G-TBI: Graph-Structured Ordinal Learning for TBI Outcome Prediction

Minh-Trang Duong[1], Thanh-Bac Nguyen[2], and Thanh-Hai Tran[1]

[1] School of Electrical and Electronic Engineering
Hanoi University of Science and Technology, Hanoi, Vietnam
[2] Department of Neurosurgery, 103 Military Hospital, Hanoi, Vietnam
hai.tranthithanh1@hust.edu.vn

Abstract. Clinical patient data are typically represented in tabular form, which poses two key challenges: missing values that require potentially unreliable imputation, and the common assumption that patients are independent, ignoring inter-patient relationships. These limitations are particularly critical in healthcare, where inaccurate representations can affect outcome prediction. To address this, we propose a graph-based patient representation, that models patients as nodes and captures structural similarity among them. Instead of imputing missing values, we explicitly mask unavailable features, allowing the model to rely only on observed data while leveraging neighborhood information to handle sparsity. In addition, since clinical outcomes are inherently ordered, we replace cross-entropy with an ordinal label-dependent loss to better reflect disease severity. Experiments on two traumatic brain injury outcome prediction datasets (TBI_103 and TBI_OCD) show that our approach improves robustness to missing data and yields more clinically consistent predictions compared to conventional tabular methods. The proposed method **G-TBI** achieved accuracy gains of 1.53% and 0.26% compared to the best-performing SOTA models on TBI_103 and TBI_ODC, respectively.

Keywords: Graph Neural Networks · Traumatic Brain Injury · Ordinal Classification · Tabular Data.

1 Introduction

According to the Global Burden of Disease (GBD) 2021 study, traumatic brain injury (TBI) remains a critical public health challenge [14]. To address this, automated and robust TBI prognosis tools are essential for providing rapid and accurate assessments, thereby supporting clinicians, especially novices, in making timely treatment decisions. Automatic prediction models in this domain primarily rely on clinical tabular data derived from patient records, including demographics, lab results, and CT scans [1][6][16]. From the early stages, conventional classification algorithms treat each patient record as an independent observation, overlooking latent similarity relationships and contextual information essential for precise prognosis. While similarity-based approaches offer greater robustness

to noisy or incomplete data than feature-based paradigms [19], effectively integrating both remains a challenge. This is particularly problematic because TBI datasets frequently suffer from missing data, where traditional imputation can introduce bias and noise, ultimately leading to erroneous clinical predictions. [13]. Finally, in medical diagnostics, especially TBI, underestimating a severe case is far more catastrophic than overestimating a mild one, as the former can lead to the omission of life-saving interventions [20]. Because standard models treat all errors as mathematically equivalent, they fail to align with clinical priorities, necessitating the use of asymmetric, label-dependent penalties.

To address these limitations, we propose G-TBI, a framework that explores inter-relationship among patients in a similarity graph based on shared features with the backbone Graph Attention Networks (GAT). This model takes patients as nodes in a structured-graph and adaptively weighs the importance of neighboring cases to learn from similar patients, handling missing data without external imputation. Furthermore, we model the severity prediction as an ordinal classification problem by integrating a label-dependent loss function. The model applies heavier weights to extreme prognostic errors such as misclassifying severe cases as mild compared to the lighter penalties assigned to minor inaccuracies between adjacent levels.

The main contributions of this work are summarized as follows: i) We propose G-TBI, a graph-based framework that captures inter-patient relationships to mitigate missing clinical data, incorporating a severity-aware loss function to reduce critical underestimations and enhance representation robustness; ii) We conduct extensive experiments on real-world datasets from Vietnam and the U.S., showing effectiveness of G-TBI compared to the existing method; iii) We perform comprehensive ablation studies to validate the individual and joint contributions of our structural components and loss constraints.

2 Related works

2.1 Traumatic Brain Injury outcome prediction using machine and deep learning

The TBI outcome prediction field is founded on large-scale, validated Logistic Regression (LR) models, notably IMPACT and CRASH, which set strong, interpretable benchmarks using datasets exceeding 10,000 patients. Subsequent research continues to confirm the efficacy of traditional ML (e.g., LR, Decision Tree, Random Forest, XGBoost) as robust baselines for tabular clinical data, often proving that architectural complexity does not inherently guarantee superior predictive performance [16][17]. Concurrently, modern deep learning approaches leverage multimodal integration, combining neuroimaging CNNs with structured clinical features like GCS and Rotterdam scores to achieve over 90% accuracy in multicenter studies. [18]. More recently, Transformer/LLM-based pipelines like viTBI-BERT [6] enable feature extraction from unstructured narratives, while TBI-TTM [5] demonstrates the potential of combining textual reports with incomplete clinical data. Despite these advances, data heterogeneity and limited

generalizability persist, motivating the development of robust multimodal AI frameworks for scalable TBI outcome prediction.

2.2 Graph Neural Networks for tabular data

Graph Neural Networks (GNNs) are increasingly applied to tabular data due to their innate capability to model complex element relationships and feature interactions beyond traditional methods [11]. Specifically, TabGNN [8] learns complex sample relationships via multiplexed graphs, while TabGSL [12] represents a unified approach, simultaneously learning instance correlation and feature interaction. These efforts underscore the key advantage of GNNs in tabular models, that their ability to effectively capture complex entity relationships and enhance structural data comprehension.

