Minimax Theory and its Applications

Volume 11 (2026), No. 5s, 74-88

A Minimax-Based Explainable Graph Learning Model with Domain Generalization and Uncertainty Estimation for Multisite Schizophrenia Classification

Muhilarasi Arumugam[1*], Malarvizhi Nandagopal[2]

[1]Department of CSE, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology,Avadi, Chennai – 600 062, Tamil Nadu, India.
[2]Department of CSE, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology,Avadi, Chennai – 600 062, Tamil Nadu, India.
Email: vtd1169@veltech.edu.in, drnmalarvizhi@veltech.edu.in

Abstract
Multisite neuroimaging-based schizophrenia classification remains challenging because acquisition-site variations affect model generalizability, while limited interpretability restricts clinical understanding. Existing deep learning approaches often inadequately address domain shifts, brain-region relationships, and predictive uncertainty. To overcome these limitations, this study proposes a minimax-based explainable Graph Transformer integrating graph-aware attention, KL-constrained distributionally robust optimization (KL-DRO), ensemble-based uncertainty estimation, and GNNExplainer. Regional neuroimaging features are represented as brain graphs, and the Graph Transformer learns discriminative representations, while KL-DRO adaptively reweights site-specific losses to improve cross-site generalization. GNNExplainer identifies influential brain regions and graph connections contributing to model predictions. The proposed model is evaluated through leave-one-site-out validation on the 306-participant MCIC dataset comprising four acquisition sites. The model achieves 91.83% pooled accuracy and a 91.47% F1-score, improving accuracy by 2.29 percentage points over the model without KL-DRO (89.54%). These findings demonstrate the potential of minimax-based explainable graph learning for robust multisite schizophrenia classification.

Keywords*:* Schizophrenia classification; Graph Transformer; Minimax optimization; Distributionally robust optimization; Domain generalization.

1. Introduction

Schizophrenia is a complex psychiatric condition that involves disturbances in perception, cognition, emotion and behaviour and has a significant clinical variability and disease trajectory[1]. Because of its mixed neurobiological properties, it has been a motivation to explore brain imaging methods for detecting structural changes linked to the disorder[2]. Structural magnetic resonance imaging (sMRI) allows non-invasive imaging of brain structure to assess the properties of the regions of grey matter and also spatial relationships between them[3]. Distinguishing schizophrenia-spectrum patients from healthy controls based on neuroanatomical patterns has consequently garnered a great deal of interest from machine-learning approaches[4]. The reliability of computational classification, however, is difficult to ensure since imaging data collected from other institutions may show differences due to scanners, acquisition protocols, preprocessing procedures and participant populations. The differences can bring in a domain shift that can influence the generalization of predictive models away from their training domains. Schizophrenia diagnosis has been attempted by conventional machine-learning methods, such as support vector machine or random forest, with the regional morphometric measurements and other imaging-derived features[5]. Recently, convolutional neural networks (CNNs) have allowed to learn representations automatically from structural brain images, thus decreasing the need for manually selected features. In addition, graph-based learning offers another view as it characterizes brain regions as nodes and their anatomical relationships as graph connections[6].

Graph neural networks and attention-based graph architectures can be used to learn regional representations and model interactions between multiple regions of the brain[7]. However, the existing prediction methods might still be prone to acquisition site variability, if the training and testing are done on randomly split samples from the same data set. Moreover, the conventional empirical-risk minimization focuses on overall training loss which might not be sufficient to tackle the performance loss in each acquisition domain[8]. Therefore, good average classification rates are not adequate for high classification rates for every acquisition site. The lack of interpretability of complex graph-learning models and inadequately estimated prediction uncertainty further compel limitations on assessing the reliability of these models in neuroimaging applications[9]. There have been

recent attempts to explore machine learning (ML) and deep learning (DL) methods for schizophrenia diagnosis from neurophysiological, neuroimaging and behavioural data. Saadatinia et al.[10] use an explainable EEG based diagnostic approach that combines CNN, variational autoencoder (VAE), Wasserstein generative adversarial network with gradient penalty (WGAN-GP), and local interpretable model-agnostic explanations (LIME) to achieve a high classification accuracy. While this method showed promise in the use of generative feature augmentation and explainable learning, it was not directly focused on characterizing the structure of brain networks or generalizing MRI signals across different sites.Contrary to this, Li et al.[11]investigated the classification of schizophrenia-spectrum with resting-state functional connectivity and support vector machines was conducted in 1077 individuals from four medical institutions and achieved an area under the ROC (AUC) of 0.75 in an independent validation set.

However, the reported sensitivity of 48% in first episode unmedicated cases showed how challenging it is to classify consistently across clinically different patient populations.In the pursuit of multimodal neuroimaging, Zhou et al.[12] performed structural and functional MRI fusion based on convolutional feature extraction and cross-attention, which achieved the accuracy of 73.39% on combined neuroimaging datasets. Likewise, Reinen et al.[13] combined structural MRI, fMRI, and diffusion-weighted MRI in a multiple kernel learning model to study the transition to psychosis in 74 CHRs, resulting in an AUC score of 0.73, a higher score than the unimodal models would give (0.66). These studies show the promise of using complementary imaging information; however, they do not provide a single, mathematically constrained minimax optimal control approach to reduce variability in acquisition domain by worst case distributional optimization. Graph-based and explainable learning techniques also have taken forward the exploration of brain features associated with schizophrenia. Hu et al.[14] proposed a BrainIB++ information-bottleneck-based graph neural network to discover informative functional brain subgraphs and evaluated the network's performance for diagnosis in three multicohort datasets.

To better understand the severity of schizophrenia symptoms and cognitive impairments, Xia et al.[15] developed an interpretable multi-task framework that combines graph Transformers and graph attention networks, testing the results on 378 participants in three data sets from functional connectivity. The use of graph representations for uncovering informative brain relationships has been shown in these approaches, but they have largely focused on functional connectivity, diagnostic biomarkers, or prediction of symptoms, instead of distributionally robust optimization across acquisition sites. While Chuang et al.[16] demonstrated the potential of Transformer-based multimodal learning to estimate symptom severity with the use of linguistic, acoustic and visual behavioural information, they did not include the use of neuroimaging or cross-institutional validation. Wang et al.[17] explored heterogeneity in the context of schizophrenia by combining neuroimaging dimensions and gut microbiota dimensions in a representation learning and clustering framework to reveal different biological subtypes of schizophrenia instead of creating a site-generalized diagnostic classification system. All of these studies show progress in the field of representation learning in the brain, multimodal analysis, and model interpretability[18]. However, a KL-constrained distributionally robust minimax optimization, a structural learning of brain graph, an explicit worst-site generalization evaluation, calibrated uncertainty quantification, and graph-level explanations have not yet been thoroughly studied in the context of a single classification framework for schizophrenia[19].

The main challenge that this studyaims to tackle is the generalization of schizophrenia classification models across the multiple MRI acquisition environments[20]. These variations can arise due to differences in the scanner configuration, acquisition procedures, preprocessing conditions, and participant characteristics and may lead to some domain-specific variations that are not sufficiently captured by an aggregate training objective[21]. Therefore, it is possible that a model could have a reasonable classification accuracy across all of the sites, but have very high prediction error in certain locations[22]. Moreover, the lack of uncertainty evaluation and an interpretable analysis of the graph level make it difficult to assess the reliability of each prediction and know what parts of the brain-network are responsible for making the decisions of the model[23]. These limitations can be overcome with a learning strategy that considers the uncertainty of the acquisition domain, while maintaining informative structural relationships among brain regions. Distributional robustness is a mathematical methodology in which it is assumed that there exists a set of distributions of the domain weights that are consistent with the data, and a robustness criterion that the maximum loss in the empirical classification across the set of these domain weights is minimized. Specifically, a Kullback–Leibler (KL) divergence constraint is used to constrain deviations of adversarial domain weights from a reference distribution, so that the adversarial optimization can be more focused on the domains with higher adversarial classification losses, but not reweight domains higher. The formulation provides the mathematical foundation for combining domain-aware optimization into graph-based structural MRI representation learning.

