Iranian Journal of Fuzzy Systems

Volume 22, Number 6, (2025), pp. 167-182

Original paper

University of Sistan and Baluchestan

FuzzyCAL: A fuzzy-logic enhanced causal attention GNN for robust cocaine use disorder classification

M. Pakravan [1]

[1] *Department of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran*

mpakravan@modares.ac.ir

Abstract

Fuzzy logic has emerged as a powerful tool for handling uncertainty and imprecision in machine learning, enabling models to make robust predictions in complex, noisy domains. In this work, we integrate fuzzy-set theory with causal attention learning to improve the classification of Cocaine Use Disorder (CUD) from resting-state functional Magnetic Resonance Imaging (fMRI) brain networks. Building on the Causal Attention Learning (CAL) framework, we introduce *FuzzyCAL*, which augments node- and edge-level causal masks with fuzzy membership functions that quantify "weak," "medium," and "strong" relevance of graph components. This fusion of fuzzy logic and causal Graph Neural Networks (GNNs) not only preserves the interpretability of causal attention but also adapts to the inherent variability of neuroimaging data. Through extensive 5-fold cross-validation on the SUDMEX CONN dataset, FuzzyCAL achieves a mean test accuracy of 86.20% (±7.18%) and F1-score of 85.87% (±7.23%), outperforming both a non-causal GNN baseline and the original CAL model. We further demonstrate how fuzzy-weighted causal masks reveal anatomically meaningful biomarkers in temporal and sensorimotor cortices. Our results suggest that embedding fuzzy reasoning into causal graph models enhances both predictive performance and neuroscientific interpretability, offering a promising direction for precision diagnostics in substance use disorders.

Keywords: Fuzzy logic, causal attention, graph neural networks, cocaine use disorder, functional connectivity.

1 Introduction

Cocaine Use Disorder (CUD) remains a major public health challenge worldwide, characterized by persistent alterations in brain functional connectivity and impaired cognitive control [2]. Recent advances in Graph Neural Networks (GNNs) have demonstrated powerful capabilities for modeling complex interactions in brain networks, yielding promising classifiers for neuropsychiatric conditions [1, 28]. However, standard GNNs often fail to disentangle causal from spurious patterns, limiting their robustness under distributional shifts such as demographic or scanner variation [26].

Distinguishing causal mechanisms from mere associations is fundamental for deploying models that remain reliable under changes in environment or intervention. Seminal works by Pearl and colleagues laid the theoretical foundations, formalizing structural causal models and do-calculus for reasoning about interventions. These ideas are extended to statistical learning, demonstrating how causal inference can be integrated into predictive models for improved robustness [20]. More recently, Invariant Risk Minimization (IRM) has been proposed, enforcing that learned predictors rely on causal features shared across training environments [15], while Schölkopf *et al.* surveyed how causality can guide representation learning in deep networks [24]. In parallel, graph-based causal discovery methods, such as GNN-based structure learning [7, 8], have shown promise in uncovering causal graph structure directly from data.

Combining structural causal modeling with fuzzy-logic interpretability, yielding a graph classifier that robustly isolates causal subgraph patterns even under distributional shifts. Concurrently, the field of fuzzy systems has offered a rich set of tools for handling uncertainty and imprecision in real-world data. There are numerous studies on

Corresponding Author: M. Pakravan
Received: August 2025; Revised: October 2025; Accepted: November 2025.
https://doi.org/10.22111/ijfs.2025.52840.9343

fuzzy-enhanced classification, including fuzzy SVMs for biomedical signals [22], fuzzy clustering for EEG-based disease detection [29], and fuzzy rule-based deep architectures for image analysis [17]. These works highlight the strength of fuzzy membership functions and rule consistency in capturing heterogeneous and noisy biological signals. A fuzzy learning–integrated graph transformer model has also been shown to improve classification performance in complex, high-dimensional domains [3].

Beyond classical fuzzy classification and clustering methods, recent research has advanced higher-order fuzzy modeling for nonlinear and uncertain systems. For example, an *Interval Type-3 Fuzzy Wiener Model* is proposed for nonlinear dynamic processes, achieving improved uncertainty representation [10]. An *indirect adaptive interval type-3 fuzzy tracking control* strategy is introduced for nonlinear networked systems under Denial-of-Service attacks [12]. An *adaptive general type-2 fuzzy model-based control* framework is presented for nonlinear systems with packet dropouts [11]. Other recent advances include fuzzy-supervised proportional-integral (PI) controllers for flow-control processes and data-driven fuzzy tuning schemes validated experimentally [25]. While these methods demonstrate the growing capability of fuzzy systems in uncertainty handling and adaptive control, they primarily operate on low-dimensional dynamic systems rather than structured graph data. In contrast, our proposed method integrates fuzzy reasoning directly within a causal graph-neural framework, extending fuzzy logic beyond traditional control domains toward interpretable brain-network analysis.

Recent developments in graph-based neuroimaging have explored either fuzzy uncertainty modeling or causal representation learning, but rarely both in a unified framework. Fuzzy graph neural networks (FuzzyGNN) [14, 28] introduce membership-based uncertainty weighting to improve robustness, yet they do not explicitly disentangle causal from spurious substructures. Conversely, causal GNN models such as CAGNN [13] and Invariant Causal Representation Learning on Graphs [16] aims to isolate invariant features across environments but treats node and edge attributes as crisp values without modeling uncertainty or linguistic interpretability.

Our work bridges these two paradigms by embedding fuzzy logic directly into the causal-attention mechanism, thus enabling soft, interpretable, and uncertainty-aware disentanglement of causal versus contextual graph patterns. This integration yields an end-to-end model that simultaneously captures causal invariance and fuzzy uncertainty, offering improved robustness and neuroscientific interpretability compared with both standard causal GNNs and fuzzy deep networks.

In this work, we propose *FuzzyCAL*, the first fuzzy-logic extension of the recently introduced Causal Attention Learning (CAL) model [26], tailored to resting-state functional Magnetic Resonance Imaging (fMRI) graphs in CUD. Whereas CAL disentangles *causal* versus *shortcut* subgraph patterns via soft attention masks on nodes and edges, we augment both the graph extraction pipeline and the internal attention mechanisms with fuzzy membership functions and rule-consistency regularization. Specifically, we (i) fuzzify edge strengths into weak/medium/strong memberships and node importance into degree- and centrality-based fuzzy sets; (ii) integrate these fuzzy features into a gating mechanism controlled by hyperparameters $\alpha_{\text{node}}, \beta_{\text{edge}}$; and (iii) impose a fuzzy-rule consistency loss weighted by λ_{rule}. Our results on the SUDMEX CONN dataset demonstrate that FuzzyCAL consistently outperforms both a standard GNN baseline and the original CAL classifier. This fusion of causal graph learning with fuzzy uncertainty modeling represents a novel paradigm for robust neuroimaging classification.

To complement our analysis on real-world cocaine use disorder data, we also perform a controlled simulation study using synthetic graphs with known causal and trivial substructures.

Our contributions are as follows:

1. We adapt and apply the CAL causal disentanglement architecture for graph classification to the domain of cocaine use disorder using functional connectivity derived from rs-fMRI, providing a structured approach to isolate causal subnetworks in perturbed brain graphs.

2. We introduce *FuzzyCAL*, a fuzzy-logic enhanced extension that endows nodes and edges with graded membership values and incorporates fuzzy rule-based regularization into the causal/object-context decomposition, improving robustness to noise and offering more interpretable attention masks.

3. We design a training objective that combines the original CAL components (uniform regularization on contextual stream, supervised object and combined readouts) with fuzzy consistency terms that encourage alignment between learned masks and soft semantic priors.