3 Proposed framework

3.1 General framework

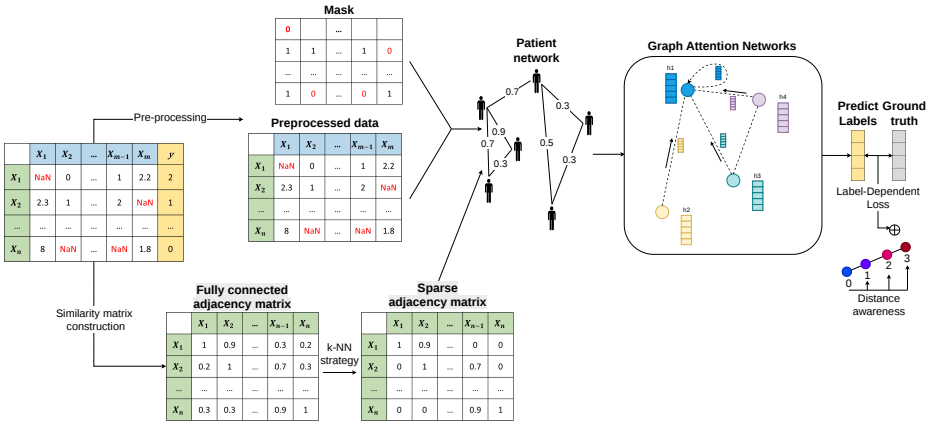

Fig. 1: Overview of the proposed G-TBI framework.

The proposed framework integrates missingness-aware preprocessing, similarity-based graph construction, and joint GNN-driven imputation and classification for robust TBI outcome prediction, which are build upon a graph-based method for the tabular data [10]. Formally, given a clinical dataset

$$D = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$$

be a clinical dataset of n patients, where d denotes the number of clinical features. Each patient i is represented by a feature vector $\mathbf{x}_i = [x_{i,1}, x_{i,2}, \ldots, x_{i,d}]$ where

$x_{i,f}$ denotes the f-th feature attribute, and y_i corresponds to their TBI severity label. These vectors are aggregated into a feature matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^\top \in \mathbb{R}^{n \times d}$. We reformulate the tabular prediction task as a graph learning problem, where each patient is modeled as a node and edges encode similarity relationships derived from clinical features. The graph structure captures both feature-level and relational dependencies among patients. Meanwhile, the model incorporates feature interactions, instance associations, and missingness-awareness to ensure robust TBI outcome prediction. The objective is to learn a predictive model f_θ that maps the partially observed features and inter-patient relations to the outcome labels, formulated as:

$$\hat{\mathbf{y}} = f_\theta(\mathbf{X}, \mathbf{M}, \mathbf{A})$$

where $\mathbf{M} \in \{0, 1\}^{n \times d}$ is a binary mask matrix, $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a similarity-based adjacency matrix encoding patient-to-patient relationships, and θ represents the learnable parameters of the network.

The proposed G-TBI framework, illustrated in Figure 1, consists of four primary stages: data pre-processing, similarity-based graph construction, graph representation learning, and outcome classification. The pipeline begins with preprocessing the tabular data to generate a missingness mask and a cleaned feature matrix. These are used to construct both fully connected and sparsified adjacency matrices based on feature similarity. The resulting graph is then processed through multiple GAT layers to learn patient-level representations, followed by a fully connected layer for TBI severity classification. The following sections detail the technical implementation of each component.

3.2 Data pre-processing

The preprocessing pipeline transforms raw tabular data $\mathbf{X}$ into a structured format suitable for graph construction and node feature representation. Features are partitioned into numerical, binary, and multiclass groups. Simultaneously, a binary mask $\mathbf{M}$ is generated, where $M_{i,f} = 1$ if the value $x_{i,f}$ is observed (valid), and $M_{i,f} = 0$ if the value is missing. To ensure numerical stability, multiclass features are One-Hot encoded, while numerical features are scaled and binary features remained in their original format for the graph input. Notably, raw numerical values are preserved for similarity computation. All missing entries in $\mathbf{X}$ are filled with a placeholder value of -1.0. This, in conjunction with the binary mask, allows the GNN to distinguish between true data and missing values during training.

3.3 Similarity matrix construction for mixed-type tabular data

The initial and most critical step in adapting tabular data for GNN processing is the construction of a meaningful graph structure. We employ Gower distance for mixed data types to jointly incorporate categorical (include Binary and Multiclass) and numerical features while inherently handling missing values. For each

pair of samples $(\mathbf{x}_i, \mathbf{x}_j)$, distances are computed exclusively over valid feature entries where both samples are present ($\delta_{ij}^{(f)} = M_{i,f} \cdot M_{j,f} = 1$). Initially, we compute local distance scores for categorical (s_{cat}) and numerical variables (s_{num}) independently.

