Predicting Ionic Liquid Toxicity with Graph Attention Networks

Safa Sadaghiyanfam[1], Hiqmet Kamberaj[2,3], Yalcin Isler[4,*]
[1]Cozumcul Technology, Department of Software Development, Menderes, Izmir, Turkey
[2]International Balkan University, Department of Computer Engineering, Skopje, North Macedonia
[3]National Institute of Physics, Fan Noli Square, Academy of Sciences of Albania, Tirana, Albania
[4]Alanya Alaaddin Keykubat University, Department of Electrical and Electronics Engineering, Alanya, Antalya, Turkey
ORCIDs: 0000-0002-6487-7501, 0000-0001-7357-7490, 0000-0002-2150-4756
E-mails: safa.sd1993@gmail.com, h.kamberaj@gmail.com, islerya@yahoo.com
*Corresponding author.

***Abstract*—Ionic liquid toxicity is a major cause of morbidity and mortality worldwide. However, determining the toxicity of ionic liquids is still a challenge for the industry. In this study, we propose a graph attention network (GAT) based framework specifically adapted for predicting ionic liquid toxicity using graph-based molecular representations. In order to improve the model's accuracy, we added physicochemical descriptors and hyperparameter optimization using the Optuna framework. In addition, Grad-CAM (gradient-weighted class activation mapping) visualizations and permutation feature importance analysis were employed to visualize the significance of the nodes and edges in the graphs. Experimental results demonstrate that the EnhancedGAT achieves lower prediction errors and stronger correlations compared to traditional baselines, indicating high predictive reliability.**

***Keywords*—Ionic liquids, Toxicity prediction, Graph Attention Network, Molecular graphs, Physiochemical descriptors, Permutation feature importance**

I. INTRODUCTION

Ionic liquids are considered novel solvents and have gained attention across various fields because of their low volatility, thermal stability, and other properties of interest [1]. These features, however, raise concerns about potential toxicity, which makes ionic liquids challenging for safe, sustainable use. Concerns about possible harm to the health of a living organism or the environment present a significant challenge for sustainability and safe use. Traditional methods are mostly inefficient, time-consuming, and often painfully depend on animal testing [2]. Historically, compound toxicity has been predicted using Quantitative Structure-Activity Relationship (QSAR) models, which calculate the toxicity with the use of chemical descriptors for the molecules. In principle, these models should work for every case, but they are overly reliant on manually crafted features and molecular descriptors, which might be critical to predict toxicity and overlook intricate model interactions [3]. Furthermore, these models suffer from losing predictive reliability from bias in the selection of descriptors. These models utilize flexible molecular fingerprints and descriptors, as opposed to QSAR, which increases their versatility [4].

Deep learning, especially its recent advancements, offers increasingly sophisticated systems with automated, predefined feature extraction, in particular, for GNNs [5]–[7]. With regards to GNNs, they have become remarkably useful in portraying chemical compounds as graphs with atoms as nodes and bonds as edges [8]. GATs, in particular, have shown strong performance in molecular property prediction tasks such as solubility, drug-likeness, and quantum properties [9], [10], [11]. GNNs also incorporate more complex features from molecular graphs, such as properties of atoms: atomic number, degree of bond's electronegativity, state of hybridization, formal charges, and types of bonds, which include single, double, and aromatic bonds. This flexibility improves predictive precision across various chemical structure types [9]. Several variants of GNNs, such as Message Passing Neural Networks (MPNNs), Graph Convolutional Networks (GCNs), as well as Graph Attention Networks (GATs), have been used in attempts to predict toxicity [8], [12]. These models utilize the information contained in the adjacent atoms residing in the molecular graphs, and these aggregates are essential to form the accurate predictions of toxicity, since they capture the interactions of different constituents of the molecule, both at the local and global levels. Among all models, GATs have shown remarkable performance due to their advantage of their attention mechanisms. GATs consider the interactions between atoms and assign attention weights based on the property of the target. These properties allow the model to focus on the most critical molecular interactions. This attention mechanism makes GATs outperform traditional ML models and other GNNs in tasks like toxicity prediction [12].

The performance of various Graph Neural Network (GNN) models, such as GCN, MPNNs, and GATs, on a variety of toxicity prediction tasks was assessed in a study by Ketkar et al. [13]. This specific GNN variation makes use of attention mechanisms to improve the predictability and interpretability of its outcomes. Using a hierarchical attention mechanism,

Attentive FP analyzes interactions at the bond and atom levels. By using this method, the model may focus on the atomic interactions that are most important for toxicity. In comparison to other models, Attentive FP was able to significantly reduce prediction errors in this way, underscoring the growing importance of attention-based GNNs in cheminformatics [8], [9], [12].

Despite their successes, GNNs still have limitations, such as the need for large and diverse datasets to generate accurate models. Furthermore, it might be challenging for GNNs to capture long-range molecular interactions that affect toxicity. Researchers are looking on hybrid models that combine GNNs with other deep learning architectures, such as transformers, to improve the collection of both local and global molecular information [14]. Additionally, studies are being conducted to enhance the interpretability of GNNs, particularly about attention mechanisms that reveal details about the molecular origins of toxicity.

In this study, we propose a task-specific GATs framework to predict the toxicity of ionic liquids. By representing ILs as molecular graphs, GATs can capture complex atomic interactions that affect toxicity. By assigning varying degrees of priority to different bonds and atomic features, the network's attention mechanism allows the model to rank the most significant molecular interactions. This approach clarifies the specific atomic properties that cause toxicity, while facilitating interpretation. It should be noted that the main goal of this study is to illustrate how the attention mechanism of GATs can outperform traditional machine learning methods, where they need hand-crafted features and lack an attention mechanism.

Although GAT-based models such as AttentiveFP [9] and ChemBERTa [15] have performed well for small-molecule problems, their application to IL toxicity is still lacking. Furthermore, most of these methods offer limited insight into the decision-making process of the model, leaving a gap in interpretability.

The main contributions of this work are:

- **Task-Specific Adaptation of GAT:** We present a graph attention network framework tailored for IL toxicity prediction. While based on the established GAT architecture, our model integrates atomic, bond-level, and global molecular descriptors to capture multi-scale chemical information relevant to toxicity.
- **Explainability Techniques:** Grad-CAM and the importance of the permutation feature, which offer crucial insights into the contributions of features and the significance of nodes, were used to improve the model's interpretability.
- **Extensive Evaluation:** We employed a variety of performance metrics and some tests to evaluate the model's effectiveness.
- **Visual Interpretability:** To increase transparency, we provided representations of the residuals, predictions, and Grad-CAM explanations.