4. We empirically validate our method on CUD neuroimaging data and simulated graphs, demonstrating superior out-of-distribution generalization and interpretability compared to baseline GNN and causal attention variants.

The remainder of this paper is organized as follows. In Section 2, we describe our data and preprocessing pipeline, including the Dataset and rs-fMRI Preprocessing, followed by Graph Extraction and Fuzzy Feature Construction, with

details on Fuzzy Edge Strengths, and Fuzzy Node Importance. Section 3 introduces our FuzzyCAL Model, starting from the underlying Structural Causal Model, through the original CAL Architecture, and culminating in our novel FuzzyCausal Extension and its Training Objective. In Section 4, we present our experimental findings, first on the real fMRI Data Study-including Extraction of Fuzzy-Causal Edges—and then on a controlled Simulation Study. We discuss the implications and limitations of our results in Section 5, and conclude in Section 6 with a summary of contributions and future directions.

2 Dataset and preprocessing

2.1 Dataset

We leveraged the publicly available SUDMEX CONN repository [2], which comprises multimodal MRI scans from 138 adults: 74 individuals meeting criteria for cocaine use disorder (CUD) and 64 demographically matched healthy controls. The CUD cohort included 9 females, while the control group included 6 females. For each participant, the dataset provides high-resolution structural T1-weighted images, diffusion-weighted images (DWI), and resting-state functional MRI (rs-fMRI). In this work, we focus exclusively on the rs-fMRI time series to characterize functional connectivity. Since fMRI effectively characterizes both cross-subject and multivoxel functional patterns in complex brain networks [18, 19], the fMRI modality was incorporated in the current study to enable comprehensive modeling of functional interactions.

2.2 rs-fMRI preprocessing

All rs-fMRI data were preprocessed in AFNI [6] via the following routine steps: Slice-timing correction, Motion correction, Coregistration and Normalization, Spatial smoothing, Temporal filtering between 0.01–0.15Hz, Nuisance regression, and Brain masking.
We also parcellated each preprocessed volume using the Harvard–Oxford cortical atlas, yielding 48 cortical regions of interest (ROIs). The resulting dataset consists of a $T \times N$ matrix, where $T = 200$ time points and $N = 48$ ROIs; each entry is the average Blood-Oxygen-Level-Dependent (BOLD) signal in that region at a given time.

2.3 Graph extraction

We begin with pre-parcellated resting-state fMRI time series for each subject, yielding a matrix $\mathbf{B} \in \mathbb{R}^{T \times N}$ where T is the number of time points and N the number of brain regions (nodes). For each subject, we construct a functional brain graph in the following steps.

2.4 Functional connectivity and sparsification

To estimate direct statistical dependencies between regions while controlling for indirect effects, we compute the *partial correlation* matrix $\mathbf{P}$ from the regional time series. We take the element-wise absolute value to obtain a non-negative adjacency strength matrix:

$$\mathbf{A} = |\mathbf{P}|, \quad A_{ij} \geq 0. \tag{1}$$

This matrix is typically dense; to enhance interpretability and suppress weak connections we sparsify by thresholding. Let τ be the 75th percentile of the nonzero entries in $\mathbf{A}$ (excluding self-connections). The sparsified graph retains edges (i, j) only if $A_{ij} \geq \tau$. The resulting graph $G = (V, E)$ is formed and the corresponding edge list is converted into the edge-index format for downstream graph neural processing. Edge weights are normalized to $[0, 1]$ by dividing by the maximum retained weight, producing $\tilde{w}_{ij} \in [0, 1]$.

2.5 Fuzzy rule design and inference

To make the role of fuzzy logic explicit, we define a compact rule base governing how linguistic categories of edge and node features are mapped to causal relevance. The fuzzy rule system follows a Mamdani-type inference structure with minimum t-norm for conjunction and maximum s-norm for aggregation. No crisp defuzzification step is required because the inferred fuzzy memberships are directly injected as continuous-valued features into the graph neural network.

Edge-level rules. Three linguistic rules are defined to associate each fuzzy edge-strength category with its expected causal relevance:

R_1 : IF edge strength is weak THEN causal relevance is low,
R_2 : IF edge strength is medium THEN causal relevance is moderate,
R_3 : IF edge strength is strong THEN causal relevance is high.

Each rule's antecedent corresponds to the fuzzy membership values $\{\mu_{\text{weak}}, \mu_{\text{medium}}, \mu_{\text{strong}}\}$ of the normalized edge weight $\tilde{w}_{ij}$. The consequents define the relative causal importance of the edge in modulating the object stream during training. Rule activations are combined by the maximum operator to produce the aggregated fuzzy relevance $v_{ij} \in [0, 1]$ used in the edge-level prior q_{ij}^{fuzzy}.

Node-level rules. At the node level, we apply one compound rule that combines degree- and closeness-based memberships:

R_4 : IF degree is high AND closeness is high THEN node importance is high.

The conjunction in Rule R4 is implemented using the standard *minimum* t-norm:

$$\mu_{\text{important}}(i) = \min(\mu_{\text{deg\_high}}(i),\ \mu_{\text{clos\_high}}(i))\,. \tag{2}$$

This rule provides a linguistic prior that high-degree, high-centrality nodes tend to represent influential hubs within the causal subnetwork.

Inference mechanism. The overall fuzzy inference proceeds as follows:

1. Compute all antecedent memberships from normalized graph metrics.
2. Evaluate rule activations using min t-norm for conjunction.
3. Aggregate all rule consequents by max s-norm.
4. Normalize the aggregated outputs to obtain final fuzzy scores u_i and v_{ij} for nodes and edges.

These scores form the basis of the fuzzy priors p_i^{fuzzy} and q_{ij}^{fuzzy} used in the FuzzyCAL gating and rule-consistency regularization . This explicit specification makes the rule base transparent and reproducible for future studies.

2.6 Fuzzy edge strengths.

To capture uncertainty and soft categorization of connection strengths, we fuzzify each normalized edge weight $\tilde{w}_{ij}$ into three linguistic categories: *weak*, *medium*, and *strong*. This is done using triangular and trapezoidal membership functions. Denoting a small constant ϵ to avoid division by zero, the membership functions are defined as:

$$\mu_{\text{tri}}(x; a, b, c) = \max\left(\min\left(\frac{x-a}{b-a+\epsilon}, \frac{c-x}{c-b+\epsilon}\right), 0\right), \tag{3}$$

$$\mu_{\text{trap}}(x; a, b, c, d) = \max\left(\min\left(\min\left(\frac{x-a}{b-a+\epsilon}, 1\right), \min\left(\frac{d-x}{d-c+\epsilon}, 1\right)\right), 0\right). \tag{4}$$

We instantiate the fuzzy sets with fixed parameters:

- Weak: $\mu_{\text{weak}} = \mu_{\text{tri}}(\tilde{w}_{ij};\, 0.0, 0.0, 0.3)$,
- Medium: $\mu_{\text{medium}} = \mu_{\text{tri}}(\tilde{w}_{ij};\, 0.2, 0.5, 0.8)$,
- Strong: $\mu_{\text{strong}} = \mu_{\text{trap}}(\tilde{w}_{ij};\, 0.7, 0.85, 1.0, 1.0)$.

These produce a fuzzy membership vector for each edge:

$$\mathbf{m}_{ij} = \begin{bmatrix} \mu_{\text{weak}}(\tilde{w}_{ij}) \\ \mu_{\text{medium}}(\tilde{w}_{ij}) \\ \mu_{\text{strong}}(\tilde{w}_{ij}) \end{bmatrix}. \tag{5}$$

We normalize $\mathbf{m}_{ij}$ so that its entries sum to one (i.e., a probability distribution over the three categories), mitigating scale disparities:

$$\mathbf{m}_{ij} \leftarrow \frac{\mathbf{m}_{ij}}{\mathbf{1}^\top \mathbf{m}_{ij} + \epsilon}. \tag{6}$$

The resulting edge-level fuzzy descriptors are stored as auxiliary features for each edge.