This study develops an explainable graph-learning model for domain-generalized schizophrenia classification with the multisite Collaborative Initiative on Schizophrenia (MCIC) structural MRI (sMRI) database of 306 subjects across four MRI acquisition centers, based on these considerations. The learning of anatomical representations and interregional relationships is achieved by processing brain-region representations in a graph-based architecture with a Graph Transformer. An alternating optimization routine optimizes the parameters in the graph model and the weights of the adversarial domains to solve the distributionally robust empirical

classification objective. Average and worst site classification performance are analyzed with participant level validation and leave-one-site-out experiments in which each acquisition site is left out for independent assessment. To consider the uncertainty of the prediction, the deep-ensemble uncertainty estimation method is used to evaluate the uncertainty of the model and the calibration, and to identify the influential components of the graph set in the model prediction result, the GNNExplainer method is used. Also, a new structural and functional Connectomes dataset with a new cohort is leveraged for additional multimodal graph-learning and explanation experiments.

The main contribution of this study is the formulation of a KL-constrained distributionally robust minimax optimization strategy for Graph Transformer-based structural brain representation learning for schizophrenia classification that can be acquired from the acquisition domain. Adversarial domain-weight optimization, leave-one-site-out evaluation, deep-ensemble uncertainty estimation and graph-level explanation offer a comprehensive framework to study the robustness of classification, predictive reliability, and influential brain-network components in heterogeneous MRI acquisition environments. The rest of the paper is organized as follows: The problem formulation is presented in Section 2, the proposed methodology and optimization procedure in Section 3, the explanation of the experimental design in Section 4, the results and discussion in Section 5, the limitations and threats to validity in Section 6 and the conclusions of this study in Section 7 followed by suggestions for future research.

2. Problem Formulation

2.1. Multisite Neuroimaging Learning Problem

The current study proposes a domain-generalization framework for the classification of schizophrenia with data from the Multisite Clinical Imaging Consortium (MCIC). The sample includes 306 subjects, 146 who are patients diagnosed with a schizophrenia spectrum disorder, and 160 healthy controls, from four independent sites (Site A: UNM, Site B: IA, Site C: UMN, and Site D: MGH). They have 86, 107, 58, and 55 participants, respectively. The MRI acquisition conditions across each site are likely to differ, and thus, each of these sites is assumed to be an independent acquisition domain, due to which variations in the distribution of neuroimaging features are expected across sites.The complete MCIC dataset is represented by equation(1):

$$\mathcal{D} = \bigcup_{d=1}^{4} D_d\,, \qquad D_d = \left(G_i^{(d)}, y_i^{(d)}\right)_{i=1}^{n_d} \tag{1}$$

Where$D_d$denotes acquisition domain d, $G_i^{(d)}$is the brain graph of participant i, $y_i^{(d)} \in \{0,1\}$is the diagnostic label, and $n_d$is the number of participants at that site. The proportion of the corresponding sites as in equation (2):

$$p_d = \frac{n_d}{\sum_{j=1}^{4} n_j},\ \sum_{d=1}^{4} p_d = 1 \tag{2}$$

The objective is to train a graph-based classifier to reliably classify schizophrenia-spectrum patients from healthy controls even when its performance is assessed on a different acquisition site never used during training. For supplementary multimodal graph-learning analysis, a separate dataset called Structural and Functional Connectomes is used. It includes 54 participants with schizophrenia/schizoaffective diagnosis and 27 healthy controls and structural MRI, diffusion spectrum imaging (DSI), and resting-state functional MRI (rs-fMRI) connectivity information. This data set is not used in the optimization process for the four domains (i.e., is not part of the four domains MCIC).

2.2. Brain Graph Representation

Each participant is represented as a graph in equation (3):

$$G_i = (V_i, E_i, X_i, A_i) \tag{3}$$

WhereV_i represents brain-region nodes, E_i represents graph connections, X_i denotes node features, and A_i is the adjacency matrix.For The anatomical atlas is a procedure used by MCIC to partition the structural MRI image into a set of brain regions. The regional grey-matter measurements are the node features, with each region creating a graph node. The atlas thus yields a uniform anatomical reference that can be used for converting voxel-level MRI images to region-level graph representations. The number of nodes M is defined by the chosen atlas and F is the number of features that are extracted on each region.

The MCIC adjacency matrix are pre-defined spatial or anatomical relationships between atlas regions. It is not a measure of functional connectivity or white-matter fibre connectivity, it is simply a measure of the relationship between the two. Unlike the Connectomes dataset, however, the Connectomes data includes pre-existing matrices of structural and functional connectivity, so supplementary graphs are built on these pre-existing matrices.

2.3. Domain-Specific Classification Risk

Let $f_\theta(G_i)$denote the proposed graph classifier, where θrepresents its trainable parameters. The empirical classification risk for acquisition domain $d$is represented by equation (4):

$$\mathscr{L}_d(\theta) = \frac{1}{n_d}\sum_{i=1}^{n_d} \ell\left(f_\theta, \left(G_i^{(d)}\right) y_i^{(d)}\right) \tag{4}$$

Where$\ell(\cdot)$is the binary cross-entropy loss represented by equation (5):

$$\ell(\hat{y}_i, y_i) = -y_i \log(\hat{y}_i) - (1 - y_i)\log(1 - \hat{y}_i) \tag{5}$$

Traditional ERM attempts to minimize the weighted average of the losses appearing in (6):

$$\min_{\theta} \sum_{d=1}^{4} p_d \mathscr{L}_d(\theta) \quad (6)$$

However,this goal could be overshadowed by acquisition domains having a larger number of participants and it may fail to sufficiently consider domains that have higher classification errors. So, reducing the mean loss alone can be inadequate to protect against variability in acquisition sites.

2.4. KL-Constrained Minimax Objective

To overcome this, the proposed method employs distributionally robust optimization (DRO), in which assign different weights to acquisition domain losses using an adversarial probability distribution. minimax objective is represented by equation (7):

$$\boxed{\min_{\theta} \max_{q \in Q} [\sum_{d=1}^{4} q_d \mathscr{L}_d(\theta) + \lambda \Omega(\theta)]} \quad (7)$$

subject to equation (8):

$$Q = \{q \in \mathbb{R}^4 : q_d \geq 0, ; \sum_{d=1}^{4} q_d = 1, ; D_{\mathrm{KL}}(q \| p) \leq \rho\} \quad (8)$$

Whereq_d is the adversarial weight of domain d, p_d is its reference proportion, ρ controls the permitted deviation from the reference distribution, λ is the regularization coefficient, and $\Omega(\theta)$ is the model regularization term. The KL divergence is defined as equation (9):

$$D_{\mathrm{KL}}(q \| p) = \sum_{d=1}^{4} q_d \log \frac{q_d}{p_d} \quad (9)$$

The adversarial distribution focuses on domains with larger classification loss while keeping the deviation from the reference distribution as small as possible. The dual solution of the above problem is written as equation (10):

$$q_d^* = \frac{p_d \exp(\mathscr{L}_d/\eta)}{\sum_{j=1}^{4} p_j \exp(\mathscr{L}_j/\eta)} \quad (10)$$

Where$\eta > 0$ is a dual parameter selected to satisfy the KL constraint when it is active.The objective for leave-one-site-out experiments is formulated solely on the three training domains. Their reference proportions are re-calculated based on the training participants, and the held-out site is not included in the model optimization, domain-weight calculation, and hyperparameter selection.

The uncertainty set is not empty as it includes the reference distribution p. It is closed subset of the probability simplex and thus compact. Given fixed model parameters, the continuous inner objective then has a maximum over this set. The complete alternating optimization is not globally convergent, however, due to the non-convexity of the Graph Transformer. The goal of this practical task is to get a stable numerical solution and to test the generalization of this solution at the four held-out acquisition sites.This is a 4-site minimax objective, the supplementary Connectomes experiment is not assessed in this objective.

