





# CardiOT: Towards Interpretable Drug Cardiotoxicity Prediction Using Optimal Transport and Kolmogorov–Arnold Networks

Xinyu Zhang, Hao Wang, Zhenya Du, Linlin Zhuo , Xiangzheng Fu , Dongsheng Cao, Boqia Xie , and Keqin Li , *Fellow, IEEE*

**Abstract**—Investigating the inhibitory effects of compounds on cardiac ion channels is essential for assessing cardiac drug safety. Consequently, researchers have developed computational models to evaluate combined cardiotoxicity (CCT) on cardiac ion channels. However, limitations in experimental data often cause issues like uneven data distribution and scarcity. Additionally, existing models primarily emphasize atomic information flow within graph neural networks (GNNs) while overlooking chemical bonds, leading to inadequate recognition of key structures. Therefore, this study integrates optimal transport (OT), structure remapping (SR), and Kolmogorov–Arnold networks (KANs) into a GNN-based CCT prediction model, CardiOT. First, the proposed CardiOT model employs OT pooling to optimize sample-feature joint distribution using expectation maximization, identifying “important” sample-feature pairs. Additionally, SR technology is used to emphasize the role of chemical bond information in message propagation. KAN technology is integrated to greatly enhance model interpretability. In summary, the model mitigates challenges related to uneven data distribution and scarcity. Multiple experiments on public datasets confirm the model’s robust performance. We anticipate that this model will provide deeper insights into compound inhibition mechanisms on cardiac ion channels and reduce toxicity risks.

Received 25 September 2024; revised 15 November 2024; accepted 27 November 2024. Date of publication 2 December 2024; date of current version 7 March 2025. (Xinyu Zhang, Hao Wang, and Zhenya Du contributed equally to this work.) (Corresponding authors: Linlin Zhuo; Xiangzheng Fu; Boqia Xie.)

Xinyu Zhang and Linlin Zhuo are with the School of Data Science and Artificial Intelligence, Wenzhou University of Technology, Wenzhou 325027, China (e-mail: 20210339@wzut.edu.cn).

Hao Wang is with the Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China.

Zhenya Du is with the Guangzhou Xinhua University, Guangzhou 510520, China.

Xiangzheng Fu is with the College of Information Science and Engineering, Hunan University, Changsha 410082, China (e-mail: fxz326@hnu.edu.cn).

Dongsheng Cao is with the Xiangya School of Pharmaceutical Sciences, Central South University, Changsha 410003, China.

Boqia Xie is with the Department of Cardiology, Cardiovascular Imaging Center, Beijing Chaoyang Hospital, Capital Medical University, Beijing 100020, China (e-mail: dr.boqiaxie@hotmail.com).

Keqin Li is with the Department of Computer Science, State University of New York, New York, NY 12561 USA.

Our code and data are accessible at: <https://github.com/2014402680/CCT>.

Digital Object Identifier 10.1109/JBHI.2024.3510297

**Index Terms**—Cardiac drug safety, compound cardiotoxicity (CCT), inhibitory effects of drugs, Kolmogorov–Arnold networks (KANs), optimal transport (OT), uneven distribution and data scarcity.

## I. INTRODUCTION

CURRENTLY, numerous drugs have been scientifically confirmed to adversely affect cardiac function, including arrhythmias [1], weakened myocardial function [2], and direct myocardial cell damage [3]. Blockade of hERG ion channels often leads to QT interval prolongation, resulting in cardiotoxicity [4]. Additionally, interference with other ion channels, such as Nav1.5 and Cav1.2, is also considered a significant cause of cardiotoxicity [5]. Therefore, studying the effects of compounds on specific ion channels is crucial for mitigating potential cardiac risks. Moreover, in the multidisciplinary field of drug development, drug safety and efficacy remain the primary considerations [6]. A thorough understanding of drug absorption, distribution, metabolism, and excretion (ADME), along with toxicological (T) properties, is critical for safety assessments, clinical trials, and eventual market access. Predicting compound cardiotoxicity is particularly critical during drug development. Compound cardiotoxicity (CCT) specifically refers to the inhibition of hERG, Cav1.2, and Nav1.5 ion channels and the potential prolongation of the QT interval. The complexity of CCT prediction has led to significant losses in the pharmaceutical industry, as failures to detect it in early, preclinical, or clinical stages [7], [8] have resulted in drug withdrawals and halted development projects [9]. Therefore, enhancing the accuracy of CCT prediction is essential to accelerate drug development, lower costs, and ensure patient safety. Accurate predictions can identify potential cardiotoxicity during early screening, prevent resource waste and late-stage failure, shorten clinical trial durations, and mitigate safety risks. Furthermore, it can reduce drug recalls and market withdrawals due to cardiotoxicity, optimize the R&D process, and further ensure patient safety.

Traditionally, following ICH S7B guidelines by the International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH) [10], CCT prediction involves two main methods: in vivo and in vitro. In vivo evaluation directly measures the QT interval in animal models and analyzes electrocardiograms to assess drug effects

on cardiac electrophysiological activity [11]. The other method evaluates the effects of compounds on Kv11.1 voltage-gated potassium channels through in vitro delayed rectifier potassium current (IKr) assays on cell lines expressing hERG [12]. While biological experiments provide reliable data on compound cardiotoxicity, they inevitably encounter efficiency bottlenecks, high costs, and ethical concerns. Fortunately, advancements in computing technology have opened new opportunities for early cardiotoxicity identification in compounds. Computer-aided drug discovery (CADD) [13] is an effective approach to reducing costs and accelerating the development of promising drug candidates. Toxicity prediction algorithms have become integral to modern CADD, with CCT prediction algorithms being particularly essential. These algorithms employ advanced data analysis and modeling techniques to accurately identify potential CCT risks early in drug development, supporting the screening and optimization of drug candidates.

In the early stages of computer-aided drug discovery, researchers enthusiastically embraced machine learning (ML) techniques, aiming to develop models capable of accurately predicting cardiac ion channel inhibition. During the early 21st century, statistical methods like Naïve Bayes, Gaussian processes, expectation maximization, and partial least squares (PLS) were widely used in ML-driven toxicity prediction, with their applications extensively discussed and validated [14]. However, most prediction models at the time were constrained by small datasets, with fewer than a thousand compounds used for training. As more advanced algorithms, such as random forests (RF), support vector machines (SVM), and deep neural networks (DNN), emerged, they became mainstream due to their superior predictive performance and generalization ability [15], [16]. Researchers also started utilizing larger datasets for model training, with some studies involving over 15,000 compounds. This expansion significantly enhanced model accuracy and reliability. This surge in data volume reflects not only the rapid expansion of medicinal chemistry databases but also the distinct advantages of ML technology in handling large-scale complex data.

The rapid advancement of computational biopharmaceutical technology has sparked unprecedented innovation in drug toxicity prediction, with statistical learning and early machine learning models widely adopted [17], [18], [19], [20], [21], [22], [23]. In 2016, Wang et al. pioneered a study that combined Naïve Bayes (NB) with support vector machine (SVM) algorithms and integrated multi-pharmacophore features using ensemble learning to accurately predict CCT [24]. The study demonstrated ensemble learning's potential to enhance model performance and opened new avenues for future research. Cai et al. later introduced the DeepERG model, bringing deep learning to CCT prediction. The model used a multi-task deep neural network to integrate operating environment and mol2vec descriptors, generating final molecular representations and significantly improving HERG-related CCT prediction accuracy [25]. In 2020, Ryu et al. expanded the scope of deep learning in CCT prediction with the introduction of the DeepHIT framework. This framework adapts to various input data types, including molecular descriptors, fingerprints, and graph structures, enabling efficient

identification of hERG blockers via customized deep learning models [26]. However, these studies primarily focus on specific ion channels (e.g., hERG), and the effects of compounds on other cardiac ion channels require further investigation. In 2023, Arab et al. addressed this gap with their research. They conducted multi-group tests to evaluate the inhibitory effects of compounds on multiple cardiac ion channels (e.g., hERG, Nav1.5, Cav1.2), offering a broader perspective for drug safety evaluation [27]. In 2024, Wang et al. introduced a novel solution leveraging progressive multimodal fusion technology. By integrating feature information from various modalities, they enhanced molecular representation while facilitating deep analysis of correlations and complementarities between modalities, further improving the model's robustness and accuracy in cardiotoxicity prediction [28].