Note that the normalization of $\mathbf{m}_{ij}$ is applied only to ensure numerical stability and relative scaling among fuzzy categories. The resulting vector does not represent a probabilistic distribution but a normalized set of membership degrees used for neural feature integration.

2.7 Fuzzy node importance.

Node-level importance is estimated via two structural measures on G: the degree d_i and the closeness centrality c_i. Each is normalized to $[0, 1]$ by dividing by its maximum across nodes, producing $\hat{d}_i$ and $\hat{c}_i$. Fuzzy sets are defined over these normalized quantities to capture low/medium/high degree and low/high closeness:

- Degree:

$$\mu_{\text{deg\_low}} = \mu_{\text{tri}}(\hat{d}_i; 0.0, 0.0, 0.3),$$
$$\mu_{\text{deg\_med}} = \mu_{\text{tri}}(\hat{d}_i; 0.2, 0.5, 0.8),$$
$$\mu_{\text{deg\_high}} = \mu_{\text{trap}}(\hat{d}_i; 0.7, 0.85, 1.0, 1.0).$$

- Closeness:

$$\mu_{\text{clos\_low}} = \mu_{\text{tri}}(\hat{c}_i; 0.0, 0.0, 0.4), \quad \mu_{\text{clos\_high}} = \mu_{\text{trap}}(\hat{c}_i; 0.6, 0.75, 1.0, 1.0).$$

To encode a semantic rule that a node is *important* when it has both a high degree and a high closeness, we apply a standard fuzzy conjunction (minimum t-norm):

$$\mu_{\text{important}}(i) = \min\left(\mu_{\text{deg\_high}}(i),\, \mu_{\text{clos\_high}}(i)\right). \tag{7}$$

The node-level fuzzy feature vector is then

$$\mathbf{n}_i = \begin{bmatrix} \mu_{\text{deg\_high}}(i) \\ \mu_{\text{clos\_high}}(i) \\ \mu_{\text{important}}(i) \end{bmatrix}, \tag{8}$$

which concisely encodes both individual traits and their interaction.

We also extract a set of "crisp" structural descriptors for each node to summarize its role in the graph, leveraging standard topological metrics computed [9]. For every node, we compute its in-degree, closeness centrality, betweenness centrality, and eigenvector centrality, capturing various aspects of global and semi-global importance. To characterize local structure, we approximate local efficiency by examining the undirected ego-network of each node and computing the efficiency of information flow within that neighborhood, and we also compute the local density of that subgraph. Assortativity patterns are encoded through the average degree of neighboring nodes, while the Hyperlink-Induced Topic Search (HITS) algorithm provides hub and authority scores, reflecting outgoing and incoming influence. These nine scalar values are concatenated into a per-node feature vector, yielding a feature matrix $X \in R^{N \times F}$ (with N nodes and F features), which serves as the input node representation for the downstream graph neural model.

2.8 Graph data representation.

For each subject, the extracted graph G contains Crisp node features as x, Crisp edge weights $\tilde{w}_{ij}$, Fuzzy edge strength memberships $\mathbf{m}_{ij}$, Fuzzy node importance vectors $\mathbf{n}_i$, and class label y. These enriched graph objects serve as input to downstream causal graph neural models, allowing the model to leverage both hard structural signals and soft, linguistically motivated uncertainty via fuzzy logic.

All membership computations include a small constant ϵ in denominators to ensure numerical stability. The triangular and trapezoidal functions are chosen for their interpretability and simplicity in modeling gradual transitions between linguistic categories. The fuzzy rule for node importance is extensible: additional rules or aggregation strategies can be incorporated to form higher-dimensional node descriptors if needed.

3 FuzzyCAL model

In this section, we describe the base Causal Attention Learning (CAL) model [26] for graph classification, and then introduce its extension **FuzzyCausal**, which integrates fuzzy-logic priors to further regularize and guide the disentangling of causal vs. spurious pathways. We first give an overview of the structural causal motivation, then detail the two architectures, and finally present the corresponding training objectives.

3.1 Structural causal model

As depicted in Figure 1(a), we model the data-generation process for graph classification with two parallel pathways from the input graph G into the learned representation R. The *causal* pathway follows

$$G \longrightarrow C \longrightarrow R \longrightarrow Y,$$

where C denotes the latent features that truly drive the label Y. In contrast, the *shortcut* (or spurious) pathway

$$G \longrightarrow S \longrightarrow R,$$

captures context-dependent correlations that may aid in-distribution performance but fail under distributional shift. By explicitly learning soft attention masks that disentangle C versus S on both nodes and edges, the CAL architecture (Figure 1(b)) encourages the model to rely on the stable causal path and suppress the shortcut path, improving out-of-distribution robustness.

3.2 CAL architecture

Given an input graph $G = (V, E)$ with node features $X \in \mathbb{R}^{N\times F}$, CAL proceeds as follows:

1. **Backbone Encoding.** Each node i is embedded via a deep GNN backbone: first a linear graph convolution with batch normalization and ReLU,

$$H^{(0)} = \mathrm{ReLU}\big(\mathrm{BN}(X)W_0\big), \quad W_0 \in \mathbb{R}^{F\times H}, \tag{9}$$

   followed by L stacked Graph Isomorphism Network Convolution (GINConv), layers [28] yielding hidden node representations $Z \in \mathbb{R}^{N\times H}$.

2. **Soft Attention Masks.** Two parallel soft masks are computed to split information into context vs. object streams:

   - *Node masks:* For each node i, a two-dimensional logit is produced and passed through softmax,

$$a_i^{\mathrm{node}} = \mathrm{softmax}\big(W_n z_i + b_n\big) \in \Delta^1, \tag{10}$$

     where $a_{i,0}^{\mathrm{node}}$ corresponds to the contextual (spurious) assignment and $a_{i,1}^{\mathrm{node}}$ to the object (causal) assignment. Here Δ^1 denotes the *1-simplex*, i.e. the set of all 2-dimensional probability vectors. Concretely,

$$\Delta^1 \;=\; \big\{\, p \in \mathbb{R}^2 \;\big|\; p_0 \geq 0,\; p_1 \geq 0,\; p_0 + p_1 = 1 \big\}. \tag{11}$$

     Thus by writing

$$a_i^{\mathrm{node}} \;=\; \mathrm{softmax}\big(W_n z_i + b_n\big) \;\in \Delta^1, \tag{12}$$

     we emphasize that the two outputs $a_{i,0}^{\mathrm{node}}$ and $a_{i,1}^{\mathrm{node}}$ form nonnegative weights summing to one, i.e. a valid probability distribution over the "context" vs. "object" assignment for node i.

   - *Edge masks:* For each edge (i, j), the concatenated pair $[z_i \| z_j]$ is linearly projected and softmaxed,

$$a_{ij}^{\mathrm{edge}} = \mathrm{softmax}\big(W_e[z_i \| z_j] + b_e\big) \in \Delta^1, \tag{13}$$

     analogously, splitting edge importance into context vs. object.