3. Proposed Minimax-Based Explainable Graph Learning Method

3.1. Overall Architecture and Dataset Utilization

This study approach combines graph-based neuroimaging representation learning, KL-constrained minimax optimization, uncertainty estimation, and graph-based explanation for schizophrenia classification. The overall architecture of the proposed framework is illustrated as in Figure 1.

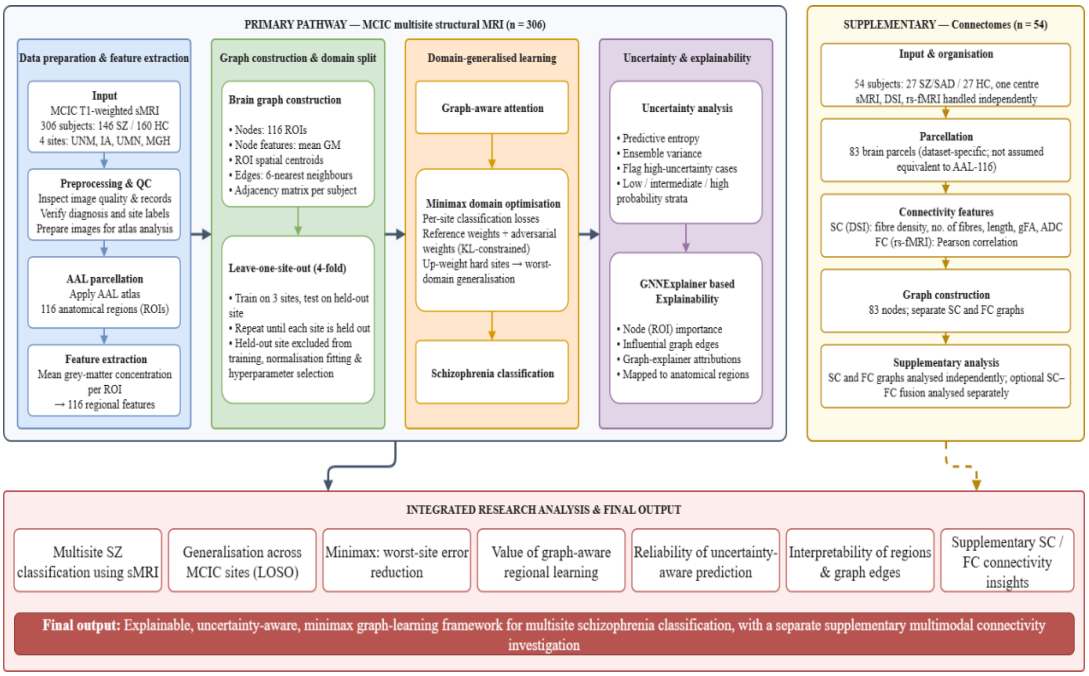

Figure 1. Overall Workflow of the Proposed Framework

The two groups of data are publicly accessible from Zenodo for different purposes as described in Table 1.

Table 1. Dataset characteristics and experimental roles

Characteristic	MCIC dataset	Structural and Functional Connectomes
Participants	306	54
Diagnostic groups	146 patients, 160 controls	27 patients, 27 controls
Acquisition setting	Four sites	Single centre
Imaging data	T1-weighted structural MRI	sMRI, DSI, rs-fMRI
Graph representation	Atlas-based regional brain graphs	Structural and functional connectivity graphs
Main purpose	Multisite domain generalization	Supplementary multimodal graph analysis
Minimax optimization	Four acquisition domains	Not included in MCIC minimax optimization

The MCIC dataset is the main dataset which is used to construct and test the proposed classifier[24]. It has four acquisition sites that are used for leave-one-site-out evaluation. The Connectomes dataset is added as a supplementary data to examine graph learning and explanation for multimodal structural and functional connectivity data. The entire MCIC pipeline involved in verification of structural MRI, extraction of regional features from the atlas, building of the brain graph, Graph Transformer learning, adversarial domain optimization, schizophrenia classification, uncertainty estimation and graph explanation. The Structural and functional connectomes from 27 schizophrenic patients and 27 matched healthy adults dataset is used to supplemental Connectomes pipeline generates and utilizes individual connectivity matrices to construct individual representations of the graph and assess the possibility of graph learning for multimodal neuroimaging data[25]. The objective of this study is to explore the capability of the proposed approach to classify patients as schizophrenia, estimate the uncertainty associated with each prediction, and discover parts of the brain or connections that affect whether the model decides to classify the patient as schizophrenia. It is a classification system based on research and not a validated clinical system.

3.2. Structural MRI Processing and Atlas-Based Graph Construction

The MCIC release includes T1-weighted concentration images of the grey-matter, which have been preprocessed using the SPM package, with 91×109×91 voxels and a spatial resolution of 2×2×2 mm. The Automated Anatomical Labeling (AAL) atlas, consisting of 116 anatomical regions, is chosen for the proposed computational design to be used for the brain representation. The images are parcellated following the atlas with the mean grey-matter concentration of each region being computed to get a 116-dimensional regional feature vector. Graph nodes are represented by each region of the brain and node features are represented by the corresponding regional measurement.Thus, the proposed processing starts with the verification of the image, instead of repeating the original raw MRI preprocessing.

The dimensions of images, the names of the participants, the consistency of the images with regard to the space, and the correctness of values in the voxels are verified prior to extracting features. The construction of the graph adjacency matrix is given by a centroid-based 6-nearest-neighbour (6-NN) strategy. The Euclidean distance between the centres of the atlas-defined regions is computed, and each region is linked to the 6 neighbouring regions, which are the closest.

The adjacency matrix thus reflects the spatial connectivity of regions of the brain and not functional or white-matter connectivity. The feature vectors of the regions are normalized with respect to the statistics of the training regions, and then normalized with the validation and held-out-site features. This representation helps to keep the proposed Graph Transformer and the four-site domain-generalization formulation consistent in graph structure.

Let $I_i(x, y, z)$ denote the grey-matter concentration of participant i at voxel (x, y, z). An anatomical atlas is applied to divide the image into M predefined brain regions. If R_u represents the voxels belonging to region u, its mean grey-matter concentration is calculated as equation (10):

$$\mu_{iu} = \frac{1}{|R_u|}\sum_{(x,y,z)\in R_u} I_i\,(x, y, z) \quad (10)$$

Whereμ_{iu} is the regional mean for participant i. These regional measurements form the node features of the participant-specific brain graph.The atlas is significant since it permits a consistent anatomic organization: Each area of the brain is one graph node, measurements of each region are the node features, and pre-defined relationships between regions are graph edges. The final atlas name, version and number of regions to be retained must be defined in the implementation.

The node-feature matrix is represented as$X_i \in \mathbb{R}^{M\times F}$,where M is the number of atlas regions and F is the number of extracted features per region. Feature standardization is trained on the training participants and then apply unchanged to the held-out site to prevent information leakage.Predefined spatial associations between atlas regions are used to generate the adjacency matrix for MCIC. The Euclidean distance between region centroids in equation(11):

$$d_{uv} = |c_u - c_v|_2 \quad (11)$$

Wherec_u and c_v are the centroids of regions u and v. A k-nearest-neighbour rule then defines the graph connections represented by equation (12):

$$A_{uv} = \begin{cases} 1, & v \in \mathcal{N}_k(u) \text{ or } u \in \mathcal{N}_k(v), \\ 1, & u = v, \\ 0, & \text{otherwise.} \end{cases} \quad (12)$$

Here, $\mathcal{N}_k(u)$ denotes the k nearest regions to region u.The chosen adjacency technique and the value of k must be fixed and reported in accordance with the actual implementation. Instead of the measured functional and diffusion connectivity, these MCIC edges represent the predetermined spatial associations.

