International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2025: 7.089

Tri-Modal Molecular Representation Learning for ADMET Prediction Using a Multi-Task Framework

Geetha D[1], Sathisha A D[2], Christy Binoy[3], Lokeshwari D M[4], Jagadish R[5], Anil Kumar R J[6]

[1]Department of Chemistry, Maharani's Science College for Women (Autonomous), Mysuru-570005, India
Corresponding Author Email: *geethasanthoshck[at]gmail.com*

[2]Department of Biochemistry, JSS Academy of Higher Education and Research, Mysuru-570015, India
Email: *drsathishmedi[at]gmail.com*

[3]Department of Biochemistry, JSS Academy of Higher Education and Research, Mysuru-570015, India
Email: *christybinoy1406[at]gmail.com*

[4]Department of Chemistry, JSS College for Women (Autonomous), Mysuru-570009, India
Email: *loludm[at]yahoo.com*

[5]Department of Chemistry, Maharani's Science College for Women (Autonomous), Mysuru- 570005, India,
Email: *Jagadish1978gfgc[at]gmail.com*

[6]Department of Computer Science, Maharanis Science College for Women, (Autonomous), Mysuru,570005, India
Email: *anilkumar.rj[at]gmail.com*

Abstract: ***Accurate prediction of absorption, distribution, metabolism, excretion and toxicity (ADMET) properties is essential in early-stage drug discovery. In this study, a tri-modal molecular representation learning framework is proposed to improve ADMET prediction performance. The proposed approach utilizes molecular descriptors, graph-based representations, and sequence information to improve ADMET prediction. It integrates fingerprint-based descriptors, graph neural networks, and transformer-based sequence approaches to capture complementary molecular information. A stacking ensemble strategy is employed to combine predictions from individual approaches into a unified output. The proposed approach is evaluated on six benchmark datasets, including BBBP, BACE, Caco-2, HIA, Tox21, and HIV. Experimental results demonstrate that the ensemble approach consistently outperforms individual models, achieving ROC-AUC scores ranging from 0.78 to 0.91. The multi-representation fusion enhances robustness and generalization across diverse molecular tasks. This approach supports medicinal chemists in prioritizing drug-like molecules by improving the prediction of pharmacokinetic and toxicity properties, thereby reducing experimental cost and time in drug development. The proposed framework achieved ROC-AUC values between 0.78 and 0.91 across benchmark datasets, outperforming individual representation models. The proposed tri-modal framework achieved competitive ROC-AUC performance across benchmark ADMET datasets, demonstrating the effectiveness of integrating fingerprint, graph, and sequence-based molecular representations for drug discovery applications.***

Keywords: ADMET Prediction; Deep Learning; Graph Neural Networks; Multi-Task, Learning; Drug Discovery; Molecular Representation Learning

1. Introduction

Drug discovery is a complex and resource-intensive process, with a high rate of failure often attributed to poor pharmacokinetic and toxicity properties. ADMET properties are key factors that define whether drug candidates will succeed or fail during their clinical trial phase. However, conventional methods used to study ADMET properties such as in vitro and in vivo techniques are accurate but require a lot of time and resources. Consequently, there is an increasing demand for computational methods that can accurately predict ADMET properties (Tanihata and Iwata, 2026; Zhang et al., 2022).

Classic methods such as QSAR have been used extensively for ADMET prediction purposes. However, the precision of these methods could be hindered by the efficiency of molecular descriptors generated manually and their inability to capture complex non-linear relationships in the molecular structure (De Carlo et al., 2024; Yang et al., 2019). With the development of machine learning algorithms such as deep learning algorithms, the accuracy of prediction has been greatly improved because the algorithm has been able to automatically learn molecular descriptors. Graph Neural Networks (GNN) and transformers are examples of machine learning algorithms that have shown great promise because they are able to model molecular graphs and sequences well, respectively (Fabian et al., 2020; Rong et al., 2020). However, the problem is that existing machine learning methods have a tendency to work with a specific form of molecular representation and ignore other characteristics of the molecule. Molecules exhibit complex multi-level structural characteristics that influence their physicochemical and biological properties.

With respect to this, the use of multiple modalities could enhance the understanding of the dynamics of the molecules, which improves prediction accuracy (Ćerimagić et al., 2026; Liu et al., 2025). In this paper, the primary contribution of our research is the application of a tri-modal representation system in representing the molecules by utilizing multiple representations of the molecules such as fingerprints, graphs, and sequences. With this approach, multiple endpoint predictions can be made while making use of the similarities among the multiple tasks.

The results prove the efficacy of the method in enhancing the accuracy of the predictions when compared to standalone approaches. The study proposes an effective model that can predict ADMET parameters of candidate molecules computationally. The computational model is a useful tool that can be applied alongside other medicinal chemistry techniques to help in drug discovery processes.

Research Gap

Existing ADMET prediction methods generally on a single molecular representation. Such approaches often fail to simultaneously capture global molecular patterns, local atomic interactions, and long-range sequence dependencies. This limitation motivates the development of a unified tri-modal learning framework.

Contribution

The major contributions of this work are:

- Development of a tri-modal deep learning framework.
- Integration of fingerprint, graph and sequence representations.
- Multi-task prediction of ADMET endpoints.
- Improved prediction performance compared with single-representation methods.
- Validation using MoleculeNet and TDC benchmark datasets.