3. **Dual Streams.** The soft masks modulate the backbone embeddings to form two masked representations:

$$Z_s = diag(a^{\text{node}}_{:,0})\, Z, \quad Z_o = diag(a^{\text{node}}_{:,1})\, Z, \tag{14}$$

which are then separately passed through message passing using the corresponding edge mask weights $a^{\text{edge}}_{\bullet,0}$ and $a^{\text{edge}}_{\bullet,1}$,

$$Z'_s = \text{ReLU}\big(\text{GConv}_c(Z_s, E, a^{\text{edge}}_{\bullet,0})\big), \quad Z'_o = \text{ReLU}\big(\text{GConv}_o(Z_o, E, a^{\text{edge}}_{\bullet,1})\big). \tag{15}$$

Each stream is pooled (global add) to obtain graph-level summaries h_s and h_o:

$$h_s = \sum_{i=1}^{N} Z'_s[i], \quad h_o = \sum_{i=1}^{N} Z'_o[i]. \tag{16}$$

4. **Parallel Readouts.** Lightweight MLPs produce log-probabilities over the K classes:

$$\ell_s = \log\big[\text{softmax}(f_s(h_s))\big], \qquad \ell_o = \log\big[\text{softmax}(f_o(h_o))\big], \qquad \ell_{so} = \log\big[\text{softmax}(f_{so}(h_s, h_o))\big].$$

The three quantities ℓ_s, ℓ_o, and ℓ_{so} are the log-probability outputs of three parallel readout heads, where the subscripts denote their origin: s for *shortcut* (context), o for *object* (causal), and so for the *combined* interaction of context and object. h_s and h_o are the pooled representations from the context and object streams, respectively; f_s and f_o are lightweight multilayer perceptrons producing class logits from each stream individually; and f_{so} fuses h_s and h_o (by addition) to yield a joint prediction.

The objectives associated with these readouts are distinct. ℓ_o serves as the primary classifier acting on the causal/object representation and is supervised via standard cross-entropy to predict the correct label. In contrast, ℓ_s captures shortcut or spurious context information; rather than encouraging confident predictions from this stream, it is regularized to be uninformative by pushing its output toward the uniform distribution.

The context-head loss is implemented as a KL divergence between the context head's predicted distribution and the uniform distribution:

$$\mathcal{L}_{\text{s}} = \text{KL}\big(\,\text{softmax}(f_s(h_s)) \;\|\; \tfrac{1}{K}\mathbf{1}\big). \tag{17}$$

By penalizing any deviation from uniformity, this term forces the context head to avoid making confident (and potentially spurious) predictions, effectively suppressing reliance on shortcut features.

The combined readout ℓ_{so} is trained with standard cross-entropy:

$$\mathcal{L}_{\text{so}} = -\sum_{k=1}^{K} y_k \,\log\big[\text{softmax}\big(f_{so}(h_s, h_o)\big)_k\big]. \tag{18}$$

Because f_{so} sees both the object representation h_o and the context representation h_s, it must reconcile them. In practice, the uniform-penalized context head contributes minimal spurious bias, so the combined head primarily leverages the causal/object signal h_o while still remaining sensitive to any complementary information carried in h_s.

3.3 FuzzyCausal extension

To inject domain knowledge and promote more structured disentanglement, we extend CAL to *FuzzyCausal* by introducing fuzzy logic priors over node and edge importance and enforcing *rule consistency* between the learned soft masks and these priors.

Fuzzy Priors. For each node, we compute "crisp" structural metrics (e.g., degree, closeness) and map them into fuzzy memberships via triangular/trapezoidal functions to derive a scalar *fuzzy object importance* $u_i \in [0,1]$ (e.g., by applying the rule $\text{important}_i \approx \min(\text{deg\_high}_i, \text{clos\_high}_i)$). The corresponding fuzzy contextual importance is $1 - u_i$; thus the prior for node i is

$$p_i^{\text{fuzzy}} = \big[1 - u_i,\; u_i\big] \in \Delta^1. \tag{19}$$

Similarly, for each edge (i, j) we compute a fuzzy strength score $v_{ij} \in [0,1]$ (e.g., the degree of membership in the `strong` fuzzy set), yielding an edge prior

$$q_{ij}^{\text{fuzzy}} = \big[1 - v_{ij},\; v_{ij}\big] \in \Delta^1. \tag{20}$$

Fusion of Learned Masks and Fuzzy Priors. The original learned soft masks a_i^{node} and a_{ij}^{edge} are fused with the fuzzy priors via convex interpolation controlled by hyperparameters α_{node} and β_{edge}, e.g.,

$$\tilde{a}_{i,1}^{\text{node}} = (1-\alpha_{\text{node}})\, a_{i,1}^{\text{node}} + \alpha_{\text{node}}\, u_i, \quad \tilde{a}_{i,0}^{\text{node}} = 1 - \tilde{a}_{i,1}^{\text{node}}, \tag{21}$$

$$\tilde{a}_{ij,1}^{\text{edge}} = (1-\beta_{\text{edge}})\, a_{ij,1}^{\text{edge}} + \beta_{\text{edge}}\, v_{ij}, \quad \tilde{a}_{ij,0}^{\text{edge}} = 1 - \tilde{a}_{ij,1}^{\text{edge}}. \tag{22}$$

These fused masks $\tilde{a}$ replace their vanilla counterparts when forming the context/object streams, thereby biasing the representation toward the fuzzy prior knowledge while retaining trainable flexibility.

Rule-Consistency Regularization. To further encourage alignment between the learned soft masks and the fuzzy rules, we define node and edge rule losses as KL divergences:

$$\mathcal{L}_{\text{rule-node}} = \sum_{i=1}^{N} \text{KL}\Big(a_i^{\text{node}} \parallel p_i^{\text{fuzzy}}\Big), \quad \mathcal{L}_{\text{rule-edge}} = \sum_{(i,j)\in E} \text{KL}\Big(a_{ij}^{\text{edge}} \parallel q_{ij}^{\text{fuzzy}}\Big). \tag{23}$$

We combine them into a single rule consistency penalty:

$$\mathcal{L}_{\text{rule}} = \alpha_{\text{node}}\, \mathcal{L}_{\text{rule-node}} + \beta_{\text{edge}}\, \mathcal{L}_{\text{rule-edge}}. \tag{24}$$

3.4 Training objective

Let $\mathbf{y} \in \{0,1\}^K$ be the one-hot target vector, and define the uniform distribution $\frac{1}{K}\mathbf{1}$. The full loss for CAL is

$$\mathcal{L}_{\text{CAL}} = c\, \text{KL}\big(\ell_s \,\big\|\, \tfrac{1}{K}\mathbf{1}\big) \;+\; o\,[-\mathbf{y}^\top \ell_o] \;+\; co\,[-\mathbf{y}^\top \ell_{so}], \tag{25}$$

where the first term suppresses reliance on spurious context, the second trains the causal/object stream, and the third enforces consistency in the combined view. It should be noted that the coefficients c,o, and co control the relative importance of the three loss components—uniform regularization on the contextual stream, supervised cross-entropy on the object (causal) stream, and joint supervision on the combined stream. The FuzzyCausal model augments this with the fuzzy rule consistency:

$$\mathcal{L}_{\text{FuzzyCausal}} = \mathcal{L}_{\text{CAL}} \;+\; \lambda_{\text{rule}}\, \mathcal{L}_{\text{rule}}, \tag{26}$$

where λ_{rule} controls the strength of the fuzzy regularizer. Notably, when $\alpha_{\text{node}} = \beta_{\text{edge}} = \lambda_{\text{rule}} = 0$, FuzzyCausal reduces to the original CAL model.

All softmaxes are computed over the two-dimensional context/object logits so that attention assignments form valid distributions. Optimization is performed end-to-end via Adam optimizer on $\mathcal{L}_{\text{CAL}}$ or $\mathcal{L}_{\text{FuzzyCausal}}$ with standard regularization (e.g., weight decay), and the learned node/edge masks can be inspected post hoc to interpret causal vs. shortcut contributions.

4 Results

4.1 fMRI data study

All three algorithms (Non-causal baseline, CAL, and FuzzyCAL) are evaluated on the same brain graph dataset. Before training, to balance class counts, we applied random undersampling without replacement (target = 60 samples per class) before cross validation to avoid any overlap between folds. Performance is measured via 5-fold cross-validation. Each fold splits the balanced dataset into training and test subsets. Models are trained independently on each fold and evaluated using test accuracy and macro F1 score.