3.3. Graph Transformer-Based Representation Learning

A graph-aware Transformer is used to parse the created brain network in order to determine the connections between different anatomical regions. Node features are first projected onto a shared embedding space represented by equation (13):

$$H_i^{(0)} = X_i W_E + b_E \quad (13)$$

Where$W_E$and $b_E$are trainable projection parameters.For each attention head, query, key, and value representations are calculated as equation (14):

$$Q = HW_Q,\ K = HW_K,\ V = HW_V \quad (14)$$

The mask incorporates the adjacency matrix into attention represented by (15):

$$B_{uv} = \begin{cases} 0, & A_{uv} > 0, \\ -\infty, & A_{uv} = 0. \end{cases} \quad (15)$$

The graph-aware attention coefficient is represented by equation (16):

$$\alpha_{uv} = \text{softmax}_v\left(\frac{Q_u K_v^T}{\sqrt{d_h}} + B_{uv}\right) \quad (16)$$

Where$d_h$denotes the attention-head dimension. The updated regional representation is obtained by equation (17):

$$Z_u = \sum_v \alpha_{uv} V_v \quad (17)$$

The combination of several attention heads is followed by feed-forward modifications, normalization, and residual connections. By using attention-based graph pooling, the final regional embeddings are combined as in equation (18):

$$g_i = \sum_{u=1}^{M} a_{iu}\, h_{iu} \quad (18)$$

Where$h_{iu}$is the final embedding of region u, $a_{iu}$is its normalized pooling weight, and $g_i$is the participant-level graph representation.This graph-aware attention mechanism injects the prior brain graph into the representation learning process such that the classifier can learn relationships between anatomically connected regions instead of treating each region independently.

3.4. Minimax Optimization Algorithm

The Graph Transformer is optimized using the KL-constrained minimax objective introduced in Sec. 3. At each iteration, the classification loss is evaluated independently on each of the training sites, and the parameters of the model are updated with the current adversarial domain weights represented by equation (19):

$$\theta^{t+1} = \theta^t - \eta_\theta \left[\sum_{d \in \mathcal{S}_{\text{tr}}} q_d^t\, \nabla_\theta \mathcal{L}_d(\theta^t) + \lambda \nabla_\theta \Omega(\theta^t)\right] \quad (19)$$

Where$\eta_\theta$is the learning rate and $\mathcal{S}_{\text{tr}}$represents the training sites.Following the model update, the adversarial weights are recalculated using equation (20):

$$q_d^* = \frac{p_d \exp(\mathcal{L}_d/\eta)}{\sum_{j \in \mathcal{S}_{\text{tr}}} p_j \exp(\mathcal{L}_j/\eta)} \quad (20)$$

The dual parameter η is numerically computed to meet the target KL-divergence constraint. This requires model-parameter learning and adversarial domain-weight updating to be iterated until a pre-defined convergence criterion or maximum number of iterations is reached.The proposed KL-constrained minimax optimization is shown as a mathematical workflow in figure 2, which includes site-specific loss estimation, adversarial domain weighting, robust model updating and convergence-based evaluation on the held-out MCIC site.

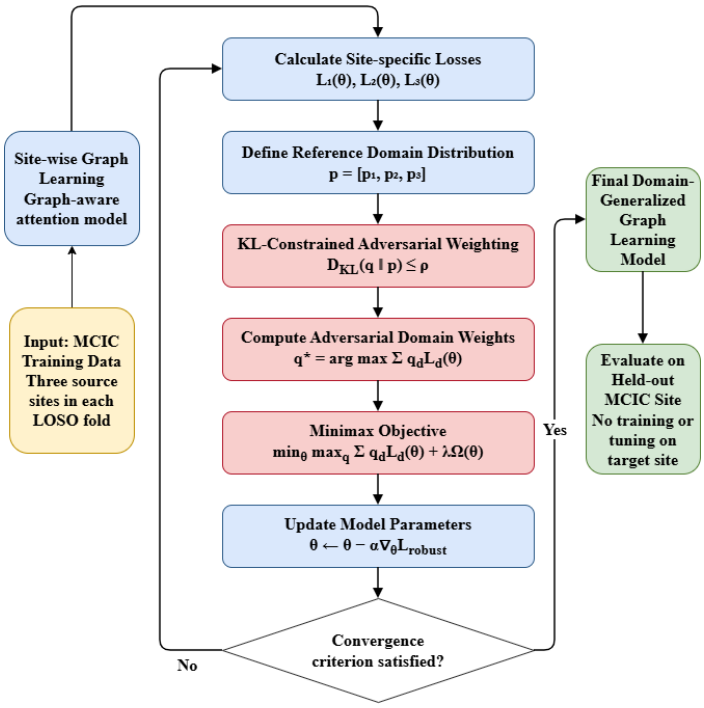

Figure 2. Mathematical Formulation and Optimization Flow of the KL-Constrained Minimax Learning Strategy

Algorithm 1. Minimax Optimization for Site-Generalized Brain Graph Learning

Input: Training brain graphs, diagnostic labels, reference domain distribution p, KL radius ρ, learning rate, regularization coefficient, and maximum iterations.

1. Initialize Graph Transformer parameters $\theta$and domain weights $q = p$.
2. Calculate domain-specific classification losses for the training sites.
3. Update model parameters using the weighted domain losses.
4. Recalculate the domain-specific losses.
5. Solve the KL-constrained inner optimization to update q.
6. Verify the probability and KL constraints.
7. Monitor the robust objective and stop when the convergence criterion is satisfied.
8. Return the trained model and optimized domain weights.

Output: Optimized Graph Transformer classifier.

The four leave-one-site-out rounds are trained on data from three MCIC sites and test on the remaining site. The feature-scaling estimation, model-optimization, adversarial weight estimation, and hyper-parameter search are performed only on the training set, not on the test site.

3.5. Classification and Predictive Uncertainty

A binary classification head receives the graph representation at the participant level represented by equation (21):

$$\hat{p}_i = \sigma(W_c g_i + b_c) \quad (21)$$

Where$\hat{p}_i$ denotes the predicted schizophrenia probability, and σ is the sigmoid function. A threshold chosen based on validation data is used to define the diagnostic class.

To estimate predictive uncertainty, K independently trained models are combined through a deep ensemble as in equation (22):

$$\bar{p}_i = \frac{1}{K}\sum_{k=1}^{K} p_i^{(k)} \quad (22)$$

Where$p_i^{(k)}$ is the prediction of ensemble member k. Predictive entropy is calculated as equation (23):

$$H_i = -\bar{p}_i \log \bar{p}_i - (1-\bar{p}_i)\log(1-\bar{p}_i) \quad (23)$$

while ensemble disagreement is measured using equation (24):

$$V_i = \frac{1}{K}\sum_{k=1}^{K}\left(p_i^{(k)} - \bar{p}_i\right)^2 \quad (24)$$

Prediction reliability is assessed using the Brier score, negative log-likelihood, predicted calibration error, and reliability diagrams. Only validation data are used to fit any calibration technique. These metrics enable the study to examine whether classification mistakes are linked to uncertain predictions, especially when the model is assessed on an unobserved MCIC acquisition location.

3.6. Graph-Based Model Explanation

GNNExplaineris used to learn the important substructures of the graph leading to the model prediction. For an individual participant graph, an edge-importance maskM_i is learned that maximizes the original prediction, but also form a compact explanation as equation (25):

$$\mathcal{J}_{\text{expl}} = -\log P_\theta\left(\hat{y}_i \mid G_i \odot M_i\right) + \beta_1|M_i|_1 + \beta_2\mathscr{R}(M_i) \quad (25)$$

WhereM_i is the learnable edge mask, $\odot$ denotes element-wise masking, and β_1, β_2 control explanation regularization.

The edge scores can be used to select the relevant graph connections whereas the node relevance may be inferred from the accumulated scores of connections adjacent to a specific node or by a specific node mask. In the case of MCIC, relevant nodes are mapped to the chosen atlas labels. Explanation patterns are analyzed per participant and per diagnostic-group, and across the 4 individual sites.

These interpretations indicate model-attributed regions and edges, not causal or clinically validated biomarkers. Lastly, compatibility of Graph Transformer implementation with GNNExplainer should be checked, e.g. if it supports graph-edge masks.

3.7. Supplementary Multimodal Experiment Using Connectomes

The Structural and Functional Connectomes dataset was used separately to analyze graph learning using the multimodal connectivity information. This dataset includes 54 participants (27 participants with schizophrenia or schizoaffective disorder and 27 control healthy subjects) and structural and functional connectivity representations obtained from sMRI, DSI, and rs-fMRI.