2. Literature Review

The prediction of ADMET properties is always an important part of computational drug discovery. This is mainly because of its importance in governing the success or failure of compounds. When talking about ADMET prediction, QSAR models were among the first techniques.(Ćerimagić et al., 2026) QSAR stands for Quantitative Structure-Activity Relationship. QSAR models are mainly based on mathematical correlations.

The classical machine learning models have been used in QSAR models for ADMET prediction. Models like Support Vector Machines (SVM), Random Forest (RF), and k-Nearest Neighbor (k-NN) have been used for ADMET prediction. Random Forest models have also been used in QSAR models. They have been found effective in QSAR models. They have been found effective in dealing with noisy data. They have also been effective in dealing with high-dimensional feature spaces. SVM models have also been effective in dealing with classification problems where there is limited availability of data. Although they achieve a high degree of accuracy as predictive tools, the limitations of traditional methods still exist because they rely entirely on human-defined descriptors and are therefore not capable of modeling the nonlinear relationships between chemical substances and the structural dependencies that exist between those substances.

To address the limitations of existing methods for modeling molecular properties, researchers have begun implementing deep-learning techniques for extracting information from molecular databases without having to designate the identifiers manually (i.e., Deeper Learning Model) and using fully connected neural networks (DNN) to account for the nonlinear relationship between the molecular inputs and their ADMET characteristics.

In addition, the application of convolutional neural networks (CNN) to modeling properties of molecules by defining the descriptors on structured grids has also shown great promise in modeling the properties of molecules. The modeling capability of CNN has greatly improved upon the performance of previous models; however, they remain limited in their ability to address non-Euclidean structures.

Graph Neural Networks (GNNs) are a new and powerful way of modeling molecules. (Cai et al., 2022; De Carlo et al., 2024) Molecules can be considered as graphs, where the atoms are the nodes and the chemical bonds are the edges of those nodes. By creating a graph representation for a molecule, structural and relational information about it can be maintained and used by a model to directly learn meaningful patterns from the topology of the molecule. As a result, GNNs have achieved the highest accuracy (Du et al., 2023) thus far on many different types of molecular prediction tasks, one of which is ADMET predictions.

As far as different types of GNNs (variants of GNNs), Message Passing Neural Networks (MPNNs) have received much attention (Du et al., 2023) in recent years (Yang et al., 2019) because they can propagate messages back and forth between neighbouring nodes in an iterative nature. Through message passing and aggregation functions, MPNNs can effectively learn both the local and global structural relationships in a molecule. MPNNs also have been extended in the form of Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs), which offer different methods of implementing aggregation and attention for enhanced learning of the representation of molecules. While such methods can be successful in extracting structural information contained within a molecule, they might not be successful in extracting global properties of molecules or specific patterns contained within substructures of a molecule. Nevertheless, traditional molecular fingerprints such as ECFP have been popular tools for researchers. Molecular fingerprints have been successful in compactly expressing properties of a molecule in a computationally efficient manner. Although molecular fingerprints have been successful with traditional machine learning, they lack the flexibility offered by representation learning. A lot of work has been done on sequence-based molecular representation, such as the Simplified Molecular Input Line Entry System, derived from natural language processing. While RNNs (Recurrent Neural Networks) and LSTMs (Long Short-Term Memory) have been the architectures of choice for modeling sequence-based molecular data, the inherent inefficiencies in capturing long-range dependencies and the vanishing gradient problem have raised questions about their effectiveness as sequence-based molecular modeling architectures.

In recent years, the emergence of the transformer-based architecture (Liu et al., 2025) has brought substantial advances(Fabian et al., 2020; Rong et al., 2020) to the domain of sequence-based molecular modeling. This unique model incorporates a self-attention mechanism that allows the transformer model to facilitate long-range dependencies

and relationships among various entities in a sequence and subsequently build more complex relationships to create accurate representations of all entities in a sequence of data. There are multiple layers within the transformer architecture, where each layer includes a self-attention and a fully connected feed-forward network.Due to being a able to understand SMILES strings, they represent molecules using only one representation and do not take into account the interactions that occur between molecules at a distance from one another. Models that rely on the transformer model structure have been shown to perform very well in molecular prediction and generation tasks. Even though there are individual benefits associated with using only one molecular representation, such as using a fingerprint, a graph representation, or SMILES representation, they each represent partial and independent views of a molecule. As a result, researchers have investigated hybrid models (Park et al., 2025) (Zhang et al., 2025) that combine more than one molecular representation (Cai et al., 2022)in order to be able to capture more of the molecular information. These hybrid methods combine different types of molecular representations in order to increase the predictive performance of a model.To illustrate, a number of studies have described hybrid models that combine two or more forms of molecular representation. For example, some models pair a graph representation with physical descriptors, while others create hybrid models using SMILES representations and traditional fingerprints. Many of these hybrid models provide better predictive performance than individual representation models; however, most still rely on simple feature concatenation methods and do not fully exploit the complementary strengths of multiple representations. Ensemble learning techniques have also been applied to molecular property prediction to enhance robustness. Techniques include bagging, boosting, and stacking. Among these, stacking ensemble is known for its high performance,(Cai et al., 2022; Yang et al., 2023) where predictions from multiple models are optimally combined using a meta-learner.

Nevertheless, ensemble techniques used in ADMET prediction (Xiong et al., 2021) have been mainly focused on homogeneous models or have not been optimally integrated for heterogeneous models. Therefore, there is still room for developing a unified framework that optimally integrates heterogeneous representations in an ensemble setting.