- Categorical variables score (s_{cat}):

$$s_{cat}(x_{i,f}, x_{j,f}) = \begin{cases} 1 & \text{if } x_{i,f} = x_{j,f} \\ 0 & \text{if } x_{i,f} \neq x_{j,f} \end{cases} \tag{1}$$

- Numerical variables score (s_{num}):

$$s_{num}(x_{i,f}, x_{j,f}) = 1 - \frac{|x_{i,f} - x_{j,f}|}{R_f} \tag{2}$$

  where R_f is the range of the feature.
- Total score combination (s_{total}): The total score is the average of all the scores for the available features

$$s_{total}(\mathbf{x}_i, \mathbf{x}_j) = \frac{\sum_{f=1}^{d} \left(s(x_{i,f}, x_{j,f}) \cdot \delta_{ij}^{(f)} \right)}{\sum_{f=1}^{d} \delta_{ij}^{(f)}} \tag{3}$$

- The distance ($dist_{total}$) between observations $\mathbf{x}_i$ and $\mathbf{x}_j$ is computed as follows:

$$dist_{total}(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{1 - s_{total}(\mathbf{x}_i, \mathbf{x}_j)} \tag{4}$$

The final combined squared distance is transformed into a similarity matrix $\mathbf{A}$ using the Radial Basis Function (RBF) kernel, providing nonlinear edge weights for graph-based learning:

$$A_{ij} = \exp\left(-\gamma \cdot dist_{\text{total}}\right) \tag{5}$$

where γ controls the sensitivity of the similarity measure. Each data instance is modeled as a node characterized by its feature vector, while the relationships between pairs of nodes are captured as weighted edges reflecting their similarity. The similarity-based adjacency matrix initially forms a fully connected graph. To enhance sparsity and relevance, weak edges with low similarity values are pruned by replacing them with zeros.

3.4 Graph sparsification

The resulting dense similarity matrix $\mathbf{A}$ is sparsified using the k-Nearest Neighbors (k-NN) strategy. For each patient node, only the k edges corresponding to the highest similarity weights A_{ij} are retained. This sparsification step prevents the GNN from aggregating noise from distant or irrelevant neighbors, ensuring that the message-passing process focuses on the most informative and contextually relevant connections. Finally, the sparse adjacency structure is formatted

into edge indices and edge weights A_{ij}. The initial node feature matrix is then constructed by concatenating the cleaned features with the binary missingness mask. This combined representation allows the GNN to leverage both observed features and missingness patterns during the message passing process.

3.5 Proposed GNN architecture

The G-TBI framework utilizes GAT as its primary backbone to model patient-level dependencies. Unlike GCN with fixed structural weights, GAT introduces learnable attention coefficients to adaptively prioritize clinically significant neighbors. Furthermore, GAT supports inductive training, enabling efficient generalization to unseen patient data. The core of the G-TBI framework is the GAT layer, which adaptively learns the importance of patient relationships. A scoring function $e : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ computes an attention score for every edge (j, i), indicating the relative importance of neighbor j's clinical features to node i:

$$e(\mathbf{h}_i, \mathbf{h}_j) = \text{LeakyReLU}\left(\mathbf{a}^T \cdot [\mathbf{W}\mathbf{h}_i \,\|\, \mathbf{W}\mathbf{h}_j]\right) \tag{6}$$

where $\mathbf{h}_i, \mathbf{h}_j \in \mathbb{R}^d$ denote the latent feature representations (embeddings) of patient i and patient j, respectively. The vectors $\mathbf{a} \in \mathbb{R}^{2d'}$ and $\mathbf{W} \in \mathbb{R}^{d' \times d}$ are learnable parameters, where d' denotes the output embedding dimension of the GAT layer and $\|$ denotes vector concatenation. These scores are normalized across the neighborhood $\mathcal{N}_i$ (which represents the set of top-k most similar patient nodes connected to node i) using the softmax function to obtain the attention coefficients α_{ij}:

$$\alpha_{ij} = \text{softmax}_j\left(e\left(\mathbf{h}_i, \mathbf{h}_j\right)\right) = \frac{\exp\left(e\left(\mathbf{h}_i, \mathbf{h}_j\right)\right)}{\sum_{v \in \mathcal{N}_i} \exp\left(e\left(\mathbf{h}_i, \mathbf{h}_v\right)\right)} \tag{7}$$

The attention mechanism allows the framework to autonomously emphasize neighbors with similar clinical characteristics (e.g., body temperature), laboratory features, or specific structural brain injury patterns. Then, these coefficients scale a linear combination of neighbor features, yielding the node's updated representation after a non-linear activation σ:

$$\mathbf{h}_i' = \sigma\left(\sum_{j \in \mathcal{N}_i} \alpha_{ij} \mathbf{W}\mathbf{h}_j\right) \tag{8}$$

To enhance model expressiveness, G-TBI employs a multi-head attention mechanism with L independent heads across two stacked GAT layers. In the first hidden layer, features from parallel attention heads are concatenated to form a rich intermediate representation. In the second and final layer, concatenation is replaced by an averaging operation to maintain consistent output dimensionality. Formally, the updated representation $\mathbf{h}_i^{(out)}$ of patient i after the final GAT layer is computed by averaging the heads, followed by the non-linear Exponential Linear Unit (ELU) activation function σ:

$$\mathbf{h}_i^{(out)} = \sigma\left(\frac{1}{L}\sum_{l=1}^{L}\sum_{j\in\mathcal{N}_i}\alpha_{ij}^{l}\mathbf{W}^{l}\mathbf{h}_j^{(1)}\right) \tag{9}$$

where $\mathbf{h}_j^{(1)}$ denotes the intermediate hidden embedding of neighbor node j from the first layer, $\mathbf{W}^l$ represents the weight matrix of the l-th attention head, and α_{ij}^l is its corresponding normalized attention coefficient. The resulting embedding $\mathbf{h}_i^{(out)}$ is subsequently passed through a Fully Connected (FC) layer and a Softmax activation to generate the probability distribution across the four TBI severity levels. The entire G-TBI framework is trained end-to-end using an ordinal-aware loss function that minimizes the discrepancy between predicted and ground-truth severity labels while respecting the inherent ordinal dependencies of the TBI prognosis task.