These contributions tackle the challenges associated with predicting the toxicity of ionic liquids by prioritizing both predictive accuracy and enhanced model interpretability.

II. MATERIALS & METHODS

This section outlines the steps involved in developing a GAT model for predicting the toxicity of ionic liquids. We describe the preprocessing techniques of the dataset, the GAT architecture, and the training process.

A. Data Acquisition and Representation

In this research, the dataset utilized for predicting the toxicity of ionic liquids was obtained from the article by [16], who gathered a thorough dataset of ionic liquids specifically for predicting their properties. The dataset encompasses 355 ionic liquid compounds expressed through their SMILES (Simplified Molecular Input Line Entry System) strings, alongside their respective toxicity values (Refer to Table I). To ensure the dataset's quality and reliability, we conducted a Tanimoto similarity assessment to analyze the chemical diversity of the compounds. Furthermore, we employed violin plots to depict the diversity and distribution of toxicity values. As shown in our previously published research [17], [18], these techniques confirmed the robustness and representativeness of the dataset used in this study.

B. Data Preprocessing and Graph Construction

We created a pipeline to convert molecular structures into graph representations by transforming SMILES strings into graph objects tailored for Graph Neural Networks. This method captures features at both the atom and bond levels, while also incorporating molecular-level descriptors to effectively represent multi-scale molecular information.

1) Atom-Level Feature Extraction: Every atom within the molecule is depicted by a feature vector generated using RDKit [19]. These features encompass:

- **Atomic Number:** A distinct identifier for the element.
- **Degree:** The quantity of neighboring atoms that are bonded to the atom.
- **Formal Charge:** The overall charge that resides on the atom.
- **Total Number of Hydrogens:** The total count of hydrogen atoms associated with the atom.
- **Aromaticity:** A binary indicator that shows if the atom is involved in an aromatic ring.
- **Hybridization State:** The atom's hybridization type (e.g., sp, sp2, sp3), represented as an integer.
- **Normalized Molecular Weight:** The atom's part of the molecular weight, adjusted by a factor of 500.
- **Isotopic Information:** Information regarding the atom's isotope, if relevant.

These features are integrated into a numerical format that represents the physicochemical characteristics of each atom.

2) Bond-Level Feature Extraction: Features at the bond level reflect the characteristics of the chemical connections between atoms. These features comprise:

- **Bond Type:** A one-hot encoded format that indicates whether the bond is single, double, triple, or aromatic.
- **Conjugation Status:** A binary feature that determines if the bond is part of a conjugated system.

Table I: Experimental $\log EC_{50}$ values of selected ionic liquids (ILs).

IL No.	Name	SMILES	Experimental $\log EC_{50}$
1	(Ethoxymethyl)ethyldimethylammonium Chloride	[N+](C)(C)(CC)COCC.[Cl-]	3.59
2	(Ethoxymethyl)ethyldimethylammonium Bis(trifluoromethylsulfonyl)amide	[N+](C)(C)(CC)COCC.[N-](S(=O)(=O)C(F)(F)F)S(=O)(=O)C(F)(F)F	3.80
3	Tetraethylammonium Bis[1,2-benzenediolato(2-)-O1,O2]borate	O1c4c(OB-[ ]12Oc3c(O2)cccc3)cccc4.CC[N+](CC)(CC)CC	1.17
4	Benzyldecyldimethylammonium Chloride	[N+](C)(C)(Cc1ccccc1)CCCCCCCCCC.[Cl-]	0.64
5	(Ethoxycarbonylmethyl)ethyldimethylammonium Bis(trifluoromethylsulfonyl)amide	[N+](C)(C)(CC)CC(=O)OCC.[N-](S(=O)(=O)C(F)(F)F)S(=O)(=O)C(F)(F)F	3.53
6	Butyltrimethylammonium Bis(trifluoromethylsulfonyl)amide	[N+](C)(C)(C)CCCC.[N-](S(=O)(=O)C(F)(F)F)S(=O)(=O)C(F)(F)F	3.61
7	Ethyl(2-ethoxyethyl)dimethylammonium Bis(trifluoromethylsulfonyl)amide	[N+](C)(C)(CC)COCC.[N-](S(=O)(=O)C(F)(F)F)S(=O)(=O)C(F)(F)F	3.27
8	Ethyl(2-hydroxyethyl)dimethylammonium Bis(trifluoromethylsulfonyl)amide	[N+](C)(C)(CC)COC.[N-](S(=O)(=O)C(F)(F)F)S(=O)(=O)C(F)(F)F	3.70

- **Stereochemistry:** Encodes the spatial arrangement of the bond to capture chirality or stereoisomerism.

Every bond is depicted by a feature vector, and bidirectional edges are created between atom pairs to form an undirected graph structure, making it appropriate for processing by GNNs.

3) SMILES to Graph Conversion: The conversion of a SMILES string into a graph representation involves three essential steps:

- **Atom and Bond Feature Encoding:** Features at the atomic and bond levels obtained from the molecule are converted into PyTorch tensors. The node attribute matrix X is filled with atomic characteristics, while the edge attribute matrix E is populated with bond properties.
- **Edge Index Construction:** The connections between atoms (nodes) are represented through an edge index tensor that logs the source and target nodes for each bond.
- **Graph-Level Descriptors:** Descriptors that pertain to the entire molecule, including molecular weight (MolWt), partition coefficient (MolLogP), and topological polar surface area (TPSA), are calculated and included as global graph features. These descriptors offer a broader understanding of molecular characteristics that may not be entirely reflected by the relationships between atoms and bonds alone.

4) Graph Representation for GAT Input: The resulting molecular graph is represented as a `PyTorch Geometric Data` object, where:

- $\mathbf{X} \in \mathbb{R}^{N \times F}$: Represents the node feature matrix, with N nodes and F features per node.
- $\mathbf{E} \in \mathbb{R}^{M \times B}$: Contains the edge features, with M edges and B features per bond.

Graph-level attributes are added as extra inputs to the model to boost prediction performance. This method guarantees that the model is presented with a comprehensive, multi-dimensional depiction of molecular structures, which in turn enhances the precision of subsequent ILs toxicity predictions.

C. Data Splitting and Preparation

In order to set up the dataset for both training and evaluation, molecular structures represented as graphs were split into training and testing groups. This procedure guarantees that the model is trained on a segment of the data, while its effectiveness is assessed on unfamiliar samples to evaluate generalization and stability. Since the goal is to predict the toxicities of continuous ILs, a random splitting method was employed. The dataset was partitioned so that 80% of the samples were allocated for training, while 20% were reserved for testing. To maintain consistency in the outcomes across different runs, the random seed was set to random state=42. This random partitioning method preserves an unbiased distribution, facilitating a fair evaluation of the model's effectiveness.

