variables detected by our CFT architecture. Those variables highlighted in yellow represent cases that were **not detected** by GPT-2 architecture.

Groupe D	Proportion of times each radiomics group features was selected
original_shape_Sphericity	**> 95%**

CFT output

Groupe C	Proportion of times each radiomics group features was selected
1. original_shape_MajorAxis, original_shape_Maximum2DDiameterRow, 2. original_shape_Sphericity, original_gldm_DependenceNonUniformity, 3. 3-0-mm-3D_glszm_GrayLevelNonUniformity, 4. original_shape_SurfaceVolumeRatio, log-sigma-3-0-mm-3D_glszm_SizeZoneNonUniformity, 5. log-sigma-3-0-mm-3D_gldm_DependenceNonUniformity, 6. log-sigma-5-0-mm-3D_glszm_GrayLevelNonUniformity, 7. LHH_glcm_ClusterShade, 8. waveletLHH_glcm_Idn, 9. wavelet-HLL_firstorder_10Percentile, 10. waveletHLL_firstorder_MeanAbsoluteDeviation, 11. wavelet-HLL_firstorder_Skewness, 12. waveletHLH_firstorder_Minimum, wavelet-HLH_glcm_JointAverage, 13. original_shape_SurfaceVolumeRatio, 14. waveletHLH_gldm_SmallDependenceHighGrayLevelEmphasis, 15. waveletHHL_gldm_LowGrayLevelEmphasis	**> 85%**

Groupe B	Proportion of times each radiomics group features was selected
1. original_shape_MajorAxis, 2. original_shape_Maximum2DDiameterRow, 3. original_shape_SurfaceVolumeRatio, 4. original_shape_Maximum3DDiameter, 5. original_shape_Sphericity, 6. original_glszm_SizeZoneNonUniformity, 7. original_gldm_DependenceNonUniformity, 3-0-mm-3D_glszm_GrayLevelNonUniformity, 8. logsigma-3-0-mm-3D_glszm_SizeZoneNonUniformity, 9. log-sigma-3-0-mm-3D_gldm_DependenceNonUniformity, 10. log-sigma-5-0-mm-3D_glszm_GrayLevelNonUniformity, 11. LHH_glcm_ClusterShade, 12. wavelet-LHH_glcm_Idn, 13. wavelet-HLL_firstorder_10Percentile, 14. wavelet-HLL_firstorder_MeanAbsoluteDeviation, 15. wavelet-HLL_firstorder_Skewness, 16. waveletHLH_firstorder_Minimum, 17. wavelet-HLH_glcm_JointAverage, 18. original_shape_SurfaceVolumeRatio, 19. waveletHLH_gldm_SmallDependenceHighGrayLevelEmphasis, 20. waveletHHL_gldm_LowGrayLevelEmphasis	**> 75%**

Groupe A	Proportion of times each radiomics group features was selected

1. original_shape_MajorAxis, original_shape_Maximum2DDiameterRow,
2. original_shape_Maximum2DDiameterSlice,
3. original_shape_Maximum3DDiameter,
4. original_shape_Sphericity, original_shape_SurfaceArea,
5. original_glszm_SizeZoneNonUniformity,
6. original_gldm_DependenceNonUniformity,
7.   log-sigma-3-0-mm-3D_glszm_GrayLevelNonUniformity,
8. log-sigma-3-0-mm-3D_glszm_SizeZoneNonUniformity,
9. log-sigma-3-0-mm-3D_gldm_DependenceNonUniformity,
10. log-sigma-5-0-mm-3D_glszm_GrayLevelNonUniformity,
11. wavelet-LLL_glszm_LargeAreaEmphasis,
12. wavelet-LLL_glszm_ZoneEntropy,
   21. wavelet-LLH_glrlm_GrayLevelNonUniformityNormalized,
   22. wavelet-LLH_glrlm_RunEntropy,
   23. wavelet-LHL_firstorder_Variance,
   24. waveletLHL_glcm_Correlation, wavelet-LHL_glrlm_LowGrayLevelRunEmphasis,
   25. waveletLHL_glrlm_RunVariance,
   26. wavelet-LHL_glszm_SmallAreaHighGrayLevelEmphasis,
   27. waveletLHL_glszm_ZoneVariance, wavelet-LHL_gldm_SmallDependenceLowGrayLevelEmphasis,
   28. wavelet-LHH_firstorder_Entropy,
   29. wavelet-LHH_firstorder_Variance,
   30. waveletLHH_glcm_ClusterShade,
   31. wavelet-LHH_glcm_Idn,
   32. wavelet-HLL_firstorder_10Percentile,
   33. wavelet-HLL_firstorder_MeanAbsoluteDeviation,
   34. wavelet-HLL_firstorder_Skewness,
   35. waveletHLH_firstorder_Minimum,
   36. wavelet-HLH_glcm_JointAverage,
   37. original_shape_SurfaceVolumeRatio,
	> 60%