3.6 Training strategy

To train the G-TBI framework, we apply an ordinal classification approach. In TBI prognosis, the primary objective is to minimize clinically dangerous misclassifications rather than just maximizing global accuracy. For instance, misclassifying a "Mild" case as "Critical" must incur a significantly higher penalty than predicting it as "Moderate". Following the decision-theoretic framework of Delgado [4], we define a label-dependent loss ω_{jm} based on the ordinal displacement between categorical tiers, rather than continuous intervals. For both datasets, clinical outcomes are mapped onto discrete ranks $j, m \in \{0, 1, 2, 3\}$. To capture asymmetric clinical risks, where underestimating TBI severity is significantly more hazardous than overestimation, the misclassification cost ω_{jm} is formulated using two independent generating functions (g_L and g_R) to ensure discrete convexity and stable decision thresholds:

$$\omega_{jm} = \begin{cases} g_L(j-m) & \text{if } j > m \\ g_R(j-m) & \text{if } j < m \end{cases} \tag{10}$$

where j is the ground-truth severity level and m is the predicted level. We define the growth functions g_L and g_R as:

$$\begin{aligned} g_L(d) &= d + \alpha_L d^2 \\ g_R(d) &= d + \alpha_R d^2 \end{aligned} \tag{11}$$

where $d = |f(j) - f(m)|$ represents the clinical distance. In our implementation, we set $\alpha_L > \alpha_R$ (for TBI prognosis) to ensure that underestimating a patient's severity is penalized with a significantly higher quadratic penalty.

Based on this cost structure, the optimal prognostic label is determined by minimizing the expected clinical risk. For a given predictive probability distribution $\mathbf{p} = (p_1, \ldots, p_r)$ generated by the GAT-backbone, the risk R_ω associated

with choosing label m is formulated as:

$$R_\omega(\mathbf{p}, m) = \sum_{j=1}^{r} \omega_{jm} p_j = \sum_{j \neq m} \omega_{jm} p_j \tag{12}$$

where p_j is the probability that the patient belongs to the true severity level j. As established by Delgado, the scoring rule for a predictive distribution $\mathbf{p}$ and true label j is given by $S_\omega(\mathbf{p}, j) = -R_\omega(\mathbf{p}, j)$. Maximizing this score during the training phase ensures that the model's probabilistic outputs are calibrated to minimize the expected clinical risk, effectively bridging the gap between feature learning and ordinal decision-making.

The second equality holds because the cost of a correct prediction is zero ($\omega_{jj} = 0$). To produce the final clinical decision, we employ the Omega-weighted Ordinal Maximum A Posteriori rules (ω-Ord-MAP). The framework selects the label $\hat{y}$ that yields the minimum expected loss, effectively balancing the probabilistic confidence of the model with the clinical safety requirements:

$$\hat{y} = \arg \min_{m \in 1,\ldots,r} R_\omega(\mathbf{p}, m) \tag{13}$$

To prevent data leakage, G-TBI operates inductively during validation and testing. When an unseen patient node $\mathbf{h}_{new}$ is introduced, it is dynamically integrated into the fixed training graph by computing similarities and retaining only the k highest-weight connections via a k-NN strategy. During the forward pass, frozen GAT layers update the new node's embedding $\mathbf{h}_{new}^{(out)}$ by aggregating features from its training neighbors. This embedding is then mapped via a Fully Connected layer and Softmax to output a probability distribution, where the ω-Ord-MAP rule determines the final prognostic label $\hat{y}$.

4 Experiments and results

4.1 Dataset

In our research, to generally evaluate the efficiency and generalizability of the proposed GNN-based framework for TBI outcome prediction, we utilize two distinct datasets collected from different sources. The first dataset is a self-collected dataset, referred to as the TBI_103 dataset and the second dataset is publicly available, named as TBI_ODC [15].

Input features for both datasets are categorized into four primary groups: demographic, clinical, laboratory and CT scan features. These features encompass a diverse range of data types, including numerical and categorical (binary and multi-class) variables. Outcome classification is structured into four categories across both datasets and labeled as 0,1,2,3: TBI_103 classifies severity as Mild, Moderate, Severe, and Critical, whereas TBI_ODC distinguishes between Uncomplicated Mild, Complicated Mild, Moderate and Severe. Both datasets exhibit notable class imbalance and data missingness across feature groups, causing challenges for reliable model training and generalization.

Table 1: Summary of the clinical datasets

Dataset	#Samples	#Features	#Num.Fs.	#Bi.Fs.	#Multi.Fs.	MR (%)	Class dist. (%)
TBI_103	504	64	15	34	15	8.15	10.9 / 58.1 / 20.0 / 10.9
TBI_ODC	2531	23	1	14	8	7.32	48.3 / 25.0 / 9.3 / 17.4

*#**Num.Fs.**, #**Bi.Fs.**, and #**Multi.Fs.** denote the number of numerical, binary, and multi-class features, respectively. **MR** represents the average missing rate across features. **Class dist.** reports the sample percentages for each target label (0/1/2/3).*

Dataset TBI_103 This study introduces the first TBI dataset specifically curated from Vietnamese clinical records, establishing a foundational resource for localized analysis and subsequent validation research. The dataset includes patient records from the 103 Military Hospital, collected in 2023, and due to security considerations [5] in Vietnam. The full dataset, containing sensitive medical and patient data, was managed using REDCap, is kept confidential and restricted in accordance with ethical and legal requirements.