The molecular structures, which were previously transformed from SMILES to graph representations, function as input features for the model, with the corresponding toxicity values of ILs extracted as labels. Each pair of graph and label constitutes a data point that is subsequently provided to the model during both training and testing

To facilitate batch processing for both training and evaluation, graph-label pairs are loaded into PyTorch Geometric's DataLoader. The training DataLoader processes data in batches of 64 samples and uses shuffling to randomize the order of graphs and labels for each epoch. This method aids in reducing overfitting and enhances generalization. Conversely, the test DataLoader maintains a consistent sequence by disabling shuffling to ensure a reliable performance evaluation.

The dataset was divided into a training set containing 284 samples and a test set with 71 samples. The specific allocation of samples is displayed during the data preparation phase to verify the division. This data preparation approach allows the model to leverage a varied collection of training samples while maintaining an impartial assessment of new data during the testing phase.

D. Graph Attention Network Architecture

Our approach is built upon the Graph Attention Network (GAT), a unique neural network designed to process graph-structured information. GATs build upon conventional GCNs by incorporating attention mechanisms at the graph's edges, which enable the model to attribute different levels of significance (or attention) to various atomic connections within the molecule. This attention mechanism allows the model to more effectively capture intricate relationships between atoms and their interactions, resulting in improved prediction accuracy. For an in-depth discussion of GATs, including their architecture and use in graph-based data, we direct readers to the study by [12].

1) Model Architecture Overview: The proposed EnhancedGAT architecture is made up of three graph

attention layers, followed by batch normalization and dropout techniques for regularization. The resulting graph representation is pooled and combined with molecular-level descriptors to create the input for the fully connected output layer.

Graph Attention Layers: Each GAT layer utilizes self-attention mechanisms to gather information from adjacent nodes. Specifically, for a node i with feature vector $\mathbf{h}_i \in \mathbb{R}^F$, and its neighboring nodes $j \in N(i)$, the attention mechanism calculates the revised representation as:

$$\mathbf{h}_i' = \sigma\left(\sum_{j \in N(i)} \alpha_{ij} \mathbf{W}\mathbf{h}_j\right) \tag{1}$$

Where:

- $\mathbf{W}$ is a learnable weight matrix.
- α_{ij} denotes the attention coefficient between nodes i and j, computed as:

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(\mathbf{a}^T\left[\mathbf{W}\mathbf{h}_i \,||\, \mathbf{W}\mathbf{h}_j\right]\right)\right)}{\sum_{k \in N(i)} \exp\left(\text{LeakyReLU}\left(\mathbf{a}^T\left[\mathbf{W}\mathbf{h}_i \,||\, \mathbf{W}\mathbf{h}_k\right]\right)\right)} \tag{2}$$

Where:

- $\mathbf{a}$ is a learnable vector.
- $||$ denotes concatenation.

This approach guarantees that the model appropriately weighs the significance of various neighbors, enabling the architecture to efficiently capture both local and global structural details.

2) Model Components: **Input Layer:** The input consists of features at the atom level for nodes and characteristics at the bond level obtained from the SMILES-to-graph transformation. Each feature vector for the atoms contains 8 dimensions, leading to an overall input dimension of:

$$\text{input\_dim} = 8 \tag{3}$$

Multi-Head Attention Mechanism: The model employs multi-head attention with four attention heads for each layer. By using multi-head attention, the learning process is stabilized and the model can gather varied information from adjacent nodes. The combined output generated by the attention heads is then normalized through batch normalization.

Three GAT Layers: The model consists of three sequential GAT layers that improve expressiveness:

- **GAT Layer 1:** Transforms the input atom features into hidden representations with 256 dimensions, using 4 attention heads, yielding a total output of 1024 dimensions.
- **GAT Layer 2:**Takes the output from the first layer while maintaining the same dimensionality for the hidden representations.
- **GAT Layer 3:** Merges the final node representations without using concatenation.

3) Graph-Level Pooling and Molecular Features: To achieve a representation at the graph level, the last node-level embeddings are combined through global mean pooling:

$$\mathbf{h}_G = \frac{1}{N}\sum_{i=1}^{N} \mathbf{h}_i \tag{4}$$

Where N is the number of nodes in the graph.

Molecular-Level Descriptors: To improve the model further, three molecular-level features (Molecular Weight, Partition Coefficient are combined with the graph-level and Topological Polar Surface Area) embeddings prior to being sent to the final output layer. The combined vector serves as the input to a fully connected linear layer that translates it to the required output dimension.

4) Fully Connected Output Layer: The output layer performs a linear transformation that translates the graph-level representation and molecular descriptors into the ultimate prediction:

$$\hat{y} = \mathbf{W}_{\text{out}}\mathbf{h}_G + b \tag{5}$$

Where:

- $\mathbf{W}_{\text{out}}$ is the weight matrix.
- b is the bias term.

5) Training and Optimization: **Loss Function:** The model was developed employing the mean squared error loss, which reduces the squared disparity between the predicted values and the actual results.

$$L = \frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{6}$$

Where:

- y_i is the ground truth value.
- $\hat{y}_i$ is the predicted value.

Optimizer and Learning Rate Scheduler: The model was optimized using the Adam optimizer, beginning with a learning rate of 0.001. A `ReduceLROnPlateau` scheduler was utilized to monitor the validation loss and will reduce the learning rate by 50% if no improvement in performance is seen for 5 consecutive epochs.

Early Stopping: To avoid overfitting, early stopping is implemented with a patience threshold of 10 epochs. During training, the model saves its best-performing version, which has the lowest validation loss, to `best_model.pt`. Table II illustrates a summary of the model parameters. This architecture successfully combines graph-level embeddings with molecular-level characteristics, improving the prediction of molecular properties while maintaining interpretability.

As presented in Figure 1, the computational graph of the completed EnhancedGAT model showcases the complex architecture utilized during inference. This visual representation complements the architectural overview shown in Figure 1, offering a clear understanding of the internal workings of the model and supporting the design decisions made during its development.

Table II: EnhancedGAT Model Architecture. The model consists of three GATConv layers with 4 attention heads, batch normalization applied after the first two layers, and a final fully connected linear layer.