Accurately predicting CCT remains a significant challenge in drug development for both the chemical and medical communities. Experimental constraints result in limited and unevenly distributed data. Additionally, current CCT prediction models focus on atomic information flow within propagation rules and overlook chemical bond information. Consequently, we introduce the CardiOT model, based on GNN and integrating OT, SR, and KAN technologies, to enhance CCT prediction performance. Initially, we transform "bonds" in the molecular graph into new nodes and "bond-atom-bond" configurations into new links, creating an updated molecular graph. Next, we apply OT technology to ascertain the joint distribution of sample-feature dimensions. Lastly, the KAN classifier is employed to predict the CCT score. The key contributions of this study are summarized as follows:

- 1) We have integrated OT, SR, and KAN techniques into GNN to develop a CCT prediction model that yields promising results.
- 2) We are the pioneers in employing OT technology for CCT prediction. By optimizing the joint distribution of samples and features to identify key sample-feature pairs, we thus mitigate issues of sample scarcity and uneven distribution.
- 3) We introduce SR techniques to convert atom-centric molecular graphs into bond-centric graphs, highlighting the critical role of chemical bonds in GNN propagation.
- 4) We introduce the KAN technique for CCT prediction, demonstrating superior performance and enhanced interpretability compared to MLP.

## II. MATERIALS AND METHODS

### A. Data Preparation

This study integrates data on hERG, Cav1.2, and Nav1.5 from multiple sources including ChEMBL Bioactivity Database [29], PubChem [30], BindingDB [31], hERGCentral [32], numerous U.S. patents, and scientific literature [33]. Activity data were categorized into IC50 type and inhibition type based on their measurement methods. The IC50 type includes quantitative indicators like IC50, EC50, ED50, Ki, and Kd, reflecting inhibitory potency, whereas the inhibition type reports the percentage inhibition rate at a specific concentration. To ensure data quality, we conducted a cross-validation with the hERGCentral database

TABLE I  
DATA STATISTICS

Datasets	hERG		Nav1.5		Cav1.2	
	CCT	Non-CCT	CCT	Non-CCT	CCT	Non-CCT
Property						
Dev. Set	13428	8818	1423	673	530	272
Eval	125	125	48	16	41	21

to correct potential errors in hERG activity values. We then used RDKit (<http://www.rdkit.org>) and MolVSMoIVS (<https://github.com/mcs07/MoIVS>) toolkits to standardize chemical structures rigorously, including selecting major fragments, removing explicit hydrogen, adjusting ionization states, and calculating stereochemical information. The processed structures were converted into SMILES strings and subsequently into InChI keys to remove duplicates and ensure unique, accurate compound identification. Detailed information about the dataset is presented in Table I.

To facilitate subsequent analysis, we converted all activity data to nanomolar (nM) units, averaged the values of duplicate compounds, and applied a 95% quantile filtering strategy to exclude potential extreme outliers. Subsequently, we calculated the pIC50 values. Activity data standardization was aligned with conventional standards in drug toxicity prediction. Under these standards, compounds with IC50 values  $\leq 10 \mu\text{M}$  (pIC50  $\geq 5$ ) were classified as CCT, while others were designated as non-CCT. We adhered to a 4:1 ratio principle, splitting the hERG, Nav1.5, and Cav1.2 datasets into training and validation sets. Additionally, we selected external test sets with structural similarity below 60% from sources like ChEMBL and PubChem, to further assess the model’s predictive performance and practical utility.

## B. Model Overview

This study introduces a CCT prediction model called CardiOT, which utilizes GNN architecture to integrate OT, SR, and KAN technologies, as shown in Fig. 1. Module (A) focuses on collecting the atomic and chemical bond features of compounds. Module (B) remaps the molecular graph bonds to new nodes, integrates adjacent atom information, forms new “bond-atom-bond” links, and constructs an updated molecular graph. Module (C) describes the architecture of KAN. Module (D) applies expectation maximization to optimize the relationship between sample indicators and feature dimensions, identifying key sample-feature pairs. Module (E) utilizes a GIN encoder and OT pooling layer from (D) to extract molecular representations. Module (F) employs KAN technology to extract the final molecular representation. Module (G) employs contrastive learning to train both the original and remapped molecular graph representations.

## C. Structure Remapping

The traditional GNN model conducts aggregation and updates by transmitting atomic information within the message propagation mechanism, which often results in the loss of chemical

bond information, treating bonds merely as conduits for message transfer. In their study on compound ADT prediction, Ma et al. redefined “bonds” in the molecular graph as new nodes and “bond-atom-bond” relationships as new links, creating an updated molecular graph [34]. Inspired by Ma et al., this study similarly incorporates chemical bond information into the message propagation mechanism of our model.

Specifically, the compound’s SMILES sequence is inputted and converted into a directed graph  $G = (X, E, C)$ .  $X$  denotes all atomic vectors in  $G$ , each encompassing details like atomic symbol, bond count, formal charge, bonding hydrogens, hybridization state, aromaticity, and atomic mass. The  $i$ -th atomic vector,  $X_i$ , belongs to  $X$ .  $E$  comprises all chemical bond vectors, with  $E_{i,j}$  in  $E$  representing the bond vector from atom  $i$  to  $j$ . Each bond vector details bond type, conjugation, and ring presence.  $A$  denotes the adjacency matrix of molecular graph  $G$ , where  $A_{i,j}$  in  $A$  indicates the presence of a chemical bond between atoms  $i$  to  $j$ .

Perform the SR operation on molecular graph  $G$  to generate a new molecular graph  $G^s = (X^s, E^s, A^s)$ , where chemical bonds are mapped to new nodes, incorporating information from directly connected atoms:

$$X_{i,j}^s = X_i \| E_{i,j} \| X_j; X_i, X_j \in X, \text{ and } E_{i,j} \in E, \quad (1)$$

where  $X_{i,j}^s$  denotes the feature vector of the new node,  $\|$  signifies the concatenation operation, and  $X_i$  and  $X_j$  are the feature vectors of atoms directly linked by bond  $E_{i,j}$ . The “bond-atom-bond” configuration is mapped to a new link:

$$E_{i,k,k,j}^s = E_{i,k} \| X_k \| E_{k,j}; E_{i,k}, E_{k,j} \in E \text{ and } X_k \in X, \quad (2)$$

where  $E^s$  represents the set of edge vectors, and the adjacency matrix  $A^s$  of graph  $G^s$  is defined as follows:

$$A_{i,k,k,j}^s = 1, A_{i,k}, A_{k,j} \in A, \quad (3)$$

Following these steps, the original molecular graph  $G$  and its structurally remapped counterpart  $G^s$  are constructed using the compound’s SMILES sequence.

## D. Molecular Graph Encoder

GNN technology, particularly the Graph Isomorphism Network (GIN) [35], is crucial for predicting compound properties as it elucidates compound structures and highlights their heterogeneous information. Consequently, this study employs GIN to extract representations of molecular graphs and their mappings. For node  $i$  in both the molecular graph and its mapping, GIN aggregates and updates information within the node’s local

