

# Knowledge-augmented Patient Network Embedding-based Dynamic Model Selection for Predictive Analysis of Pediatric Drug-induced Liver Injury

Linjun Huang, Zixin Shi, Fei Tang, and Haolin Wang

**Abstract**—Objective: To address the challenges of developing machine learning frameworks for Electronic Health Records (EHRs)-based predictive tasks, such as the intricate occurrence mechanism of clinical events, patient diversity, and the inherent limitations of real-world data like data incompleteness and class imbalance, we propose the Knowledge-augmented Patient Network embedding-based Dynamic model Selection (KPNDS) framework, focusing on two key aspects: dynamically selecting the most suitable model for each individual and integrating biomedical knowledge into the framework.

Methods: KPNDS utilizes graph machine learning algorithms to generate patient embeddings from a knowledge-augmented network which integrates data from a diverse range of data sources including EHRs, drug-related information, toxicogenomics data and other relevant information to enrich the understanding of patients. A meta-learning based framework is adopted to dynamically select the optimal classifiers based on the latent patient representations to perform individualized risk prediction. Multi-Layer Perceptron, Transformer and Kolmogorov-Arnold Networks are used as meta-classifiers to enhance the selection of the optimal classifiers for each patient.

Results: The KPNDS framework was validated for the early prediction of drug-induced liver injury (DILI) in pediatric patients. Experimental results show that it outperforms common baseline models and dynamic ensemble selection methods.

Conclusion: The KPNDS framework effectively integrates domain knowledge, graph-based machine learning and dynamic model selection strategies, thereby enhancing predictive performance.

Significance: The KPNDS framework seamlessly integrates knowledge-augmented networks with dynamic model selection techniques, which has the potential to enable more accurate risk assessment and personalized medicine in complex scenarios, highlighting a novel approach to integrating external knowledge with data-driven models.

**Index Terms**—Dynamic Classifiers Selection, Kolmogorov-Arnold Networks, Knowledge-augmented Patient Network, Meta-learning, Pediatric Drug-Induced Liver Injury, Graph Machine Learning.

## I. INTRODUCTION

**A**DVERSE drug reactions (ADRs) are defined as harmful and unintended effects occurring during normal drug usage for disease prevention, diagnosis, treatment, or physiological regulation [1]. ADRs represent a significant clinical concern, being one of the leading causes of morbidity and mortality globally. In Asia, the incidence of ADRs is estimated

This work was supported by the National Natural Science Foundation of China under Grant 72101040.

Linjun Huang is with College of Medical Informatics, Chongqing Medical University, Chongqing, China (e-mail: LinjunHuang414@163.com).

Zixin Shi is with College of Medical Informatics, Chongqing Medical University, Chongqing, China (e-mail: zixin\_plus@163.com).

Fei Tang is with College of Medical Informatics, Chongqing Medical University, Chongqing, China (e-mail: tangfei@stu.cqmu.edu.cn).

Haolin Wang is with College of Medical Informatics, Chongqing Medical University, Chongqing, China (e-mail: haolinwang@cqmu.edu.cn).

to be 1.8 to 4.2 cases per 1,000 hospitalized patients [2], [3]. Children are particularly vulnerable to ADRs due to their dynamic physiological development and the lack of pediatric-specific data during drug development [4]. Off-label drug use is common in pediatrics, often driven by limited therapeutic options [5]. However, this practice creates significant gaps in understanding the pharmacological effects of these drugs in children, emphasizing the need for early ADR prediction in this population [6]. Drug-induced liver injury (DILI) is a rare but severe form of ADR that can lead to acute liver failure and fatal outcomes [7], [8]. Antibiotics are among the primary drug categories associated with allergic reactions and even fatalities due to ADRs [9].

In recent years, Machine Learning (ML) has increasingly been employed to predict ADRs [10], [11]. However, several limitations exist in these applications. First, most existing models rely solely on clinical data for prediction , neglecting the potential value of integrating non-clinical features [12]. Second, research often focuses on single-center, single-department datasets, limiting generalizability and failing to address data-related issues like incompleteness that are prevalent in real-world multi-center setting. Third, studies frequently compare conventional ML models (e.g., artificial neural networks, logistic regression, random forests, and support vector machines) in isolation, overlooking the complementary strengths of diverse models when applied to different data features [13]. Finally, ADRs are rare events, resulting in highly imbalanced datasets. Many current studies use relatively balanced, small datasets, which can bias results toward models that perform well on smaller datasets but may not generalize effectively to real-world scenarios [14], [15].

In an effort to overcome the limitations inherent in existing prediction models, we propose the KPNDS framework. The framework capitalizes a knowledge-augmented graphical structure to uncover potential associations among patients. Specifically, we construct a heterogeneous network based on fundamental medication information and integrate additional knowledge to strengthen the network structure. Graph embedding methods are applied to learn robust patient representations within the patient knowledge network. Additionally, we introduce a dynamic classifier selection approach to identify the similarities among patients and dynamically select a tailored subset of models for each individual patient. This strategy is founded on the concept that each model may be an expert in a specific feature space and capable of accurately identifying patterns within that space for classification tasks. By leveraging diverse classifiers, this method tackles the difficulties posed by the heterogeneous, diverse, and incomplete patient data prevalent in real-world clinical settings. Given that different patients may have distinct modeling requirements, and the

use of a Multiple Classifier System (MCS) [16] allows for the aggregation of the complementary advantages of various classifiers, thereby enhancing the accuracy and flexibility of the framework. Furthermore, we employ the Multilayer Perceptron, Transformer and Kolmogorov-Arnold Network [17] as meta-classifiers for models selection based on the patients' embeddings. These methods play a crucial role in selecting the most appropriate single classifier or a subset of classifiers for each patient by capitalizing on the latent relationships among patients, enabling a more effective handling of specific clinical tasks. In summary, the key contributions of this study are as follows:

- We present a dynamic classifier selection framework that integrates domain knowledge to assess patient similarities, identify patient subgroups, and predict the risk of DILI in children receiving antibiotics.
- We enrich the network with extra relevant knowledge and explore several graph embedding methods like Node2Vec, Metapath2Vec, GCN and GAT to capture latent information within the KPN.
- We employ MLP, Transformer and KAN to select the optimal classifier subset for each test sample. Experimental results demonstrate that the proposed KPNDS outperforms traditional machine learning models and dynamic classifier selection methods, particularly by leveraging knowledge-augmented network embedding to dynamically select the classifier subsets for each patient.

## II. RELATED WORKS

### A. Machine learning for ADRs prediction

Recently, machine learning has gained significant attention in ADR prediction due to its ability to extract features from high-dimensional, complex, and heterogeneous data, uncovering hidden patterns and relationships. Various studies have demonstrated its potential in this field. For example, Hu et al. [18] constructed a multicenter cohort of critically ill children and developed interpretable prediction models for AKI using 11 machine learning algorithms. Their optimal model achieved excellent performance in external validation (AUC = 0.910) and has been translated into a practical clinical tool to enhance its application in clinical settings. Attayeb Mohsen et al. [14] developed 14 predictive models using deep neural networks across different datasets, achieving an average accuracy of 89.4%. Their findings highlighted the models' outstanding performance in predicting ADRs in the contexts of duodenal ulcer and fulminant hepatitis, demonstrating the effectiveness of deep learning approaches for ADR prediction. Anca Mirela Chiriac et al. [19] constructed predictive models for  $\beta$ -lactam allergy in a retrospective cohort of 1,991 patients and a prospective cohort of 200 patients using logistic regression and decision tree methods. Although the models performed similarly across the two cohorts, the sensitivity for logistic regression was 51% (vs. 60%), and for the decision tree, it was 29.5% (vs. 43.5%). This finding indicates that the accuracy in predicting positive cases remains relatively low, underscoring the need to explore alternative approaches to improve the performance of machine learning models in identifying positive ADR cases.

### B. Network-based machine learning

Network-based machine learning has gained increasing attention in recent years for its potential to integrate and

analyze complex interconnections among diseases, patients, and clinical outcomes [20]–[22]. Specifically, a network-centric perspective was adopted by Liya Wang et al. [23], who constructed multimorbidity networks to explore gender- and age-specific differences in chronic disease patterns, identifying complex interactions that varied significantly across demographic groups. Tao Wang et al. [24] employed temporal bipartite network modeling to analyze the evolving multimorbidity trajectories in patients with severe mental illnesses (SMI). In addition, the combination of network theory with machine learning and deep learning methods to enhance disease risk assessment has garnered increasing attention. For instance, Ping Yang et al. [25] introduced a novel framework for the early prediction of high-cost inpatients with ischemic heart disease by combining traditional clinical features with network-based metrics derived from a phenotype comorbidity network and a cost-distance network. Shahadat Uddin et al. [26], [27] and colleagues have conducted multiple studies constructing patient networks or disease networks based on comorbidity relationships and extracting network-level features, such as three variants of PageRank score and attributes like degree centrality, to improve the prediction of comorbidity and multimorbidity risks using different machine learning and deep learning classifiers. They also utilized node embeddings for link prediction in chronic disease comorbidity networks, comparing the performance of traditional similarity metrics and embedding techniques such as node2vec and graph neural networks. Further advancing the predictive capabilities of network-based models, Qiu et al. [28] combined multimorbidity and patient similarity networks to enhance length-of-stay predictions, while Yang et al. [29] applied protein-protein interaction networks to develop GraphSynergy, a graph convolutional model for anticancer drug synergy prediction. Expanding the scope of network-based learning, frameworks like DCNeT integrate temporal and structural features to predict cardiovascular risks in patients with mental disorders. By incorporating disease embeddings, attention mechanisms, and time-aware LSTM models, Qiu et al. [30] demonstrated the potential of hybrid architectures to capture both static and dynamic aspects of patient trajectories. Collectively, these studies illustrate the versatility and effectiveness of network-based machine learning in addressing a wide range of clinical challenges, spanning from early disease prediction to personalized treatment strategies. These studies strongly inspire us to explore the integration of domain knowledge in constructing networks, while leveraging advanced embedding techniques and deep learning frameworks to better capture the rich structural information and latent patterns.