Layer	Configuration
GATConv (gat1)	Input: 8, Output: 256, Heads: 4
BatchNorm1d (bn1)	Features: 1024, eps=1e-5, momentum=0.1
GATConv (gat2)	Input: 1024, Output: 256, Heads: 4
BatchNorm1d (bn2)	Features: 1024, eps=1e-5, momentum=0.1
GATConv (gat3)	Input: 1024, Output: 256, Heads: 4
Linear (fc)	Input: 259, Output: 1, bias=True

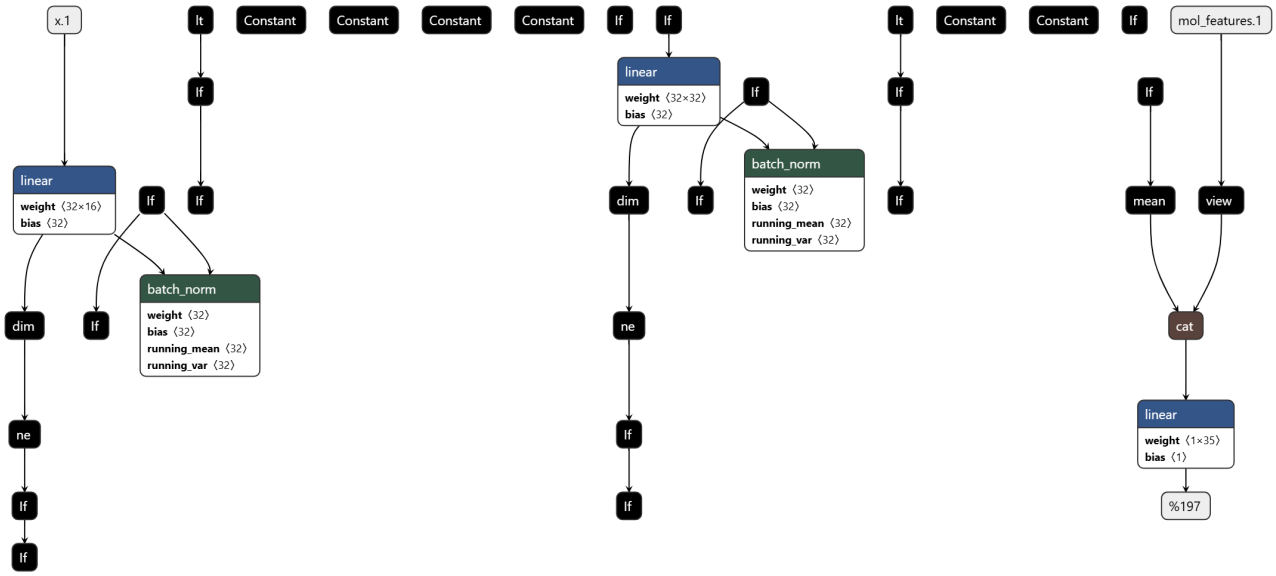

Figure 1: The computational graph of the EnhancedGAT model, presented using Netron, illustrates the flow of data throughout the network. It highlights the linear layers, batch normalization processes, transformations of dimensions, and the final phase where graph-level features are integrated with molecular descriptors prior to reaching the output layer. This visualization provides a clear insight into the internal architecture of the model and underscores its modular design.

E. Hyperparameter Optimization with Optuna

In order to enhance the performance of the proposed EnhancedGAT, hyperparameter optimization was conducted using Optuna [20], a sophisticated framework intended for automated tuning of hyperparameters. The objective was to minimize the validation error of the model, with a specific emphasis on the MSE, by systematically investigating a predefined search space.

The optimization process consisted of defining an objective function that assesses the model's performance for each set of hyperparameters. During each trial, the objective function instantiated a new version of EnhancedGAT with a selected set of hyperparameters, trained the model for 50 epochs while implementing early stopping after 10 epochs of no improvement.

The defined search space for the optimization included these hyperparameters:

- **Hidden Dimension (D_{hidden}):** Possible values are $\{64, 128, 256, 512\}$, which determine the size of the intermediate node embeddings following each GAT layer.
- **Number of Attention Heads (H):** An integer within the range of 2 to 8, which governs the number of independent attention mechanisms that operate in parallel to capture various types of neighborhood information.
- **Dropout Rate (p_{dropout}):** Samples are drawn uniformly within the range of 0.1 to 0.5, aiding in the reduction of overfitting by randomly eliminating certain node embeddings during the training phase.
- **Learning Rate (lr):** Values are logarithmically sampled between 1×10^{-4} and 1×10^{-2}, allowing for an extensive exploration of both low and high learning rate options.
- **Batch Size (B):** The potential values are $\{32, 64, 128\}$, which represent the number of graph samples that are processed simultaneously throughout the training.

A total of 30 experiments were performed in which Optuna explored various combinations of these hyperparameters, built and trained a model, and assessed its performance. The optimal hyperparameters identified after 30 trials were as follows:

- **Hidden Dimension:** 512

- **Number of Attention Heads:** 5
- **Dropout Rate:** 0.166
- **Learning Rate:** 0.00168
- **Batch Size:** 32

The final EnhancedGAT model was configured using these parameters, leading to better predictive performance. Utilizing Optuna for hyperparameter optimization greatly improved the model's ability to generalize among different molecular structures, resulting in increased stability, quicker convergence, and lower validation error.

F. Retraining with Optimized Hyperparameters

After identifying the most effective hyperparameters through the Optuna optimization process, the EnhancedGAT was relaunched with these optimal settings to enhance prediction accuracy. The selected parameters were utilized to establish the final model architecture and training framework, ensuring that the EnhancedGAT functioned under the best conditions for improved performance. Afterward, a new instance of EnhancedGAT was established using the optimized hyperparameters. The model underwent training with the Adam optimizer at a suitable learning rate of 0.00168 to reduce the MSE loss function. The choice of the Adam optimizer was due to its adaptive learning rate and momentum features, which enhance convergence speed and provide greater stability. The retraining process involved 100 epochs. By retraining the EnhancedGAT using optimal hyperparameters, the predictive capacity and reliability of the model were further improved, making it highly effective for the precise prediction of molecular properties.

III. Model Evaluation Metrics

The following metrics were used to evaluate the model's performance:

1. Mean Squared Error (MSE)

MSE measures the average of the squared differences between predicted and actual values:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{7}$$

Where:

- y_i is the actual value.
- $\hat{y}_i$ is the predicted value.

A lower MSE indicates better model performance.

2. Root Mean Squared Error (RMSE)

RMSE is the square root of MSE and provides an interpretable measure of error in the same unit as the target variable:

$$\text{RMSE} = \sqrt{\text{MSE}} \tag{8}$$

RMSE penalizes large errors more than small errors, making it useful for identifying significant deviations in predictions.