Dataset TBI_ODC This dataset originated for the manuscript Nelson et al [15] and was published in Open Data Commons for Traumatic Brain Injury (ODC-TBI). The sample in raw datasets comprises 2545 humans age 17 or older who were enrolled with traumatic brain injury (TBI) in the multicenter, U.S.-based Transforming Research and Clinical Knowledge in TBI (TRACK-TBI) study. In our preprocessing stage, we excluded records with missing target labels, resulting in a final dataset.

The significant scale disparity between the large-scale TBI_ODC and the smaller TBI_103 provides a rigorous test for the generalizability of G-TBI. Despite sharing diagnostic commonalities, these datasets exhibit distinct demographic and pathological variations specific to their respective regions. Validating our model across both datasets demonstrates its adaptability and effectiveness in overcoming challenges related to sample scale and data heterogeneity.

4.2 Experimental setup

The model is implemented using the PyTorch Library and Python 3.10. The dataset was partitioned using a 10-fold cross-validation strategy. To ensure robust evaluation, the data was divided into training, validation, and testing subsets with a ratio of 8:1:1. Moreover, the experimental process was conducted three times using different random seeds. The optimal hyperparameters were determined through a systematic search across following ranges: k-neighbors in $[15, 40]$, RBF γ in $[0.25, 1.5]$, dropout rate in $[0.1, 0.3]$, temperature τ in the range of $[0.1, 1]$. For the GAT-backbone, the number of attention heads was selected from $[2, 8]$. The hidden layer sizes were explored within $[32, 128]$, and the models were trained for up to 1000 epochs with an early stopping patience of 50. The model was optimized using Adam with a learning rate 0.001 and weight decay

5×10^{-4}. Specifically, the asymmetric penalty coefficient α_L was set up as 2 and α_R as 1 to prioritize clinical safety.

4.3 Results

The experimental results are presented below, rigorously evaluating the G-TBI framework. We perform these experiments: (1) Evaluate the benchmark classification performance with machine learning, deep learning methods and our previous work; (2) Analyses and assess the effectiveness of the proposed architecture through ablation studies.

Benchmark classification performance We compared multiple models using the tabular modality. All traditional machine learning models, including Linear Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF) and XGBoost (XGB), as well as the deep learning-based TabNet [2], TabPFN [9], NAIM [3], FT-Transformer [7] and T2G-Former [12]. Most of the traditional machine learning methods required missing value imputation. Based on our prior experiments [5], we selected the best-performing imputation strategy K-Nearest Neighbors (KNN) with k=5, and applied it to the data at the pre-processing step. The hyperparameters for some baseline models were optimized and showed in Table 2. For the remaining models, we adopted the default parameters provided by their respective libraries or followed the specific configurations recommended in their original manuscripts.

Table 2: Hyperparameter search space for baseline models

Model	**Configuration Details**
Random Forest	$n\_estimators = 100$, $min\_samples\_split \in \{2, 5\}$
Decision Tree	$min\_samples\_split \in \{2, 5\}$
XGBoost	$n\_estimators \in \{100, 300\}$, $max\_depth = 3$, $LR = 0.1$
TabNet	$max\_epochs = 1000$, $batch\_size \in \{64, 512\}$
FT-Transformer	$n\_layers = 3$, $d\_token = 192$, $attention\_dropout = 0.2$
T2G-Former	$n\_layers = 2$, $n\_heads = 4$, $learning\_rate \in \{1e-4, 5e-4\}$

To validate the proposed framework, Table 3 compares classification outcomes between a primary private dataset and a larger-scale public dataset. This evaluation highlights the framework's flexibility in leveraging graph attention mechanisms to effectively capture the complex, heterogeneous patterns typical of large-scale public data, while simultaneously adapting to smaller, more focused clinical cohorts without risking overfitting.

The experimental results in Table 3 demonstrate that the proposed G-TBI framework consistently outperforms all baseline models across both datasets. On the smaller TBI_103 cohort, G-TBI achieves a peak accuracy of 85.45%, notably outperforming powerful tree-based ensembles like XGBoost (83.80%) and

Table 3: Performance (%) comparison of classification models on TBI outcome prediction across two datasets. **Bold** indicates the best value, while an underlined value indicates the second-best result.