### C. Dynamic Classifier Selection

MCS aims to leverage the complementary strengths of different classifiers by combining their predictions to improve recognition performance. These systems, built using the same or different models and/or datasets, aim to address Wolpert's "No Free Lunch" theorem: "no single classifier can be universally optimal for all pattern recognition tasks, as each model has its specific capability range" [16], [31]. The core idea of dynamic selection (DS) is to identify the position of a new sample within its local space and select the optimal model or subset of models for prediction. This is categorized as dynamic classifier selection (DCS) when choosing a single model or dynamic ensemble selection (DES) when selecting a subset of

models for prediction. DS generally involves three stages: (a) generation, (b) selection, and (c) integration. The classifier pool generation stage forms the foundation of MCS, with extensive research focusing on improving the quality of classifierS pools. For instance, Marcos Monteiro Jr. et al. [32] proposed guiding classifier pool generation using data complexity and classifier decision space diversity. Hesam Jalalian et al. [33] constructed meta-models based on dataset meta-features to predict the most suitable pool generation strategies and DES methods for given datasets. Additionally, Mariana A. Souza et al. [34] generated locally accurate classifier pools to address the limitations of globally generated pools in selecting the most capable classifiers for challenging regions in the feature space. Rafael M. O. Cruz et al. [35] extracted five groups of meta-features that measure classifier performance for a given input sample, demonstrating that meta-classifiers can achieve results comparable to an Oracle system. Reza Davtalab et al. [36] proposed using fuzzy hyper-boxes to generate capability and incapability maps for each classifier, addressing the limitations of KNN-based DES methods in large-scale problems and imbalanced data. Mariana A. Souza et al. [37] introduced a multi-label graph neural network (GNN) to learn dynamic classifier combination rules by modeling instance-instance and classifier-classifier dependencies, simplifying the embedding space for dynamic selection tasks. Che Xu et al. [38] developed a data-driven decision model for dynamic classifier selection by considering the historical similarity of alternative models and their prediction accuracy, successfully applying this approach to thyroid nodule diagnosis. Overall, DS methods serve as an effective ensemble learning approach, particularly for handling the challenges within the local feature space. They have consistently demonstrated superiority over baseline models, thus representing a promising research avenue worthy of in-depth exploration.

### III. MATERIALS AND METHODS

#### A. Basic problem setup

Our prediction task was formulated as a binary classification problem aimed at assessing whether pediatric patients receiving antibiotics are at risk of developing DILI. We began by extracting structured features from pediatric patients' EHRs to create a structured dataset. Inspired by DS techniques, we used this dataset to train multiple base classifiers, denoted as  $c_{ij}$ , to form a classifier pool  $C = \{c_{11}, c_{12}, \dots, c_{1m}, c_{21}, c_{22}, \dots, c_{2r}\}$ . Also, we incorporated supplementary information from a patient heterogeneous knowledge network, enabling the extraction of unique patient embeddings. These embeddings—representing each patient's specific characteristics and latent similarities to others—were then used as inputs to a deep neural network (MLP, Transformer or KAN). The model aimed to select the optimal classifier or subset of classifiers for each test sample, enabling more tailored predictions based on individual patient profiles.

#### B. Data source and study cohort

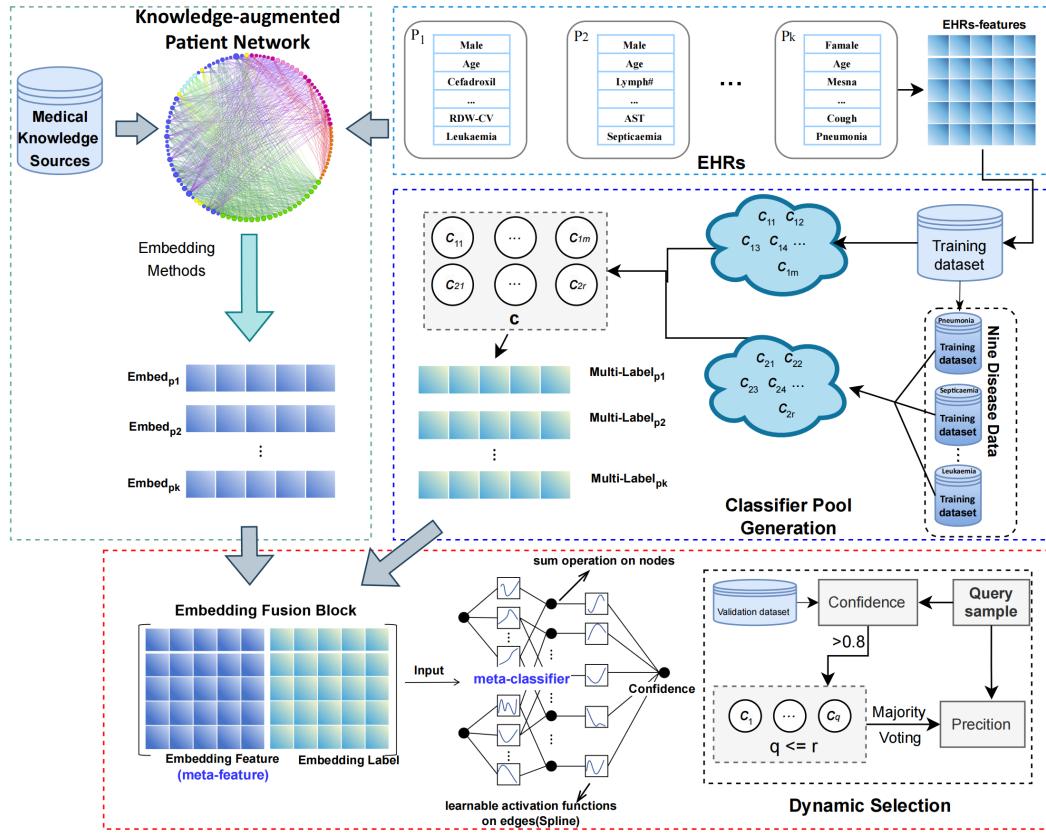
This study is a retrospective prognostic investigation that utilizes a pediatric administrative dataset from six tertiary hospitals in Chongqing, China, covering the years 2013 to 2023 to develop a risk prediction model. All the included patients had received antibiotic treatment, the leading cause of severe ADRs in pediatric population, though DILI wasn't always antibiotic-induced. Whether a patient developed DILI

was marked according to the discharge diagnosis or the medical course records. Given the pronounced imbalance in ADR incidence, systematic random sampling was used to select control samples, resulting in a final dataset of 12,353 patients. The cohort includes 10,065 children (ages 0-10) and 2,288 adolescents (ages 10-19), defined according to the World Health Organization's age criteria. We used the first relevant data records of the patients after their hospital admission to develop the predictive model, and we confirmed that the patients had not been found to have DILI at that time based on the medical course records.

Each sample in the dataset includes both structured EHRs features and embedding features (Embed) extracted from the KPN, incorporating both clinical data and representations enriched by external domain knowledge sources. Due to the pronounced imbalance in the dataset (1,005 positive samples and 11,348 negative samples), we ensured an equal proportion of positive and negative cases in each fold during stratified five-fold cross-validation, enabling the model to effectively learn the characteristics of positive cases. For each fold, we randomly split the training set in an 8:2 ratio to create separate training and validation sets, which were used to train and validate the proposed KPNDS framework. Quantitative predictors were assessed using the Mann-Whitney U test, while binary predictors were evaluated based on Chi-square tests ( $\chi^2$  test) to calculate the p-value. A detailed description of the dataset is provided in Table 1 of the supplement.