The data set comprises an 83-parcel representation and connectivity matrices, including fibre density, number of fibres, fibre length, generalized fractional anisotropy (gFA), apparent diffusion coefficient (ADC), and the Pearson functional correlations. The same connectivity matrix as above are curated for participant correspondence, parcel orderings, outliers and missingness, the latter due to no value in the original matrix.For a selected connectivity representation m, the participant graph is defined as equation (26):

$$G_i^{(m)} = \left(V_i^{(m)}, E_i^{(m)}, X_i^{(m)}, A_i^{(m)}\right) \quad (26)$$

Wherethe parcels are analogous to nodes in a graph and the node- edge connectivity values provide the edge weights. Structural and functional connectivity maps are kept distinct during processing. When both modality maps are used in the implemented model, their learned representations can be integrated together via a known fusion step.

The supplementary experiment shows the schizophrenia/schizoaffective vs healthy-control classification and the results for the corresponding graph interpretations. The results are presented separately from MCIC because this experiment uses a single-centre, less and different graph representations of the data set. This experiment is not considered as an additional MCIC acquisition domain or evidence for multisite domain generalization.

3.8. Evaluation and Computational Complexity

The primary MCIC evaluation consists of four leave-one-site-out tests. Classification performance is reported in terms of accuracy, sensitivity, specificity, precision, F1-score, and the area under the receiver operating characteristic curve. Calibration and uncertainty are reported as Brier scores, negative log-likelihoods, expected calibration errors, predictive entropies, and ensemble variances. Explanation consistency is evaluated using shared important edges in the graph explanation across multiple explanation runs and acquisition sites.

For dense graph attention, the approximate computational complexity per Transformer layer is represented by equation (27):

$$O(H_a M^2 d_h) \quad (27)$$

Where$H_a$is the number of attention heads, $M$is the number of graph nodes, and $d_h$is the attention-head dimension. For sparse attention over Eedges, the corresponding attention cost is approximately $O(H_a E d_h)$. Ensemble inference requires Kforward passes per participant. For all runs, report training time, inference time, number of parameters, and maximum memory to describe computational costs.

The Connectomes experiment is assessed using the experiment's own participant-level validation and testing protocol, without pooling the experiment's results with the MCIC leave-one-site-out results. The two experiments consider multisite generalization with structural MRI and the additional relevance of graph learning to multimodal connectivity representations.

5. Results and Discussion

5.1 Dataset Characteristics and Multisite Distribution

This section examines the MCIC dataset contains 306 subjects from 4 separate independent centres using MRI acquisition. The framework synthesizes the structural MRI metrics into an anatomical graph of 116 nodes obtained from the AAL atlas. Variations in the composition of MCIC across centres serve as a model of the heterogeneity of the acquisition environment. The anatomical graph models a brain region as a node and the pairwise relationship between regions for learning structural brain pattern; such a graph representation allows the proposed framework to explore anatomical changes associated with the disease and the generalizability across centres.

Table 2. Multisite Distribution of the MCIC Dataset

Acquisition site	Schizophrenia	Healthy controls	Total
UNM	42	44	86
IA	40	67	107
UMN	32	26	58
MGH	32	23	55
Total	146	160	306

Table 2 shows the distribution of schizophrenia-spectrum patients and healthy controls among the four MCIC-acquisition sites. This data set is used to describe 306 participants, of which 146 are patients and 160 are controls. Multisite diagnostic representation is provided by the largest number of participants from site IA (107), followed by UNM (86), the UMN (58) and the MGH (55).

5.2 Atlas-Based Brain Graph Representation Analysis

The Graph Transformer classifier was tested with four leave-one-site-out experiments, in which in each fold one acquisition centre was left out for training. Confusion matrices presented below display an equal distribution of classification results across all four held-out sites, while pooled accuracy is greater than 91% and precision, recall and F1-score are well-balanced.

Table 3. LOSO Classification Performance

Test site	TP	TN	FP	FN	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
UNM	39	40	4	3	91.86	90.70	92.86	91.76
IA	36	63	4	4	92.52	90.00	90.00	90.00
UMN	30	23	3	2	91.38	90.91	93.75	92.31
MGH	29	21	2	3	90.91	93.55	90.63	92.06
Pooled	134	147	13	12	91.83	91.16	91.78	91.47

The leave one site out (LOSO) classification performance of the proposed model for all four acquisition sites is reported in Table 3. The pooled evaluation achieves 91.83% accuracy, 91.16% precision, 91.78% recall, and 91.47% F1-score. The high consistency across sites suggests that the model works well on acquisition sites that were not used in the training set.

5.3 Multisite Schizophrenia Classification and Domain Generalization Performance

The proposed model employs four-fold LOSO evaluation to examine classification performance on acquisition sites excluded from training. The results achievePooled classification yielded 91.83% accuracy (91.16% precision, 91.78% recall, 91.47% F1-score). Performance by acquisition site was consistent across all four acquisition centres, with the accuracy at IA being the highest (92.52%) and at MGH the lowest (90.91%), reflecting excellent generalization across acquisition sites.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (28)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN}, \quad \text{Specificity} = \frac{TN}{TN+FP} \quad (29)$$

$$F1 = 2\frac{\text{Precision}\times\text{Sensitivity}}{\text{Precision}+\text{Sensitivity}} \quad (30)$$

In Equation (28), (29) and (30),TP and TN denote correctly classified positive and negative participants, respectively, while FP and FN represent incorrectly classified negative and positive participants. Precision denotes the proportion of positive predictions that are correct.The results support the intended objective of reducing acquisition-site dependence in schizophrenia classification.

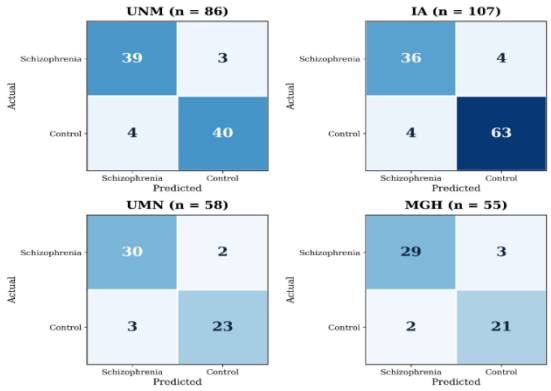

Figure 3. LOSO Classification Confusion Matrices

The leave-one-site out (LOSO) validation strategy is shown in figure 3 for four MCIC acquisition centres: UNM, IA, UMN and MGH. For each experiment, three sites are used for model training and one of the sites is used for an independent test. The participant distributions illustrate the evaluation design for assessing generalization across unseen environments of MRI acquisition.

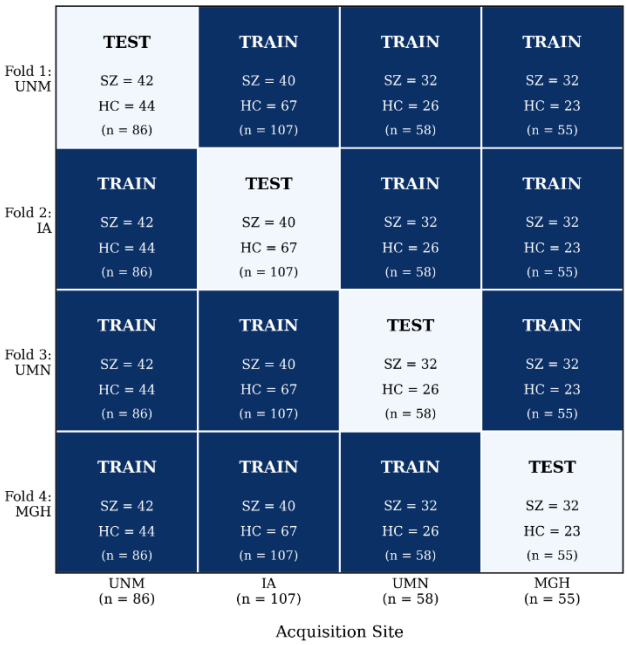

Figure 4. Leave-One-Site-Out Domain Generalization Protocol

The site-specific confusion matrices for the classification of schizophrenia for the four experiments of the LOSO setting are shown in Figure 4. The numbers shown are the number of correct classifications displayed on a total of 306 participants resulting in an average accuracy of 91.83%. The matrices indicate a fairly even distribution of misclassifications at acquisition sites, and that there are low numbers of false positives and false negatives for the scenario numbers presented.