	TBI_103				TBI_ODC			
Model	**Acc.**	**Prec.**	**Sens.**	**Spec.**	**Acc.**	**Prec.**	**Sens.**	**Spec.**
LR	$76.33_{\pm 5.17}$	$70.00_{\pm 7.00}$	$66.77_{\pm 6.75}$	$89.39_{\pm 2.28}$	$88.23_{\pm 1.76}$	$80.94_{\pm 1.84}$	$79.28_{\pm 1.96}$	$95.87_{\pm 0.58}$
SVM	$77.97_{\pm 3.46}$	$69.98_{\pm 7.71}$	$66.81_{\pm 6.10}$	$89.57_{\pm 1.99}$	$92.06_{\pm 2.01}$	$86.81_{\pm 3.15}$	$84.65_{\pm 2.67}$	$97.31_{\pm 0.63}$
DT	$80.62_{\pm 6.39}$	$78.25_{\pm 8.37}$	$\mathbf{76.10_{\pm 7.30}}$	$91.66_{\pm 2.57}$	$94.19_{\pm 1.39}$	$91.21_{\pm 1.98}$	$90.80_{\pm 2.35}$	$98.02_{\pm 0.52}$
RF	$83.06_{\pm 3.25}$	$74.17_{\pm 11.92}$	$69.12_{\pm 4.64}$	$91.03_{\pm 1.54}$	$95.85_{\pm 1.49}$	$93.92_{\pm 2.76}$	$92.55_{\pm 2.91}$	$98.61_{\pm 0.46}$
XGB	$83.80_{\pm 4.81}$	$76.92_{\pm 9.26}$	$73.06_{\pm 5.51}$	$91.24_{\pm 2.56}$	$\underline{96.96_{\pm 0.81}}$	$95.67_{\pm 1.59}$	$95.41_{\pm 1.17}$	$98.94_{\pm 0.28}$
CB	$\underline{83.92_{\pm 3.32}}$	$78.27_{\pm 10.64}$	$72.03_{\pm 4.84}$	$91.58_{\pm 1.64}$	$96.21_{\pm 1.00}$	$95.03_{\pm 1.66}$	$94.12_{\pm 2.02}$	$98.66_{\pm 0.34}$
TabNet	$74.07_{\pm 4.56}$	$65.98_{\pm 8.42}$	$56.97_{\pm 5.22}$	$87.71_{\pm 1.76}$	$93.45_{\pm 1.55}$	$91.20_{\pm 2.10}$	$90.95_{\pm 2.05}$	$97.25_{\pm 0.65}$
TabPFN	$82.75_{\pm 4.15}$	$72.80_{\pm 7.90}$	$71.20_{\pm 5.30}$	$90.85_{\pm 1.85}$	$97.12_{\pm 1.75}$	$\underline{95.80_{\pm 1.25}}$	$\underline{95.65_{\pm 1.10}}$	$\underline{98.90_{\pm 0.35}}$
NAIM	$82.94_{\pm 4.64}$	$77.30_{\pm 7.81}$	$73.61_{\pm 5.80}$	$91.51_{\pm 2.14}$	$96.77_{\pm 1.27}$	$95.64_{\pm 1.95}$	$94.69_{\pm 2.39}$	$98.87_{\pm 0.46}$
FT-Trans	$80.49_{\pm 4.99}$	$76.38_{\pm 5.67}$	$70.77_{\pm 6.96}$	$91.55_{\pm 2.25}$	$96.12_{\pm 0.85}$	$94.85_{\pm 1.40}$	$94.30_{\pm 1.35}$	$98.45_{\pm 0.35}$
T2G-Former	$82.16_{\pm 6.14}$	$\underline{78.44_{\pm 7.49}}$	$\underline{74.81_{\pm 6.60}}$	$\underline{92.37_{\pm 2.49}}$	$95.45_{\pm 4.15}$	$93.80_{\pm 1.60}$	$93.55_{\pm 2.45}$	$97.45_{\pm 0.60}$
G-TBI	$\mathbf{85.45_{\pm 3.25}}$	$\mathbf{86.56_{\pm 4.10}}$	$73.56_{\pm 3.90}$	$\mathbf{92.44_{\pm 2.65}}$	$\mathbf{97.22_{\pm 1.80}}$	$\mathbf{96.11_{\pm 2.27}}$	$\mathbf{96.51_{\pm 2.13}}$	$\mathbf{99.04_{\pm 0.64}}$

advanced tabular deep learning models like TabPFN (82.75%). Similarly, on the larger-scale TBI_ODC dataset, G-TBI establishes a superior accuracy baseline of 97.22%. This sustained performance highlights the advantage of integrating graph attention mechanisms with label-dependent margins, which adaptively weight patient similarity and effectively handle class imbalance.

Ablation study To rigorously validate the necessity and quantify the specific contribution of the novel components within the G-TBI framework, we conducted a detailed ablation study through comparison of the performance of the full G-TBI model against several reduced variants: (1) impact of the loss function; (2) missing data handling, and (3) sensitivity analysis.

Impact of the loss function For the first experiment, we compared the full G-TBI framework, which utilizes the ordinal-aware loss to preserve the natural progression of traumatic brain injury severity, against different loss functions. These include standard Cross-Entropy (CE), Focal Loss (FL) representing cost-sensitive learning for hard examples, Class-Distance Weighted Cross-Entropy (CDW-CE) loss as a baseline for distance-awareness, and Label-Dependent (LD) loss representing data-dependent margin maximization.

In Table 4, conventional CE and FL loss provide competitive baselines, but the proposed LD loss yields a consistent performance lift across both cohorts. Specifically, LD increases accuracy by 2.31% on TBI_103 and 1.53% on TBI_ODC compared to standard CE. As illustrated in the 3-seed accumulated confusion matrices (Figure 2), LD shows a denser diagonal of correct predictions and fewer critical underestimation errors, confirming its alignment with clinical priorities.