#### C. The Proposed KPNDS framework

Figure1 illustrates the proposed KPNDS framework with three key components: network-enhanced patient representation learning, classifier pool generation, and dynamic selection. The first component involves developing a basic patient-drug network through extracting information from EHRs data and enriching it with external domain knowledge. The EHRs data incorporates demographic information, medication history, laboratory indicators, and disease status. The integration of external knowledge sources, along with the exploration of various graph embedding methods, play a pivotal role in facilitating the identification of clinically relevant patient similarity groups. During the classifier pool generation phase, we employed 11 common machine learning methods, including Logistic Regression (LR), Extra Trees, Random Forests (RF), Linear Discriminant Analysis (LDA), Gradient Boosting Classifier, Support Vector Machine (SVM), Decision Tree, Adaptive Boosting (AdaBoost), Quadratic Discriminant Analysis (QDA), Light Gradient Boosting Machines (LightGBM), XGBoost, TabNet-Classifier [39], and ExplainableBoostingClassifier [40] (EBM), to construct a heterogeneous pool of classifiers. This diversified approach leverages different perspectives towards classification problems, ultimately enhancing overall performance. Finally, the meta-classifier (the KAN model in this study) processes meta-feature (patient embeddings) through nonlinear transformations and aggregates node features using summation operations to establish edge connections. Along these edges, the model employs learnable activation functions (spline functions) to adjust connection strengths and capture complex relationships among patient features. This mechanism empowers KAN to effectively model the intricate dependencies within patient embeddings. The output of KAN represents the confidence scores of each base classifier for every patient embedding, indicating the suitability and preference of each classifier for specific patients.



**Fig. 1.** The proposed Patient Knowledge Network-based Dynamic Selection (KPNDS) framework

In the testing phase, confidence scores for all classifiers are computed based on the patient embeddings. The framework dynamically selects the most suitable classifier or subset of classifiers for each test sample. The predictions of the selected classifiers are aggregated using majority voting to yield the final classification result.

### 1) Network-enhanced Patient Representation Learning

To identify patient subgroups characterized by latent similarities and select appropriate subsets of classifiers for each test sample, we constructed an undirected network, referred to as the KPN. In this study, the network we developed consists of 19,396 nodes, categorized into seven types: Patient (12,353, 63.69%), Phenotype (5,159, 26.6%), ADR (1,492, 7.69%), Gene (330, 1.7%), Drug (48, 0.25%), Disease (9, 0.05%), and Symptom (5, 0.03%). Data for Patient, Drug, Disease, and Symptom nodes were sourced from EHRs, while Phenotype, ADR, and Gene data were obtained from the Comparative Toxicogenomics Database (CTD)(<https://ctdbase.org/>) and the Side Effect Resource (SIDER) databases(<http://sideeffects.embl.de/drugs/>). The CTD provides information on drugs and chemicals associated with DILI, including curated associations (as markers, mechanisms, and/or therapeutics) or inferred links through curated gene interactions. Gene nodes relevant to the study were included and linked to drug nodes, indicating that certain drugs may influence specific genes leading to DILI. Additionally, genes associated with DILI are linked to specific phenotypes, leading to the addition of edges between gene and phenotype nodes. The SIDER database contains information on marketed drugs and their associated ADRs, prompting

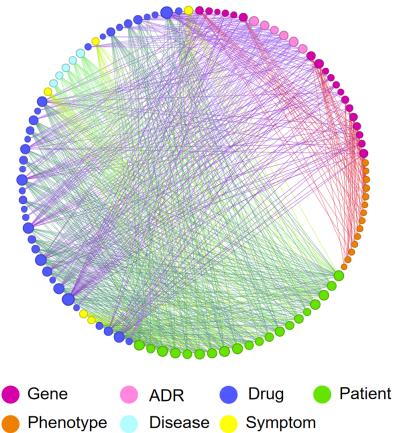
the introduction of ADR nodes for the drugs under study, with edges representing the correlation between drugs and their ADRs. Consequently, the final constructed KPN includes an extra 6,981 nodes and 20,159 edges. Figure2 presents a simplified example of the KPN, with an average degree of 16.55, a network diameter of 6, a network density of 0.01, and an average path length of 2.885. Considering node and edge representation learning is essential for visualization and downstream machine learning tasks, we employed several state-of-the-art graph representation learning methods and explored their performance variations in this task.

#### a) Random Walk-based Graph Machine Learning

Node2Vec captures contextual information of nodes by performing random walks within the graph. Its flexible random walk strategy effectively captures structural information, although it is typically applied to single-type graphs. In this study, we treat all node types as a single type (except for patient nodes) and project them into a single-type graph, enabling us to obtain patient embeddings with node2vec.

Metapath2Vec generates node embeddings in heterogeneous graphs using random walk strategies, with the key distinction being its ability to predefine metapaths that specify the relationship paths between nodes. For this study, we defined several metapaths to capture relationships between patients and other nodes:

- $\mathcal{P}_1$  : Patient  $\rightarrow$  Drug  $\rightarrow$  Patient
- $\mathcal{P}_2$  : Patient  $\rightarrow$  Symptom  $\rightarrow$  Patient
- $\mathcal{P}_3$  : Patient  $\rightarrow$  Drug  $\rightarrow$  Gene  $\rightarrow$  Drug  $\rightarrow$  Patient
- $\mathcal{P}_4$  : Patient  $\rightarrow$  Drug  $\rightarrow$  ADR  $\rightarrow$  Drug  $\rightarrow$  Patient



**Fig. 2.** A simplified example illustrating the KPN

- $\mathcal{P}_5 : \text{Patient} \rightarrow \text{Drug} \rightarrow \text{Gene} \rightarrow \text{Phenotype} \rightarrow \text{Gene} \rightarrow \text{Drug} \rightarrow \text{Patient}$

### b) Neural Network-based Graph Machine Learning

Graph Convolutional Networks (GCN) are a type of neural network that operate directly on graph structures, updating node embeddings by aggregating information from neighboring nodes. GCN consists of three graph convolution layers, each implemented with GCNConv operations, followed by a batch normalization layer to stabilize the learning process. The first two convolution layers have hidden dimensions of 256 and 128, respectively, capturing progressively complex features. The final layer projects the embeddings to a fixed output dimension, matching the required embedding dimension for downstream tasks.

Graph Attention Networks (GAT) extend GCN by introducing self-attention mechanisms, assigning different weights to each node's neighbors for more flexible aggregation of neighbor information. After two convolutional layers, these features progressively capture structural information from neighboring nodes, leading to effective embedding generation. GAT consists of two graph attention layers, each implemented with the GATConv operation.

Given that the KPN constructed in this study lacks node attributes, in both models, we initialized each node with a set of features sampled from a Gaussian distribution. ReLU activation functions are used to introduce non-linearity, which is crucial for capturing complex graph patterns. We also used the Adam optimizer, Dropout, and a weight decay of 5e-4 to prevent overfitting. We use the L2 norm of the output as the loss function, which serves as a regularization term. The L2 norm is defined as:

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n x_i^2} \quad (1)$$

Where  $x_i$  denotes each element of the output tensor. L2-norm allows the output features to be more regularly distributed in space, thus improving the model's performance on similarity learning.

### 2) Classifier Pool Generation

The construction of the base classifier pool is decoupled from patient representation and consists of two primary components. The first component involves training a heterogeneous set of classifiers on the entire training dataset, including

13 distinct models, resulting in a classifier pool denoted as  $C1 = \{c_{11}, c_{12}, \dots, c_{1m}\}$ . The second component addresses the inclusion of pediatric patients with DILI across nine different diseases during hospitalization in the dataset. In order to group homogeneous patients for more targeted classification, we created nine smaller training subsets for additional classifiers, resulting in a second classifier pool denoted as  $C2 = \{c_{21}, c_{22}, \dots, c_{2r}\}$ . These two sets of base classifiers are then merged into an expanded pool  $\{C = C1 + C2\}$ . This approach enhances the diversity of model combinations, allowing the framework to prioritize the selection of disease-specific classifiers when test samples belong to a specific disease subset. This strategy improves overall classification performance and substantially enhances the robustness of the classification system, ensuring more accurate and reliable results across various scenarios.

### 3) Dynamic Selection

Prior to utilizing the MLP, Transformer or KAN model for classifier selection, we compute the labels for each patient's embedding features, resulting in a multi-label matrix. Each row in this matrix corresponds to a patient, while each column represents a base classifier. If the predicted label  $\hat{y}$  from a base classifier matches the original label  $y$  of the patient, the corresponding entry in the matrix is set to 1; otherwise, it is set to 0. Next, we combine the patient embedding features with the corresponding label matrix. The training dataset  $Embed_{Train}$  is fed into the MLP, Transformer or KAN for training, while  $Embed_{valid}$  is used to validate the model's performance. We use BCEWithLogitsLoss as the loss function. It is appropriate because the model outputs logits, and BCEWithLogitsLoss combines the computation of Sigmoid and binary cross-entropy in a single step, allowing efficient and numerically stable training. The loss is computed by comparing the predicted logits with the true labels in the multi-label matrix, as follows.

$$\text{Loss} = -\frac{1}{N} \sum_i^N [y_i \cdot \log(\sigma(\hat{y}_i)) + (1 - y_i) \cdot \log(1 - \sigma(\hat{y}_i))] \quad (2)$$

where  $\sigma(\hat{y}_i)$  is the Sigmoid function applied to predicted logits  $\hat{y}_i$ .