3. Mean Absolute Error (MAE)

MAE computes the average of the absolute differences between predicted and actual values:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{9}$$

MAE provides an easily interpretable measure of average error and is less sensitive to outliers compared to MSE.

4. R-Squared ($\mathbf{R^2}$) Score

The R^2 score, also known as the coefficient of determination, quantifies how well the predicted values explain the variance in the actual values:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \tag{10}$$

Where:

- $\overline{y}$ is the mean of the actual values.

An R^2 score closer to 1 indicates a strong correlation between the predictions and true values.

5. Pearson Correlation Coefficient (r)

The r measures the linear relationship between predicted and actual values:

$$r = \frac{\sum_{i=1}^{N}(y_i - \overline{y})\left(\hat{y}_i - \overline{\hat{y}}\right)}{\sqrt{\sum_{i=1}^{N}(y_i - \overline{y})^2 \sum_{i=1}^{N}\left(\hat{y}_i - \overline{\hat{y}}\right)^2}} \tag{11}$$

A Pearson correlation coefficient close to 1 signifies a robust positive relationship, while values nearing 0 indicate a lack of correlation [21]. When the target values do not change, the correlation is adjusted to zero to avoid computational complications. The use of different performance metrics allowed for a comprehensive evaluation of the model's predictive capabilities. MSE and RMSE measured the overall size of the error, MAE highlighted the average difference, and R^2, along with the Pearson correlation, assessed the strength of the relationship between predicted and actual values. This comprehensive evaluation framework guaranteed that the EnhancedGAT model underwent a thorough examination and validation with unseen data.

IV. RESULTS AND DISCUSSION

A. Model Performance Evaluation

The performance of the EnhancedGAT was evaluated using multiple metrics to ensure a comprehensive assessment of predictive accuracy. The model's results on the validation dataset are summarized in Table III.

Table III: Performance Metrics of the EnhancedGAT Model

Metric	Value
MSE	0.2145
RMSE	0.4632
MAE	0.3766
R^2 Score	0.8233
Pearson Correlation	0.9134

The model demonstrated a MSE of 0.2145 and a RMSE of 0.4632, as shown in Table III, indicating that the predicted values were generally accurate when compared to

the actual outcomes. RMSE serves as a more sensible metric for prediction error that aligns with the units of the target variable, since it is derived from taking the square root of MSE. The model effectively reduced prediction errors, which is crucial for dependable IL toxicity forecasts, as evidenced by the relatively low RMSE.

The MAE value of 0.3766 in Table III, shows the average magnitude of the prediction error. The low number suggests that the model consistently generated predictions with small errors because MAE is less impacted by significant discrepancies than MSE. This represents how the model may accurately estimate toxicity for a range of IL structures.

With an R^2 score of 0.8233, the predictions of the model can explain about 82.33% of the variance in the target values. The ability of the EnhancedGAT model to generalize is confirmed by a strong R^2 value, which is displayed in Table III and demonstrates that it effectively captures the basic correlations between structure and attributes in the ILs data.

The Pearson correlation coefficient of 0.9134, shown in Table III, signifies a very strong positive linear correlation between the predicted and actual values. A correlation value nearing 1 implies that the model's predictions are in close agreement with the actual values, further confirming its strength and dependability. This elevated correlation suggests that the EnhancedGAT model successfully understands intricate molecular interactions and converts them into precise predictions.

B. Comparison with Baseline Models

To further assess the effectiveness of the GAT model, we conducted a performance comparison with traditional machine learning techniques, such as Random Forest (RF) and Support Vector Machine (SVM), which are commonly used in toxicity prediction and QSAR modeling. RF has shown impressive results in the classification of compounds and property prediction due to its robustness and capability to handle high-dimensional datasets [22]. Likewise, SVMs have been effective in binary classification and regression tasks, particularly within cheminformatics [23]. These methods were chosen due to their interpretability and widespread historical use in QSAR modeling. The comparison results are presented in Table IV, highlighting the performance metrics of the EnhancedGAT model against RF and SVM.

Although the GAT model outperformed RF and SVM in terms of RMSE, R^2, and Pearson correlation, a more comprehensive comparison against recent deep learning-based models will be an important focus of future work to assess generalization performance more rigorously.

1) Residual Analysis: Residuals, which represent the difference between the observed and predicted values, were examined to evaluate how errors are distributed in the model. A Kernel Density Estimation (KDE) plot of the residuals (Figure 2) indicates that most residuals cluster around zero, signifying that the predictions of the model are unbiased. The shape of the residuals' distribution is fairly symmetrical, implying that prediction errors are distributed evenly throughout the dataset. The main important insights are:

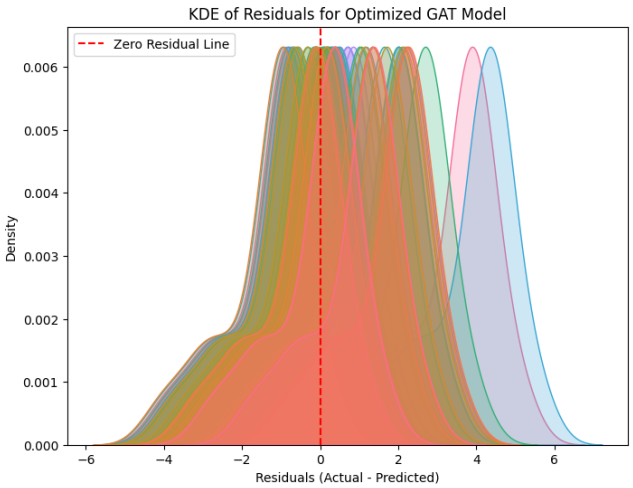

Figure 2: KDE of residuals for the optimized GAT model. Most residuals are clustered around zero, indicating minimal bias in predictions.

- The residuals show a distribution that is close to normal, suggesting that there is little to no systematic error present.
- The existence of a zero-residual line (indicated by the red dashed line) emphasizes the symmetry of the residuals, demonstrating that the deviations from the actual values are small and concentrated around zero.

C. Actual vs. Predicted Values

A scatter plot comparing actual and predicted values (Figure 3) provides a visual assessment of the model's predictive accuracy. The plot shows a strong linear relationship between the actual and predicted values, with the points closely following the ideal line ($y = x$).