Table 4: Ablation study results on loss components. **Bold** indicates the best value, while an underlined value indicates the second-best result

	TBI_103				TBI_ODC			
G-TBI (loss)	**Acc.**	**Prec.**	**Sens.**	**Spec.**	**Acc.**	**Prec.**	**Sens.**	**Spec.**
CE	83.13	74.32	69.17	91.05	95.69	93.88	92.13	98.55
FL	83.53	81.23	69.97	91.38	94.88	92.58	91.78	98.24
CDW-CE	84.12	79.85	73.14	91.84	95.14	92.60	91.99	98.34
LD (our)	**85.45**	**86.56**	**73.56**	**92.44**	**97.22**	**96.11**	**96.51**	**99.04**

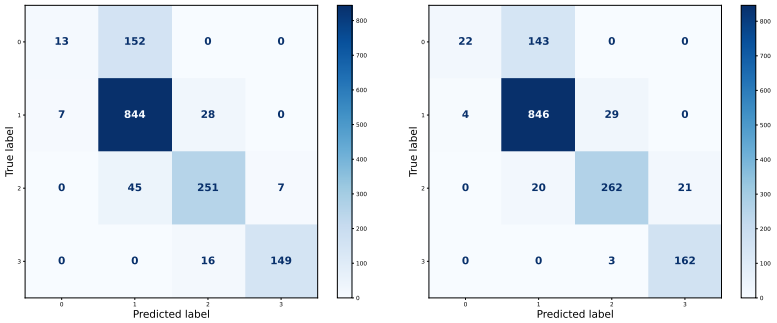

Fig. 2: Confusion matrix of TBI_103 with CE loss (left) and LD loss (right)

Table 5: Robustness evaluation under varying missing data ratios (10%–50%) on TBI_ODC dataset. The best performance metrics highlighted in **bold**.

Model	**Metric**	**10%**	**20%**	**30%**	**40%**	**50%**
RF	Acc. (%)	95.61	92.22	88.20	85.34	80.89
	F1 (%)	92.88	87.97	82.60	77.90	71.71
XGB	Acc. (%)	96.39	93.02	89.08	**86.42**	81.59
	F1 (%)	**94.81**	**90.13**	84.51	**80.41**	73.86
TabPFN	Acc. (%)	95.83	92.59	88.58	85.26	80.81
	F1 (%)	93.83	89.38	83.85	78.19	72.04
NAIM	Acc. (%)	94.17	91.24	85.34	81.36	75.27
	F1 (%)	91.18	87.10	78.91	73.08	64.69
G-TBI	Acc. (%)	**96.44**	**93.28**	**90.12**	85.38	**81.82**
	F1 (%)	94.17	89.52	**86.75**	78.27	**74.07**

Missing data handling Table 5 evaluates framework robustness against simulated data sparsity from 10% to 50% on the TBI_ODC dataset. While performance predictably degrades across all architectures as the missing ratio escalates, the proposed G-TBI framework maintains highly resilient baselines, establishing top accuracy scores in nearly all configurations (peaking at 96.44% and 81.82% under 10% and 50% sparsity). G-TBI demonstrated superior stability under extreme data omission (50%), leveraging its graph attention topological structure to effectively mitigate severe features loss.

Sensitivity analysis Finally, a sensitivity analysis evaluated the impact of hyperparameters—attention heads (L), network layers (D), and neighborhood size (k)—on performance, using ($L = 4, D = 2$) as the baseline. To isolate individual effects, each variable was altered independently while keeping others fixed at their optimal values. Specifically, when varying L, the depth D was maintained at 2, and conversely, L remained frozen at 4 while varying D. The empirical results are consolidated in Table 6 and Figure 3.

Table 6: Comparison of attention heads and network layers

Structural component	**Variants**	**TBI_103**		**TBI_ODC**	
		Acc. (%)	**F1 (%)**	**Acc. (%)**	**F1 (%)**
Attention heads L	2	83.33	70.89	91.98	85.12
	8	76.59	73.68	88.62	81.21
Network layers D	3	83.53	70.45	96.48	94.56
	4	82.54	69.92	93.64	90.13
	5	79.17	70.55	92.26	85.74
Proposed framework	L=4, D=2	85.44	79.46	97.22	96.31

Table 6 showed that our proposed configuration achieves the highest performance across both datasets. Deviations through adding more attention heads or network layers lead to performance degradation, likely due to over-smoothing and noise accumulation. Extending the network beyond the optimal configuration degrades accuracy, as deeper layers trigger over-smoothing that blurs node representation distinctness. For graph connectivity, Figure 3 highlights a scale-driven divergence in neighborhood boundaries. The smaller TBI_103 cohort peaks early at $k = 25$ before larger sizes inject non-informative noise, whereas the expanded sample space of TBI_ODC requires a denser connectivity $k = 60$ to effectively stabilize feature propagation.

5 Conclusion

In this paper, we proposed G-TBI, a graph-based framework for TBI severity prediction from tabular data under missing data conditions. By incorporating a

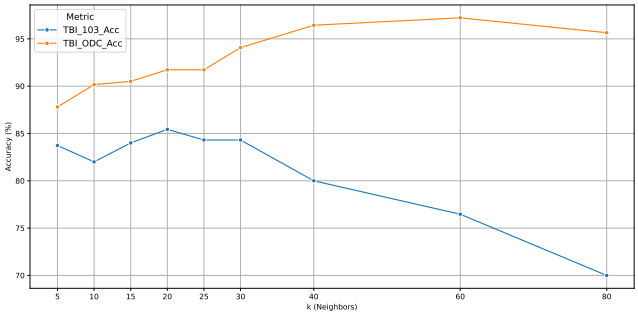

Fig. 3: Classification accuracy across varying neighborhood k

Label-Dependent loss function, the framework enforces class penalization margins, successfully minimizing critical underestimation errors and preserving the continuous ordinal progression of injury severity. Furthermore, the inductive learning setup is highly suitable for clinical deployment, as it enables instant evaluation of new patients without requiring computationally expensive graph retraining. Experimental results demonstrate G-TBI's robustness, achieving an accuracy of 85.45% on the TBI_103 dataset and 97.22% on the TBI_ODC, consistently outperforming traditional machine learning and deep tabular baselines. Ablation analyses confirm that graph attention-based message passing combined with ordinal regularization significantly improves representation quality. Future research will focus on evolving G-TBI into a transparent clinical decision support tool and validating it on larger, multi-center datasets to ensure global reliability.