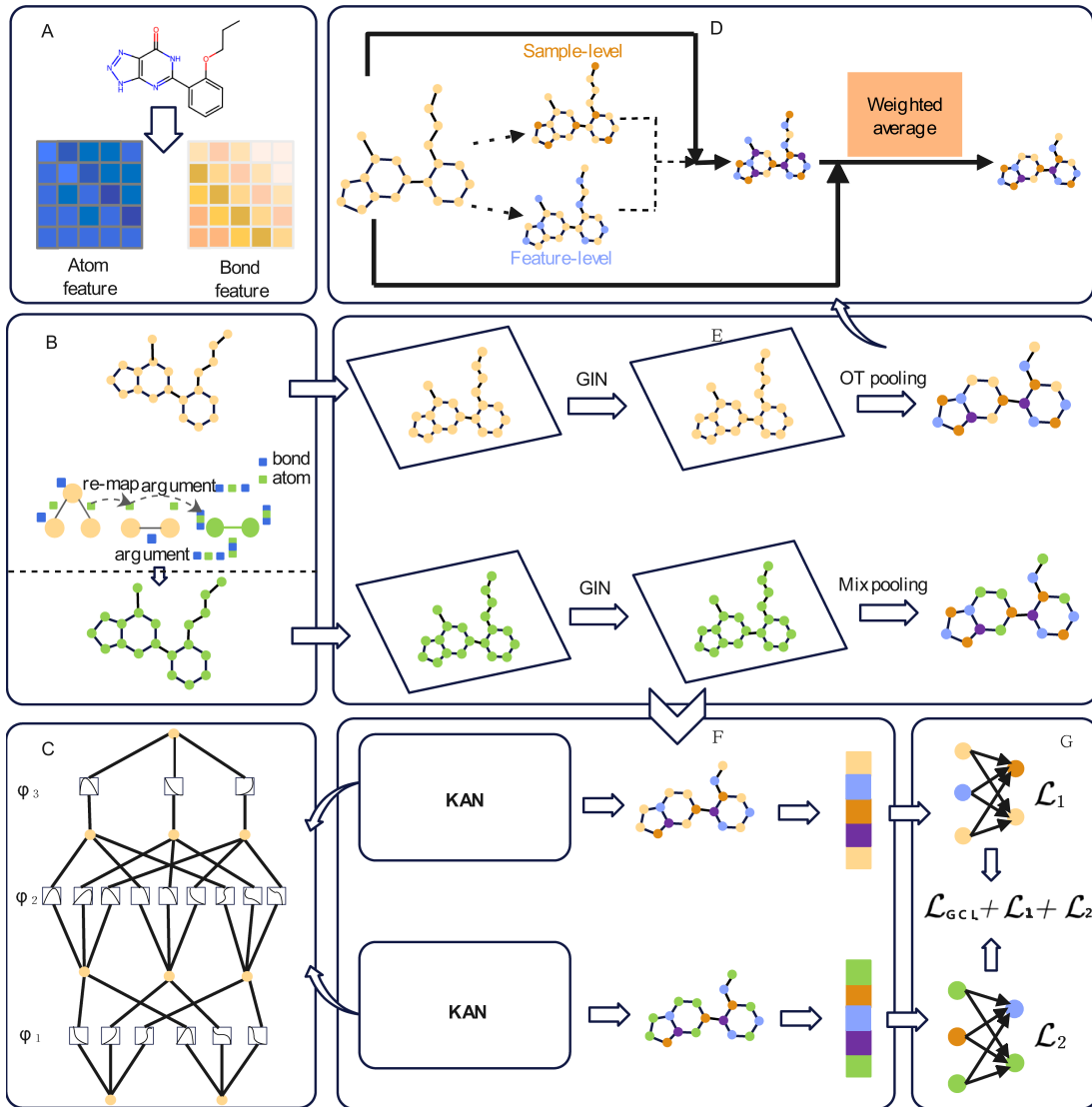


Fig. 1. The architecture of the CardiOT model comprises four main modules: (A) data preparation, (B) structure remapping, (C) KAN, (D) OT pooling layer, (E) GIN encoder, (F) graph contrastive learning, and (G) Optimization.

neighborhood:

$$h_i^t = MLP \left( (1 + \gamma^t) \cdot h_i^{t-1} + \sum_{j \in N(i)} h_j^{t-1} \right), \quad (4)$$

where,  $h_i^t$  is the embedding of node  $i$  in GIN's  $t$ -th layer,  $\gamma$  denotes the weight parameter, and  $N(i)$  is the set of node  $i$ 's neighboring nodes.  $h_i^t$  is subsequently input into the optimal transmission layer.

### E. Optimal Transfer-Based Pooling

Global pooling methods, including average pooling, maximum pooling [36], and attention pooling [37], are widely utilized across various domains like image classification, where they excel. However, these methods depend on high data density to either highlight crucial features or smooth out features,

mitigating the effects of noisy data. In CCT prediction tasks, available data is often sparse and typically suffers from uneven distribution. Inspired by the previous work [38], we employ OT technology in the pooling layer to address these issues. Following the expectation maximization principle, we optimize the joint distribution of samples and feature dimensions to identify key sample-feature pairs, enhancing the reliability of feature dimensionality reduction. This enhancement boosts the model's generalization capabilities in handling scarce data.

1) *Optimal Transfer and Pooling Theory*: In the sample feature matrix  $X$ ,  $x_{kl}$  denotes the  $k$ -th feature of the  $l$ -th sample. In statistical signal processing, a signal corresponds to a pair of sample and feature dimensions. Given all samples and feature dimensions, we define the joint distribution  $P = [p_{kl}] \in [0, 1]^{K \times L}$ , element  $p_{kl}$  to represent the significance of the signal  $x_{kl}$ . Typical global pooling operations are unified into an interpretable algorithm framework based on the expectation

maximization principle:

$$f(X) = (X \odot \text{diag}^{-1}(P1_K)P), 1_K = \prod_{k=1}^K E_{l \sim p_{l|k}}[x_{kl}] \quad (5)$$

where  $\hat{P} = [p_{l|k}] = \text{diag}^{-1}(P1_K)P$ . In the above equation, pooling operations are determined by the joint distribution  $P$ , utilizing the OT technique to maximize the expectation of this distribution [34]:

$$\begin{aligned} P^* &= \operatorname{argmax}_{P \in \Omega(u,v)} \sum_{k=1}^K u_k E_{l \sim p_{l|k}}[x_{kl}] \\ &= \operatorname{argmax}_{P \in \Omega(u,v)} E_{(k,l) \sim P}[x_{kl}], \end{aligned} \quad (6)$$

where  $E_{(k,l) \sim P}[x_{kl}]$  is represented by the vector inner product  $\langle X, P \rangle$ , where  $u = [p_k] \in \Delta^{K-1}$  and  $v \in \Delta^{L-1}$  are predefined distributions for feature dimensions and sample indices, respectively. Consequently, the marginal distribution of  $P$  is constrained to  $u$  and  $v$ , thus defining  $P \in \Omega(u, v) = \{P \geq 0 | P1_L = u, P^T 1_K = v\}$ .

However, this process encounters challenges such as significant sparsity, difficulties in obtaining accurate marginal distributions, and overlooking the structural relationships between samples and features. Therefore, we extend the above equation to address the regularized optimal transport (ROT) problem:

$$\begin{aligned} P_{OT}^*(X; \Phi) &= \operatorname{argmin}_{P \in \Omega} \langle -X, P \rangle + \tau_1 \langle C(X, P), P \rangle \\ &\quad + \tau_2 R(P) + \tau_3 KL(P1_L | u_0) + \tau_4 KL(P^T 1_L | v_0), \end{aligned} \quad (7)$$

where  $\langle \cdot, \cdot \rangle$  represents the vector inner product,  $KL(\cdot)$  denotes the KL divergence, and  $\Omega = \{P > 0 | 1_K^T P 1_L = 1\}$ . The first term is the optimal transport (OT) term, the second term is the structural regularization term, and together, they form the fused Gromov-Wasserstein Discrepancy. The third term is the smoothing regularization term, while the fourth and fifth terms are marginal regularization terms. Structural regularization utilizing the Gromov-Wasserstein difference enhances the correlation between feature and sample covariances, preserving their structural relationship. Consequently, we construct the following structural cost function:

$$C(X, P) = -Z_1 P Z_2^T, \quad (8)$$

where  $X$  denotes the embedding matrix of the compound extracted by the GIN encoder.  $Z_1$  denotes the feature-level covariance matrix  $\frac{1}{N}(X - \mu_1 1_N^T)(X - \mu_1 1_N^T)^T$  derived from  $X$ , while  $Z_2$  represents the sample-level covariance matrix  $\frac{1}{D}(X - 1_D \mu_2^T)(X - 1_D \mu_2^T)^T$ . And  $\mu_1 = \frac{1}{N} X 1_N$ ,  $\mu_2 = \frac{1}{D} X^T 1_D$ , where  $N$  and  $D$  represent the number of samples and feature dimensions respectively.  $P$  denotes the pairing method between the source and target sample spaces, aiming to minimize structural mismatch. Equation (8) measures the structural relationship between features and samples, enhancing the pooling effect.

**2) OT Pooling Using Sinkhorn:** In this study,  $X$  is defined as the embedding matrix of the compound, extracted using the GIN encoder. The objective is to employ OT technology to optimize the joint distribution of samples (nodes) and embedding

dimensions, thus facilitating the pooling of the molecular graph. Specifically, the near point method [39] is applied iteratively to solve (7), based on the compound's joint distribution in the current feedforward module:

$$\begin{aligned} P^{t+1} &= \operatorname{argmin}_{P \in \Omega} \langle -X, P \rangle + \tau_1 \langle C(X, P^t), P \rangle + \tau_2 R(P) \\ &\quad + \tau_3 KL(P1_L | u_0) + \tau_4 KL(P^T 1_L | v_0) + \kappa KL(P | P^t), \end{aligned} \quad (9)$$

where the weight  $\kappa$  dictates the significance of its proximal term. The equation  $R(P) = \langle P, \log P - 1 \rangle$  leads to the entropy imbalance optimal transfer (EUOT) problem:

$$\begin{aligned} \min_{P \in \Omega} &\langle C^t, P \rangle + (\tau_1 + \kappa) \langle \log P, P \rangle \\ &+ \tau_3 KL(P1_L | u_0) + \tau_4 KL(P^T 1_L | v_0). \end{aligned} \quad (10)$$

At this stage,  $C^t$  is defined by the input data and the current variable  $P^t$ . As per [40], it transitions into the Fenchel dual form of the EUOT problem, suitable for resolution via the sinkhorn-scaling algorithm [41].

Specifically, this study extends to the ROT framework, implemented with the sinkhorn algorithm. The choice of the sinkhorn algorithm is primarily due to its advantages in stability and efficiency. First, the sinkhorn algorithm stabilizes the optimization process by introducing a regularization term, enabling effective handling of larger datasets. Compared to other optimal transport methods, the sinkhorn algorithm achieves superior convergence and computational efficiency. Secondly, the sinkhorn algorithm preserves marginal distribution balance, which is critical for processing complex distributions. Additionally, the interpretability of the sinkhorn algorithm enhances the transparency of the model's decision-making process. Furthermore, the algorithm's wide application in various machine learning tasks provides extensive empirical support for its effectiveness and practicality. Finally, subsequent experimental results confirm the appropriateness of this strategy.