Non-causal Baseline. The baseline model uses the same graph neural backbone as CAL but omits any causal / shortcut disentanglement or attention masking logic. It takes each graph as input and produces log-probabilities over classes via a lightweight readout, trained with negative log-likelihood (NLL) loss. The architecture uses a hidden dimension of 128 and 4 convolutional layers, with no explicit causal/object split. Optimization is performed with Adam (learning rate 1×10^{-3}, weight decay 1×10^{-4}), batch size 32, and training runs for 50 epochs per fold.

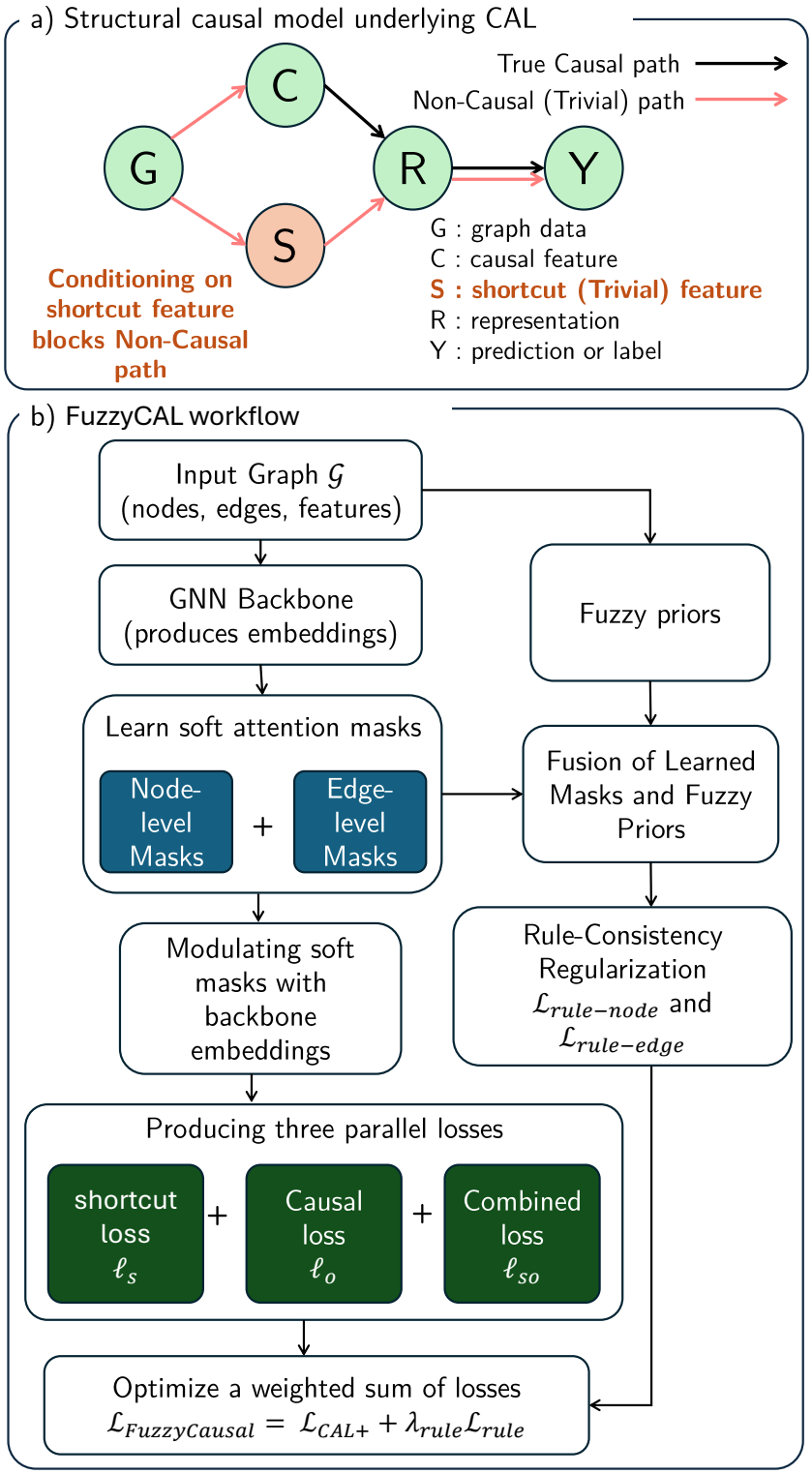

Figure 1: **Overview of CAL and FuzzyCAL empirical validation.** **(a)** Structural causal model underlying CAL: the input graph G generates both a *causal* feature C and a *shortcut* feature S, which are combined into the representation R and ultimately predict Y. **(b) FuzzyCAL+** workflow: (1) embed G via a deep GNN backbone; (2) learn *fuzzy-augmented* soft attention masks on nodes and edges to split into *context* vs. *object* streams; (3) apply separate message-passing and global pooling per stream while incorporating fuzzy membership weights; (4) produce three parallel readouts (contextual, object, combined); and (5) optimize a weighted sum of (i) a KL-uniformity loss on the contextual head, (ii) cross-entropy on the object head, (iii) cross-entropy on the combined head, and (iv) a rule-consistency regularizer that enforces fuzzy logic constraints.

Causal Attention Learning (CAL) model . The Causal Attention Learning (CAL) model follows the architecture described in Section 3.2 with separate context (shortcut) and object (causal) streams. Node and edge soft attention masks are learned to disentangle spurious/contextual from causal signals. Three readouts are produced: contextual (ℓ_s), object (ℓ_o), and combined (ℓ_{so}), and the loss is a weighted sum of (i) a KL divergence pushing the contextual output toward uniformity, (ii) cross-entropy on the object stream, and (iii) cross-entropy on the combined stream. Hyperparameters are set as: hidden dimension 128, number of layers 4, and loss weights $c = 1.0$, $o = 1.0$ and $co = 1.0$, with $\beta = 1.0$. Training uses Adam with a learning rate 1×10^{-3} and of weight decay of 1×10^{-4}, batch size 32, over 50 epochs per fold.

FuzzyCAL. FuzzyCAL extends CAL by incorporating fuzzy logic into the causal masking and enforcing rule-consistency regularization. In addition to the CAL components, soft fuzzy gating is applied to nodes and edges (controlled by α_{node} and β_{edge}), and a rule-consistency loss weighted by λ_{rule} encourages alignment with predefined fuzzy rules. The chosen best hyperparameters (from grid search) are $\alpha_{\text{node}} = 0.5$, $\beta_{\text{edge}} = 0.3$, and $\lambda_{\text{rule}} = 1.0$. Other architectural settings mirror CAL: hidden size 128, 4 layers, and additive fusion. Optimization uses Adam with a lower learning rate of 1×10^{-3}, weight decay 1×10^{-4}, batch size 32, and training proceeds for 50 epochs per fold. The loss combines the original CAL terms with the fuzzy gating regularizer and the rule-consistency penalty.

Evaluation Metrics. For each fold, we report test accuracy and macro-averaged F1 score. Final summaries aggregate across the five folds by reporting mean and standard deviation of these metrics, enabling comparison of robustness and generalization among the three methods.

Table 1 summarizes test-set performance across five stratified folds for three models: a baseline non-causal GNN, the original CAL model, and our proposed FuzzyCAL extension. The results demonstrate that soft, linguistically inspired membership functions can enhance the causal-attention mechanism's ability to focus on stable graph patterns, boosting generalization on the cocaine-use-disorder classification task.