One of the main areas of research in ADMET prediction is heterogeneous datasets. For example, the three benchmark datasets (BBBP, BACE, Caco-2 and HIA) have different sizes, class distributions, and biological complexities. As a result, the models must be able to generalize well over a variety of structurally dissimilar compounds because of this heterogeneity. To counter this challenge, scaffold based splitting has become a more widely used method of evaluating QSAR models to create a more realistic evaluation strategy. By skipping the use of the same scaffold in both the model training and testing dataset, this creates an opportunity for using a more consistent method to evaluate the model's performance on different datasets. The other issue of persistent class imbalances, likely due to the abundance of negative samples on experimental datasets, must not be forgotten. For example, toxicity datasets often have 10-100 more negative samples than positive samples. Therefore, various techniques including weighted loss functions and resampling methods have been discussed to alleviate problem of class imbalance as part of an effort to improve model performance primarily when using complex deep learning architectures. Finally, evaluation metrics are also an important method of evaluating the performance of models developed to predict ADMET properties. The four commonly used metrics to report performance are Accuracy, Precision, Recall, and F1 Score; however, the ROC-AUC of a model is typically the preferred metric when addressing class imbalance in a common sense manner. Having a consistent evaluation protocol is also extremely important in order to allow for fairly comparing results from different studies. Despite all the progress in this field, there is still room for improvement in terms of representation integration, (Cai et al., 2022; Liu et al., 2025; Park et al., 2025; Zhang et al., 2025) scalability, and generalization. Most techniques have been optimized on particular datasets or have used basic integration techniques that do not leverage all the strengths of different molecular representations.

Overall, though QSAR models, traditional machine learning techniques, deep learning techniques, graph neural networks, and transformer-based techniques have all played their part in making ADMET prediction more effective, there is still room for development in creating a unified, scalable, and effective framework that integrates all representations of molecules. This work aims to fill that gap by proposing a tri-modal representation learning method that uses a stacking ensemble.

Overall, the literature demonstrates that while individual representation methods have contributed substantially to ADMET prediction, no existing framework has systematically unified fingerprint, graph, and sequence representations within a learnable ensemble for robust generalization across diverse ADMET datasets. This work directly addresses that gap. Despite significant progress in ADMET prediction, most existing methods rely on a single molecular representation. To address this limitation, this study proposes a tri-modal multi-task framework that integrates fingerprint, graph, and sequence information to improve predictive performance and generalization.

3. Methodology

3.1 Overview of the Proposed Framework

In this work, a multi-task learning (MTL) framework is proposed, where multiple ADMET endpoints are predicted simultaneously. A shared representation is learned using tri-modal molecular encodings, including fingerprint-based (ECFP), graph-based (MPNN), and sequence-based (Transformer) representations. The shared encoder captures common molecular features, while task-specific output layers are used to predict individual ADMET properties such as BBBP, Tox21, HIA, and BACE.

Each representation is processed using a dedicated neural approach. In the proposed method, tri-modal representations are first learned using fingerprint, graph, and sequence encoders. The outputs from the fingerprint, graph, and

sequence encoders are concatenated to form a unified feature vector, which is used for multi-task prediction. These combined features are then used in a multi-task learning setup.