Besides the sinkhorn-scaling algorithm, alternatives like Bregman ADMM [42] and other proximal algorithms can also be used to solve the EOUT problem. However, the sinkhorn-scaling algorithm demonstrates clear advantages in computational efficiency and accuracy, as confirmed by subsequent experimental results and analysis.

**3) Running Rules:** As depicted in Fig. 2,  $X$  denotes the feature matrix produced by GIN. Initially, similarities at both the feature and sample levels are computed from the feature matrix, leading to the extraction of specific samples and features based on these similarities. Subsequently, the ROT method is utilized to optimize the joint distribution, identifying key "sample-feature" pairs that significantly contribute to the data representation. Following this, the significant "sample-feature" pairs within the optimized joint distribution are weighted against the feature matrix to achieve the pooling result.

## F. Kolmogorov-Arnold Networks (KANs)

Multilayer perceptrons (MLPs) are extensively utilized across deep learning fields due to their ability to approximate non-linear functions. However, MLPs encounter challenges such as

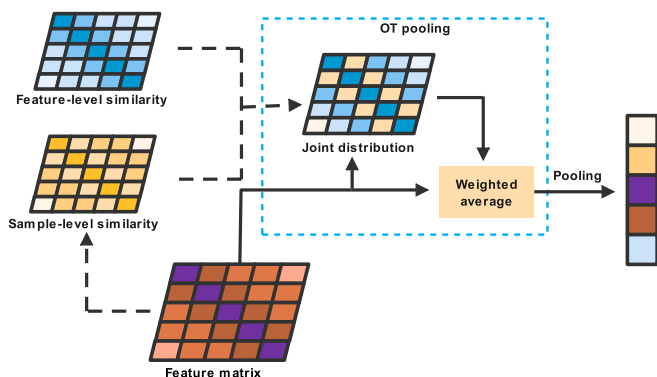


Fig. 2. The general principles of OT pooling operation.

a large parameter count and limited flexibility. Consequently, KAN technology was developed to address these limitations. The fundamental principle of Kolmogorov-Arnold theory posits that any multivariate continuous function can be expressed as a combination of continuous one-dimensional functions:

$$f(x_1, x_2, \dots, x_n) = \sum_{m=1}^{2n+1} \varphi_m \left( \sum_{c=1}^n \phi_{m,c}(x_c) \right). \quad (11)$$

This theorem guarantees that any multivariate continuous function can be represented as the sum of continuous one-dimensional functions. This structure allows each weight parameter in KAN to be represented by a single-variable function. These functions are typically parameterized as spline functions, offering high flexibility to model complex functions with fewer parameters, thereby enhancing model interpretability. Consequently, KAN demonstrates potential for efficient approximation of complex functions. Here,  $\phi_{m,c}$  represents a univariate function mapping each input variable ( $x_c$ ) following  $\phi_{m,c} : [0, 1] \rightarrow R$  and  $\varphi_m : R \rightarrow R$ . Given that all target functions are univariate, each can be parameterized as a  $b$ -spline curve with coefficients from learnable local  $b$ -spline basis functions. By integrating MLP architecture, a deeper KAN is achieved by combining  $L$  layers:

$$KAN(x) = (\varphi_{L-1} \circ \varphi_{L-2} \dots \varphi_1 \circ \varphi_0)x, \quad (12)$$

Consequently, KAN inherits the MLP framework, enabling deep feature extraction due to its structural similarity to MLP. By flexibly adjusting KAN's number of layers or the spline curve's dimensions, it achieves greater accuracy and adaptability. Consequently, this allows for effective control of the model's parameters and complexity, thereby reducing the risk of overfitting.

### G. Graph Contrastive Learning

Graph contrastive learning is an unsupervised method for learning representations of graph data, effectively enhancing node representations by mining structural similarities within graphs. In this study, the original molecular graph was constructed from SMILES sequences, and a structural remapping graph was developed using SR technology. Contrastive training was performed by identifying similarities between the two graph

types and differences among various compounds. Assuming a total of  $N$  compounds in the training set, for compound  $v$ ,  $h_v$  and  $h_v^s$  denote the molecular and structural remapping graphs, respectively, with contrastive learning loss calculated using the InfoNCE function [43]:

$$\mathcal{L}_{GCL} = -\frac{1}{N} \sum_{v=1}^N \log \frac{e^{h_v \cdot h_v^s / \tau}}{\sum_{v=1}^N e^{h_v \cdot h_{v-}^s / \tau}} \quad (13)$$

where  $\tau$  represents the temperature parameter, set by default to 0.5. Given the extensive calculations required for the denominator, this study implements a batch sampling strategy.

### H. Optimization Objectives

The optimization objective for the CardiOT model is to minimize both classification and contrastive learning losses. This study employs the Binary Cross-Entropy (BCE) function to calculate classification losses for both the molecular graph and the structural remapping graph:

$$\mathcal{L}_G = -\sum_{v=1}^N y_v \cdot \log \sigma(\hat{y}_v) + (1 - y_v) \cdot \log \sigma(1 - \hat{y}_v) \quad (14)$$

$$\mathcal{L}_G^s = -\sum_{v=1}^N y_v^s \cdot \log \sigma(\hat{y}_v^s) + (1 - y_v^s) \cdot \log \sigma(1 - \hat{y}_v^s) \quad (15)$$

where  $\mathcal{L}_G$  denotes the classification loss for the molecular graph, while  $\mathcal{L}_G^s$  denotes the loss for the structural remapping graph.  $N$  denotes the number of molecules, and  $\sigma$  symbolizes the sigmoid function. For the  $v$ -th compound,  $\hat{y}_v$  and  $\hat{y}_v^s$  represent the CCT prediction scores derived from the molecular and structural remapping graphs, respectively, corresponding to its true label. Classification and contrastive learning losses are integrated as follows:

$$\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2 + \alpha \mathcal{L}_{GCL}, \quad (16)$$

where  $\alpha$  is an adjustable parameter.

## III. RESULTS

### A. Experimental Settings

To assess the CardiOT model and a comparator model on the CCT prediction task, we utilized six standard metrics: accuracy (ACC), sensitivity (SEN), specificity (SPE), F1-score (F1), correct classification rate (CCR), and Matthews correlation coefficient (MCC) [44], [45], [46], [47]. Furthermore, the hERG, Cav1.2, and Nav1.5 datasets were split into training and validation sets at a 4:1 ratio, using cross-validation to identify the optimal model. Subsequently, we evaluated the model's predictive performance on externally collected datasets. To guarantee fairness, all models were configured identically, and the online platform operated under default settings.

The proposed CardiOT model is implemented and executed on the autoDL platform.<sup>1</sup> The hardware configuration includes an NVIDIA RTX 4090 GPU, a Xeon(R) Platinum 8352V CPU,

<sup>1</sup><https://www.autodl.com/>

**TABLE II**  
RESULTS OF ALL MODELS ON THE HERG EXTERNAL TEST SET (%)

Models/Metrics	ACC	F1	SEN	SPE	CCR	MCC
CardPred [15]	53.9	45.4	37.9	70.2	54.1	8.6
admetSAR 2.0 [48]	66.4	70.6	80.8	52.0	66.4	34.3
ADMETlab 2.0 [49]	68.0	67.5	66.4	69.6	68.0	36.0
CToxPred-HERG [27]	71.2	72.9	77.6	64.8	71.2	42.8
MultiCTox [28]	73.2	73.9	76.0	70.4	73.2	46.4
CardiOT	74.7	76.4	82.2	68.2	75.2	50.1

and 32 GB of RAM. In the experiment, the GIN encoder has 2 layers, with a batch size of 1024, 300 training epochs, a learning rate of 0.0005, and the ADAM optimizer. In the predictor, the input layer dimension is 512, the hidden layer dimension is 256, the output layer dimension is 1, and the contrastive learning loss weight ( $\alpha$ ) is set to 0.3.

### B. Performance Comparison

In this section, we assess the CardiOT model against three cutting-edge online prediction platforms, CardPred [15], ADMETsa2.0 [48], ADMETlab 2.0 [49], as well as the CToxPred [27] and MultiCBlo [28] models. For the CardiOT, MultiCTox, and CToxPred models, we identified the optimal configurations via a 5-fold cross-validation experiment and subsequently evaluated the hERG external test set using these models. For the other online platforms, we utilized default parameters and summarized the performance comparisons in Table II.