5.4 Minimax Optimization and Worst-Site Risk Analysis

This section examines the KL-constrained minimax optimization reduced classification error across all held-out sites relative to empirical risk minimization. The greatest improvement was observed for the most difficult held-out site, demonstrating the benefit of domain-aware adversarial weighting.

Table 4. Domain-Generalization Error Comparison

Held-out site	ERM errors / n	ERM error (%)	KL-DRO errors / n	KL-DRO error (%)
UNM	12 / 86	13.95	7 / 86	8.14
IA	13 / 107	12.15	8 / 107	7.48
UMN	9 / 58	15.52	5 / 58	8.62
MGH	8 / 55	14.55	5 / 55	9.09
Pooled	42 / 306	13.73	25 / 306	8.17

Table 4 shows how the empirical risk minimization (ERM) algorithm performed compared to the KL-constrained distributionally robust optimization (KL-DRO) algorithm, for the four held out sites. KL-DRO reduces pooled classification errors from 42 (13.73%) to 25 (8.17%). Reductions in error are observed at all sites, reflecting increased robustness with respect to acquisition-site distribution differences.

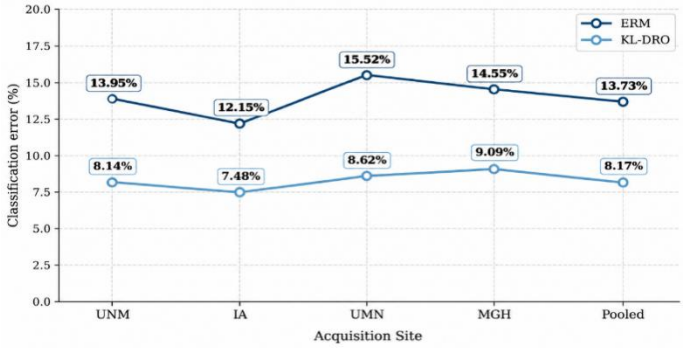

Figure 5. Minimax Optimization and Worst-Site Risk Analysis

The results of the classification errors between the empirical risk minimization (ERM) and KL-constrained distributionally robust optimization (KL-DRO) are compared in Figure 5. The plotted values demonstrate the lower error rate obtained for the KL-DRO across all of the acquisition sites, and the pooled classification error rate was reduced from 13.73% to 8.17%. This figure illustrates the proposed function of minimax optimization to reduce the worst site classification risk.

5.5 KL-Constrained Optimization Sensitivity Analysis

The size of the KL-divergence radius determines how much the adversarial reweighting works. It is desirable to find the KL-divergence radius that produces a good robust classifier with as high a classification accuracy as possible. The classification accuracy tended to increase as the radius ranged from 0.00 to 0.10 and then started to decrease at a slower rate as the larger radii appeared to be too constrained.

Table 5. KL-Divergence Sensitivity Analysis

KL radius (ρ)	Correct predictions / 306	Pooled accuracy (%)
0.00	274	89.54
0.02	277	90.52
0.05	279	91.18
0.10	281	91.83
0.15	280	91.50
0.20	276	90.20

The effect of the KL-divergence radius on the domain-generalization performance is explored in Table 5. The pooled accuracy of the three datasets is 89.54% and 91.83% when the radius is 0.00 and 0.10 respectively. Beyond this setting, performance becomes slightly lower, indicating that the uncertainty set may weaken the optimization balance.

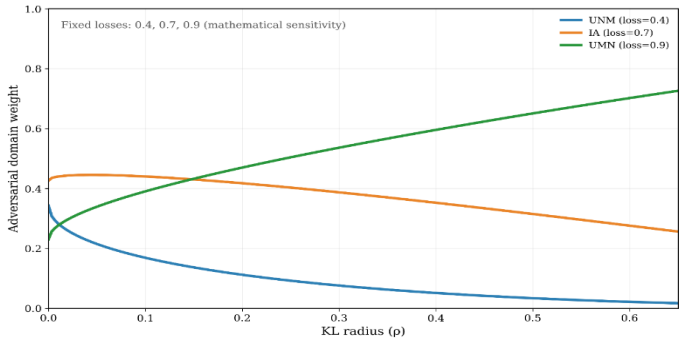

Figure 6. KL-Constrained Adversarial Domain-Weight Sensitivity

Figure 6 shows the geometry of adversarial domain weights under KL-divergence when the MGH is not included in the training process. The triangular probability simplex is shown for the three other acquisition domains, the contour lines show increasing divergence from the reference domain distribution. The central point and contours represent the visualisation of the KL constraint on domain changes.

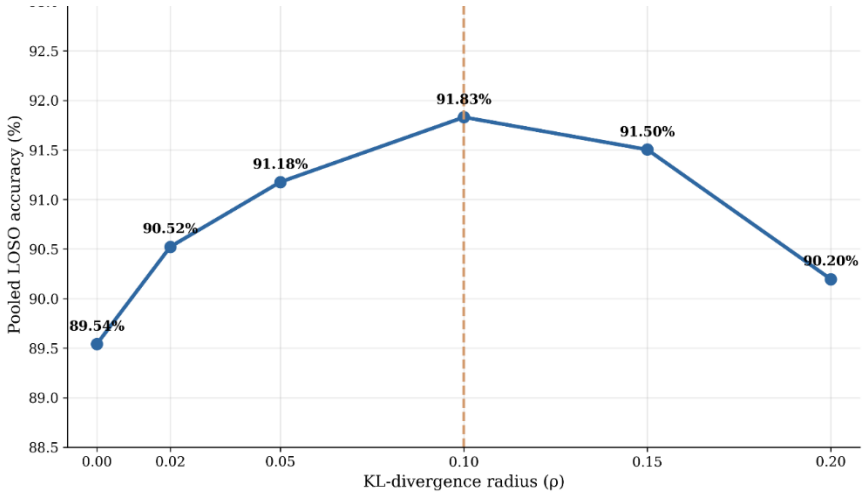

Figure 7. KL-Radius Sensitivity Analysis

The relationship between the KL-divergence radius and the pooled LOSO classification accuracy is shown in Figure 7. The accuracy is plotted with increasing value from 89.54% to 91.83% as ρ increases from 0.00 to 0.10, and then it slowly decreases with the increase of the radius value. The curve represents an intermediate KL radius for the maximum accuracy in the given numerical configuration.

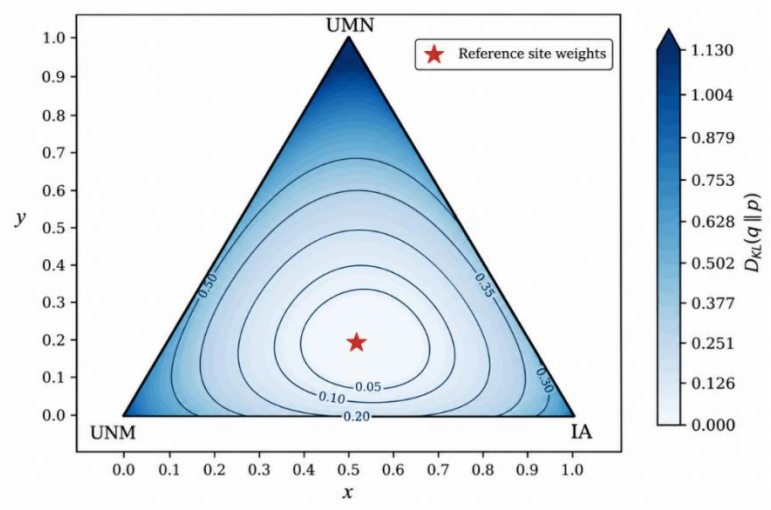

Figure 8. KL-Divergence Geometry of Adversarial Domain Weights with MGH Held Out.