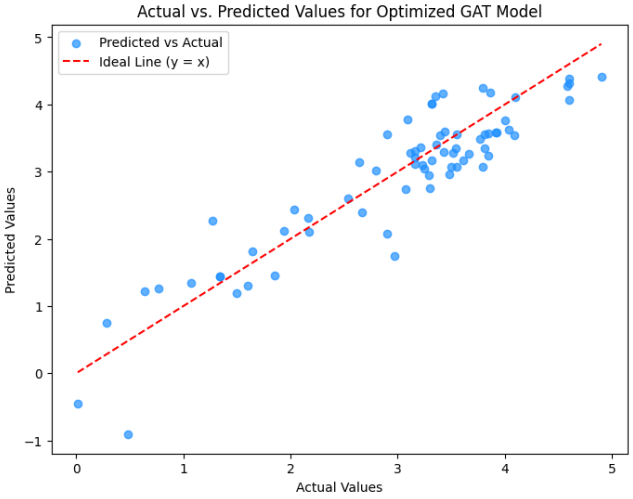

Figure 3: Actual vs. predicted values for the optimized GAT model. The red dashed line represents the ideal line ($y = x$), indicating perfect predictions.

Table IV: Comparison of EnhancedGAT, RF, and SVM models based on RMSE, R-squared (R^2), and Pearson Correlation metrics.

Model	RMSE	R^2	Pearson Correlation
EnhancedGAT (Proposed Model)	0.4632	0.8233	0.9134
Random Forest (RF)	0.6752	0.6540	0.8231
Support Vector Machine (SVM)	0.6928	0.6402	0.8119

Figure 3 is representing:

- Most of the predictions lie near the ideal line, confirming the model's ability to predict molecular properties with high precision.
- The scatter plot highlights minimal dispersion, indicating that the model effectively captures the underlying trends in the data.

D. Statistical Analysis of Model Predictions

To assess the dependability of the predictions in greater detail, a 95% confidence interval (CI) was established for the model's forecasts. The CI range of (2.6637, 3.1842) indicates that the true mean of the predictions is likely to be within this interval, providing strong evidence of the model's accuracy.

In addition, a paired t-test was performed to determine whether the difference between the actual and predicted values was statistically significant. The test yielded the following results:

- **T-statistic:** -1.4406
- **P-value:** 0.1541

Since the p-value is greater than the usual significance threshold of 0.05, we fail to reject the null hypothesis, suggesting that there is no statistically significant difference between the actual and predicted values. This result enhances the reliability of the model's predictions and suggests that the differences observed between the actual and predicted values are not statistically significant.

The interpretation of the statistical results is as follows:

- **Confidence Interval:** The 95% confidence interval indicates that the average predicted values are closely related to the true mean, thereby increasing the model's reliability.
- **t-Test Significance:** The absence of significance in the paired t-test further suggests that the differences between actual and predicted values are minimal, underscoring the stability of the model.

To summarize, the EnhancedGAT model, fine-tuned with optimal hyperparameters, exhibits exceptional predictive performance with low error rates, unbiased residuals, and statistically insignificant differences between observed and predicted values. These findings validate the effectiveness and dependability of the proposed model for forecasting IL toxicity.

V. Interpretability and Feature Importance

To improve the interpretability of enhanced GAT, Grad-CAM (gradient-weighted class activation mapping) was integrated to visualize the significance of the nodes and edges in the molecular graphs. Grad-CAM offers a post hoc explainability framework that emphasizes the role of specific nodes and edges in the prediction process, allowing for a deeper comprehension of the model's decision-making activities.

A. Model Interpretability with Grad-CAM

Grad-CAM [24] was utilized to assess the importance of individual nodes, emphasizing those that had the most significant impact on the model's predictions. The importance of each node was calculated by averaging the gradients of the target output with respect to the final GAT layer activations. The scores of importance obtained were normalized to allow for comparisons across various graphs. Key Observations include the following:

- The visual representations of node importance for three selected molecular graphs (Figure 4, Figure 5, and Figure 6) demonstrate that the nodes with the greatest contribution to predictions are typically situated in areas with high connectivity or near functional groups within the molecular structures.
- Nodes exhibiting higher importance scores are represented in red, indicating a more substantial influence on the model's decision, whereas those with lower scores are illustrated in blue, signifying a lesser impact.

1) Edge Importance Analysis: Beyond interpreting individual nodes, Grad-CAM has been adapted to evaluate edge significance by averaging the importance scores of interconnected nodes. This method assigns a score to each edge by considering the average importance of its neighboring nodes, facilitating the recognition of vital connections and associations within the molecular graph. Main observations are discussed as follows:

- Visualizations of edge importance demonstrate that the most critical edges generally correspond to bonds linking high-importance nodes.
- Edges that receive higher importance scores are illustrated with darker shades, signifying stronger connections that are essential in determining the molecular property.

2) Identification of Critical Substructures: Through the examination of node and edge significance across various graphs, crucial molecular substructures and areas that play a major role in the predictions were recognized. The top-ranking nodes and edges are frequently linked to chemically important substructures, like aromatic rings, functional groups, and heteroatoms, underscoring the model's capability to effectively encapsulate intricate molecular characteristics.

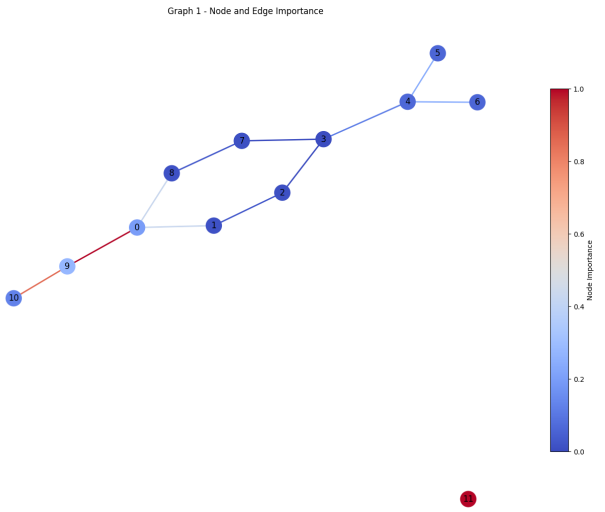

Figure 4: Graph 1 highlights nodes 11, 9, and 0 as the most influential nodes, with node 11 demonstrating the highest importance (importance = 1.0143).