The CardiOT model outperforms others in nearly all metrics, trailing only slightly behind the MultiCTox model in SPE score. The MultiCTox model effectively employs progressive multimodal technology to integrate data from various sources, demonstrating strong performance. Conversely, the CToxPred model, relying solely on the Fingerprints modality, likely captures limited information. This underscores the importance of utilizing multiple modalities to enhance the accuracy of predictions for molecular properties like CCT. Currently, there is no effective solution for integrating various types of information. Notably, the proposed model solely utilizes SMILES to construct molecular graphs, yet achieves superior performance. This indicates that OT technology can be an effective alternative to address data scarcity and distribution issues in the absence of robust multimodal fusion techniques.

### C. Comparison on Other Datasets

To assess the effectiveness of the CardiOT model in CCT prediction, we conducted comparisons using two datasets, Nav 1.5 and Cav 1.5. We selected the advanced models, CToxPred and MultiCTox, for comparison due to their support for local deployment and ease of replication. Each model underwent 5-fold cross-validation to identify the optimal settings, followed by performance evaluation on an external test set.

Table III shows “CToxPred-Nav” where CToxPred was trained, verified, and evaluated on the Nav dataset using Fingerprints and Descriptors. “CToxPred-Cav” indicates that CToxPred was trained, verified, and evaluated on the Cav dataset

with Fingerprints and Descriptors. The CardiOT model demonstrates high performance across various datasets, indicating strong generalization capabilities. The results confirm the superior performance of the CardiOT model in CCT prediction, likely due to the KAN and OT techniques it employs. KAN technology effectively learns weight parameters and enhances model flexibility, improving adaptability. OT technology aims to emphasize crucial “sample-feature dimension” pairs, addressing issues related to data scarcity and uneven distribution.

### D. Ablation Experiments

We conducted a series of ablation studies to assess the impact of various modules on the CardiOT model’s performance using the hERG dataset. Table IV displays the outcomes of these ablation experiments. ‘w/o GCL’ indicates the removal of graph contrast learning, ‘w/o KAN’ denotes the elimination of the KAN component, ‘w/o SR’ refers to the removal of the structural remapping, and ‘w/o OT’ signifies the exclusion of the optimal transport layer in the CardiOT model. The table results demonstrate that all modules positively enhance the model’s performance. Notably, the CardiOT model performs poorest without the OT technology.

Additionally, we conducted a set of experiments to examine model performance when only the original molecular graph or the structure remapping graph is used. Results are displayed in Table V, where “OG” denotes the model using only the original molecular graph, and “SRG” indicates the use of only the structure remapping graph. Observations indicate that using only the original molecular graph or the structure remapping graph results in lower performance. When both the original molecular graph and structure remapping graph are used in comparative learning, the model’s performance significantly improves. This indicates that the original molecular graph and structure remapping graph achieve complementary information, thereby enhancing model performance.

### E. Parameter Experiments

1) *Testing Different Pooling Techniques*: Essentially, CCT prediction involves graph classification tasks on molecular graphs of compounds. Graph pooling techniques are commonly employed in graph classification tasks to extract crucial topological structures or nodes. Consequently, we designed experiments to assess the impact of various pooling techniques on CCT prediction efficacy. Table VI illustrates that “Mean” corresponds to average pooling, “Max” to maximum pooling, and “Mix” to the combination of global maximum and average pooling, with vector concatenation following these operations. “OT” indicates the use of optimal transfer-based pooling for both graphs. “MO” denotes Mix pooling for the molecular graph and OT pooling for the remap, while “OM” reverses these roles. “No” indicates no pooling was performed in the model.

The CardiOT model exhibits optimal performance when OT pooling is applied to the molecular graph and Mix pooling to the remap graph. The performance falters when OT pooling is applied solely to the remap graph, possibly due to changes in the molecular structure that affect OT pooling’s applicability.

**TABLE III**  
RESULTS OF THE MODELS ON THE NAV 1.5 AND CAV 1.5 EXTERNAL TEST SETS (%)

Datasets	Models/Metrics	ACC	F1	SEN	SPE	CCR	MCC
Nav 1.5	CToxPred-Nav [27]	73.4	80.5	82.4	52.8	67.6	33.9
	MultiCTox [28]	69.3	78.8	77.4	45.6	61.5	26.4
	CardiOT	74.6	81.1	79.3	57.3	68.3	35.4
Cav 1.5	CToxPred-Cav [27]	68.9	75.4	72.1	63.3	67.7	33.8
	MultiCTox [28]	68.3	73.5	69.5	62.3	65.9	25.6
	CardiOT	70.5	76.4	73.2	64.8	69.0	37.1

**TABLE IV**  
RESULTS OF ABLATION EXPERIMENTS (%)

Models/Metrics	ACC	F1	SEN	SPE	CCR	MCC
w/o GCL	72.7	74.2	78.7	66.7	72.7	45.8
w/o KAN	73.3	74.3	77.7	69.9	73.8	46.8
w/o SR	71.7	72.1	76.1	65.3	70.7	41.7
w/o OT	70.6	71	76.4	62.8	69.6	40.7
CardiOT	74.7	76.4	82.2	68.2	75.2	50.1

**TABLE V**  
RESULTS OF THE MODEL USING THE ORIGINAL MOLECULAR GRAPH OR THE STRUCTURE REMAPPING GRAPH (%)

Datasets	Metrics	ACC	F1	SEN	SPE	CCR	MCC
HERG	OG	71.7	72.1	76.1	65.3	70.7	41.7
	SRG	71.4	72.5	77.4	65.2	71.3	42.5
	CardiOT	74.7	76.4	82.2	68.2	75.2	50.1
Nav1.5	OG	72.0	75.9	74.4	54.0	64.2	29.7
	SRG	71.8	76.4	74.6	54.2	64.4	34.9
	CardiOT	74.6	81.1	79.3	57.3	68.3	35.4
Cav1.5	OG	67.6	71.9	66.8	59.6	63.2	34.3
	SRG	66.7	71.6	66.4	60.2	63.3	33.9
	CardiOT	70.5	76.4	73.2	64.8	69.0	37.1

**TABLE VI**  
EVALUATING MODEL PERFORMANCE WITH VARIOUS POOLING TECHNIQUES (%)

Methods/Metrics	ACC	F1	SEN	SPE	CCR	MCC
No	70.6	71	76.4	62.8	69.6	40.7
Mean	70.4	72.9	80.3	60.5	70.4	41.8
Max	72.8	72.1	76.1	65.3	70.7	41.7
Mix	71.6	73.5	79.2	64	71.6	43.7
OT	73.9	75.1	81.9	64.3	73.1	46.7
MO	72.5	74.9	74.8	70.2	72.5	45.8
OM	74.7	76.4	82.2	68.2	75.2	50.1

Nevertheless, this approach outperforms the Mean, Max, or Mix methods, affirming the efficacy of OT technology in pooling layers.

**2) Testing Different Feedforward Module Numbers:** The number of feedforward modules serves as a crucial hyperparameter in the OT pooling layer. Increasing the number of feedforward modules can enhance the precision of solutions to (7), albeit at the cost of extended time for calculations and backpropagation. Consequently, we investigated how varying numbers of feedforward modules affect model performance on hERG, Cav and Nav datasets, respectively. Specifically, we determined the optimal configuration for feedforward modules within the range of 4 to 12, as illustrated in Fig. 3(A)–(C). Results

**TABLE VII**  
RESULTS OF CARDIOT MODEL USING DIFFERENT GNN ENCODERS ON HERG, NAV1.5, AND CAV1.5 DATASETS (%)

Datasets	Metrics	ACC	F1	SEN	SPE	CCR	MCC
HERG	GCN	73.3	75.3	80.5	66.7	73.6	47.9
	GAT	73.5	74.2	81.6	67.8	74.7	48.2
	GIN	74.7	76.4	82.2	68.2	75.2	50.1
Nav1.5	GCN	73.9	80.0	78.4	55.2	66.8	34.2
	GAT	74.2	80.8	79.0	56.8	67.9	34.9
	GIN	74.6	81.1	79.3	57.3	68.3	35.4
Cav1.5	GCN	68.9	75.5	72.1	63.3	67.7	35.9
	GAT	69.7	76.0	72.7	64.5	68.6	36.6
	GIN	70.5	76.4	73.2	64.8	69.0	37.1

indicated improved performance with an increase in feedforward modules from 4 to 8. However, performance declined when the number of modules rose from 9 to 12. Optimal performance was achieved with eight feedforward modules. This suggests that while appropriately increasing the number of feedforward modules can enhance performance, excessive modules may overload computations and degrade performance.

**3) Testing Different GNN Encoders:** The aggregation method used by GIN effectively captures heterogeneous information between nodes, enhancing the representational capacity for complex molecular structures. This design enables each node's embedding to depend not only on the properties of its immediate neighbors but also on higher-order adjacent nodes, thereby more comprehensively capturing the molecular structural information. Additionally, it highlights the compound's heterogeneous information. Recent research and applications indicate that GIN performs well in fields like drug discovery and toxicity prediction. GIN effectively identifies compound activity and potential toxicity, supporting decision-making in drug development. However, models like GCN [50] and GAT [51] are less effective than GIN at capturing compound structure.