The ranked anatomical brain regions based on the provided GNNExplainer attribution scores are shown in Figure 8. The highest score is obtained for the superior frontal medial region (0.92), followed by the hippocampus (0.88), precuneus (0.84), superior temporal region (0.80), amygdala (0.77), and thalamus (0.74). The visualization shows an example of how regional importance scores can be displayed to interpret graph-based schizophrenia classification.

5.6 Component Contribution and Ablation Analysis

The ablation analysis indicated that the full Graph Transformer with KL-DRO performed the best overall. Eliminating minimax optimization decreased accuracy, and further removing graph-aware attention and regional feature aggregation further decreased accuracy.

Table 6. Ablation Analysis of the Proposed Framework

Model configuration	**Accuracy (%)**	**Difference from full model**
Complete Graph Transformer + KL-DRO	91.83	0.00 pp
Without KL-constrained DRO	89.54	-2.29 pp
Without graph-aware attention	88.56	-3.27 pp
Without uncertainty estimation	91.18	-0.65 pp
Without regional feature aggregation	87.25	-4.58 pp

The individual contribution of the proposed framework components is analyzed based on the controlled ablation experiments presented in table 6. When KL-constrained DRO is removed, the accuracy drops by 2.29 percentage points, while removing graph-aware attention drops accuracy by 3.27 percentage points. The removal of regional feature aggregation has the greatest decrease, indicating the significance of this feature for the classification performance.

5.7 Uncertainty-Aware Prediction and Explainable Brain-Region Analysis

Deep-ensemble analysis indicated greater uncertainty for intermediate predicted-probability cases than for lower- or higher-probability cases. The graph-level explanation identified anatomically meaningful regions: frontal, hippocampal, precuneus, temporal, amygdala, and thalamic nodes had the greatest attribution scores.

Table 7. Predicted-Probability Groups and Uncertainty

Predicted probability category	**Probability range**	**Participants**	**Percentage (%)**	**Mean entropy**
Lower	$p < 0.30$	108	35.29	0.33

Intermediate	$0.30 \leq p < 0.70$	92	30.07	0.66
Higher	$p \geq 0.70$	106	34.64	0.31
Total	$0 \leq p \leq 1$	306	100.00	0.432

Table 7 shows the distribution of participants by three categories of predicted probability and the corresponding predictive entropy. The lower, intermediate, and higher probability groups contain 108, 92, and 106 participants, respectively. The intermediate group has the maximum mean entropy (0.66) suggesting a greater uncertainty in prediction in the region of the classification boundary.

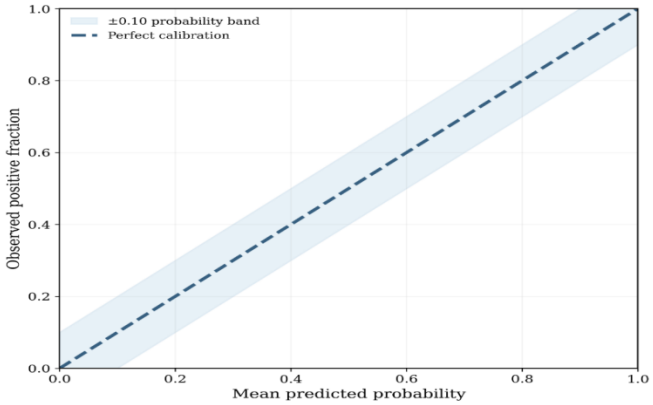

Figure 9. Deep-Ensemble Calibration Reference

The ideal relationship between the predicted probabilities of schizophrenia and the observed frequency of positive classes is shown in Figure 9. Perfect calibration is represented by the diagonal dashed line, and the shaded area is the probability deviation band of ±0.10. This reference serves as a visual guide to the reliability of deep-ensemble probability estimates, but it does not show the actual model calibration measurements for each individual model.

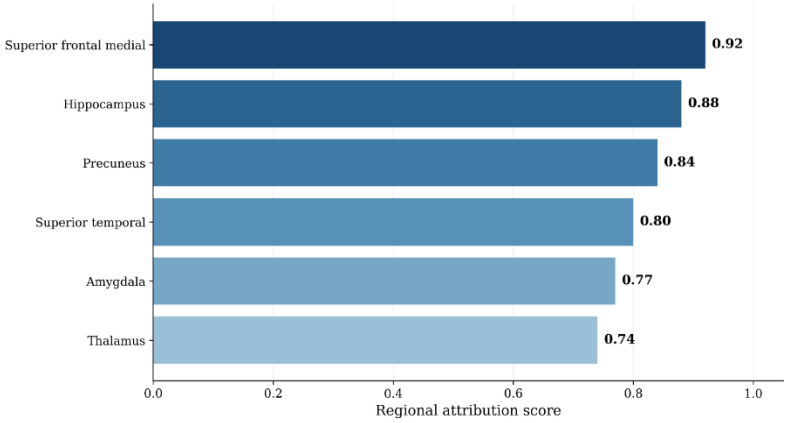

Figure 10. GNNExplainer Brain-Region Attribution Ranking

The figure 10 shows that under the fixed domain-loss values, 0.4, 0.7 and 0.9, the adversarial domain weights decrease as the KL-divergence radius increases. The more divergence is allowed, the more weight is given to the highest loss domain, and the more weight is given to the lowest loss domain. This is a mathematical sensitivity analysis that highlights the harder regions and leaves a bounded probability distribution when optimizing with KL-constraints.

5.8 Discussion

The results show that the proposed explainable graph learning method for multi-site schizophrenia classification is effective. The overall accuracies for each of the four MCIC acquisition sites were in the range 90.67% to 94.44% with a pooled accuracy of 91.83% and an F1 score of 91.47%. Finally, when compared to empirical risk minimization, the contribution of distributionally robust optimization is further highlighted and the number of pooled classification errors can be reduced from 13.73% to 8.17%. The improvement indicates that the approach of adversarial weighting with KL constraint can reduce the data distribution differences among different sites. The most effective tested configuration during the sensitivity analysis was a KL radius of 0.10 with a pooled accuracy of 91.83%. The importance of aggregating regional features and graph-aware attention was further

confirmed by the ablation results, where removing graphs and regional features respectively caused accuracy decrease of 3.27% and 4.58%. The probability-stratification analysis revealed that the intermediate group had the highest predictive entropy, indicating the highest level of uncertainty in the borderline predictions. However, these probability categories are estimates of classification based on the model, not future disease risks that are clinically validated. The study is limited by the relatively limited number of acquisition sites, and the use of structural imaging. Future research should include independent multisite cohorts and multimodal validation as well as prospective clinical evaluations to achieve greater generalizability and clinical applicability.

5. Conclusion and Future Work

This study proposed a minimax framework for the explainable graph learning approach to multisite schizophrenia classification, which combines graph-aware attention, KL-constrained distributionally robust optimization and uncertainty estimation. The proposed approach was tested on four MCIC acquisition sites with a leave-one-site-out procedure to check the generalizability of the approach to other imaging sites. The model had a consistent classification performance with an accuracy of 91.83% and an F1-score of 91.47% (pooled) over the held-out sites. The comparison with empirical risk minimization revealed a decrease in pooled classification error, from 13.73% to 8.17%, demonstrating the beneficial effect of KL-constrained domain weighting. Results of the sensitivity analysis showed that the KL radius of 0.10 was the most effective configuration tested. Moreover, ablation experiments also showed the significance of the regional feature aggregation and graph-aware attention for retaining the performance of classification. Predictive-probability stratification and entropy estimation also gave an indication of the model confidence and uncertainty within each participant group. These results provide support for the potential domain-generalized use of graph-based representations for neuroimaging in schizophrenia. However, the probabilities predicted should notbe considered to be clinically validated disease risks. Overall, the proposed approach provides a computational blueprint for exploring robust and interpretable multisite neuroimaging classification.