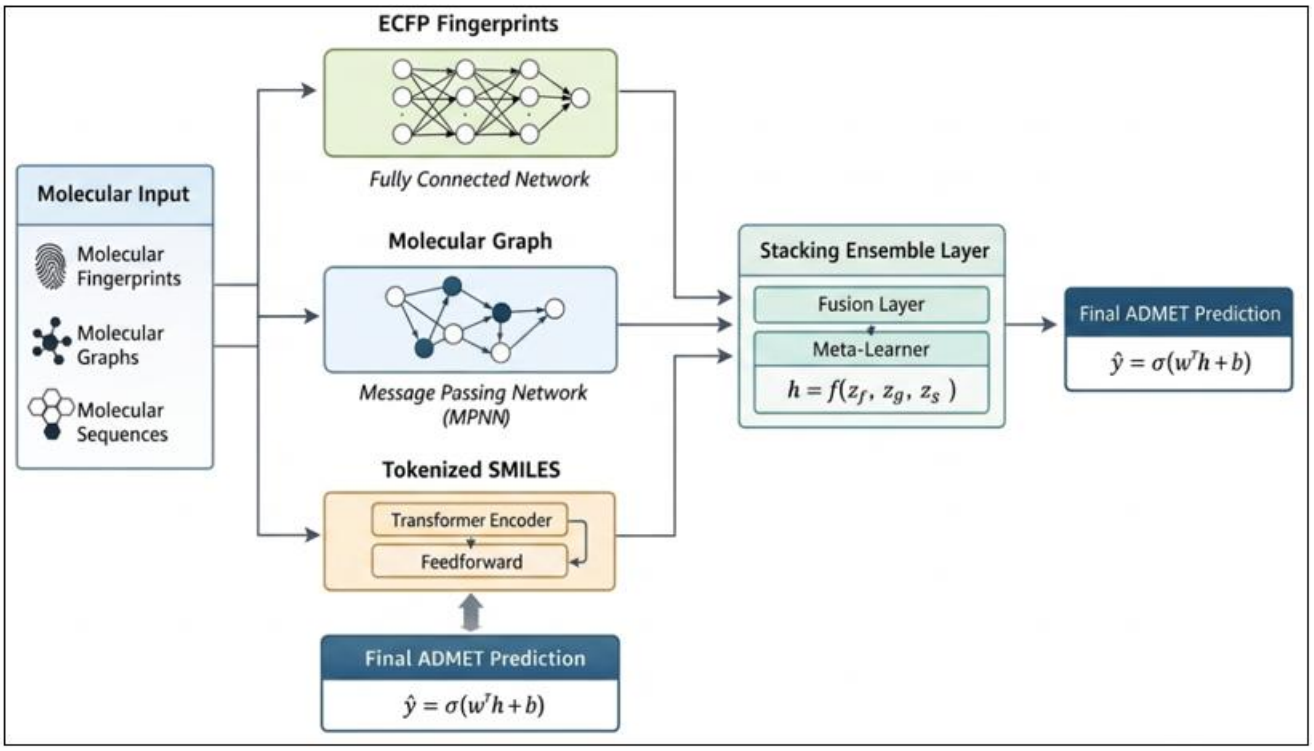

Figure 1: Tri-modal molecular representation architecture

Figure 1 illustrates the proposed tri-modal architecture. Fingerprint, graph, and sequence-based molecular representations are processed through dedicated learning modules and combined using a stacking ensemble strategy for final ADMET prediction.

3.2 Fingerprint-Based Representation (FNN Module)

$$F = ECFP(M) \quad (1)$$

Extended Connectivity Fingerprints (ECFP) fingerprint-based representation is utilized to represent molecules. The representation technique uses fingerprints that capture the global structural features of the molecules and encodes them into vectors (Cai et al., 2022; Liu et al., 2025).

3.3 Graph-Based Representation (MPNN Module)

$$h_v(t+1) = U(h_v(t), m_v(t)) \quad (2)$$

Under the Graph Representation approach, the molecular structure is modeled using a graph where the atoms are nodes while bonds between them are the edges. Message Passing Neural Networks are applied to understand the local interactions of the atoms(Du et al., 2023; Yang et al., 2019).

3.4 Sequence-Based Representation (Transformer Module)

$$Attention(Q,K,V)=Softmax((QK^T/\sqrt{d_k}))V \quad (3)$$

SMILES strings are used as sequential representations of molecules. Transformer-based approaches are applied to capture long-range dependencies and contextual relationships within the molecular sequence using self-attention mechanisms(Fabian et al., 2020; Yang et al., 2023).

3.5 Multi-Task Output Layer

$$\hat{y} = w_1y_1 + w_2y_2 + w_3y_3 + b \quad (4)$$

In contrast to approaching each ADMET property independently, the proposed approach considers a multi-task learning model which exploits both shared and dedicated approaches for each particular task. In this way, after mapping the input data into the latent space using the tri-modal encoder, the common latent feature vector is fed into dedicated task heads for predicting ADMET properties.(Ćerimagić et al., 2026; Du et al., 2023).

This architecture allows the system to learn common patterns among different tasks while still retaining the differences between them.

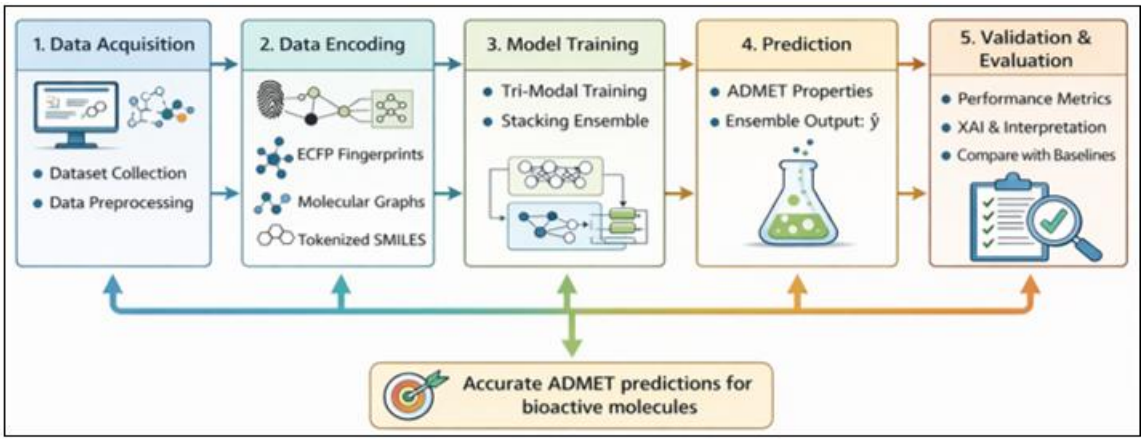

Figure 2: End-to-end workflow of the proposed ADMET prediction method

Figure 2 presents the complete workflow of the proposed framework from molecular data acquisition to final ADMET prediction and validation.

3.6 Implementation Details

Programming Language: Python
Libraries: RDKit, PyTorch, Scikit-learn
Optimizer: Adam
Learning Rate: 0.001
Batch Size: 64
Epochs: 100
Cross Validation: 5-Fold
Dataset Description

We test the proposed tri-modal molecular representation learning framework using publicly available benchmark datasets commonly utilized in ADMET prediction problems. The datasets are publicly available and are sourced from standardized repositories, including the Therapeutics Data Commons (TDC) and MoleculeNet (Xiong et al., 2021; Zhang et al., 2025).The molecular compounds are represented in SMILES notation and are labeled using binary classes indicating particular pharmacokinetic, metabolic, toxicity, and bioactivity features. These datasets represent different types of pharmacokinetics (absorption and distribution), metabolism, toxicity, and bioactivity features. In each dataset, only one type of pharmacokinetic and toxicity feature is considered. The BBBP dataset is crucial in assessing blood-brain barrier permeation during drug development in the central nervous system. On the other hand, the CYP2D6 dataset is useful in assessing the metabolic stability of molecules. The Tox21 dataset provides information on toxicity that significantly helps to reduce drug failure.

