IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

Drug Property Prediction Using Graphical Neural Networks

[1]P. Jashwanth Reddy, [2]M. Manisha, [3]K. Venkateshwarlu and [4]Mr P. Ramu
Dept of CSE, UG Students[1-3]
[4]Assistant Professor, Dept of CSE[4]
Sreenidhi Institute of Science and Technology (SNIST), Hyderabad
[1]23315a0529@cse.sreenidhi.edu.in, [2]23315a0531@cse.sreenidhi.edu.in,
[3]23315a0528@cse.sreenidhi.edu.in, [4]pasamramu39@gmail.com

Abstract: *The identification of effective drug compounds is considered an intricate and labor-intensive process in the domain of pharmaceutical research. The experimental method of testing and evaluating the effectiveness of chemical compounds is considered a laborintensive process. However, advent of and information, the use computerbased tools has come up as an effective way to speed up the initial stages of drug research. The research paper proposes an approach to predict the properties of effective drug compounds by utilizing the Graph Neural Network. The proposed approach utilizes the graphical representation of the molecular structure of the compounds. This approach helps the deep learning method learn the relationships between the atoms of the compounds and identify the features of the compounds. The proposed framework uses the available dataset of the compounds. system predicting the the effective compounds. Experiments were conducted using the proposed framework. The experiments proved that the proposed system is effective in learning the structures of the compounds using the graph-based method.*

Keywords: Drug property prediction, Graph Neural Networks, molecular graphs, deep learning, computational drug discovery, bioinformatics

I. INTRODUCTION

importance of in modern healthcare is significant, as it allows for the development of drugs to cure various diseases. The process of discovering compounds that can interact with a biological system in a safe and efficient manner is a complex problem in science. Traditionally, various experiments are carried out on compounds identified for drug development. Although these methods are reliable, they are associated with considerable financial costs. Recently, computer methods have emerged as an important feature in modern drug research. Using machine learning methods, molecular data can be analyzed to determine possible efficacy without conducting costly experiments. Traditionally, machine learning methods rely on molecular descriptors to analyze data. These descriptors identify certain features of a molecule. However, these descriptors are insufficient to identify the complex relationship between various structures in a molecule.

The recent developments in deep learning have also introduced novel techniques in the analysis of complex structured data.. Such nodes and edges help GNNs in understanding the interaction between the atoms in the molecule and the patterns associated with the molecule's behavior. A novel system has been proposed in this research using GNNs with the aim of predicting essential drug characteristics.

II. RELATED WORK

Several research studies have been conducted on the application of graph-based deep learning approaches in molecular structure analysis.

Duvenaud et al. carried out pioneering research on the to graph- based molecular data.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

The research introduced the concept of using neural networks to learn chemical features from molecular structures. Gilmer et al. carried out research on the application of a framework called Message Passing Neural Networks to graphbased molecular data. The framework allows nodes in a molecular graph to pass information to other nodes or atoms in the structure.

For improving graph neural networks, Graph Attention Networks have been proposed.

Benchmark studies using various datasets, including the MoleculeNet dataset, have shown graph neural networks' potential in predicting drug-related properties competitively or even better than traditional machine learning methods.

However, further research is necessary in developing an integrated system with a user interface for drug discovery applications using graph analysis of molecules. The system proposed in this work aims at contributing to the development of a GNN-based drug property prediction framework.

III. DATASET AND DATA PREPARATION

The data set that was utilized in this research consists of molecular compounds that were retrieved from publicly accessible databases. These data sets have information on molecular structures, as well as related pharmacological properties, including toxicity, solubility, and bioactivity.

It that represents the molecular structures, including the composition and the bonds that are a part of it. In order to implement the graph neural network, it is necessary to transform it into a graph form.

The data pre-processing includes several stages. At first, the SMILES representations have to be parsed in order to validate the molecular structures. This also includes obtaining information related to atomic compositions.

Secondly, molecular graphs have to be constructed, with atoms and chemical bonds being represented in the form related features.

