IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 28 April 2024, accepted 28 May 2024, date of publication 31 May 2024, date of current version 16 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3407877

RESEARCH ARTICLE

Detecting Side Effects of Adverse Drug Reactions Through Drug-Drug Interactions Using Graph Neural Networks and Self-Supervised Learning

OMKUMAR CHANDRAUMAKANTHAM, SRINITISH SRINIVASAN, AND VARENYA PATHAK
Vellore Institute of Technology, Chennai 600127, India
Corresponding author: OmKumar ChandraUmakantham (omkumar.cu@vit.ac.in)

ABSTRACT Adverse Drug Reactions(ADRs) due to drug-drug interactions present a public health problem worldwide that deserves attention due to its impact on mortality, morbidity, and healthcare costs. There have been major challenges in healthcare with the ever-increasing complexity of therapeutics and an aging population in many regions. At present, no standard method to detect such adverse drug reactions exists until otherwise reported by patients after the drug is released to the market. Further, several studies show that it is extremely challenging to detect these rare cases during clinical trials held before the drug is released. Therefore, a reliable and efficient technique to predict such side effects before the release of the drug to the market is the need of the hour. Through the power of Graph Neural Networks and the knowledge representation abilities of self-supervised learning, we designed an effective framework to model drug-drug interactions by leveraging the spatial and physical properties of drugs by representing them as molecular graphs. Through this approach, we developed a technique that resembles the dynamics of a chemical interaction. On training and testing this approach on the TwoSIDES Polypharmacy Dataset by Therapeutic Data Commons(TDC), we achieve a state of the art results by obtaining a precision of 75% and an accuracy of 90% on the test dataset. Further, we also perform a case study on the DrugBank dataset and compare our results on the interaction type prediction task to validate our approach on the drug-drug interaction domain and achieve excellent results with precision, F1, and accuracy of 99%. Our study and experimental approaches lay the groundwork for further research on side-effect prediction through drug-drug interaction and the use of Graph Neural Networks in the field of Molecular Biology.

INDEX TERMS Adverse drug reaction, drug-drug interaction, side effect prediction, graph neural network, self-supervised learning, scientific machine learning.

I. INTRODUCTION

An Adverse Drug Reaction(ADR) can be defined as a significantly harmful or unpleasant reaction usually attributed to the use of medicines and may warrant treatment, prevention, alteration of dosage, or withdrawal of the usage of that drug [1]. They are a major threat to the healthcare system since they contribute to mortality, morbidity, extended hospital stays, and increased healthcare costs [2]. Many of the side effects are not observed during clinical trials but are mostly identified only after the drug has reached the market [3].

Several studies also show that reports of Adverse Drug Reactions tend to be skewed based on sex, geographic region of origin, and country of origin. For instance, according to [4], women are at more risk of adverse drug reactions due to

The associate editor coordinating the review of this manuscript and approving it for publication was Gina Tourassi.

differences in pharmacokinetic and pharmacodynamic effects of drugs in addition to their higher dosage concerning body weight. Further, factors such as access to healthcare based on country of residence and their healthcare quality play a role as well. According to [5], the median proportion of preventable ADRs in developed and developing countries is 71.6% and 59.6% respectively, and the median proportions of ADRs relating in mortality were 1.7% and 1.8% in developed and developing countries respectively. The majority of these ADRs were preventable in both situations signifying the importance of early prediction of such ADRs and improving medical use in developing countries.

Accurate identification of ADRs is difficult considering it depends on the expertise of the attending physician and the quality of information available. Sometimes even experienced physicians have difficulty in determining causality. In addition to this, the usage of multiple medicines, underlying conditions of the patient, and assumptions such as the active principle of the medication cause adverse drug effects make the task of accurately identifying them extremely challenging. Further, they are sufficiently rare since they are tested amongst a small sample size during clinical trials, they tend to elude detection during the drug development process.

A primary cause of Adverse Drug Reactions is Drug-Drug Interactions. Drug-drug interactions occur when drugs compete for the same target. They also involve drug metabolizing enzymes and influx-efflux drug transporters which determine the adsorption, distribution, metabolism, and excretion of drugs [6]. They are one of the most common causes of medication error particularly among elderly people due to poly-therapy [7]. Drug-drug interactions can render a clinically prescribed drug ineffective and may cause adverse drug reactions which can be fatal.