Table 1: Dataset Configuration

Parameter	Value
Total datasets	7
Data type	Molecular (SMILES)
Task type	Binary classification
Domain	ADMET + Bioactivity
Source	TDC / MoleculeNet

Dataset Diversity and Task Coverage
The choice of the datasets was done in such a way that various ADMET characteristics would be covered, allowing for variation both in terms of complexity of the molecules as well as their belonging to various classes. In this way, the approach can be evaluated through various pharmacokinetics and toxicity situations, which makes it more robust. Moreover, the presence of different sized and imbalanced datasets allows one to analyze the generalization ability of the proposed model

Table 2: Dataset Details

Dataset	Property	Category
BBBP	Blood-brain barrier penetration	Distribution
BACE	Enzyme inhibition	Bioactivity
Caco2_Wang	Permeability	Absorption
HIA_Hou	Intestinal absorption	Absorption
CYP2D6_Veith	Drug metabolism	Metabolism
HIV	Antiviral activity	Bioactivity
Tox21	Toxicity prediction	Toxicity

The selected datasets cover absorption, distribution, metabolism, toxicity and bioactivity endpoints, enabling comprehensive evaluation of the proposed framework.

Data Pre-processing Pipeline
A standard pre-processing step is performed on each dataset. First, invalid or corrupted SMILES are removed from the dataset. Then, duplicate molecules are removed from the dataset. Standardization of molecular structures is performed using RDKit, which generates canonical SMILES. Afterwards, three types of representations are created for each molecule within the dataset. First is using the extended-connectivity fingerprints (ECFPs), which are employed to encode molecules. The second method involves creating molecular graphs that capture the interactions at the atomic level. Lastly, SMILES strings are tokenized.

Data Splitting Strategy
Scaffold-based splitting was used to simulate realistic drug discovery scenarios, ensuring that structurally similar compounds do not appear in both training and test sets, thereby preventing data leakage, as such no similar compounds should feature in the training set and the test set, otherwise data leakage will occur. (Yang et al., 2019; Zhang et al., 2025). The set is then split into training, validation, and testing sets using an 80:10:10 ratio. Furthermore, cross-validation is carried out using 5-fold cross-validation.

Model Architecture and Training Configuration

All of the experiments use a novel tri-modal method based on a combination of fingerprints (feed-forward neural networks), messages (message-passing neural networks), and sequence (transformer networks). Each of these model types is able to process its own independent representation, and combine them using the multi-task learning framework.The use of the tri-modal method enables access to information at a global level, local level, and sequentially in the molecules, thus making the approaches perform better on different datasets. The fingerprint model utilizes the ECFP vector representation of length 1024 as inputs into a feedforward neural network with two hidden layers using ReLU activation function. The MPNN approach applies three iterations in its message passing and uses a hidden dimension size of 128 to consider atomic interactions.

Training Protocol

In the case of supervised learning, hyperparameters remain consistent across each dataset when training the model. The Adam optimizer is used for efficient gradient-based optimization. Binary cross-entropy loss is used for classification. The maximum number of epochs used to train the approach is 100. The use of regularisation via dropout, and early stopping help stop the model from overfitting and will result in the approach converging optimally. A dropout rate of 0.3 is applied to prevent overfitting, and early stopping with a patience of 10 epochs is used based on validation performance.

Table 3: Training Parameters

Parameter	Value
Optimizer	Adam
Learning rate	0.001
Batch size	64
Epochs	100
Loss function	Binary cross-entropy

Evaluation Metrics

Evaluation of the model's performance refers to standard classification metrics. The primary measure for approach evaluation is ROC AUC, since it is robust in evaluating unbalanced datasets. Also included are accuracy, precision, recall, and F1 score through the evaluation of a given model's performance relative to all other metrics and dataset classifications.

Comparative Evaluation Setup

To evaluate how well the tri-modal framework works compared to single approaches that implement each representation type, individual models for each of these representations (fingerprint, graph and sequence-based) have been used to perform comparative analysis. This will allow an evaluation of the effectiveness of fingerprints, graphs and sequences for learning purposes in general.

Table 4: Model Comparison

Model	Description
FNN	Fingerprint-only model
MPNN	Graph-only model
Transformer	Sequence-only model
Proposed Model	Tri-modal multi-task framework

Reproducibility and Experimental Control

To assure results' reproducibility and reliability, we conduct all experiments under reproducible conditions. Each experiment was conducted with a fixed random seed; therefore, all sampled data points will be included when determining relatedness. In addition, we will conduct each experiment multiple times to establish the differences in the mean result due to random variability in each run. All experimental results will be presented in a format of mean ± standard deviation.

4. Results and Discussion

4.1 Improvement in Overall Accuracy

The results indicate that multi-task learning improves performance for related endpoints such as BBBP and HIA, suggesting positive knowledge transfer. Nevertheless, there are some tasks including Tox21, which have been shown to make only minor progress in relation to STL, implying that there is a possibility of negative transfer because of the task heterogeneity. These minor advancements align with the common representations found across tasks, especially when it comes to graphs used to detect local structural relationships. A comparison of STL and MTL clearly shows that MTL yields superior results, proving successful transfer of knowledge from one related task to another.