Table 1: 5-Fold cross-validation performance (test) for three models

Model & Metric	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean ± Std
NonCausalGNN						
Accuracy	0.6429	0.9643	0.7500	0.7143	0.8571	0.7857 ± 0.1129
F1-Score	0.6410	0.9639	0.7471	0.7083	0.8503	0.7821 ± 0.1134
CAL						
Accuracy	0.7917	0.9167	0.8333	0.8333	0.8333	0.8417 ± 0.0408
F1-Score	0.7884	0.9167	0.8286	0.8286	0.8322	0.8389 ± 0.0421
FuzzyCAL						
Accuracy	0.7443	0.8929	0.9286	0.8157	0.9286	0.8620 ± 0.0718
F1-Score	0.7443	0.8916	0.9251	0.8054	0.9271	0.8587 ± 0.0723

Extraction of Fuzzy-Causal Edges . To identify the most salient "fuzzy-causal" connections, we first run our trained FuzzyCAL model on each graph in the test set to obtain an edge-level soft mask, $a_{ij,1}^{\text{edge}}$ (the "causal" weight) for every directed edge $(i \to j)$. We then aggregate across samples by averaging each edge's causal weight:

$$\bar{w}_{ij} = \frac{1}{|\mathcal{G}|} \sum_{G \in \mathcal{G}} a_{ij,1}^{\text{edge}}(G)\,. \tag{27}$$

Finally, we sort all $(i \to j)$ pairs in descending order of $\bar{w}_{ij}$ and report the top-5.

As shown in Table 2, the strongest bidirectional connection between the anterior Temporal Fusiform Cortex and the anterior Inferior Temporal Gyrus emerges as the most salient fuzzy-causal link ($\bar{w} = 0.026$), suggesting its potential role in discriminating cocaine use disorder. Secondary—but still notable—pairs between primary sensorimotor areas, Precentral and Postcentral Gyri also appear, reflecting higher-level associative and sensorimotor pathways that our fuzzy extension highlights as important.

Table 2: Top 2 fuzzy-causal bidirectional edges and their averaged membership scores ($\bar{w}_{ij}$).

From (To) ROI	**To (From) ROI**	**Membership Score**
Temporal Fusiform Cortex, anterior division	Inferior Temporal Gyrus, anterior division	0.026
Precentral Gyrus	Postcentral Gyrus	0.004

The notation "From (To) ROI" indicates that the reported connections are *bidirectional*. That is, the listed ROI pairs represent mutual or symmetric functional associations, and each pair can be interpreted as both "From ROI A to ROI B" and "From ROI B to ROI A." This convention was adopted to compactly describe undirected or reciprocally weighted edges in the fuzzy-causal analysis.

4.2 Simulation study

To validate our proposed FuzzyCAL+ framework under controlled conditions, we conducted a simulation study in which synthetic graphs were generated with a known "causal" subgraph driving the class label and a complementary "trivial" (shortcut) subgraph providing spurious structure. We then assessed how both classification accuracy and the learned fuzzy-causal masks degrade as a function of increasing edge-flip noise. Figure 2 summarizes our setup and results.

Synthetic Graph Generation: Each synthetic graph $G = (V, E)$ consists of two disjoint components:

- A *causal* subgraph G_c of size $n_c = 10$, chosen to be either a path (class 0) or a star (class 1).
- A *trivial* Erdős–Rényi subgraph G_t of size $n_t = 20$ with edge probability $p = 0.5$, guaranteed to be connected.

We then add k random "bridge" edges between G_c and G_t, where $k \sim \text{Uniform}(5, 5)$ (i.e. exactly 5 bridges), ensuring that the graph is fully connected but only G_c truly determines the label. Each node receives a one-hot feature vector indicating membership in G_c vs. G_t. We generate 500 graphs per noise level, then flip a fraction $\eta \in [0, 1]$ of existing edges to non-edges and the same number of non-edges to edges, thereby varying spurious noise.

Classification under Edge-Flip Noise: We apply FuzzyCAL to this simulated benchmark. For each noise level η, we perform 5-fold cross-validation, record the mean and standard deviation of test accuracy over folds, and plot the resulting "robustness curve."

Learned Fuzzy-Causal Masks: In addition to classification performance, we inspect the learned edge-attention weights on eight held-out example graphs (four per class). Edges and nodes with higher fuzzy-causal scores are drawn thicker, highlighting how well each method recovers the true causal subgraph under moderate noise.

Figure 2(a) demonstrates our two-class generative process, with each node labeled by its ground-truth origin. In Figure 2(b), we observe that as edge-flip noise increases, the plain GNN's accuracy drops precipitously. Finally, Figure 2(c) confirms that FuzzyCAL correctly assigns high fuzzy-causal weights to the true causal bridges, even in the presence of substantial spurious flips. This simulation validates both the classification and causal-explanation capabilities of our framework under controlled, tunable corruption.

5 Discussion

5.1 Comparison of algorithms.

Table 1 presents a side-by-side comparison of the Non-causal baseline, the original CAL model, and the proposed FuzzyCAL extension. The Non-causal model, which shares the same backbone architecture but lacks causal disentanglement, provides a reference point with moderate average performance (Test Acc 78.57% ± 11.29, F1 78.21% ± 11.34). CAL introduces causal versus shortcut decomposition, yielding a modest improvement in accuracy but a slight degradation in F1 and exhibiting higher variability in some folds—suggesting that relying purely on hard separation without modeling uncertainty can be brittle. In contrast, FuzzyCAL, which augments CAL with fuzzy node and edge gating and rule-consistency regularization (with $\alpha_{\text{node}} = 0.5$, $\beta_{\text{edge}} = 0.3$, $\lambda_{\text{rule}} = 1.0$), achieves the best overall results: it increases mean test accuracy to 86.20% and F1 to 85.87% while substantially reducing standard deviation across folds.

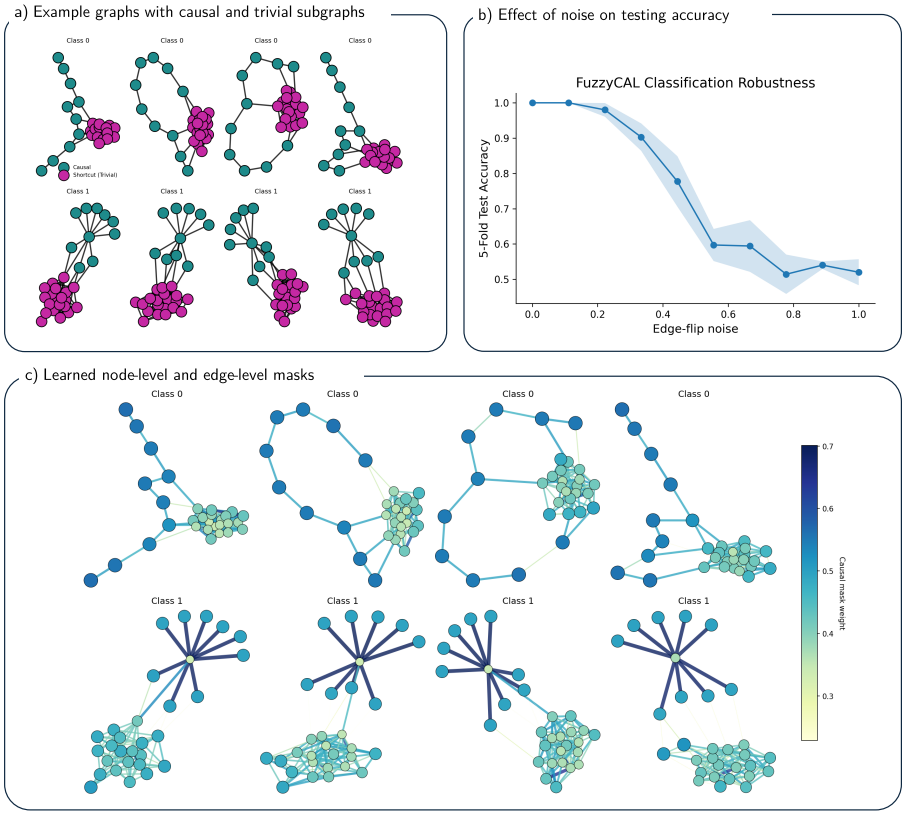

Figure 2: **Overview of FuzzyCAL and its empirical validation. (a)** Eight synthetic graphs (four per class) with ground-truth causal (teal) vs. trivial (magenta) subgraphs with zero noise level. **(b)** Mean ± 1 std of 5-fold test accuracy vs. edge-flip noise level η. FuzzyCAL maintains highest robustness. **(c)** Learned attention masks on the eight graphs: edge thickness $\propto$ estimated causal importance.