IV. PROPOSED METHODOLOGY

The proposed drug property prediction system utilizes a graph neural network to learn the the drug molecules and predict their pharmacological properties.

The first step in the drug property prediction process using is the representation molecule. Each molecule in the drug property prediction process

The During the message passing phase, nodes are able to share information with their respective neighboring nodes. This is an iterative process that allows the model to update its respective nodes based on the chemical environment surrounding them. Once this process is complete, a graph readout function is applied to aggregate information from individual nodes to obtain a single vector representing the entire molecule.

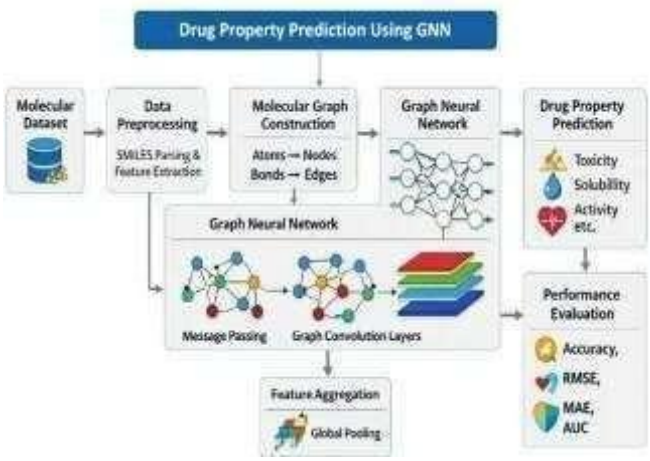


Fig. 1. Methodology Flow Diagram

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

The generated molecular representation is then passed through several layers of fully connected neural networks to carry out the actual prediction task. The output layer is decided depending upon the property to be predicted. The parameters are optimized using gradient-based learning algorithms such as Adam.

V. SYSTEM ARCHITECTURE

proposed drug property prediction model includes different stages of processing, as discussed below:
The first stage of the drug property prediction model is the data input stage, where !After the completion of the data input stage, the next stage processing cheminformatics tools. After the completion of the processing of the molecular structure of different molecules, the system needs to obtain different types of information at the atomic and bond levels.
The second stage of the drug property prediction model includes the graph construction stage, where different types of molecular structures are given as graphs. In the proposed model, the different atoms of different molecules and the different bonds of different molecules are represented as nodes and edges, respectively. In addition, different types of information are given at the atomic and bond levels. The third stage of the drug property prediction model includes the Graph Neural Network layers. These layers are used to perform the message passing operations.
layers are used to enable the system to learn the relationships of the molecules. The message passing operation is performed using the Graph Neural Network. Finally, the graph pooling or readout operation is performed. This process involves the repeated use of the graph convolution operation. This results in a fixed-size vector that represents the entire molecule. This fixed-size vector represents the features of the molecule. These features include the local interactions between the atoms, as well as the global features. After this, the fully connected neural network layers are used. These are used to make the predictions for the important drug properties.

VI. MODEL TRAINING AND IMPLEMENTATION

Finally, the model, after being trained, is used to make the predictions for the unknown molecular compounds. As a result, the properties of the molecular compounds, which may include toxicity, solubility, and bioactivity, can be estimated. Additionally, the performance of the proposed model is evaluated by using the most important performance metrics. These include Accuracy, Area Under the Curve (AUC), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), etc. The proposed system for predicting the drug property has been designed with an architecture that includes processing

VII. RESULT AND DISCUSSION

In the first stage of this system, molecular compound data has been acquired from publicly available along with their experimentally known properties such as toxicity, solubility, bioactivity, etc.
The second stage is the construction of the molecular graph. As stated, the molecules have a natural tendency to form a graphlike structure. Therefore, the molecules are converted to a graph-like structure. In the conversion of the molecules to a graph-like structure, the RD Kit cheminformatics library is used. This library string format to the graph format. The atomic features, i.e., atomic number, degree, valence, and hybridization, of the molecules are extracted using the RD Kit cheminformatics library.
The Geometric library is used to implement the operations of the graph. In the proposed system, the graph convolution
The proposed system, i.e., the drug property prediction system has been implemented and tested on a dataset that includes various chemical compounds. The proposed system, as described in this research, includes an interface that allows users to interact with the system for the analysis of molecules, prediction of pharmacological properties, and assessment of safety properties for various chemical compounds.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