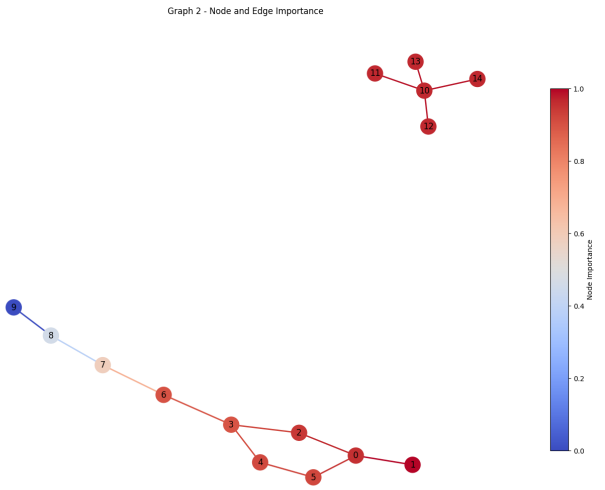

Figure 5: Graph 2 emphasizes nodes 1, 10, and 6 as the most critical nodes, indicating that these nodes contribute significantly to the model's prediction.

B. Permutation Feature Importance Analysis

A permutation feature importance analysis [25], [26] was conducted to assess how each distinct atomic-level feature contributes to the predictions made by the model, thereby enhancing the interpretability of the EnhancedGAT. This approach involves shuffling the values of each feature, thereby disrupting its connection to the target variable; the resulting increase in MSE is then calculated to evaluate the impact of each feature. An increased MSE following permutation suggests that the particular feature holds greater significance. The analysis was conducted in a three-step process:

1) **Initial MSE Measurement:** The model's original MSE was calculated using the dataset without any modifications.
2) **Feature Permutation:** The MSE was recalibrated after separately shuffling each feature.
3) **Importance Assessment:** The significance score for each feature was derived by comparing the permuted MSE to the baseline MSE.

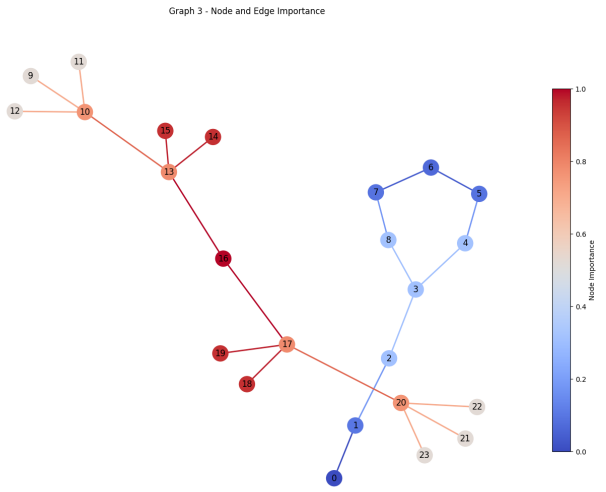

Figure 6: Graph 3 highlights nodes 16, 13, and 17, with node 16 playing the most pivotal role in the model's prediction.

The results of the permutation feature importance analysis are shown in Figure 7, which highlights the relative significance of each attribute.

The most important features were identified as:

- **H-bonds:** Changing the amount of hydrogen bonds increased the MSE the most, making this aspect the most crucial for the model's predictions.
- **Atomic Number:** The atomic number underscored the importance of elemental identity in shaping molecule properties, ranking as the second most significant attribute.
- **Charge:** The formal charge of the atoms places third in terms of relevance, underscoring the importance of electrostatic properties in affecting the desired attribute.

Conversely, features such as Aromaticity, Hybridization, and Normalized Molecular Weight (MW) exhibited minimal influence on the model's performance during permutation, indicating that these elements are less critical for the prediction task of the dataset.

The bar plot in Figure 7 presents a ranked visualization of feature importance. This visualization shows that hydrogen bonds, atomic number, and charge play a significant role in the model's predictions. The permutation feature significance analysis offers several crucial insights:

- **High Sensitivity to Hydrogen Bonding:** The importance of characteristics connected to hydrogen bonds is consistent with the information that IL harmfulness is essentially impacted by hydrogen bonding.

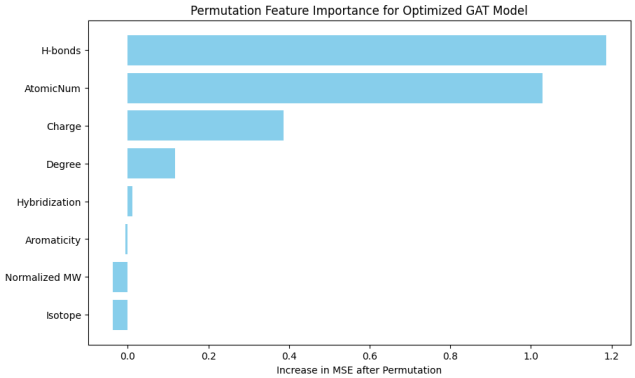

Figure 7: Permutation feature importance for the EnhancedGAT model. The length of each bar represents the increase in MSE after permuting the corresponding feature, indicating the feature's relative importance.

- **Significance of Atomic Number:** The importance of the atomic number is crucial because it highlights the model's reliance on elemental information to distinguish between atoms with different chemical properties.
- **Lower Contribution of Aromaticity and Hybridization:** Despite their chemical relevance, these traits may not significantly affect the toxicity within the dataset in question.

VI. Limitations and Future Work

Despite the promising results obtained with the EnhancedGAT for the prediction of IL toxicity, certain limitations remain that need to be addressed in future research. The main limitation of our model is the relatively small and narrow dataset. We'd love to add more diverse ILs to make it more generalizable. Also, while GATs do great with local interactions, they sometimes miss long-range effects.

Future steps might include:

- Building larger, more varied datasets
- Combining GATs with transformer models
- Improving interpretability tools
- Trying transfer learning or multitask learning for broader predictions
- Integrating hybrid models such as those that combine GATs with Message Passing Neural Networks (MPNNs), which offer a promising way to better capture long-range molecular interactions. By addressing both local and global patterns within molecular structures, these models could notably enhance the accuracy of toxicity predictions, particularly for complex ionic liquids (ILs).

VII. Conclusions

This study presents an adapted graph attention network (EnhancedGAT) for the prediction of IL toxicity using graph-based molecular representations. The model incorporates atom-, bond-, and molecule-level features to represent IL structures comprehensively and capture relevant interactions. This helps the system decide what parts are important for figuring out the toxicity. By leveraging multi-head attention mechanisms, the model prioritizes chemically relevant substructures that contribute significantly to toxicity. This not only makes it guess the toxicity better but also helps us understand how it made its guess. To check how good it is, we compared it with some older systems and used different evaluation measurements. Experimental results demonstrate that the EnhancedGAT achieved lower prediction errors and stronger correlations compared to traditional baselines, indicating high predictive reliability. To understand how the system made its decisions, we used special tools that show what parts of the ILs were most important in making the prediction. Despite these promising outcomes, certain limitations remain, particularly the limited diversity of the dataset and the model's ability to capture long-range molecular effects. How well it works sort of depends on the variety of ILs that it has seen before. And, it's a bit unreliable in figuring out how far away parts of the ILs affect each other. So, in the future, we should try systems that can fill these gaps, look at more types of ILs, and get clearer about how decisions are made. Including other ways of learning could make EnhancedGAT useful for different things. In conclusion, EnhancedGAT offers a scalable, interpretable, and accurate approach to IL toxicity prediction and serves as a promising framework for safer ionic liquid design. Future research will focus on expanding the dataset, improving long-range interaction modeling, and benchmarking against state-of-the-art deep learning methods.