Additionally, we conducted experiments on the HERG, Nav1.5, and Cav1.5 datasets to assess the impact of different GNN encoders on model performance, with results presented in Table VII. Results indicate that the model using GIN as an encoder outperforms models using other GNN encoders. Similarly, to better capture compound structure, the CardiOT model employs GIN as the GNN encoder.

**4) Testing Different Optimization Algorithms:** This study selected the sinkhorn-scaling algorithm to solve the EOUT problem, among options such as Bregman ADMM. To verify the



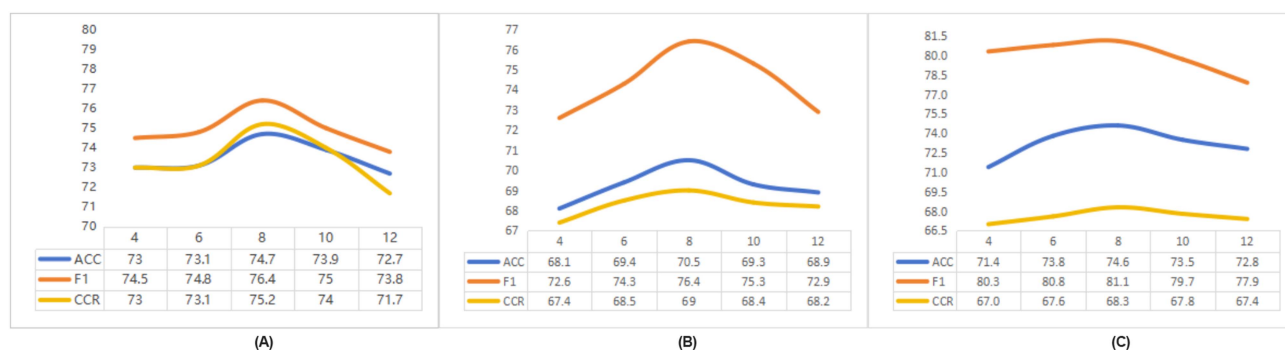


Fig. 3. Evaluating model performance with different feedforward module numbers on (A) hERG, (B) Cav and (C) Nav datasets, respectively.

TABLE VIII

RESULTS OF USING DIFFERENT ALGORITHMS FOR THE RTP PROCESS IN THE MODEL (%)

Algorithms/Metrics	ACC	F1	SEN	SPE	CCR	MCC
badmm-e	72.7	74.1	78.5	66.7	72.6	45.6
badmm-q	73.7	75.3	80.4	67.0	73.7	48.0
sinkhorn	74.7	76.4	82.2	68.2	75.2	50.1

effectiveness of the sinkhorn-scaling algorithm, a set of experiments was conducted, with results displayed in Table VIII. In this model, “badmm-e” indicates that the ROT process employs the BADMM algorithm with entropy smoothing regularization, “badmm-q” represents the BADMM algorithm with quadratic smoothing regularization, and “sinkhorn” denotes the ROT process using the sinkhorn-scaling algorithm.

The ROT process in the model performs optimally with the sinkhorn-scaling algorithm. This is likely due to the sinkhorn-scaling algorithm’s ability to maintain marginal distribution balance, a crucial feature for handling complex distribution data. Additionally, the algorithm’s interpretability enhances model transparency in decision-making. Furthermore, the computational complexity of the ROT process with the sinkhorn-scaling algorithm is  $O(T(N^2D + D^2N + KND))$ , where  $(N^2D + D^2N)$  represents  $K$  iterations within each sinkhorn-scaling module. The sinkhorn-scaling algorithm introduces an entropy regularization term to transform the original linear programming problem into a differentiable optimization problem, significantly reducing computational complexity.

## F. Interpretability Analysis

We employed the EdgeSHAPer method [52] to analyze how the CardiOT method relies on key structures (“bonds”) for CCT prediction tasks. In the compound’s structure remapping graph, nodes represent “bonds.” Calculating each node’s Shapley value quantifies the contribution of these bonds to the model’s predictive accuracy. Building on this, the study examined 125 CCT and non-CCT compounds in the hERG test set. Bonds exhibiting Shapley values less than -0.005 merit special attention. Consequently, we identified chemical bonds and their adjacent structures that impact model performance. We documented the

ten most frequent chemical bonds and their adjacent structures, detailed in Table IX.

Fig. 4(A)–(C) illustrate the chemical structures of two CCT compounds ( $pIC_{50} \geq 5$ ), while Fig. 4(D)–(F) depict those of two non-CCT compounds ( $pIC_{50} < 5$ ). Blue areas signify negative impacts on CCT prediction, whereas red areas indicate positive impacts. The intensity of the color correlates with the impact’s magnitude. Utilizing the EdgeSHAPer method, this study elucidated the contribution of key chemical bonds to model and CCT prediction performance. This enhances understanding of the link between chemical structure and biological activity, potentially guiding innovation in drug research and development.

## G. Visualization Analysis

A pharmacophore is a specific structural arrangement within a molecule that interacts with a biological target, such as enzymes or receptors. Typically comprising specific atoms or functional groups, it plays a crucial role in determining a molecule’s pharmacological and toxicological properties. Identifying pharmacophores or toxicophores is vital for elucidating drug mechanisms, optimizing designs, and predicting activity. We employed t-SNE technology to visualize molecular features within the Cav, Nav, and hERG datasets. In each dataset, we selected pairs of molecules, from both CCT and non-CCT compounds, that contain highly correlated pharmacophores or toxicophores. The results are presented in Fig. 5. It is evident that compounds sharing similar pharmacophores or toxicophores do not significantly aggregate. These findings indicate a significant uneven data distribution across the three datasets. This poses a substantial challenge for molecular prediction tasks, including CCT.

## IV. DISCUSSION

This study integrates GNN technology with OT, SR, and KAN methods to propose an innovative CCT prediction model. This integration strategy addresses the challenges of traditional GNN models, such as data scarcity and uneven distribution, while enhancing the representation of complex chemical structures. Using OT technology, we optimized the joint distribution of sample and feature dimensions, accurately identifying “critical” sample-feature pairs. This approach not only mitigates data scarcity but also reduces bias from data imbalance through

TABLE IX  
TEN MOST FREQUENT CHEMICAL BONDS AND THEIR ADJACENT BONDS

Nums	973	669	397	211	106
Bonds	c-c	H-c	c-n	C-C	C-H
Adjacent bonds	c-c, H-c, H-c	c:c, c:c	c-c, H-c, H-c	C-N, C-C, C-H	C-N, C-C, C-H
Nums	95	80,	75	63	52
Bonds	c-o	C-N	C-c	C-n	O-c
Adjacent bonds	c-c, H-c, H-c	C-C, C-H, C-H	C-N, C-H, C-H	C-H, C-H, C-H	c-c, c-c, c-c

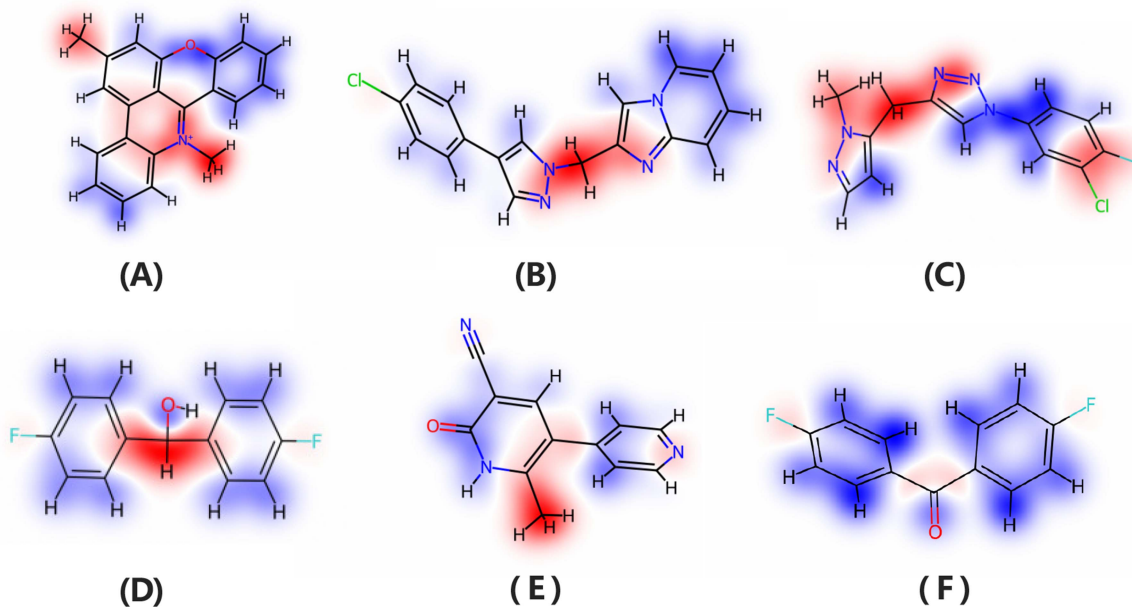


Fig. 4. Key substructures influence the CCT prediction of compounds, with blue areas indicating negative influence and red areas indicating positive influence.