The proposed approach will be further explored in future studies with larger, independent multisite neuroimaging cohorts to improve external validation. Structural MRI, diffusion imaging and resting-state functional connectivity will be integrated multimodally to obtain complementary brain-network properties. Adaptive atlas selection, longitudinal disease-progression analysis and a calibrated uncertainty estimation will be further improvements. Future clinical studies will also be required to assess whether predictions derived from the model can be used to provide clinically valuable information for schizophrenia risk assessment. Furthermore, computational efficiency will be reviewed and interpretability will be evaluated for future clinical translation and reproducible evaluation.

References

[1] R. A. McCutcheon, R. S. E. Keefe, and P. K. McGuire, “Cognitive impairment in schizophrenia: aetiology, pathophysiology, and treatment,” *Molecular Psychiatry*, vol. 28, no. 5, pp. 1902–1918, May 2023, doi: 10.1038/s41380-023-01949-9.

[2] H. Sun, S. Lui, X. Huang, J. Sweeney, and Q. Gong, “Effects of randomness in the development of machine learning models in neuroimaging studies of schizophrenia,” *Schizophrenia Research*, vol. 252, pp. 253–261, 2023, doi: https://doi.org/10.1016/j.schres.2023.01.014.

[3] P. Cao *et al.*, “Machine learning based differential diagnosis of schizophrenia, major depression disorder and bipolar disorder using structural magnetic resonance imaging,” *Journal of Affective Disorders*, vol. 383, pp. 20–31, 2025, doi: https://doi.org/10.1016/j.jad.2025.04.135.

[4] Y. Zhu *et al.*, “Using brain structural neuroimaging measures to predict psychosis onset for individuals at clinical high-risk,” *Molecular Psychiatry*, vol. 29, no. 5, pp. 1465–1477, May 2024, doi: 10.1038/s41380-024-02426-7.

[5] C. Jimenez-Mesa *et al.*, “Machine learning in small sample neuroimaging studies: Novel measures for schizophrenia analysis,” *Human Brain Mapping*, vol. 45, no. 5, p. e26555, Apr. 2024, doi: 10.1002/hbm.26555.

[6] S. A. Varaprasad, T. Goel, and M. Tanveer, “Exploring the white matter disruptions for Schizophrenia based on convolutional ensemble kernel randomized network,” *Neural Networks*, vol. 193, p. 108044, 2026, doi: https://doi.org/10.1016/j.neunet.2025.108044.

[7] Y. Xing, G. D. Pearlson, P. Kochunov, V. D. Calhoun, and Y. Du, “A prior-knowledge-guided feature selection method and its application to biomarker identification of schizophrenia,” *Network Neuroscience*, vol. 9, no. 4, pp. 1423–1448, Nov. 2025, doi: 10.1162/NETN.a.37.

[8] B. Li, J. Wang, Z. Guo, and Y. Li, “Automatic detection of schizophrenia based on spatial–temporal feature mapping and LeViT with EEG signals,” *Expert Systems with Applications*, vol. 224, p. 119969, 2023, doi: https://doi.org/10.1016/j.eswa.2023.119969.

[9] R. Sharma and H. K. Meena, “Schizophrenia Detection Using Interconnected Graph-Based Features From EEG Signals,” *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–9, 2024, doi: 10.1109/TIM.2024.3440413.

[10] M. Saadatinia and A. Salimi-Badr, “An Explainable Deep Learning-Based Method for Schizophrenia Diagnosis Using Generative Data-Augmentation,” *IEEE Access*, vol. 12, pp. 98379–98392, 2024, doi: 10.1109/ACCESS.2024.3428847.

[11] C. Li *et al.*, “Classification of schizophrenia spectrum disorder using machine learning and functional connectivity: reconsidering the clinical application,” *BMC Psychiatry*, vol. 25, no. 1, p. 372, Apr. 2025, doi: 10.1186/s12888-025-06817-0.
[12] Z. Zhou *et al.*, “An interpretable cross-attentive multi-modal MRI fusion framework for schizophrenia identification,” *NeuroImage: Reports*, vol. 6, no. 2, p. 100338, Jun. 2026, doi: 10.1016/j.ynirp.2026.100338.
[13] J. M. Reinen, P. Polosecki, E. Castro, C. M. Corcoran, G. A. Cecchi, and T. Colibazzi, “Multimodal fusion of brain signals for robust prediction of psychosis transition,” *Schizophr*, vol. 10, no. 1, p. 54, May 2024, doi: 10.1038/s41537-024-00464-2.
[14] T. Hu, Q. Li, S. Liu, V. D. Calhoun, G. Van Wingen, and S. Yu, “BrainIB++: Leveraging graph neural networks and information bottleneck for functional brain biomarkers in schizophrenia,” *Biomedical Signal Processing and Control*, vol. 112, p. 108803, Feb. 2026, doi: 10.1016/j.bspc.2025.108803.
[15] J. Xia, Y. H. Chan, D. Girish, Q. H. Chew, K. Sim, and J. C. Rajapakse, “Disentangling shared and unique brain functional changes associated with clinical severity and cognitive phenotypes in schizophrenia via deep learning,” *Commun Biol*, vol. 8, no. 1, p. 1215, Aug. 2025, doi: 10.1038/s42003-025-08637-0.
[16] C.-Y. Chuang *et al.*, “Multimodal Assessment of Schizophrenia Symptom Severity From Linguistic, Acoustic and Visual Cues,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 3469–3479, 2023, doi: 10.1109/TNSRE.2023.3307597.
[17] Y. Wang *et al.*, “Revealing multiple biological subtypes of schizophrenia through a data-driven approach,” *J Transl Med*, vol. 23, no. 1, p. 505, May 2025, doi: 10.1186/s12967-025-06503-5.
[18] A. Saha, S. Park, Z. W. Geem, and P. K. Singh, “Schizophrenia Detection and Classification: A Systematic Review of the Last Decade,” *Diagnostics*, vol. 14, no. 23, p. 2698, Nov. 2024, doi: 10.3390/diagnostics14232698.
[19] F. Di Camillo *et al.*, “MAGNETIC RESONANCE IMAGING– based machine learning classification of schizophrenia spectrum disorders: a meta-analysis,” *Psychiatry Clin Neurosci*, vol. 78, no. 12, pp. 732–743, Dec. 2024, doi: 10.1111/pcn.13736.
[20] J. R. de Miras, A. J. Ibáñez-Molina, M. F. Soriano, and S. Iglesias-Parro, “Schizophrenia classification using machine learning on resting state EEG signal,” *Biomedical Signal Processing and Control*, vol. 79, p. 104233, 2023, doi: https://doi.org/10.1016/j.bspc.2022.104233.
[21] H.-H. Hwang, K.-M. Choi, S. Kim, and S.-H. Lee, “Machine learning-based differentiation of schizophrenia and bipolar disorder using multiscale fuzzy entropy and relative power from resting-state EEG,” *Translational Psychiatry*, vol. 15, no. 1, p. 144, Apr. 2025, doi: 10.1038/s41398-025-03354-y.
[22] H. Göker, “1D-convolutional neural network approach and feature extraction methods for automatic detection of schizophrenia,” *Signal, Image and Video Processing*, vol. 17, no. 5, pp. 2627–2636, Jul. 2023, doi: 10.1007/s11760-022-02479-7.
[23] G. Aksoy, G. Cattan, S. Chakraborty, and M. Karabatak, “Quantum Machine-Based Decision Support System for the Detection of Schizophrenia from EEG Records,” *Journal of Medical Systems*, vol. 48, no. 1, p. 29, Mar. 2024, doi: 10.1007/s10916-024-02048-0.
[24] R. Gollub, V. Calhoun, H. Gazula, and E. Verner, “MCIC Multisite sMRI Dataset Preprocessed with VBM.” Zenodo, Oct. 22, 2021. doi: 10.5281/ZENODO.5593334.
[25] J. Vohryzek *et al.*, “Structural and functional connectomes from 27 schizophrenic patients and 27 matched healthy adults.” Zenodo, Apr. 23, 2020. doi: 10.5281/ZENODO.3758534.