For the test dataset  $Embed_{test}$ , we input it into the trained MLP, Transformer or KAN and apply the Sigmoid activation function to convert the logits into probabilities, which are interpreted as the confidence scores of each base classifier. These scores quantify each base classifier's propensity to correctly classify the test sample.

### 4) Prediction

We used a grid-search strategy to optimize the confidence threshold hyperparameter across the range of 0.5 to 0.99 (in steps of 0.01) for the three meta-learners. A classifier is considered capable of correctly classifying a given sample if its confidence score exceeds the selected threshold—0.8 for both MLP and KAN networks, and 0.7 for the Transformer model in our setup. The appropriate classifier or a subset of classifiers is then used to make predictions. The final label and probability are calculated using a majority voting mechanism as:

$$\text{count}(k) = \sum_{i=1}^q \delta(\text{votes}[i], k) \quad (3)$$

- $k$  is a possible class label ( $k = 0$  or  $1$ ),
- $q$  is the number of classifiers selected for each test sample,
- $\text{votes}[i]$  represents the predicted class from the  $i$ -th classifier,

- $\delta(\text{votes}[i], k)$  is an indicator function that returns 1 if  $\text{votes}[i] = k$ , and 0 otherwise.

$$\hat{y} = \underset{k}{\operatorname{argmax}} \text{count}(k) \quad (4)$$

Where  $\operatorname{argmax}$  function selects the class label  $k$  with the highest count as the final predicted label  $\hat{y}_i$ .

## IV. EXPERIMENT AND RESULTS

### A. Baseline models and implementation details

The proposed framework was evaluated by comparing it to commonly used machine learning methods and dynamic ensemble selection techniques. The heterogeneous base classifier pool was constructed using 13 baseline models, with hyperparameters determined through grid search. Additionally, we employed several dynamic selection (DS) techniques referenced in DES.lib [41], ultimately selecting the two best-performing methods: DESP and KNORAU. The KAN model was compared with the MLP and Transformer to explore the performance disparities between these three neural network architectures for this specific task. We have provided the core code associated with the proposed framework. Please refer to the link: <https://github.com/wanghaolin/KPNDS>.

The five-fold cross-validation strategy was employed to evaluate the performance of the KPNDS framework, with 20% of the training set selected as the validation set to optimize hyperparameters. Dropout, weight decay, learning rate scheduling, and early stopping were employed to prevent model overfitting. The evaluation metrics considered in this study included Matthews Correlation Coefficient (MCC), F1 Score, ROC\_AUC, Precision, and Recall. The ROC\_AUC metric assesses the overall performance of the model across various thresholds. MCC considers all values in the confusion matrix and is commonly used to evaluate imbalanced datasets. The F1 Score balances precision and recall, making it particularly suitable for situations where correctly identifying positive cases is essential. These metrics were calculated using the following formulas:

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (5)$$

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

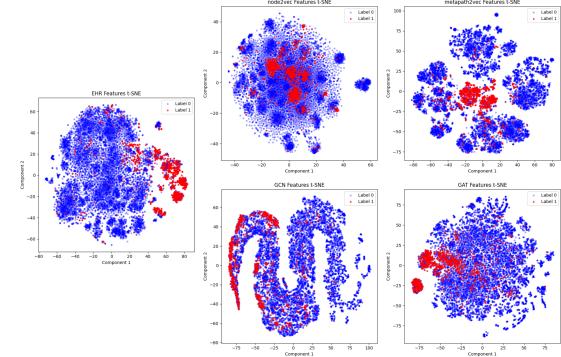
$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (8)$$

Where TP is true positive, TN is true negative, FP is false positive and FN is false negative.

### B. Performance comparison with baseline

To address performance degradation caused by severe class imbalance, we applied the SMOTE technique to augment the baseline models, resulting in noticeable improvements compared to the original versions (see Supplementary Table 2). Table I presents the performance of the proposed KPNDS framework compared to SMOTE-augmented baseline models and dynamic ensemble selection (DES) techniques on this dataset. For MCC, the best baseline model, RF, achieves a



**Fig. 3.** Visualized both conventional features (EHRs) and 4 graph-based embeddings

score of 0.723. Both DES techniques and the proposed KPNDS framework show improvements in this metric. Specifically, the MLP-based KPNDS, using GCN to obtain embeddings, achieves the highest MCC value of 0.758, representing a 3.5% improvement. In the case of LDA, while the model achieves the highest Recall of 0.948 in the entire experiment, it exhibits one of the lowest Precision scores at 0.475. Regarding the F1 score, RF achieves a value of 0.745, while KNORAU achieves 0.748, and the MLP-based KPNDS achieves the highest score of 0.777, representing an improvement of 3.2%. For ROC\_AUC, the baseline model demonstrates high performance at 0.981. In comparison, DES achieves 0.966, and the MLP-based KPNDS achieves 0.980, showing a slight decrease. However, the KAN-based KPNDS and Transformer-based KPNDS which using GAT to obtain embeddings, achieve an ROC\_AUC of 0.982, representing a slight improvement of 0.1%.

As shown in Table I, our results do not suggest that any specific graph embedding method combined with MLP or KAN model consistently outperforms across all metrics. Each graph embedding approach has unique advantages based on its underlying principles, and variations in results are to be expected. However, overall, any implementation of the KPNDS framework yields significant improvements. Furthermore, the KAN-based and Transformer-based KPNDS demonstrate superior performance in ROC\_AUC, a key metric for binary classification tasks.

### C. Feature Representation and Visualization

To assess the effectiveness of various feature representations, we applied t-SNE for dimensionality reduction to visualize both conventional EHRs features and graph-based embeddings, including node2vec, metapath2vec, GCN, and GAT. The visualizations in Figure 3 illustrate the clustering patterns of samples with Label 0 (negative) and Label 1 (positive).

The t-SNE plot of EHRs features shows significant overlap between the two classes, highlighting the limitations of conventional features in distinguishing between positive and negative samples. In contrast, graph-based embeddings exhibit varying levels of class separation, demonstrating their potential to capture the underlying structure of the data. Among the graph-based methods, metapath2vec shows clear clustering with distinct groupings, reflecting its ability to leverage the heterogeneous paths within the graph. GCN embeddings, on the other hand, demonstrate the most pronounced separation between the two classes, with well-defined boundaries, indicating that GCN effectively integrates both the graph topology

TABLE I  
PERFORMANCE COMPARISON

Method	Metric					
	MCC	Precision	Recall	F1	ROC_AUC	
SMOTE Augmented Baseline	LogisticRegression	0.661±0.011	0.541±0.023	0.893±0.019	0.673±0.013	0.974±0.003
	DecisionTree	0.582±0.024	0.471±0.027	0.832±0.080	0.598±0.017	0.953±0.010
	RandomForest	0.723±0.029	0.742±0.041	0.749±0.026	0.745±0.026	0.981±0.003
	SVM	0.688±0.016	0.770±0.032	0.659±0.025	0.709±0.014	0.976±0.004
	LDA	0.633±0.025	0.475±0.030	<b>0.948 ± 0.012</b>	0.632±0.027	0.974±0.004
	GradientBoosting	0.703±0.013	0.738±0.031	0.716±0.023	0.726±0.011	0.980±0.003
	AdaBoost	0.686±0.018	0.699±0.028	0.726±0.012	0.712±0.017	0.975±0.003
	QDA	0.647±0.020	0.509±0.026	0.914±0.016	0.654±0.021	0.965±0.005
	Extratree	0.720±0.019	0.732±0.028	0.754±0.012	0.743±0.018	0.980±0.003
	LGBM	0.701±0.027	0.754±0.044	0.697±0.019	0.724±0.024	0.980±0.003
DES	XGBoost	0.712±0.027	0.750±0.038	0.721±0.017	0.735±0.024	0.981±0.003
	TabNet	0.647±0.042	0.550±0.061	0.847±0.037	0.664±0.043	0.962±0.002
	EBM	0.673±0.026	0.753±0.044	0.648±0.013	0.696±0.022	0.978±0.003
	DESKNN	0.711±0.022	0.769±0.035	0.700±0.026	0.732±0.020	0.883±0.017
	APosteriori	0.642±0.037	0.654±0.050	0.692±0.035	0.671±0.034	0.893±0.025
	KNORAU	0.729±0.022	<b>0.797 ± 0.038</b>	0.706±0.028	0.748±0.020	0.966±0.006
	KNORAE	0.697±0.022	0.741±0.038	0.703±0.022	0.720±0.020	0.862±0.020
	DESP	0.724±0.030	0.794±0.044	0.700±0.032	0.743±0.027	0.960±0.006
	MCB	0.650±0.040	0.658±0.018	0.703±0.016	0.679±0.004	0.911±0.014
	MLA	0.648±0.028	0.747±0.027	0.607±0.033	0.669±0.027	0.948±0.016
MLP-based KPNDS	Rank	0.677±0.031	0.714±0.044	0.694±0.025	0.703±0.028	0.938±0.023
	KNOP	0.726±0.020	0.786±0.034	0.711±0.018	0.746±0.018	0.929±0.023
	LCA	0.648±0.028	0.747±0.027	0.607±0.033	0.669±0.027	0.948±0.016
	OLA	0.690±0.028	0.744±0.043	0.686±0.018	0.713±0.025	0.941±0.023
	APriori	0.671±0.012	0.692±0.028	0.705±0.024	0.698±0.010	0.904±0.032
	METADES	0.706±0.017	0.751±0.028	0.710±0.018	0.729±0.015	0.902±0.031
	node2vec	0.750±0.012	0.737±0.027	0.809±0.022	0.770±0.012	0.973±0.010
	metapath2vec	0.745±0.021	0.734±0.034	0.801±0.020	0.765±0.020	0.978±0.004
	GCN	<b>0.758 ± 0.028</b>	0.750±0.039	0.808±0.028	<b>0.777 ± 0.026</b>	0.973±0.008
	GAT	0.750±0.020	0.741±0.030	0.804±0.028	0.770±0.018	0.980±0.001
Transformer-based KPNDS	node2vec	0.733±0.014	0.697±0.038	0.821±0.025	0.753±0.014	0.975±0.010
	metapath2vec	0.740±0.015	0.713±0.036	0.818±0.060	0.759±0.013	0.979±0.004
	GCN	0.734±0.014	0.737±0.046	0.778±0.047	0.754±0.013	0.977±0.005
	GAT	0.753±0.018	0.739±0.045	0.811±0.025	0.772±0.017	<b>0.982 ± 0.002</b>
KAN-based KPNDS	node2vec	0.750±0.016	0.733±0.036	0.812±0.022	0.770±0.016	<b>0.982 ± 0.002</b>
	metapath2vec	0.749±0.015	0.734±0.037	0.810±0.023	0.769±0.015	<b>0.982 ± 0.002</b>
	GCN	0.756±0.022	0.737±0.041	0.821±0.025	0.776±0.021	<b>0.982 ± 0.002</b>
	GAT	0.754±0.019	0.737±0.043	0.816±0.022	0.773±0.018	<b>0.982 ± 0.002</b>