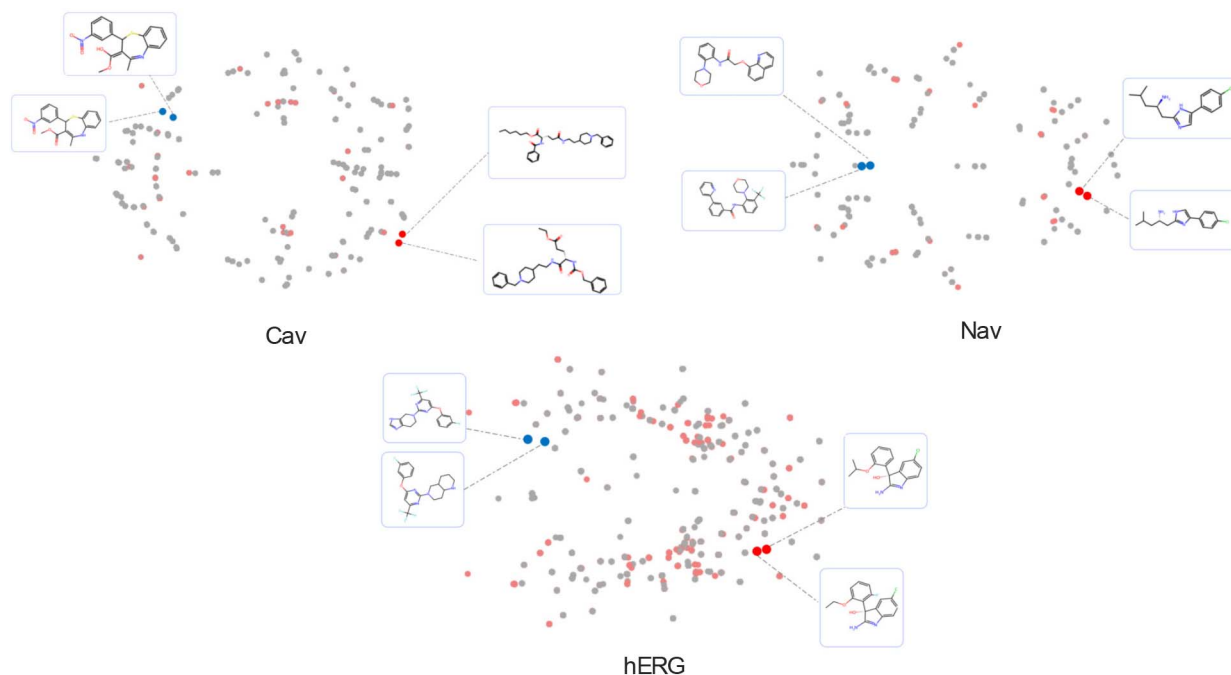


Fig. 5. Visualization of CardiOT model prediction results on Cav, Nav, and hERG datasets.

distribution optimization, creating a more robust foundation for model training. Traditional GNN models typically emphasize atomic-level information flow during transmission. This study employs SR technology to shift the focus of the molecular graph from atoms to chemical bonds, emphasizing the critical role of bonds in compound structure and function. This shift provides richer information for the model, enhancing its understanding of chemical structure complexity. Additionally, we replaced the traditional multi-layer perceptron (MLP) classifier with KAN technology. This modification streamlines the model structure and significantly enhances prediction accuracy and interpretability through its kernel function and attention mechanism. KAN's strength lies in its self-learning ability to assign accurate weights to features, capturing chemical attributes critical to CCT prediction.

The experimental results are promising. The CardiOT model demonstrated excellent performance in CCT prediction, particularly under poor data conditions. It maintained stable performance even under these challenging conditions. This stability is primarily attributed to the seamless integration of OT, SR, and KAN technologies, which are crucial for data distribution, structural characterization, and prediction accuracy. Notably, with the support of OT and KAN technologies, our model achieved a level of interpretability unmatched by other models. The results further indicate that our model accurately predicts cardiac ion channel inhibition. This breakthrough offers a new perspective on compound toxicity research, provides robust support for environmental assessment and drug development, and is expected to play a significant role in reducing health risks.

Given the growing concerns about drug abuse and safety, evaluating the effects of free compounds on the heart is critically important. The proposed CCT prediction model developed in this study significantly enhances the accuracy of compound CCT prediction through the integration of multiple advanced technologies, profoundly impacting clinical medicine. First, the model accurately identifies potential cardiac toxicity early in drug development, aiding drug screening and optimization, reducing clinical trial risks, and ensuring patient safety. Second, by swiftly identifying low-toxicity compounds and optimizing toxic drug structures, the model is expected to accelerate drug approval and enhance overall efficiency. In summary, the proposed CCT prediction model demonstrates substantial value and potential in improving drug safety and advancing research, and is expected to play a crucial role in protecting human health.

## V. CONCLUSION

Accurate CCT prediction can elucidate the compound's mechanism of action, enabling proactive management of CCT-related risks in drug development. While current advanced models excel in CCT prediction, they struggle with data scarcity and uneven distribution. Consequently, this study introduces the CardiOT model, which leverages GNN integrated with OT, SR, and KAN technologies to address these issues. Specifically, the CardiOT model utilizes OT-based pooling and expectation maximization to optimize the joint distribution of sample and feature dimensions, identifying key "sample-feature" pairs to mitigate

data scarcity and uneven distribution. Additionally, the model employs SR technology to enhance the role of chemical bond information within the GNN message propagation process. Finally, the replacement of traditional MLP with KAN technology markedly enhances the model's accuracy and generalizability. Experimental results demonstrate that the CardiOT model excels in CCT prediction. This model is expected to clarify the inhibitory mechanisms of compounds on cardiac ion channels, reducing risks in cardiac drug development.

Looking ahead, accurate CCT prediction will focus on three key strategies: First, employing multimodal fusion technology to integrate diverse knowledge sources and enhance model generalizability. Second, integrating auxiliary tasks like conformational changes, metabolic networks, and interaction predictions to aid CCT prediction. Third, leveraging transfer learning, pre-training, and large-scale models to tackle the challenges of scarce or uneven data distribution.

## REFERENCES

- [1] E. G. Kruchinin et al., "Cardiac arrhythmias and conduction disorders in acute poisoning with cardiotoxic substances," *Pharm. Formulas*, vol. 4, pp. 20–34, 2022.
- [2] S. C. Gad, "Cardiovascular toxicology and its evaluation," in *Mammalian Toxicology*. Hoboken, NJ, USA: Wiley, 2015, pp. 425–452.
- [3] H. H. D. Tsai et al., "Informing hazard identification and risk characterization of environmental chemicals by combining transcriptomic and functional data from human-induced pluripotent stem-cell-derived cardiomyocytes," *Chem. Res. Toxicol.*, vol. 37, no. 8, pp. 1428–1444, 2024.
- [4] Z. Zequn et al., "Off-label use of chloroquine, hydroxychloroquine, azithromycin and Lopinavir/Ritonavir in COVID-19 risks prolonging the QT interval by targeting the hERG channel," *Eur. J. Pharmacol.*, vol. 893, 2021, Art. no. 173813.
- [5] M. L. Ponte, G. A. Keller, and G. Di Girolamo, "Mechanisms of drug induced QT interval prolongation," *Curr. Drug Saf.*, vol. 5, pp. 44–53, 2010.
- [6] E. M. Tosca et al., "Modeling approaches for reducing safety-related attrition in drug discovery and development: A review on myelotoxicity, immunotoxicity, cardiovascular toxicity, and liver toxicity," *Expert Opin. Drug Discov.*, vol. 16, pp. 1365–1390, 2021.
- [7] B. O. Villoutreix and O. Taboureau, "Computational investigations of hERG channel blockers: New insights and current predictive models," *Adv. Drug Del. Rev.*, vol. 86, pp. 72–82, 2015.
- [8] A. Mitra, C. H. Lee, and K. Cheng, *Advanced Drug Delivery*. Hoboken, NJ, USA: Wiley, 2013.
- [9] E. Gorelik et al., "The cardiovascular safety of antiobesity drugs—Analysis of signals in the FDA adverse event report system database," *Int. J. Obesity*, vol. 44, pp. 1021–1027, 2020.
- [10] ICH Harmonised Tripartite Guideline, "The non-clinical evaluation of the potential for delayed ventricular repolarization (QT interval prolongation) by human pharmaceuticals," S7B, 2005. [Online]. Available: <http://www.ich.org/products/guidelines/safety/article/safety-guidelines.html>
- [11] T. Brennan, M. Fink, and B. Rodriguez, "Multiscale modelling of drug-induced effects on cardiac electrophysiological activity," *Eur. J. Pharmacol.*, vol. 36, pp. 62–77, 2009.
- [12] A. P. Larsen, B. H. Bentzen, and M. Grunnet, "Differential effects of kV11.1 activators on kV11.1A, kV11.1B and kV11.1A/kV11.1B channels," *Brit. J. Pharmacol.*, vol. 161, pp. 614–628, 2010.
- [13] L. Pu et al., "eToxPred: A machine learning-based approach to estimate the toxicity of drug candidates," *BMC Pharmacol. Toxicol.*, vol. 20, pp. 1–15, 2019.
- [14] S. Wang et al., "Recent developments in computational prediction of hERG blockage," *Curr. Topics Med. Chem.*, vol. 13, pp. 1317–1326, 2013.
- [15] A. Karim et al., "Cardiotox Net: A robust predictor for hERG channel blockade based on deep learning meta-feature ensembles," *J. Cheminformatics*, vol. 13, pp. 1–13, 2021.
- [16] X. Zhang et al., "HergSPred: Accurate classification of hERG blockers/nonblockers with machine-learning models," *J. Chem. Inf. Model.*, vol. 62, pp. 1830–1839, 2022.