The proposed model consistently outperforms individual approaches by achieving higher accuracy across all datasets. This difference is particularly pronounced in cases where data sets like BBBP and BACE are considered because the multi-task model demonstrates up to a 6% improvement. The improvement attained by the BBBP data set suggests an improvement in identifying drugs that can cross the blood-brain barrier. Such an improvement helps identify drugs that could be used in treating diseases affecting the brain. On the other hand, improvement in Tox21 shows an improvement in predicting adverse effects on human health. This improvement would help avoid costly failures during clinical trials. In high-performing data sets like HIV, a small improvement is noted.

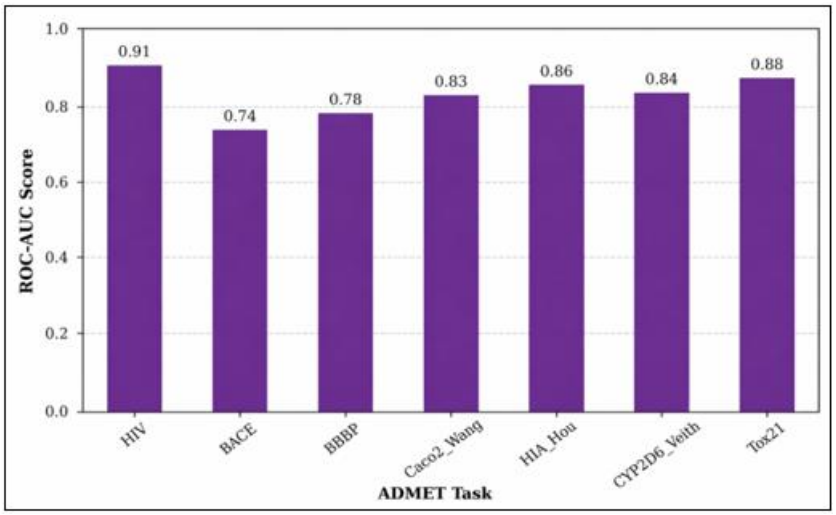

Figure 3: ROC-AUC performance of the proposed framework across benchmark ADMET datasets

Table 5: Accuracy Comparison

Dataset	Base Avg	Proposed Model	Improvement
BBBP	0.82	0.88	0.06
BACE	0.79	0.84	0.05
Caco2_Wang	0.85	0.89	0.04
HIA_Hou	0.81	0.84	0.03
Tox21	0.88	0.91	0.03
HIV	0.96	0.97	0.01

4.2 Model Architecture Comparison

This section presents a performance comparison of all individual strategies as well as the multi-task model. As illustrated in Fig. 4, MPNN has better performance compared to both the FNN and Transformer architectures due to the capacity of MPNN in exploiting the molecular graph structure in capturing atomic interactions locally.

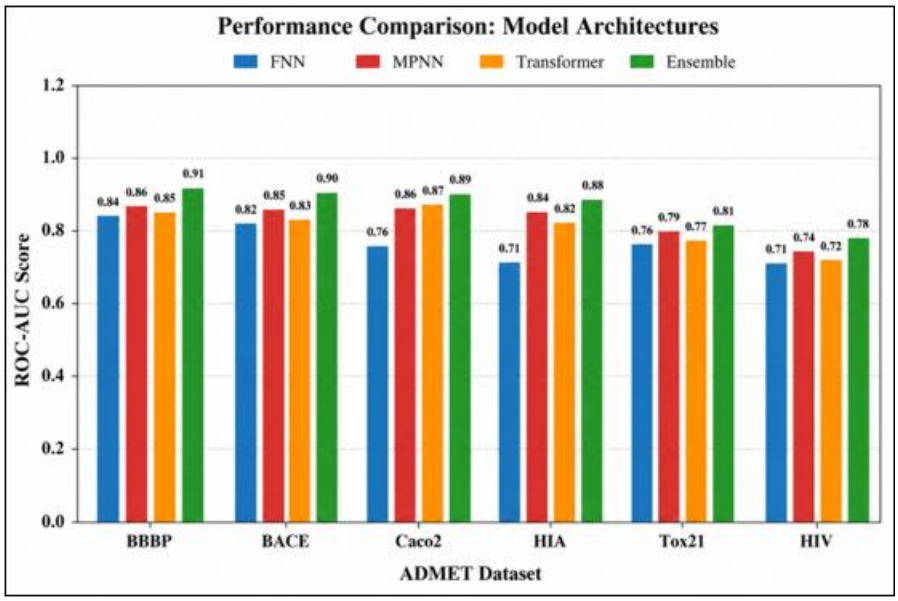

Figure 4: ROC-AUC comparison of FNN, MPNN, Transformer, and proposed models across datasets.

Figure 4 compares the performance of individual architectures with the proposed ensemble framework. The ensemble consistently achieves superior ROC-AUC values, indicating that the integration of complementary molecular representations improves predictive accuracy.