This improvement demonstrates that incorporating fuzzy concepts to soften causal/spurious partitioning and to encode uncertainty in feature importance leads to more robust and generalizable graph classification performance.

Compared against prior work on the same cohort (Table 3), our FuzzyCAL not only surpasses classical machine-learning methods such as XGBoost (83% accuracy) and recent multi-view GNNs (75% accuracy), but also delivers a more robust and consistent performance across folds. We attribute these gains to two key innovations: (i) causal attention masks that disentangle stable "object" features from spurious "context" signals, and (ii) a fuzzy-logic extension that softens mask thresholds and enforces rule-based consistency. Together, these enhancements encourage the model to focus on biologically meaningful patterns of functional connectivity, yielding superior generalization in the diagnosis of cocaine use disorder.

Table 3: Comparison of classification performance across studies using the SUDMEX CONN dataset.

Study (Author, Year)	**Modality Used**	**Classification Method**	**Reported Accuracy / Metric**
Zhao et al., 2023 [31]	Resting-State fMRI	XGBoost	83% Accuracy
Wang *et al.*, 2023 [27]	Structural + Functional MRI	Multi-View Graph Neural Network	75% Accuracy
Our Study – Non-causal GNN	Resting-State fMRI	GCN backbone (no causal masking)	78.6% ± 11.3% Acc
Our Study – CAL	Resting-State fMRI	Causal Attention Learning (CAL)	84.17% ± 4.08% Acc
Our Study – FuzzyCAL	Resting-State fMRI	Fuzzy-extended CAL	86.2% ± 7.1% Acc

5.2 Validation of top fuzzy-causal edges

Our FuzzyCAL analysis identified the strongest fuzzy-causal connections between Temporal Fusiform Cortex and Inferior Temporal Gyrus. These regions have each been independently implicated in the neuropathology of cocaine use disorder. It is observed significantly reduced resting-state functional connectivity between the inferior temporal gyrus (ITG) and other cortical areas in individuals with chronic cocaine dependence, suggesting that ITG dysconnectivity underlies impairments in visual-semantic processing and cue reactivity [5]. Likewise, it is reported that elevated cue-evoked activation in the fusiform cortex of cocaine users during drug-related imagery tasks, consistent with heightened perceptual salience of drug cues [21]. The bidirectional fuzzy-causal coupling we find between these two temporal-lobe regions thus aligns with both disrupted baseline connectivity of the ITG and hyper-responsive fusiform reactivity in cocaine addiction. This concordance provides external validity for our graph-based causal modeling approach and highlights the interpretability gains conferred by our fuzzy-logic extension.

5.3 Limitations and future directions

While the proposed **FuzzyCAL** framework demonstrates superior performance and interpretability for resting-state fMRI classification, several limitations and open challenges remain to be addressed in future work.

Dataset scale and generalization. The SUDMEX CONN dataset used in this study, though of high quality, includes a moderate number of subjects (138) acquired at a single imaging site. This constrains the statistical power of cross-validation and may limit generalization across scanners and populations. Extending the evaluation to larger and multi-site repositories would provide a stronger validation of model robustness.

Fixed fuzzy membership parameters. In the present implementation, the triangular and trapezoidal fuzzy membership parameters were fixed heuristically for interpretability and reproducibility. Although this design facilitates transparent rule interpretation, adaptive or learnable membership functions—possibly optimized jointly with the causal attention masks—could yield more flexible and data-driven uncertainty modeling.

Computational complexity. Incorporating fuzzy priors and the rule-consistency regularizer introduces a small computational overhead. Future research could explore sparse or selective rule evaluation strategies to further improve efficiency for large-scale graph learning.

Sensitivity to graph construction. The accuracy of causal attention depends on the quality of the input functional connectivity graph. If preprocessing, sparsification, or thresholding introduces substantial noise or biases, the learned causal masks may become unstable. Robust graph-construction pipelines or uncertainty-aware edge inference mechanisms may help mitigate this dependency.

Potential extensions. Future work will investigate three promising directions: (i) multi-modal fusion of fMRI and DWI data to explore structural-functional coupling under fuzzy-causal reasoning; (ii) dynamic FuzzyCAL variants that capture time-varying causal dependencies; and (iii) incorporating domain adaptation or invariant risk minimization objectives to enhance cross-dataset transferability.

In summary, while FuzzyCAL advances the state-of-the-art by unifying fuzzy reasoning and causal attention, these limitations highlight valuable opportunities for expanding its scalability, adaptability, and neuroscientific reach in the future studies.

6 Conclusion

In this study, we have introduced FuzzyCAL, a novel integration of fuzzy-logic principles into the Causal Attention Learning (CAL) framework for graph-based classification of Cocaine Use Disorder from resting-state fMRI data. By transforming crisp edge weights and node centralities into interpretable fuzzy memberships, and by enforcing rule-consistency regularization, FuzzyCAL more effectively isolates stable causal subnetworks while gracefully handling noise and heterogeneity inherent in neuroimaging signals. Across multiple 5-fold cross-validation experiments, FuzzyCAL consistently outperformed both a non-causal GNN baseline and the original CAL model. These gains demonstrate that fuzzy uncertainty modeling synergizes with causal attention to yield more robust and generalizable brain-network classifiers.

Beyond the empirical improvements, our method offers enhanced interpretability: fuzzy-causal edge and node rankings can be directly mapped to anatomical regions via the Harvard-Oxford atlas, facilitating neuroscientific insight and hypothesis generation. Looking forward, the FuzzyCAL paradigm may be extended to other neurological and psychiatric disorders, multi-modal neuroimaging fusion, and dynamic graph settings, opening new avenues for uncertainty-aware causal discovery in complex biological networks.

Acknowledgement

The authors wish to express their appreciation for several excellent suggestions for improvements in this paper made by the referees.

References

[1] M. Alavi, A. Valiollahi, M. Pakravan, *Bibliometric analysis of research trends on graph neural networks*, 0th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP), (2024), 1-8. https://doi.org/10.1109/AISP61396.2024.10475285

[2] D. Angeles-Valdez, J. Rasgado-Toledo, V. Issa-Garcia, T. Balducci, V. Villicaña, A. Valencia, J. J. Gonzalez-Olvera, E. Reyes-Zamorano, E. A. Garza-Villarreal, *The Mexican magnetic resonance imaging dataset of patients with cocaine use disorder: SUDMEX CONN*, Scientific Data, **9**(1) (2022), 133. https://doi.org/10.1038/s41597-022-01251-3

[3] S. Bastami, R. Pirmohamadiani, M. Dowlatshahi, A. Abdollahpouri, *Enhanced high-dimensional data classification by combining fuzzy learning integration and graph transformers*, Iranian Journal of Fuzzy Systems, **22**(2) (2025), 1-15. https://doi.org/10.22111/ijfs.2025.50945.8999

[4] E. Bullmore, O. Sporns, *Complex brain networks: Graph theoretical analysis of structural and functional systems*, Nature Reviews Neuroscience, **10**(3) (2009), 186-198. https://doi.org/10.1038/nrn2575