- [17] X. Yang et al., "Multi-task aquatic toxicity prediction model based on multi-level features fusion," *J. Adv. Res.*, 2024.
- [18] L. Liu, Y. Wei, Q. Zhang, and Q. Zhao, "SSCRB: Predicting circRNA-RBP interaction sites using a sequence and structural feature-based attention model," *IEEE J. Biomed. Health Inform.*, vol. 28, no. 3, pp. 1762–1772, Mar. 2024.
- [19] Z. Chen, L. Zhang, J. Sun, R. Meng, S. Yin, and Q. Zhao, "DCAMCP: A deep learning model based on capsule network and attention mechanism for molecular carcinogenicity prediction," *J. Cellular Mol. Med.*, vol. 27, no. 20, pp. 3117–3126, 2023.
- [20] J. Wang et al., "Predicting drug-induced liver injury using graph attention mechanism and molecular fingerprints," *Methods*, vol. 221, pp. 18–26, 2024.
- [21] T. Wang, J. Sun, and Q. Zhao, "Investigating cardiotoxicity related with hERG channel blockers using molecular fingerprints and graph attention mechanism," *Comput. Biol. Med.*, vol. 153, 2023, Art. no. 106464.
- [22] W. Liu, T. Tang, X. Lu, X. Fu, Y. Yang, and L. Peng, "MPCLCDA: Predicting circRNA–disease associations by using automatically selected meta-path and contrastive learning," *Brief. Bioinf.*, vol. 24, no. 4, 2023, Art. no. bbad227.
- [23] J. Wei et al., "DrugRealign: A multisource prompt framework for drug repurposing based on large language models," *BMC Biol.*, vol. 22, no. 1, 2024, Art. no. 226.
- [24] S. Wang et al., "ADMET evaluation in drug discovery. 16. predicting hERG blockers by combining multiple pharmacophores and machine learning approaches," *Mol. Pharmaceutics*, vol. 13, pp. 2855–2866, 2016.
- [25] T. T. Van Tran et al., "Artificial intelligence in drug toxicity prediction: Recent advances, challenges, and future perspectives," *J. Chem. Inf. Model.*, vol. 63, pp. 2628–2643, 2023.
- [26] J. Y. Ryu et al., "DeepHIT: A deep learning framework for prediction of hERG-induced cardiotoxicity," *Bioinformatics*, vol. 36, pp. 3049–3055, 2020.
- [27] I. Arab et al., "Benchmarking of small molecule feature representations for hERG, Nav1. 5, and Cav1. 2 cardiotoxicity prediction," *J. Chem. Inf. Model.*, vol. 64, pp. 2515–2527, 2023.
- [28] T. Wang et al., "MultiCBlo: Enhancing predictions of compound-induced inhibition of cardiac ion channels with advanced multimodal learning," *Int. J. Biol. Macromolecules*, vol. 276, 2024, Art. no. 133825.
- [29] A. Gaulton et al., "ChEMBL: A large-scale bioactivity database for drug discovery," *Nucleic Acids Res.*, vol. 40, pp. D1100–D1107, 2012.
- [30] S. Kim et al., "PubChem in 2021: New data content and improved web interfaces," *Nucleic Acids Res.*, vol. 49, pp. D1388–D1395, 2021.
- [31] T. Liu et al., "BindingDB: A web-accessible database of experimentally determined protein–ligand binding affinities," *Nucleic Acids Res.*, vol. 35, pp. D198–D201, 2007.
- [32] F. Du et al., "hERGcentral: A large database to store, retrieve, and analyze compound-human ether-a-go-go related gene channel interactions to facilitate cardiotoxicity assessment in drug development," *Assay Drug Develop. Technol.*, vol. 9, pp. 580–588, 2011.
- [33] L. S. Kumar Konda, S. K. Praba, and R. Kristam, "hERG liability classification models using machine learning techniques," *Comput. Toxicol.*, vol. 12, 2019, Art. no. 100089.
- [34] X. Ma et al., "GraphADT: Empowering interpretable predictions of acute dermal toxicity with multi-view graph pooling and structure remapping," *Bioinformatics*, vol. 40, 2024, Art. no. btae438.
- [35] K. Xu et al., "How powerful are graph neural networks?," in *Proc. 7th Int. Conf. Learn. Representations*, New Orleans, LA, USA, 2019.
- [36] H. You et al., "MC-Net: Multiple max-pooling integration module and cross multi-scale deconvolution network," *Knowl.-Based Syst.*, vol. 231, 2021, Art. no. 107456.
- [37] Z. Zhou and M. Li, "Semi-supervised regression with co-training," in *Proc. 19th Int. Joint Conf. Artif. Intell.*, L. P. Kaelbling and A. Saffiotti, Eds., Edinburgh, Scotland, U.K., Jul. 30, 2005, pp. 908–916.
- [38] H. Xu and M. Cheng, "Regularized optimal transport layers for generalized global pooling operations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 12, pp. 15426–15444, Dec. 2023.
- [39] H. Xu, D. Luo, and L. Carin, "Scalable Gromov-Wasserstein learning for graph partitioning and matching," in *Proc. 33rd Int. Conf. Neural Inf. Process. Syst.*, 2019, pp. 3052–3062.
- [40] L. Chizat et al., "Scaling algorithms for unbalanced optimal transport problems," *Math. Comput.*, vol. 87, pp. 2563–2609, 2018.
- [41] M. Cuturi, "Sinkhorn distances: Lightspeed computation of optimal transport," in *Proc. 26th Int. Conf. Adv. Neural Inf. Process. Syst.*, 2013, vol. 26, pp. 2292–2300.
- [42] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, "Bregman iterative algorithms for backlash  $\ell_1$ -minimization with applications to compressed sensing," *SIAM J. Imag. Sci.*, vol. 1, no. 1, pp. 143–168, 2008.
- [43] K. He et al., "Momentum contrast for unsupervised visual representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 9729–9738.
- [44] R. Wang et al., "An effective plant small secretory peptide recognition model based on feature correction strategy," *J. Chem. Inf. Model.*, vol. 64, pp. 2798–2806, 2023.
- [45] L. Cai, Y. He, X. Fu, L. Zhuo, Q. Zou, and X. Yao, "AEGNN-M: A 3D graph-spatial co-representation model for molecular property prediction," *IEEE J. Biomed. Health Inform.*, early access, Feb. 22, 2024, doi: 10.1109/JBHI.2024.3368608.
- [46] R. Wang et al., "Diff-amp: Tailored designed antimicrobial peptide framework with all-in-one generation, identification, prediction and optimization," *Brief. Bioinf.*, vol. 25, 2024, Art. no. bbae078.
- [47] J. Wei et al., "Efficient deep model ensemble framework for drug-target interaction prediction," *J. Phys. Chem. Lett.*, vol. 15, pp. 7681–7693, 2024.
- [48] H. Yang et al., "admetSAR 2.0: Web-service for prediction and optimization of chemical ADMET properties," *Bioinformatics*, vol. 35, pp. 1067–1069, 2019.
- [49] G. Xiong et al., "ADMETlab 2.0: An integrated online platform for accurate and comprehensive predictions of ADMET properties," *Nucleic Acids Res.*, vol. 49, pp. W5–W14, 2021.
- [50] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Representations*, 2022.
- [51] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *Proc. Int. Conf. Learn. Representations*, 2018.
- [52] A. Mastropietro et al., "EdgeSHAPer: Bond-centric shapley value-based explanation method for graph neural networks," *iScience*, vol. 25, 2022, Art. no. 105043.