and feature information. GAT embeddings provide moderate separation, capturing relational dependencies, but with less distinction compared to GCN. Node2vec embeddings exhibit limited separation, suggesting that while random walks can capture local graph structures, they may not fully represent the global graph semantics relevant to this task. These results highlight the advantages of using graph-based embedding methods, particularly GCN, in extracting discriminative features for identifying similarities between patients.

#### D. Ablation study

To evaluate the contribution of external biomedical knowledge to overall model performance, we conduct an ablation study. Table II shows the performance of the KPNDS after removing external knowledge nodes and edges. Compared to Table I, while Recall slightly increased by 0.4%, the scores for MCC, Precision, F1, and ROC\_AUC dropped by 1.9%, 2%, 1.8%, and 0.3%, respectively. The knowledge nodes help bridge patients who are otherwise sparsely or weakly connected, allowing the graph embeddings to capture richer biological context and better reflect patient similarity. These results demonstrate that incorporating external domain knowledge can

effectively support the model in identifying similar patient groups, ultimately improving its predictive performance.

#### E. Interpretability

Our experiments demonstrate that KPNDS improves several metrics discussed earlier, and use a heatmap to explain how each patient selects the most optimal classifier through their embedding representations to predict DILI effectively. As shown in Figure4, the heatmap clearly illustrates the classifier confidence outputs for each test sample after learning patient embeddings with the KAN model. In this study, a classifier is considered effective for a sample if its confidence score exceeds 0.8, with the final prediction obtained through majority voting across the selected classifiers. This heatmap visually demonstrates the dynamic selection process, highlighting how the confidence scores for different classifiers vary across samples.

To further understand context-specific risk patterns for pediatric DILI, we conducted stratified SHAP analyses across distinct patient subgroups. SHAP values provide a detailed insight into the influence of individual features, helping us understand the rationale behind the model's selections and predictions. Figure5 displays SHAP summary plots for DILI prediction in

TABLE II  
ABLATION STUDY OF KPNDS

Method	Metric				
	MCC	Precision	Recall	F1	ROC_AUC
MLP-based PNDS	node2vec	0.725±0.018	0.711±0.029	0.788±0.040	0.746±0.017
	metapath2vec	0.730±0.019	0.703±0.032	0.811±0.065	0.750±0.018
	GCN	0.739±0.010	0.730±0.033	0.794±0.034	0.759±0.009
	GAT	0.734±0.029	0.719±0.063	0.800±0.041	0.754±0.029
Transformer -based PNDS	node2vec	0.734±0.015	0.706±0.025	0.813±0.028	0.755±0.014
	metapath2vec	0.725±0.013	0.681±0.025	0.825±0.037	0.745±0.012
	GCN	0.736±0.015	0.704±0.032	0.820±0.024	0.757±0.014
	GAT	0.735±0.016	0.726±0.046	0.792±0.053	0.755±0.014
KAN-based PNDS	node2vec	0.731±0.021	0.706±0.031	0.808±0.061	0.751±0.019
	metapath2vec	0.730±0.019	0.703±0.032	0.811±0.065	0.750±0.018
	GCN	0.729±0.021	0.687±0.020	0.825±0.044	0.749±0.019
	GAT	0.726±0.019	0.695±0.026	0.810±0.050	0.747±0.017

two cohorts: (A) patients with bone marrow suppression and (B) patients with pneumonia. These plots highlight both shared and unique risk factors for hepatotoxicity. As shown in Figure 5 (A) and (B), the use of drugs such as Glutathione and Oleanolic acid, along with elevated lab values—ALT, AST, GGT, and AST/ALT ratio—are strongly associated with increased DILI risk. In both subgroups, age consistently emerged as a key predictor, aligning with prior findings [42] that suggest even small age differences in children can significantly affect the likelihood of adverse drug reactions (ADRs). In Figure 5 (A), electrolyte imbalances (hypopotassemia, K+) are also strongly associated with higher DILI risk. Moreover, the wider spread of SHAP values observed in pneumonia patients reflects the model's heightened sensitivity to extreme feature values. Both plots further underscore the critical role of comorbidities in the model's decision-making process, suggesting that pediatric DILI is more likely to occur in children with multiple coexisting conditions.

#### F. External Validation

There are no public datasets supporting ADR-related model validation, which restricts the introduction of external data and prevents us from fully demonstrating the model's superiority. Therefore, in order to rigorously validate the generalizability of KPNDS, we have conducted additional external validation on an independent public dataset. Zhang et al. [43] collected data from 2,008 heart failure patients, each described by 166 clinical features. Based on their medication records, we constructed a patient-drug bipartite network and generated patient embeddings using the same graph embedding method described earlier. All data preprocessing steps—including cohort splitting and missing value handling—followed the exact protocol used in the pediatric DILI cohort. Table III presents the performance of KPNDS in predicting 28-day readmission among heart failure patients. KPNDS demonstrates notable improvements across nearly all evaluation metrics. Since the dataset includes multiple clinical outcomes, as noted by Zhang et al., we further applied KPNDS to predict 3-month and 6-month readmission, with results provided in Supplementary Table 3 and Supplementary Table 4, respectively. These results suggest that KPNDS exhibits strong generalizability across different patient populations and clinical prediction tasks.

#### V. DISCUSSION

In this study, we introduce a novel Knowledge-augmented Patient Network-based Dynamic Selection framework, KPNDS, for assessing the risk of DILI in pediatric patients receiving antibiotics. Traditional machine learning models often face challenges in predicting ADR risk due to intricate occurrence mechanism of DILI, the wide diversity among patients and the prevalence of incomplete and low-quality clinical data. To address this, the KPN was constructed using patient-drug relationships from EHRs as the foundational graph structure. Additional medical knowledge resources, including toxicogenomics, drug information, and side effects, were incorporated to enrich the graph structure. By leveraging graph embedding algorithms, the framework generates patient embeddings that reveal potential connections and similarities among patients. A dynamic selection module was designed to exploit the complementary characteristics of a diverse pool of base classifiers. By dynamically selecting and aggregating models based on latent patient representations, this module improves the model's accuracy and adaptability to individual patients. Furthermore, both traditional MLP models and the advanced KAN were used to learn relationships among patient embeddings, guiding the selection of the optimal classifier or classifier subsets for each case. The integration of the KPN and the innovative DS approach significantly improved the framework's performance. The proposed framework outperforms 13 baseline models and 13 dynamic selection techniques, demonstrating its effectiveness in identifying high-risk pediatric patients at potential risk of DILI caused by antibiotic use.