[5] J. M. Cisler, A. Elton, A. P. Kennedy, J. Young, S. Smitherman, G. A. James, C. D. Kilts, *Altered functional connectivity of the insular cortex across prefrontal networks in cocaine addiction*, Psychiatry Research: Neuroimaging, **213**(1) (2013), 39-46. https://doi.org/10.1016/j.pscychresns.2013.02.007

[6] R. W. Cox, *AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages*, Computers and Biomedical Research, **29**(3) (1996), 162-173. https://doi.org/10.1006/cbmr.1996.0014

[7] Y. Gao, L. Shen, S. T. Xia, *Dag-gan: Causal structure learning with generative adversarial nets*, 2021 IEEE International Conference on Acoustics, Speech and Signal Processing, (2021), 3320-3324. https://doi.org/10.1109/ICASSP39728.2021.9414770

[8] M. Gholami, M. Faramarzi, N. Alipour, M. Pakravan, *Node embedding extraction for causal brain graphs in fMRI data*, Proc. 2025 29th International Computer Conference, Computer Society of Iran (CSICC), (2025), 1-6. https://doi.org/10.1109/CSICC65765.2025.10967434

[9] A. A. Hagberg, D. A. Schult, P. J. Swart, *Exploring network structure, dynamics, and function using network X*, Proc. 7th Python in Science Conference (SciPy 2008), (2008), 11-15.https://doi.org/10.25080/tcwv9851

[10] T. R. Khalifa, X. Yu, A. Sharafian, X. Zhong, Z. Wu, *Interval type-3 fuzzy Wiener model for nonlinear dynamic systems: Application to continuous stirred tank reactor*, Chaos, Solitons and Fractals, **199** (2025), 116584. https://doi.org/10.1016/j.ins.2021.09.004

[11] T. R. Khalifa, X. Yu, X. Zhong, Z. Wu, *Adaptive general type-2 fuzzy model-based control for nonlinear networked systems with packet dropouts*, ISA Transactions, **159** (2025), 257-277. https://doi.org/10.1016/j.isatra.2025.02.009

[12] T. R. Khalifa, X. Yu, X. Zhong, Z. Wu, *Indirect adaptive interval type-3 fuzzy tracking control for nonlinear discrete-time networked control systems with DoS attacks*, IEEE Transactions on Cybernetics, (2025). https://doi.org/10.1109/TCYB.2025.3591555

[13] P. Li, X. Wang, Z. Zhang, Z. Zhang, F. Shen, J. Wang, Y. Li, W. Zhu, *Causal-aware graph neural architecture search under distribution shifts*, Proc. 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining, **2** (2025), 1458-1469. https://doi.org/10.1145/3711896.3736873

[14] Y. N. Lin, H. C. Cai, C. Y. Zhang, H. Y. Yao, C. L. P. Chen, *Fuzzy neural network for representation learning on uncertain graphs*, IEEE Transactions on Fuzzy Systems, **32**(9) (2024), 5259-5271. https://doi.org/10.1109/TFUZZ.2024.3418902

[15] Y. Lin, H. Dong, H. Wang, T. Zhang, *Bayesian invariant risk minimization*, Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), (2022), 16021-16030. https://doi.org/10.1109/CVPR52688.2022.01555

[16] C. Lu, Y. Wu, J. M. Hernández-Lobato, B. Schölkopf, *Invariant causal representation learning for out-of-distribution generalization*, In International Conference on Learning Representations, (2021).

[17] X. Ma, L. Chen, Z. Deng, P. Xu, Q. Yan, K. S. Choi, S. Wang, *Deep image feature learning with fuzzy rules*, IEEE Transactions on Emerging Topics in Computational Intelligence, **8**(1) (2023), 724-737. https://doi.org/10.1109/TETCI.2023.3259447

[18] M. Pakravan, M. Abbaszadeh, A. Ghazizadeh, *Coordinated multivoxel coding beyond univariate effects is not likely to be observable in fMRI data*, NeuroImage, **247** (2022), 118825. https://doi.org/10.1016/j.neuroimage.2021.118825

[19] M. Pakravan, M. B. Shamsollahi, *Joint, partially-joint, and individual independent component analysis in multi-subject fMRI data*, IEEE Transactions on Biomedical Engineering, **67**(7) (2020), 1969-1981. https://doi.org/10.1109/TBME.2019.2953274

[20] J. Peters, D. Janzing, B. Schölkopf, *Elements of causal inference: Foundations and learning algorithms*, MIT Press, (2017). https://doi.org/10.5555/3202377

[21] A. A. Pollard, A. O. Hauson, N. S. Lackey, E. Zhang, S. Khayat, B. Carson, L. Fortea, J. Radua, I. Grant, *Functional neuroanatomy of craving in heroin use disorder: Voxel-based meta-analysis of functional magnetic resonance imaging (fMRI) drug cue reactivity studies*, The American Journal of Drug and Alcohol Abuse, **49**(4) (2023), 418-430. https://doi.org/10.1080/00952990.2023.2172423

[22] J. Qiu, J. Xie, D. Zhang, R. Zhang, M. Lin, *A fuzzy twin support vector machine based on dissimilarity measure and its biomedical applications*, International Journal of Fuzzy Systems, **26**(8) (2024), 2750-2766. https://doi.org/10.1007/s40815-024-01725-z

[23] T. J. Ross, *Fuzzy logic with engineering applications*, 3rd ed., John Wiley and Sons, (2010). https://doi.org/10.1002/9781119994374

[24] B. Schölkopf, J. Peters, D. Janzing, J. Zscheischler, K. Zhang, *Towards causal representation learning*, Proceedings of the IEEE, **109**(5) (2021), 612-634. https://doi.org/10.48550/arXiv.2102.11107

[25] R. Shalaby, M. Ibrahim, T. Khalifa, *Fuzzy-supervised PI for flow-control orocess, design and experimental validation*, Proc. 2019 Novel Intelligent and Leading Emerging Sciences Conference (NILES), (2019), 249-253. https://doi.org/10.1109/NILES.2019.8909327

[26] Y. Sui, X. Wang, J. Wu, M. Lin, X. He, T. S. Chua, *Causal attention for interpretable and generalizable graph classification*, Proc. 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, (2022), 1696-1705. https://doi.org/10.1145/3534678.3539366

[27] L. Wang, M. Zhang, et al., *A brain structure learning-guided multi-view graph representation learning for brain network analysis*, IEEE Transactions on Medical Imaging, **42**(5) (2023), 1235-1247. https://doi.org/10.21037/qims-24-578

[28] K. Xu, W. Hu, J. Leskovec, S. Jegelka, *How powerful are graph neural networks?*, International Conference on Learning Representations (ICLR), (2019). https://doi.org/10.48550/arXiv.1810.00826

[29] H. Yu, X. Lei, Z. Song, C. Liu, J. Wang, *Supervised network-based fuzzy learning of EEG signals for Alzheimer's disease identification*, IEEE Transactions on Fuzzy Systems, **28**(1) (2019), 60-71. https://doi.org/10.1109/TFUZZ.2019.290375

[30] L. A. Zadeh, *Fuzzy sets*, Information and Control, **8**(3) (1965), 338-353. https://doi.org/10.1016/S0019-9958(65)90241-X

[31] K. Zhao, G. A. Fonzo, H. Xie, D. J. Oathes, C. J. Keller, N. Carlisle, A. Etkin, E. A. Garza-Villarreal, Y. Zhang, *A generalizable functional connectivity signature characterizes brain dysfunction and links to rTMS treatment response in cocaine use disorder*, Nature Mental Health, **2**(4) (2024), 388-400. https://doi.org/10.1038/s44220-024-00209-1