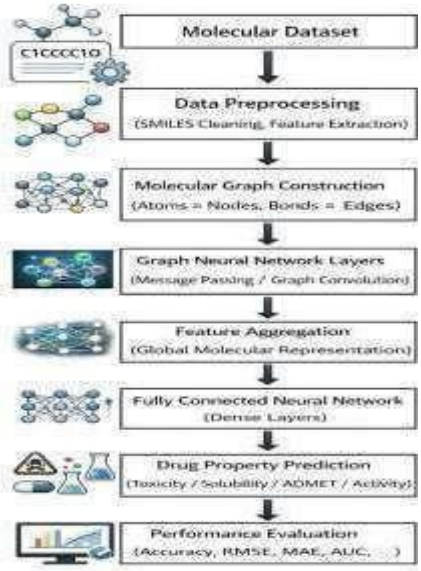


Fig. 2. System Flow Diagram molecular data, graph representation, deep learning-based drug property prediction, and evaluation. architecture four the data acquisition module, molecular graph construction module, Graph Neural Network Model, and prediction evaluation module.

The proposed system includes an interface through which the user can input the structure of the molecules in SMILES notation. After that, the analysis of the molecules is carried out by utilizing the structure of the molecules is first converted to a graph, in which the atom in the structure of the molecules is considered a node, and the chemical bond in the structure of the molecules is considered an edge. After that, the graph neural network model is used for the analysis of the structure of the molecules, and important properties related to the drug are predicted.

Fig. 3 illustrates the relationship between the actual solubility values and the predicted values generated by the Graph Neural Network (GNN) model. The scatter plot shows that most data points lie close to the ideal reference line (y = x), indicating a strong correlation between experimental and predicted LogS values. This demonstrates that the model effectively captures the underlying patterns in molecular structures to predict