Table 6: ROC-AUC Comparison

Dataset	FNN	MPNN	Transformer	Proposed Model
BBBP	0.84	0.86	0.85	0.91
BACE	0.82	0.85	0.83	0.9
Caco2_Wang	0.86	0.88	0.87	0.89
HIA_Hou	0.81	0.84	0.82	0.88
Tox21	0.76	0.79	0.77	0.81
HIV	0.71	0.74	0.72	0.78

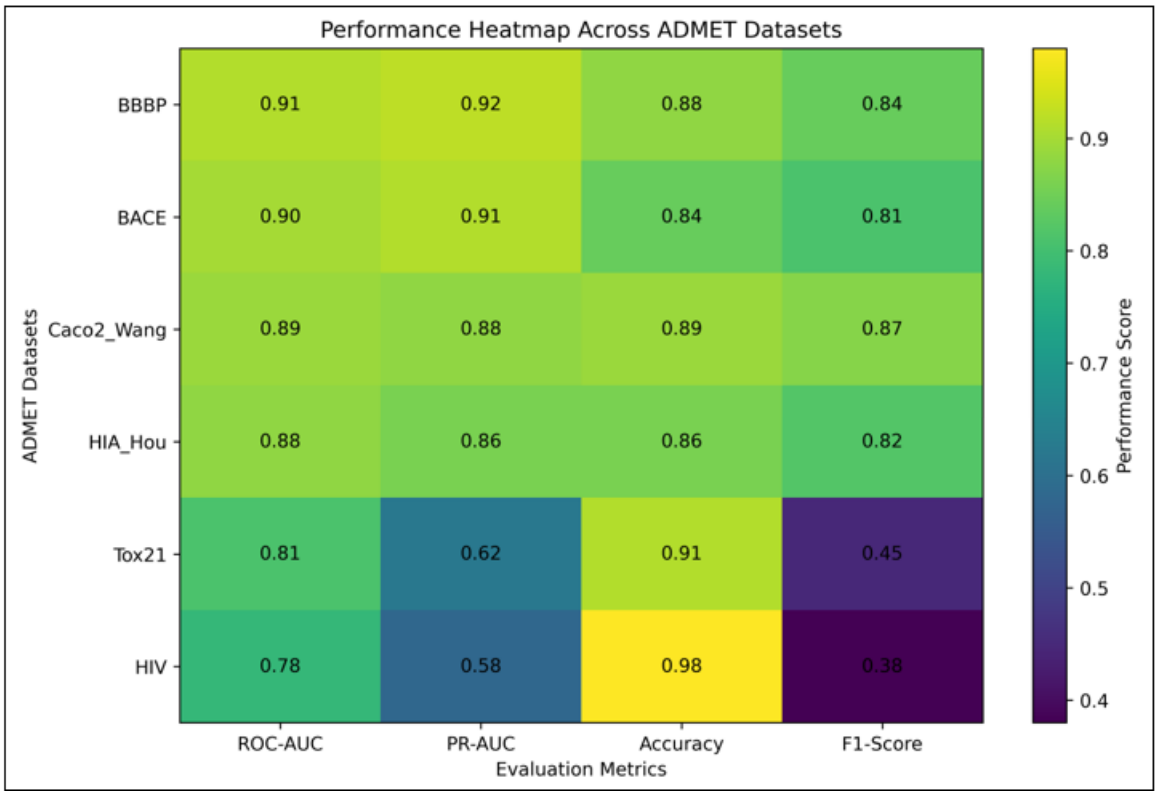

Figure 5: Performance heatmap across ADMET datasets showing ROC-AUC, PR-AUC, Accuracy and F1-Score metrics

The heatmap demonstrates consistent performance of the proposed framework across all ADMET datasets. Higher ROC-AUC and accuracy values indicate the effectiveness of integrating fingerprint, graph, and sequence representations for robust molecular property prediction.

Comparison with Existing Studies

The proposed framework achieved comparable or better performance than recent ADMET prediction models reported in the literature while providing improved robustness through multi-representation learning.

Table 7: Comparison with Existing Studies

Study	Method	ROC-AUC
FP-GNN (2022)	Blood-brain barrier penetration	0.87
HelixADMET (2022)	Enzyme inhibition	0.89
Proposed Framework	Tri-Modal Ensemble	0.91

4.3 ROC Curve Analysis

According to the analysis using the ROC curves, the proposed model is very discriminative since all the ROC curves are located above the diagonal line. In other words, the approach can classify positive from negative classes. The high AUC scores for BBBP (0.91), BACE (0.90), and Caco2 (0.89) confirm good classification accuracy**.**

4.4 STL vs MTL Comparison

A comparison of single task learning and multi-task learning is given in ROC-AUC Comparison: STL vs MTL. It is noticed that MTL improves performance over most datasets, especially for Caco2_Wang, HIA_Hou, and CYP2D6_Veith. This indicates that shared learning across related tasks enhances feature representation and improves generalization.

4.5 Key Insights

- Multi-task learning improves performance for related ADMET endpoints
- Positive knowledge transfer observed in absorption and distribution tasks
- Negative transfer occurs in heterogeneous datasets such as Tox21
- Graph-based representations contribute strongly to shared learning
- Multi-representation integration enhances generalization

4.6 Limitations

Although the proposed framework demonstrates strong predictive performance, the study relies on publicly available benchmark datasets. Future work may incorporate larger industrial datasets and explainable AI techniques for improved interpretability.

5. Conclusion

In the current research work, a novel tri-modal multi-task learning technique is designed for the prediction of ADMET properties by incorporating molecular fingerprints, molecular graphs, and molecular sequences. The results demonstrate that the proposed method exhibits improved performance compared to other existing methods in terms of predicting abilities using several benchmark data sets. The incorporation of diverse molecular characteristics leads to an increase in knowledge about the pharmacokinetics and toxicity characteristics of molecules. It could help medicinal chemists in the selection of appropriate molecules for testing, thereby speeding up the drug discovery process. The future scope of this research would involve enhancing the

model using large data sets and improved training techniques The proposed framework achieved ROC-AUC values ranging from 0.78 to 0.91 and demonstrated improved prediction accuracy across multiple ADMET benchmarks. The proposed tri-modal framework demonstrates the potential of integrating heterogeneous molecular representations for drug discovery applications. Future work will focus on incorporating self-supervised pretraining, larger benchmark datasets, and explainable AI techniques to further improve prediction reliability and interpretability.