Acknowledgment

This research is funded by Hanoi University of Science and Technology (HUST) under project number T2025-PC-072.

References

1. Acosta, J.N., Falcone, G.J., Rajpurkar, P., Topol, E.J.: Multimodal biomedical ai. Nature medicine **28**(9), 1773–1784 (2022)
2. Arik, S.Ö., Pfister, T.: Tabnet: Attentive interpretable tabular learning. In: Proceedings of the AAAI conference on artificial intelligence. vol. 35, pp. 6679–6687 (2021)
3. Caruso, C.M., Soda, P., Guarrasi, V.: Not another imputation method: a transformer-based model for missing values in tabular datasets. arXiv preprint arXiv:2407.11540 (2024)
4. Delgado, R.: Ordinal classification with label-dependent loss. Machine Learning **115**(3), 60 (2026)

5. Doan, D.K., Nguyen, T.K., Duong, M.T., Nguyen, T.B., Tran, T.H.: Tbi-ttm: Traumatic brain injury prognosis with textual data under missing tabular conditions. In: 2025 24th International Symposium on Communications and Information Technologies (ISCIT). pp. 122–127 (2025). https://doi.org/10.1109/ISCIT67082.2025.11231549
6. Duc-Khiem, D., Thanh-Hai, T., Kien, T., Lan, L.T., Vu, H., Huu-Khanh, N., Van-Mao, C., Thanh-Bac, N.: viTBI-BERT: A vietnamese language model for prediction of traumatic brain injury. ICT Research p. 7 (March 2025). https://doi.org/10.32913/mic-ict-research.v2024.n2.1303
7. Gorishniy, Y., Rubachev, I., Khrulkov, V., Babenko, A.: Revisiting deep learning models for tabular data. Advances in neural information processing systems **34**, 18932–18943 (2021)
8. Guo, X., Quan, Y., Zhao, H., Yao, Q., Li, Y., Tu, W.: Tabgnn: Multiplex graph neural network for tabular data prediction. arXiv preprint arXiv:2108.09127 (2021)
9. Hollmann, N., Müller, S., Eggensperger, K., Hutter, F.: Tabpfn: A transformer that solves small tabular classification problems in a second. arXiv preprint arXiv:2207.01848 (2022)
10. Lee, S., Park, M., Ahn, Y., Jung, G.B., Kim, D.: Analysis of tabular data based on graph neural network using supervised contrastive loss. Neurocomputing **570**, 127137 (2024)
11. Li, C.T., Tsai, Y.C., Chen, C.Y., Liao, J.C.: Graph neural networks for tabular data learning: A survey with taxonomy and directions. ACM Computing Surveys **58**(1), 1–51 (2025)
12. Liao, J.C., Chiu, J.W., Li, C.T.: Tabgsl: Graph structure learning for tabular data predictions. In: 2025 IEEE International Conference on Big Data (BigData). pp. 4155–4164. IEEE (2025)
13. Lin, W.C., Tsai, C.F.: Missing value imputation: a review and analysis of the literature (2006–2017). Artificial Intelligence Review **53**(2), 1487–1509 (2020)
14. Liu, J., Xu, A., Zhao, Z., Fang, D., Lv, W., Li, Y., Wang, P., Wang, Y., Dai, Y., Zheng, X., et al.: The burden of traumatic brain injury, its causes, and future trend predictions in 204 countries and territories (1990–2021): Results from the global burden of disease study 2021. Neuroepidemiology (2025)
15. Nelson, L., Magnus, B., Yue, J., Balsis, S., Patrick, C., Temkin, N., Diaz-Arrastia, R., Manley, G.: Data-driven characterization of traumatic brain injury severity from clinical, neuroimaging, and blood-based indicators. Research Square pp. rs–3 (2024)
16. Ngo, T.H., Tran, M.H., Nguyen, H.B., Hoang, V.N., Le, T.L., Vu, H., Tran, T.K., Nguyen, H.K., Can, V.M., Nguyen, T.B., et al.: E-tbi: explainable outcome prediction after traumatic brain injury using machine learning. Medical & Biological Engineering & Computing pp. 1–18 (2025)
17. Salvi, M., Loh, H.W., Seoni, S., Barua, P.D., García, S., Molinari, F., Acharya, U.R.: Multi-modality approaches for medical support systems: A systematic review of the last decade. Information Fusion **103**, 102134 (2024)
18. Sebastian, J., King, G.G.: Fusion of multimodality medical images-a review. 2021 Smart Technologies, Communication and Robotics (STCR) pp. 1–6 (2021)
19. Yan, J., Luo, L., Deng, C., Huang, H.: Adaptive hierarchical similarity metric learning with noisy labels. IEEE Transactions on Image Processing **32**, 1245–1256 (2023)
20. Zhou, Z.H., Liu, X.Y.: On multi-class cost-sensitive learning. Computational Intelligence **26**(3), 232–257 (2010)