38. wavelet-HLH_gldm_SmallDependenceHighGrayLevelEmphasis, waveletHHL_gldm_LowGrayLevelEmphasis	

In the following we show GPT-2's output:

Groupe D	Proportion of times each radiomics group features was selected
Null	**> 90%**

Groupe C	Proportion of times each radiomics group features was selected
1. original_shape_MajorAxis, original_shape_Maximum2DDiameterRow, 2. 3-0-mm-3D_glszm_GrayLevelNonUniformity, 3. original_shape_SurfaceVolumeRatio, 4. log-sigma-3-0-mm-3D_glszm_SizeZoneNonUniformity, 5. waveletHLH_firstorder_Minimum, wavelet-HLH_glcm_JointAverage, 6. original_shape_SurfaceVolumeRatio, 7. waveletHLH_gldm_SmallDependenceHigh GrayLevelEmphasis,	• **> 85%**

Groupe B	Proportion of times each radiomics group features was selected
1. original_shape_SurfaceArea, 2. original_glszm_SizeZoneNonUniformity, 3. original_gldm_DependenceNonUniformity, 4. wavelet-LLL_glszm_ZoneEntropy, 5. wavelet-LLH_glrlm_GrayLevelNonUniformityNormalized, 6. waveletLHL_glcm_Correlation, wavelet-LHL_glrlm_LowGrayLevelRunEmphasis, 7. wavelet-LHL_glszm_SmallAreaHighGrayLevelEmphasis, 8. waveletLHL_glszm_ZoneVariance, wavelet-LHL_gldm_SmallDependenceLowGrayLevelEmphasis, 9. wavelet-LHH_glcm_Idn, 10. wavelet-HLL_firstorder_10Percentile, 11. wavelet-HLL_firstorder_MeanAbsoluteDeviation, 12. waveletHLH_firstorder_Minimum, 13. wavelet-HLH_gldm_SmallDependenceHighGrayLevelEmphasis,	> 75%

• **Groupe B**	• Proportion of times each radiomics group features was selected
14. original_shape_MajorAxis, 15. original_shape_Maximum2DDiameter Row, 16. original_shape_SurfaceVolumeRatio, 17. original_shape_Maximum3DDiameter, 18. original_shape_MajorAxis, 19. waveletLHL_glrlm_RunVariance, 20. wavelet-LHH_firstorder_Entropy, 21. wavelet-LHH_firstorder_Variance, 22. wavelet-HLH_glcm_JointAverage, 23. original_shape_SurfaceVolumeRatio, 24. original_shape_Maximum2DDiameter Slice, 25. log-sigma-5-0-mm-3D_glszm_GrayLevelNonUniformity, 26. wavelet-LLH_glrlm_RunEntropy, 27. wavelet-LHL_firstorder_Variance, 28. original_gldm_DependenceNonUniformity, 29. log-sigma-3-0-mm-3D_gldm_DependenceNonUniformity, 30. log-sigma-5-0-mm-3D_glszm_GrayLevelNonUniformity, 31. log-sigma-3-0-mm-3D_gldm_DependenceNonUniformity, 32. original_shape_SurfaceVolumeRatio, 33. waveletHLH_gldm_SmallDependence HighGrayLevelEmphasis,	• **> 60%**