To our knowledge, this is the first study to integrate domain knowledge in constructing a patient knowledge network to identify patient similarities and generate embedding representations for dynamic classifier selection, specifically aimed at predicting the risk of DILI in pediatric patients receiving antibiotics. No prior research has attempted to combine these methods for dynamic selection, making this a novel dynamic ensemble approach. Furthermore, no previous studies have employed MCS modeling for DILI prediction. As shown in Table I, the proposed KPNDS framework outperforms 13 existing DES methods in this task, highlighting the meaningful and beneficial role of external features in dynamic selection. The framework incorporates four advanced embedding methods to enhance patient similarity representations, including node2vec, metapath2vec, GCN, and GAT. Among these, GCN

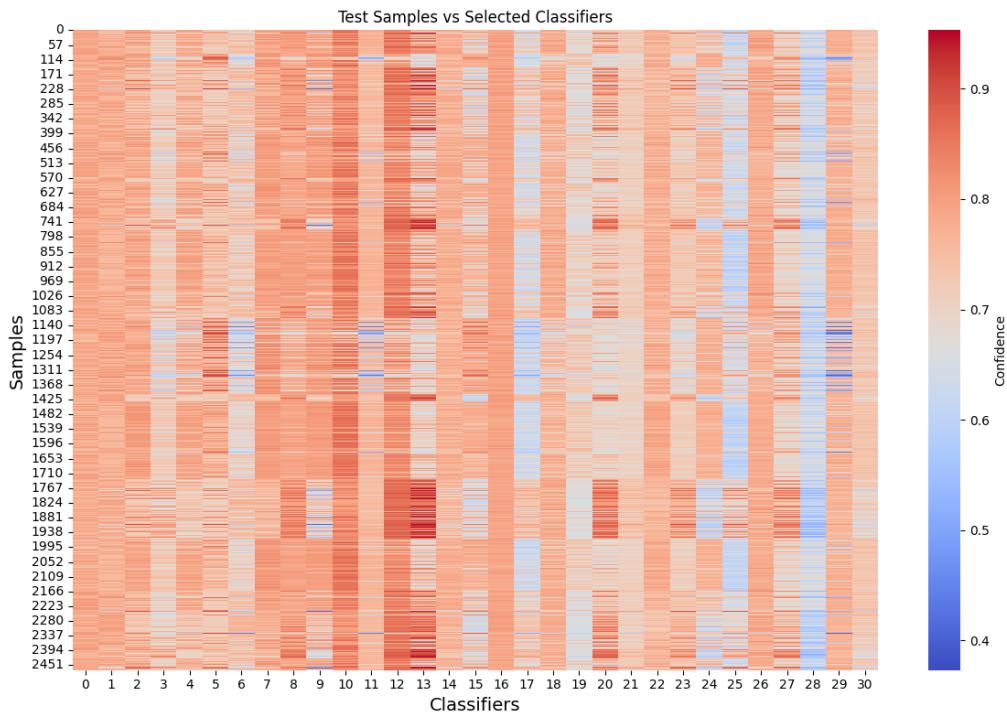


Fig. 4. Dynamic Classifier Selection Based on Confidence Scores for Pediatric DILI Prediction: A Heatmap Visualization.

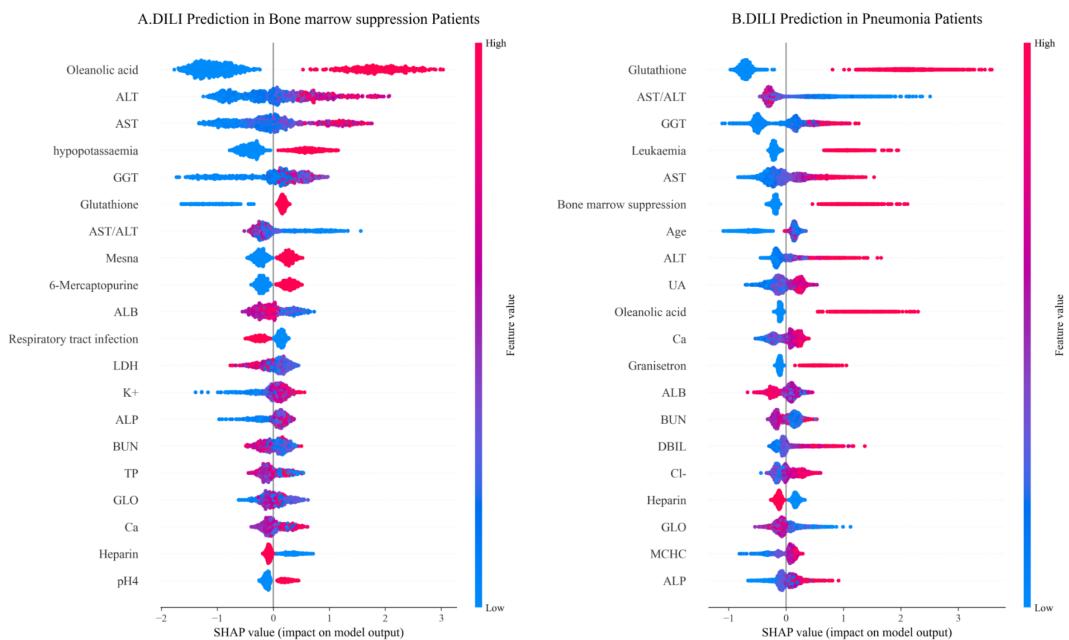


Fig. 5. Stratified SHAP Analysis of DILI Risk Factors in two Patient Subgroups.

TABLE III  
THE PERFORMANCE OF KPNDS IN PREDICTING 28-DAY READMISSION AMONG HEART FAILURE PATIENT

Method	Metric				
	MCC	Precision	Recall	F1	ROC_AUC
Baseline	LogisticRegression	0.101±0.060	0.137±0.043	0.257±0.069	0.178±0.052
	DecisionTree	0.033±0.030	0.084±0.013	0.279±0.127	0.126±0.025
	RandomForest	0.105±0.040	0.130±0.022	0.307±0.066	0.182±0.032
	SVM	0.104±0.060	0.116±0.030	0.421±0.089	0.181±0.044
	LDA	0.057±0.049	0.094±0.021	0.343±0.095	0.148±0.034
	GradientBoosting	0.091±0.046	0.123±0.030	0.293±0.042	0.172±0.036
	AdaBoost	0.084±0.063	0.105±0.027	0.407±0.103	0.166±0.042
	QDA	0.022±0.105	0.080±0.046	<b>0.464 ± 0.234</b>	0.133±0.068
	Extratree	0.101±0.060	0.137±0.043	0.257±0.069	0.178±0.052
	LGBM	0.074±0.036	0.101±0.015	0.393±0.068	0.160±0.024
MLP-based KPNDS	XGBoost	0.081±0.046	0.106±0.022	0.379±0.074	0.165±0.033
	TabNet	0.031±0.084	0.133±0.135	0.329±0.224	0.115±0.035
	EBM	0.082±0.046	0.118±0.027	0.271±0.058	0.164±0.037
	node2vec	0.204±0.075	0.281±0.139	0.279±0.069	0.252±0.052
	metapath2vec	0.189±0.053	0.255±0.096	0.257±0.027	0.243±0.037
	GCN	0.183±0.083	0.267±0.130	0.236±0.083	0.231±0.064
	GAT	<b>0.231 ± 0.066</b>	<b>0.315 ± 0.130</b>	0.279±0.057	<b>0.275 ± 0.045</b>
	node2vec	0.192±0.0906	0.278±0.144	0.243±0.042	0.241±0.068
	metapath2vec	0.195±0.089	0.279±0.014	0.250±0.045	0.245±0.066
	GCN	0.1977±0.086	0.278±0.143	0.257±0.042	0.247±0.063
Transformer-based KPNDS	GAT	0.206±0.081	0.289±0.136	0.257±0.042	0.254±0.060
	node2vec	0.142±0.035	0.178±0.074	0.329±0.097	0.205±0.021
	metapath2vec	0.168±0.065	0.244±0.151	0.307±0.107	0.217±0.041
	GCN	0.153±0.064	0.210±0.106	0.307±0.117	0.208±0.054
	GAT	0.174±0.031	0.231±0.093	0.293±0.092	0.227±0.024
	node2vec	0.192±0.0906	0.278±0.144	0.243±0.042	0.241±0.068
	metapath2vec	0.195±0.089	0.279±0.014	0.250±0.045	0.245±0.066
	GCN	0.1977±0.086	0.278±0.143	0.257±0.042	0.247±0.063
	GAT	0.206±0.081	0.289±0.136	0.257±0.042	0.254±0.060
	node2vec	0.142±0.035	0.178±0.074	0.329±0.097	0.205±0.021
KAN-based KPNDS	metapath2vec	0.168±0.065	0.244±0.151	0.307±0.107	0.217±0.041
	GCN	0.153±0.064	0.210±0.106	0.307±0.117	0.208±0.054
	GAT	0.174±0.031	0.231±0.093	0.293±0.092	0.227±0.024

embeddings yielded the best overall performance, as illustrated in figure3, where positive samples are clearly clustered in elongated regions with distinct boundaries from the negative samples. A possible explanation for this is that GCN training directly incorporates the graph's topological structure and node features, allowing it to capture complex graph patterns and effectively separate different classes. In contrast, GAT introduces an attention mechanism that, as seen in figure3, captures important local patterns within the graph. While it demonstrates better separation in certain local regions, its overall discriminative power needs further enhancement. KPNDS also employed three deep learning models—MLP Transformer and KAN—to learn from the embedding features. As shown in Table I, the MLP-based KPNDS using GCN embeddings achieved the best performance in terms of MCC and F1 score. While it demonstrated excellent precision, recall was slightly lower compared to the KAN or Transformer-based KPNDS. However, in the context of imbalanced medical datasets, recall is a more critical metric, as early identification of ADRs in pediatric patients can help clinicians adjust medication doses or modify treatment plans, ultimately improving patient outcomes and care quality.