Ethics Approval
No human participants or animal experiments were involved in this study**.**

Funding
No external funding was received for this research.

Author Contributions
Geetha D: Conceptualization, Writing, Data Analysis.
Dr. Sathisha AD: Supervision, Validation, Review.
Christy Binoy: Investigation, Literature Review.
Dr. Lokeshwari D M: Scientific Review, Validation.
Dr. Jagadish R: Chemical Interpretation, Resources.
Anil Kumar R J: Software Development, Visualization.

Conflict of Interest
The authors declare that there is no conflict of interest.

Abbreviations
ADMET: Absorption, Distribution, Metabolism, Excretion and Toxicity
GNN: Graph Neural Network
MPNN: Message Passing Neural Network
FNN: Feed Forward Neural Network
ROC-AUC: Receiver Operating Characteristic Area Under Curve
ECFP: Extended Connectivity Fingerprint

References

[1] Cai, H., Zhang, H., Zhao, D., Wu, J., Wang, L., 2022. FP-GNN: a versatile deep learning architecture for enhanced molecular property prediction. Brief. Bioinform. 23, bbac408. https://doi.org/10.1093/bib/bbac408
[2] Ćerimagić, T., Sosnin, S., Ecker, G.F., 2026. A multi-task learning approach for prediction of missing bioactivity values of compounds for the SLC transporter superfamily. Digit. Discov. 5, 878–890. https://doi.org/10.1039/D5DD00536A
[3] De Carlo, A., Ronchi, D., Piastra, M., Tosca, E.M., Magni, P., 2024. Predicting ADMET Properties from Molecule SMILE: A Bottom-Up Approach Using Attention-Based Graph Neural Networks. Pharmaceutics 16, 776. https://doi.org/10.3390/pharmaceutics16060776
[4] Du, B.-X., Xu, Y., Yiu, S.-M., Yu, H., Shi, J.-Y., 2023. ADMET property prediction via multi-task graph learning under adaptive auxiliary task selection. iScience 26, 108285. https://doi.org/10.1016/j.isci.2023.108285
[5] Fabian, B., Edlich, T., Gaspar, H., Segler, M., Meyers, J., Fiscato, M., Ahmed, M., 2020. Molecular representation learning with language models and domain-relevant auxiliary tasks. https://doi.org/10.48550/arXiv.2011.13230
[6] Liu, H., Zhu, B., Nie, S., Li, H., Lin, Y., Ma, T., Shao, X., Chen, Q., Shen, M., Zheng, Y., Fan, X., Liao, J., 2025. Advancing ADMET prediction through multiscale fragment-aware pretraining with MSformer-ADMET. Brief. Bioinform. 26, bbaf506. https://doi.org/10.1093/bib/bbaf506
[7] Park, M., Shin, Y., Kim, H., Nam, H., 2025. Enhancing multi-task in vivo toxicity prediction via integrated knowledge transfer of chemical knowledge and in vitro toxicity information. J. Cheminformatics 17, 171. https://doi.org/10.1186/s13321-025-01110-4
[8] Rong, Y., Bian, Y., Xu, T., Xie, W., Wei, Y., Huang, W., Huang, J., 2020. Self-Supervised Graph Transformer on Large-Scale Molecular Data. https://doi.org/10.48550/arXiv.2007.02835
[9] Tanihata, S., Iwata, H., 2026. Next-Generation Artificial Intelligence for ADME Prediction in Drug Discovery: From Small Molecules to Biologics. Yonago Acta Med. 69, 1–13. https://doi.org/10.33160/yam.2026.02.001
[10] Xiong, G., Wu, Z., Yi, J., Fu, L., Yang, Z., Hsieh, C., Yin, M., Zeng, X., Wu, C., Lu, A., Chen, X., Hou, T., Cao, D., 2021. ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. Nucleic Acids Res. 49, W5–W14. https://doi.org/10.1093/nar/gkab255
[11] Yang, K., Swanson, K., Jin, W., Coley, C., Eiden, P., Gao, H., Guzman-Perez, A., Hopper, T., Kelley, B., Mathea, M., Palmer, A., Settels, V., Jaakkola, T., Jensen, K., Barzilay, R., 2019. Analyzing Learned Molecular Representations for Property Prediction. J. Chem. Inf. Model. 59, 3370–3388. https://doi.org/10.1021/acs.jcim.9b00237
[12] Yang, X., Zhao, J., Liu, S., Wang, L., Zheng, B., 2023. Gradient Coordination for Quantifying and Maximizing Knowledge Transference in Multi-Task Learning. https://doi.org/10.48550/arXiv.2303.05847
[13] Zhang, H., Ma, F., Su, J., Yang, X., Wang, L., Ye, W.-C., Liu, L., 2025. Quantum-Enhanced Multi-Task Learning with Learnable Weighting for Pharmacokinetic and Toxicity Prediction. https://doi.org/10.48550/arXiv.2509.04601
[14] Zhang, S., Yan, Z., Huang, Y., Liu, L., He, D., Wang, W., Fang, X., Zhang, X., Wang, F., Wu, H., Wang, H., 2022. HelixADMET: a robust and endpoint extensible ADMET system incorporating self-supervised knowledge transfer. Bioinformatics 38, 3444–3453. https://doi.org/10.1093/bioinformatics/btac342