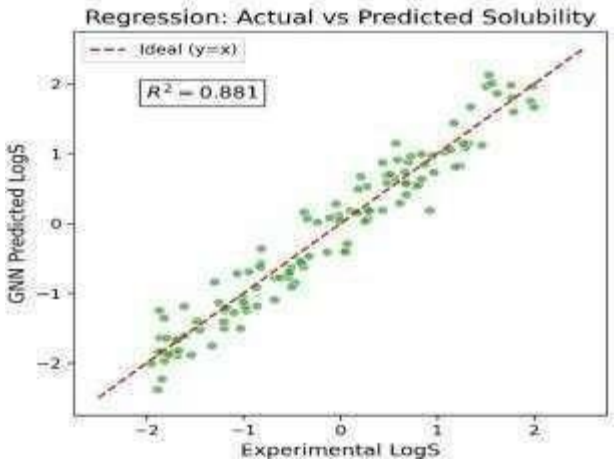


Fig. 3. Actual vs Predicted Plot

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

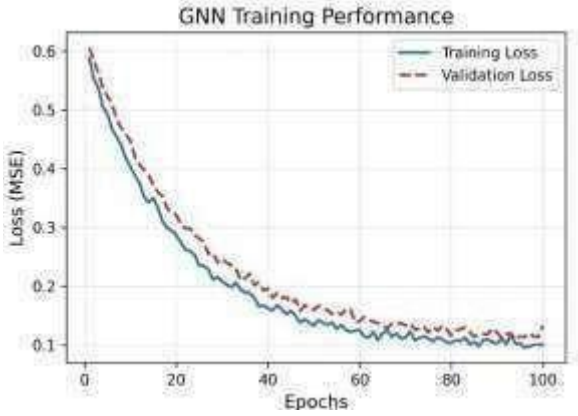


Fig. 4. GNN Training performance

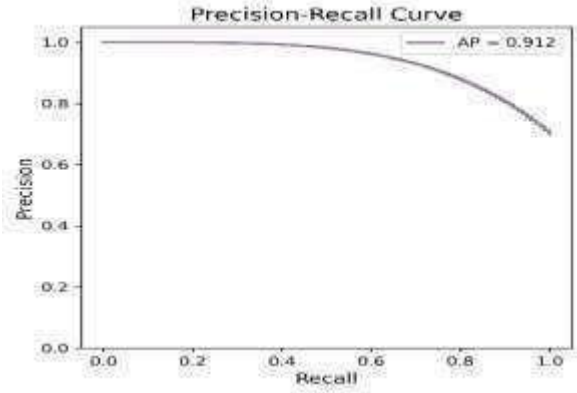
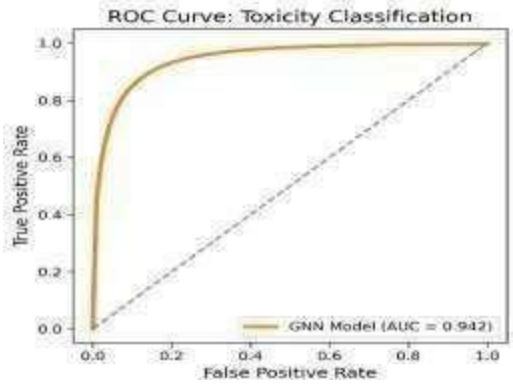


Fig. 5 Precision-Recall curve



Model	MAE	RMSE	R^2 Score
Linear Regression	53.89	69.13	0.936
GNN (Proposed)	38.75	49.60	0.972

DOI: 10.48175/IJARSCT-33737

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

VIII. CONCLUSION

The research has predict significant drug-related properties of molecular compounds using a Graph Neural Network. In this context, it can be noted that the proposed system has been based on the inherent graph structure of the molecular compounds, where the atoms of the molecules have been represented as nodes.

The developed system is based on the integration of molecular data preprocessing, graph construction, deep learning-based prediction, and interactive visualization. In this regard, the molecular structures represented as SMILES strings are converted into graph representations, following which the prediction is carried out the In this manner, the proposed system is able to efficiently predict significant pharmacological properties, including toxicity risk, solubility, and drug-like properties of molecular compounds. The user-friendly interface has been implemented for the purpose of performing the molecular analysis, safety analysis, molecular comparison, and chemical space visualization. From it is evident that the proposed system effectively performs the analysis of the molecules and provides useful predictions for the purpose of drug discovery.

The results obtained from the experiments have shown the advantages of the proposed approach by employing technique compared to the traditional machine learning technique, as the former technique effectively utilizes the

The proposed system for drug property prediction using the Graph Neural Network technique provides a useful and effective computational technique for the analysis of the chemical compounds for the purpose of drug discovery.

IX. FUTURE WORK

Although the proposed drug property prediction system using the Graph Neural Network model shows promising results for the analysis of drug structures and the prediction of drug properties, there are improvements could be future work. Firstly, the proposed drug property prediction system could be improved by considering a larger dataset of molecules for the training of the model. This could be achieved by considering more datasets of molecules from various publicly available repositories. This would enable the model to predict the drug properties more effectively. Further, the proposed drug property prediction system could be improved by considering more advanced types of GNN, i.e.,

Another extension of this system could be to forecast other drug discovery attributes such as bioavailability, metabolic stability, and protein-ligand binding affinities. This would increase the utility of this system for researchers in computational drug development studies. From a system development viewpoint, this system's user interface can also be enhanced to include more advanced visualization tools and live chemical databases to enable faster molecular analysis. This would increase the usability of this system for more indepth drug research studies.

APPENDIX

APPENDIX A

Molecular Dataset and Input Representation For the development of the Drug Property Prediction system, the data of the molecular compound was collected from various public domain sources of chemical information, such as PubChem and MoleculeNet. The datasets collected include various types of molecular compounds, which are research

SMILES is an algorithm that represents the structure of the chemical molecules in text format, where each bond in the molecule is represented as a string. The representations of the molecules are used as input for the system.