Interestingly, Table I also shows that baseline models achieved very high ROC\_AUC scores, which may be attributed to data imbalance, causing model bias toward predicting negative samples and thereby increasing accuracy by labeling more samples as negative. While DES methods can effectively select classifiers for different classes, issues such as classifier selection mechanisms and weight distribution may hinder the prediction of positive samples in some cases. In contrast, training a dedicated MLP to learn embeddings allows for better capture of positive sample features, thus improving ROC\_AUC, although it still lags behind the best SMOTE-Augmented baseline model, RF. The KAN model, a nonlinear

architecture, emphasizes feature extraction through nonlinear transformations of nodes and adjusts the output through learnable activation functions of edges. In cases of class imbalance, KAN may capture finer-grained differences between positive and negative samples, resulting in higher ROC\_AUC on the test set.

This study has several limitations. First, due to data limitations, we used data from only seven hospitals in a southwestern city of China covering the past decade. We plan to expand our dataset to incorporate a wider variety of patient samples. Second, while considering the relationships between patients, medications, and diagnosis, we did not incorporate more heterogeneous data sources. But our approach is adaptable to incorporating more diverse data and knowledge sources. Future research could benefit from integrating information such as disease comorbidity or multimorbidity networks [25], [30]. Meanwhile, the KPN assigns equal importance to all knowledge types (e.g., drug, gene, ADR), which may introduce noise and distort their actual clinical relevance. Future work can incorporate domain-specific expertise to weight these knowledge types differentially, obtaining superior knowledge embeddings. To address the limitations of purely data-driven machine learning methods, researchers are exploring ways to enhance these models by incorporating prior knowledge. This has led to the emergence of approaches that combine data-driven and knowledge-driven learning, such as Informed Machine Learning (IML) [44]. The integration of knowledge and medicine is an increasingly prominent and rapidly growing field, known as Med IML [45]. In this study, we introduced external knowledge to enrich the patient knowledge network, combining both data-driven and knowledge-driven approaches to enhance prediction accuracy. This fusion approach represents a promising direction for future research.

## VI. CONCLUSION

In this study, we introduced a novel framework named KPNDS, which utilizes graph-based machine learning algorithms to extract latent patient representations from a medical knowledge enriched patient network. With MLP, Transformer and KAN, KPNDS captures patient subgroups characterized by latent similarity and dynamically selects the optimal subset of classifiers for each test sample to predict the risk of DILI in pediatric patients receiving antibiotics. Experimental results demonstrate that KPNDS outperforms 13 baseline models and 13 DES methods. This study highlights a novel and effective approach to integrating data and knowledge, demonstrating the potential of combining external knowledge sources with data-driven models to enhance predictive accuracy. It also emphasizes the significance and effectiveness of MCS, a crucial branch of ensemble learning. Furthermore, KPNDS holds practical potential for identifying high-risk populations who could benefit from timely preventive strategies, potentially improving healthcare outcomes.

## CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Linjun Huang: Methodology, Investigation, Formal analysis, Validation, Writing - Original Draft. Zixin Shi: Methodology, Investigation, Formal analysis. Fei Tang: Investigation, Validation. Haolin Wang: Conceptualization, Methodology, Writing - Review & Editing, Supervision, Funding acquisition.

## DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## DATA AND CODE AVAILABILITY

The source code for this study is publicly accessible at: <https://github.com/wanghaolin/KPNDS>. Owing to the highly sensitive nature of patient data, the patient data used in this study will not be made public.

## ETHICAL APPROVAL STATEMENT

The study was approved by the ethics committees of Chongqing Medical University (Reference Number: 2023096, Date Approved: Dec 20, 2023), and due to its retrospective nature, this study required no informed consent and represented minimal risk to participants.

## REFERENCES

- [1] J. Lazarou et al., "Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies," *Jama*, vol. 279, no. 15, pp. 1200–1205, 1998, doi:[10.1001/jama.279.15.1200](https://doi.org/10.1001/jama.279.15.1200).
- [2] Y. Chongpison et al., "Ifn- $\gamma$  elispot-enabled machine learning for culprit drug identification in nonimmediate drug hypersensitivity," *Journal of Allergy and Clinical Immunology*, vol. 153, no. 1, pp. 193–202, 2024, doi:[10.1016/j.jaci.2023.08.026](https://doi.org/10.1016/j.jaci.2023.08.026).
- [3] B. Y.-H. Thong and T.-C. Tan, "Epidemiology and risk factors for drug allergy," *British journal of clinical pharmacology*, vol. 71, no. 5, pp. 684–700, 2011, doi:[10.1111/j.1365-2125.2010.03774.x](https://doi.org/10.1111/j.1365-2125.2010.03774.x).
- [4] E. Napoleone, "Children and adrs (adverse drug reactions)," *Italian journal of pediatrics*, vol. 36, no. 4, pp. 1–5, 2010, doi:[10.1186/1824-7288-36-4](https://doi.org/10.1186/1824-7288-36-4).
- [5] M. M. Rumore, "Medication repurposing in pediatric patients: teaching old drugs new tricks," *The Journal of Pediatric Pharmacology and Therapeutics*, vol. 21, no. 1, pp. 36–53, 2016, doi:[10.5863/1551-6776-21.1.36](https://doi.org/10.5863/1551-6776-21.1.36).
- [6] A. A. Elzagallaai et al., "Adverse drug reactions in children: the double-edged sword of therapeutics," *Clinical Pharmacology & Therapeutics*, vol. 101, no. 6, pp. 725–735, 2017, doi:[10.1002/cpt.677](https://doi.org/10.1002/cpt.677).
- [7] J. Wang et al., "Landscape of dili-related adverse drug reaction in china mainland," *Acta Pharmaceutica Sinica B*, vol. 12, no. 12, pp. 4424–4431, 2022, doi:[10.1016/j.apsb.2022.04.019](https://doi.org/10.1016/j.apsb.2022.04.019).
- [8] T. Shen et al., "Incidence and etiology of drug-induced liver injury in mainland china," *Gastroenterology*, vol. 156, no. 8, pp. 2230–2241, 2019, doi:[10.1053/j.gastro.2019.02.002](https://doi.org/10.1053/j.gastro.2019.02.002).
- [9] J. Y. Roger et al., "Emerging causes of drug-induced anaphylaxis: a review of anaphylaxis-associated reports in the fda adverse event reporting system (faers)," *The Journal of Allergy and Clinical Immunology: In Practice*, vol. 9, no. 2, pp. 819–829, 2021, doi:[10.1016/j.jaip.2020.09.021](https://doi.org/10.1016/j.jaip.2020.09.021).
- [10] X. Ji et al., "Improved adverse drug event prediction through information component guided pharmacological network model (ic-pnm)," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 3, pp. 1113–1121, 2019, doi:[10.1109/TCBB.2019.2928305](https://doi.org/10.1109/TCBB.2019.2928305).
- [11] M. Liu et al., "Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs," *Journal of the American Medical Informatics Association*, vol. 19, no. e1, pp. e28–e35, 2012, doi:[10.1136/amiajnl-2011-000699](https://doi.org/10.1136/amiajnl-2011-000699).
- [12] M. Moreno-Torres et al., "Novel clinical phenotypes, drug categorization, and outcome prediction in drug-induced cholestasis: Analysis of a database of 432 patients developed by literature review and machine learning support," *Biomedicine & Pharmacotherapy*, vol. 174, p. 116530, 2024, doi:[10.1016/j.biopha.2024.116530](https://doi.org/10.1016/j.biopha.2024.116530).
- [13] S.-H. Huang et al., "How platinum-induced nephrotoxicity occurs? machine learning prediction in non-small cell lung cancer patients," *Computer Methods and Programs in Biomedicine*, vol. 221, p. 106839, 2022, doi:[10.1016/j.cmpb.2022.106839](https://doi.org/10.1016/j.cmpb.2022.106839).
- [14] A. Mohsen et al., "Deep learning prediction of adverse drug reactions in drug discovery using open tg-gates and faers databases," *Frontiers in Drug Discovery*, vol. 1, p. 768792, 2021, doi:[10.3389/fddsv.2021.768792](https://doi.org/10.3389/fddsv.2021.768792).
- [15] H. Niu et al., "Prior drug allergies are associated with worse outcome in patients with idiosyncratic drug-induced liver injury: a machine learning approach for risk stratification," *Pharmacological Research*, vol. 199, p. 107030, 2024, doi:[10.1016/j.phrs.2023.107030](https://doi.org/10.1016/j.phrs.2023.107030).
- [16] M. Woźniak et al., "A survey of multiple classifier systems as hybrid systems," *Information Fusion*, vol. 16, pp. 3–17, 2014, doi:[10.1016/j.inffus.2013.04.006](https://doi.org/10.1016/j.inffus.2013.04.006).
- [17] Z. Liu et al., "Kan: Kolmogorov-arnold networks," *arXiv preprint arXiv:2404.19756*, 2024, doi:[10.48550/arXiv.2404.19756](https://doi.org/10.48550/arXiv.2404.19756).
- [18] J. Hu et al., "Identification and validation of an explainable prediction model of acute kidney injury with prognostic implications in critically ill children: a prospective multicenter cohort study," *EClinicalMedicine*, vol. 68, 2024, doi:[10.1016/j.eclinm.2023.102409](https://doi.org/10.1016/j.eclinm.2023.102409).
- [19] A. M. Chiriac et al., "Designing predictive models for beta-lactam allergy using the drug allergy and hypersensitivity database," *The Journal of Allergy and Clinical Immunology: In Practice*, vol. 6, no. 1, pp. 139–148, 2018, doi:[10.1016/j.jaip.2017.04.045](https://doi.org/10.1016/j.jaip.2017.04.045).
- [20] H. Qiu et al., "Comorbidity patterns in depression: A disease network analysis using regional hospital discharge records," *Journal of affective disorders*, vol. 296, pp. 418–427, 2022, doi:[10.1016/j.jad.2021.09.100](https://doi.org/10.1016/j.jad.2021.09.100).
- [21] F. Saberi-Movahed et al., "Nonnegative matrix factorization in dimensionality reduction: A survey," *arXiv preprint arXiv:2405.03615*, 2024, doi:[10.48550/arXiv.2405.03615](https://doi.org/10.48550/arXiv.2405.03615).
- [22] K. Berahmand et al., "A comprehensive survey on spectral clustering with graph structure learning," *arXiv preprint arXiv:2501.13597*, 2025, doi:[10.48550/arXiv.2501.13597](https://doi.org/10.48550/arXiv.2501.13597).
- [23] L. Wang et al., "Age-and sex-specific differences in multimorbidity patterns and temporal trends on assessing hospital discharge records in southwest china: network-based study," *Journal of medical Internet research*, vol. 24, no. 2, p. e27146, 2022, doi:[10.2196/27146](https://doi.org/10.2196/27146).
- [24] T. Wang et al., "Patient-centric characterization of multimorbidity trajectories in patients with severe mental ill-