References

1. C. R. Miller and A. Perry, "Glioblastoma: morphologic and molecular genetic diversity," *Archives of pathology & laboratory medicine,* vol. 131, no. 3, pp. 397-406, 2007.
2. H. Ohgaki, "Epidemiology of brain tumors," *Cancer Epidemiology: Modifiable Factors,* pp. 323-342, 2009.
3. M. M. Zaki *et al.*, "Genomic landscape of gliosarcoma: distinguishing features and targetable alterations," *Scientific Reports,* vol. 11, no. 1, p. 18009, 2021.
4. Z. Qian *et al.*, "Machine learning-based analysis of magnetic resonance radiomics for the classification of gliosarcoma and glioblastoma," *Frontiers in Oncology,* vol. 11, p. 699789, 2021.
5. J. Fan and J. Lv, "Sure independence screening for ultrahigh dimensional feature space," *Journal of the Royal Statistical Society Series B: Statistical Methodology,* vol. 70, no. 5, pp. 849-911, 2008.

6.   D. Tang, D. Kong, W. Pan, and L. Wang, "Ultra-high dimensional variable selection for doubly robust causal inference," *Biometrics,* vol. 79, no. 2, pp. 903-914, 2023.

7.   S. M. Shortreed and A. Ertefaie, "Outcome-adaptive lasso: variable selection for causal inference," *Biometrics,* vol. 73, no. 4, pp. 1111-1122, 2017.

8.   I. Baldé, Y. A. Yang, and G. Lefebvre, "Reader reaction to "Outcome-adaptive lasso: Variable selection for causal inference" by Shortreed and Ertefaie (2017)," *Biometrics,* vol. 79, no. 1, pp. 514-520, 2023.

9.   U. Faghihi, S. Robert, P. Poirier, and Y. Barkaoui, "From Association to Reasoning, an Alternative to Pearl's Causal Reasoning," *In Proceedings of AAAI-FLAIRS 2020. North-Miami-Beach (Florida),* 2020.

10. C. Kalantarpour, U. Faghihi, E. Khelifi, and F.-X. Roucaut, "Clinical Grade Prediction of Therapeutic Dosage for Electroconvulsive Therapy (ECT) Based on Patient's Pre-Ictal EEG Using Fuzzy Causal Transformers," presented at the International Conference on Electrical, Computer, Communications and Mechatronics Engineering, ICECCME 2023, Tenerife, Canary Islands, Spain, 2023.

11. J. Pearl and D. Mackenzie, *The book of why: the new science of cause and effect*. Basic Books, 2018.

12. U. Faghihi and A. Saki, "Probabilistic Variational Causal Effect as A new Theory for Causal Reasoning," *arXiv preprint arXiv:2208.06269,* 2023.

13. S. Robert, U. Faghihi, Y. Barkaoui, and N. Ghazzali, "Causality in Probabilistic Fuzzy Logic and Alternative Causes as Fuzzy Duals," *ICCCI 2020,* vol. Ngoc-Thanh Nguyen, et al, 2020.

14. D. Janzing, L. Minorics, and P. Blöbaum, "Feature relevance quantification in explainable AI: A causal problem," presented at the International Conference on Artificial Intelligence and Statistics, 2020.

15. D. Janzing, D. Balduzzi, M. Grosse-Wentrup, and B. Schölkopf, "Quantifying causal influences," 2013.

16. J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Computers & geosciences*, vol. 10, no. 2-3, pp. 191-203, 1984.

17. U. Faghihi, C. Kalantarpour, and A. Saki, "Causal Probabilistic Based Variational Autoencoders Capable of Handling Noisy Inputs Using Fuzzy Logic Rules," presented at the Science and Information Conference, 2022.

18. A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," *OpenAI blog,* vol. 1, no. 8, p. 9, 2019.

19. F. Jelinek, R. L. Mercer, L. R. Bahl, and J. K. Baker, "Perplexity—a measure of the difficulty of speech recognition tasks," *The Journal of the Acoustical Society of America,* vol. 62, no. S1, pp. S63-S63, 1977.

20. I. Baldé, and D. Ghosh, " Ultra-high dimensional confounder selection algorithms comparison with application to radiomics data", arXiv preprint arXiv:2310.06315, 2023.