Acknowledgments

This study was also supported by Izmir Katip Celebi University Scientific Research Council Agency, as project number 2024-TDR-FEBE-0024 for Safa Sadaghiyanfam's doctoral thesis studies.

Author Contributions

S.S. is a PhD student, Y.I. is her principal advisor, and H.K. is her co-advisor. H.K. and Y.I. contributed to the conception and design of the study. S.S. implemented all pattern recognition stages in Python. S.S. wrote the first draft of the manuscript, and all authors contributed to the reading, revision, and approval of it.

References

[1] N. V. Plechkova and K. R. Seddon, "Applications of ionic liquids in the chemical industry," *Chemical Society Reviews*, vol. 37, no. 1, pp. 123–150, 2008.

[2] G. A. Tabaaza, B. N. Tackie-Otoo, D. B. Zaini, D. A. Otchere, and B. Lal, "Application of machine learning models to predict cytotoxicity of ionic liquids using volsurf principal properties," *Computational Toxicology*, vol. 26, p. 100266, 2023.

[3] N. Abramenko, L. Kustov, L. Metelytsia, V. Kovalishyn, I. Tetko, and W. Peijnenburg, "A review of recent advances towards the development of qsar models for toxicity assessment of ionic liquids," *Journal of hazardous materials*, vol. 384, p. 121429, 2020.

[4] C. N. Cavasotto and V. Scardino, "Machine learning toxicity prediction: latest advances by toxicity end point," *ACS omega*, vol. 7, no. 51, pp. 47 536–47 546, 2022.

[5] A. Narin and Y. Isler, "Detection of new coronavirus disease from chest x-ray images using pre-trained convolutional neural networks," *Journal of the Faculty of Engineering and Architecture of Gazi UniversityKnowledge-Based Systems*, vol. 36214, pp. 2095–2 107 106 685, 2021.

[6] M. Surucu, Y. Isler, M. Perc, and R. Kara, "Convolutional neural networks predict the onset of paroxysmal atrial fibrillation: Theory and applicationsexplainability in deep reinforcement learning," *Chaos: An Interdisciplinary Journal of Nonlinear ScienceKnowledge-Based Systems*, vol. 31214, p. 11311906685, 2021.

[7] M. Surucu, Y. Isler, and R. Kara, "Diagnosis of paroxysmal atrial fibrillation from thirty-minute heart rate variability data using convolutional neural networks," *Turkish Journal of Electrical Engineering & Computer Science*, vol. 2914, pp. 2886–2 900 106 685, 2021.

[8] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in *International conference on machine learning*. PMLR, 2017, pp. 1263–1272.

[9] Z. Xiong, D. Wang, X. Liu, F. Zhong, X. Wan, X. Li, Z. Li, X. Luo, K. Chen, H. Jiang *et al.*, "Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism," *Journal of medicinal chemistry*, vol. 63, no. 16, pp. 8749–8760, 2019.

[10] L. Xu, T. Zhang, K. Lin, and Y. Wang, "Comparative evaluation of graph neural networks in molecular property prediction," *Biomolecules*, vol. 13, no. 2, p. 298, 2023.

[11] Y. Cai, H. Liu, and W. Chen, "Gat-based interpretable learning for biomedical molecular property prediction," *IEEE Journal of Biomedical and Health Informatics*, vol. 29, no. 1, p. 112–123, 2025.

[12] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio *et al.*, "Graph attention networks," *stat*, vol. 1050, no. 20, pp. 10–48 550, 2017.

[13] R. Ketkar, Y. Liu, H. Wang, and H. Tian, "A benchmark study of graph models for molecular acute toxicity prediction," *International Journal of Molecular Sciences*, vol. 24, no. 15, p. 11966, 2023.

[14] B. Chen, R. Barzilay, and T. Jaakkola, "Path-augmented graph transformer network," *arXiv preprint arXiv:1905.12712*, 2019.

[15] W. Ahmad, Z. Wang, X. Liu, F. Xu, B. Chang, and et al., "Chemberta: Large-scale self-supervised pretraining for molecular property prediction," *arXiv preprint*, vol. arXiv:2010.09885, 2021.

[16] G. Chen, Z. Song, Z. Qi, and K. Sundmacher, "Generalizing property prediction of ionic liquids from limited labeled data: a one-stop framework empowered by transfer learning," *Digital Discovery*, vol. 2, no. 3, pp. 591–601, 2023.

[17] S. Sadaghiyanfam, H. Kamberaj, and Y. Isler, "Enhanced prediction of ionic liquid toxicity using a meta-ensemble learning framework with data augmentation," *Artificial Intelligence Chemistry*, vol. 3, no. 1, p. 100087, 2025.

[18] S. Sadaghiyanfam, H. Kamberaj, and Y. Isler, "Leveraging chemberta and machine learning for accurate toxicity prediction of ionic liquids," *Journal of the Taiwan Institute of Chemical Engineers*, vol. 171, p. 106030, 2025.

[19] G. Landrum, "RDKit: Open-source cheminformatics," 2006–, http: //www.rdkit.org.

[20] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 2623–2631.

[21] Y. Isler, U. Ozturk, and E. Sayilgan, "A new sample reduction method for decreasing the running time of the k-nearest neighbors algorithm to diagnose patients with congestive heart failure: backward iterative elimination," *Sadhana*, vol. 48, p. 35, 2023.

[22] V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B. P. Feuston, "Random forest: a classification and regression tool for compound classification and qsar modeling," *Journal of chemical information and computer sciences*, vol. 43, no. 6, pp. 1947–1958, 2003.

[23] E. Byvatov, U. Fechner, J. Sadowski, and G. Schneider, "Comparison of support vector machine and artificial neural network systems for drug/nondrug classification," *Journal of chemical information and computer sciences*, vol. 43, no. 6, pp. 1882–1889, 2003.

[24] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[25] L. Breiman, "Random forests," *Machine learning*, vol. 45, pp. 5–32, 2001.

[26] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.