- nesses: A temporal bipartite network modeling approach," *Journal of Biomedical Informatics*, vol. 127, p. 104010, 2022, doi:[10.1016/j.jbi.2022.104010](https://doi.org/10.1016/j.jbi.2022.104010).
- [25] P. Yang et al., "Early prediction of high-cost inpatients with ischemic heart disease using network analytics and machine learning," *Expert Systems with Applications*, vol. 210, p. 118541, 2022, doi:[10.1016/j.eswa.2022.118541](https://doi.org/10.1016/j.eswa.2022.118541).
- [26] S. a. o. Uddin, "Comorbidity and multimorbidity prediction of major chronic diseases using machine learning and network analytics," *Expert Systems with Applications*, vol. 205, p. 117761, 2022, doi:[10.1016/j.eswa.2022.117761](https://doi.org/10.1016/j.eswa.2022.117761).
- [27] M. E. Hossain et al., "Network analytics and machine learning for predictive risk modelling of cardiovascular disease in patients with type 2 diabetes," *Expert Systems with Applications*, vol. 164, p. 113918, 2021, doi:[10.1016/j.eswa.2020.113918](https://doi.org/10.1016/j.eswa.2020.113918).
- [28] Z. Hu et al., "Network analytics and machine learning for predicting length of stay in elderly patients with chronic diseases at point of admission," *BMC Medical Informatics and Decision Making*, vol. 22, no. 1, p. 62, 2022, doi:[10.1186/s12911-022-01802-z](https://doi.org/10.1186/s12911-022-01802-z).
- [29] J. Yang et al., "Graphsynergy: a network-inspired deep learning model for anticancer drug combination prediction," *J Am Med Inform Assoc*, vol. 28, no. 11, pp. 2336–2345, 2021, doi:[10.1093/jamia/ocab162](https://doi.org/10.1093/jamia/ocab162).
- [30] H. Qiu et al., "Dcnet: A disease comorbidity network-based temporal deep learning framework to predict cardiovascular risk in patients with mental disorders," *Expert Systems with Applications*, vol. 254, p. 124312, 2024, doi:[10.1016/j.eswa.2024.124312](https://doi.org/10.1016/j.eswa.2024.124312).
- [31] A. S. Britto Jr et al., "Dynamic selection of classifiers—a comprehensive review," *Pattern recognition*, vol. 47, no. 11, pp. 3665–3680, 2014, doi:[10.1016/j.patcog.2014.05.003](https://doi.org/10.1016/j.patcog.2014.05.003).
- [32] M. Monteiro Jr et al., "Exploring diversity in data complexity and classifier decision spaces for pool generation," *Information Fusion*, vol. 89, pp. 567–587, 2023, doi:[10.1016/j.inffus.2022.09.001](https://doi.org/10.1016/j.inffus.2022.09.001).
- [33] H. Jalalian and R. M. Cruz, "Mlrs-pds: A meta-learning recommendation of dynamic ensemble selection pipelines," in *2024 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2024, pp. 1–9, doi:[10.1109/IJCNN60899.2024.10651199](https://doi.org/10.1109/IJCNN60899.2024.10651199).
- [34] M. A. Souza et al., "Online local pool generation for dynamic classifier selection," *Pattern Recognition*, vol. 85, pp. 132–148, 2019, doi:[10.1016/j.patcog.2018.08.004](https://doi.org/10.1016/j.patcog.2018.08.004).
- [35] R. M. Cruz et al., "Meta-des: A dynamic ensemble selection framework using meta-learning," *Pattern recognition*, vol. 48, no. 5, pp. 1925–1935, 2015, doi:[10.1016/j.patcog.2014.12.003](https://doi.org/10.1016/j.patcog.2014.12.003).
- [36] R. Davtalab et al., "A scalable dynamic ensemble selection using fuzzy hyperboxes," *Information Fusion*, vol. 102, p. 102036, 2024, doi:[10.1016/j.inffus.2023.102036](https://doi.org/10.1016/j.inffus.2023.102036).
- [37] M. A. Souza et al., "A dynamic multiple classifier system using graph neural network for high dimensional overlapped data," *Information Fusion*, vol. 103, p. 102145, 2024, doi:[10.1016/j.inffus.2023.102145](https://doi.org/10.1016/j.inffus.2023.102145).
- [38] C. Xu et al., "Data-driven decision model based on dynamical classifier selection," *Knowledge-Based Systems*, vol. 212, p. 106590, 2021, doi:[10.1016/j.knosys.2020.106590](https://doi.org/10.1016/j.knosys.2020.106590).
- [39] S. Ö. Ariki and T. Pfister, "Tabnet: Attentive interpretable tabular learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, no. 8, 2021, pp. 6679–6687, doi:[10.1609/aaai.v35i8.16826](https://doi.org/10.1609/aaai.v35i8.16826).
- [40] H. Nori, S. Jenkins, P. Koch, and R. Caruana, "Interpretml: A unified framework for machine learning interpretability," *arXiv preprint arXiv:1909.09223*, 2019, doi:[10.48550/arXiv.1909.09223](https://doi.org/10.48550/arXiv.1909.09223).
- [41] R. M. Cruz et al., "Deslib: A dynamic ensemble selection library in python," *Journal of Machine Learning Research*, vol. 21, no. 8, pp. 1–5, 2020, doi:[10.48550/arXiv.1802.04967](https://doi.org/10.48550/arXiv.1802.04967).
- [42] G. L. Kearns et al., "Developmental pharmacology—drug disposition, action, and therapy in infants and children," *New England Journal of Medicine*, vol. 349, no. 12, pp. 1157–1167, 2003, doi:[10.1056/NEJMra035092](https://doi.org/10.1056/NEJMra035092).
- [43] Z. Zhang et al., "Electronic healthcare records and external outcome data for hospitalized patients with heart failure," *Scientific data*, vol. 8, no. 1, p. 46, 2021, doi:[10.1038/s41597-021-00835-9](https://doi.org/10.1038/s41597-021-00835-9).
- [44] L. Von Rueden et al., "Informed machine learning—a taxonomy and survey of integrating prior knowledge into learning systems," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 1, pp. 614–633, 2021, doi:[10.1109/TKDE.2021.3079836](https://doi.org/10.1109/TKDE.2021.3079836).
- [45] F. Leiser et al., "Medical informed machine learning: A scoping review and future research directions," *Artificial Intelligence in Medicine*, vol. 145, p. 102676, 2023, doi:[10.1016/j.artmed.2023.102676](https://doi.org/10.1016/j.artmed.2023.102676).