With the growing popularity of deep learning models due to their ability to recognize underlying patterns in various types of data and the advent of Graph Neural Networks that can represent the physical and chemical properties of a drug, models based on such concepts can be used to effectively correlate drug-drug interactions to their side effects. Further techniques such as self-supervised learning and ensemble learning have proven to increase the overall model performance since self-supervised learning can capture and learn the distribution of data on an upstream task with very few labeled samples. The knowledge base obtained can be transferred to a downstream task [8] while ensemble techniques combine multiple deep learning models to improve the overall performance [9].

In consideration of such improvement and the growing concern about Adverse Drug Reactions, we propose a deep learning technique that leverages the spatial characteristics of graph neural networks and models the dynamics of a chemical interaction to predict the side effects caused by Adverse Drug Reactions(ADRs). We leverage the underlying feature learning capabilities of self-supervised learning and the complexity of Dual Path Graph Neural Networks to effectively predict side effects caused by drug-drug interactions. Our study and experiments lay the groundwork to begin further research into the field of Adverse Drug Reactions from a deep learning and mathematical perspective.

A. MOTIVATION

Adverse Drug Reactions and the side effects they cause are very challenging to detect and depend on various factors such as the quality of the physician, gender, quality of healthcare, etc, and subjectivity of the reports of adverse drug reactions. Since a framework dedicated to modeling drug-drug interactions by using the molecular structure of drugs for the prediction of side effects is yet to be established and considering the ability of deep learning with recent developments in graph neural networks, we propose a novel framework to model drug-drug interactions using Graph Neural Networks by modeling the properties of drug molecules and leveraging the pattern recognition abilities of self-supervised learning and the potential of ensemble learning. Since chemical molecules can be represented as a molecular graph, we can leverage the complex representations and operations offered by Graph Neural Networks to model their atomic and bond-level properties. Recent developments through Graph Convolution Network, Graph Attention Network [10], and SAGE [11] and their outstanding results on graph-level tasks have increased the popularity of Graph Neural Networks thus making it a reliable option in various fields [12], [13]. Despite some of the pitfalls of Graph Neural Networks such as under-fitting and the need to generate high-quality vertex embeddings, fine-tuning pre-trained weights and ensemble approaches raise the performance of Deep Learning models by a great extent unlocking the door for further research in Graph Neural Networks, and their applications on drug research.

Main Contributions:

- The study aims to reduce the number of clinical hospitalizations due to Adverse Drug Effects. Our proposed framework predicts the side effects of drug-drug interactions based on the molecular structure of the drug and their interaction dynamics. This enables a holistic and early determination of side effects caused by such adverse reactions and ensures that variables such as quality of the physician, subjectivity of the reports, and quality of healthcare are eliminated. Instead, a scientific and probabilistic view is provided to predict side effects. Our study is among the first to explore the underlying mathematical relations between drug-drug interaction and side effect prediction through deep learning.
- We present a novel approach to side effect prediction due to drug-drug interaction using Graph Neural Networks by representing the drugs involved in the reaction as molecules. The nodes of the graph are represented by atoms while the edges are represented by chemical bonds, thus making use of the spatial and physical properties of molecules by their graphical representation. Approaches that have been designed so far to predict

side effects by drug-drug interaction neither use the structure of the chemicals involved nor leverage Graph Neural Networks to represent the physical and bonding properties of chemical compounds.

- To realize the dynamics of a chemical reaction and effectively leverage the spatial and physical features of both the reactants, We develop a Dual Input Graph Neural Network Hybrid Model with a 2-stage training phase. The Dual Input framework is designed to model the dynamics of a chemical reaction by sharing features between both reactants. We achieve a stable model training curve despite the complexity of the system. Further, each reactant is pre-trained to ensure a knowledge base derived from the properties of the reacting drugs is transferred to the side effect prediction task. Such a framework despite being only in the initial stages of research shows state-of-the-art results in precision and accuracy.
- The dataset that we used for our study is the TwoSides poly-pharmacy side effects dataset from Therapeutic Data Commons(TDC). Since this is among the only datasets for side-effect prediction through Drug-Drug interaction, our experiments act as a baseline benchmark. Therefore, we aim for our study to be a base for side effect prediction and aim to open doors to further research into the field of Molecular Graph Neural Networks and Drug-Drug Interactions.
- Since the TwoSides poly-pharmacy side effects dataset is the only available dataset for side effects prediction and owing to a lack of previous benchmarks on the same, we provide an experimental validation of our proposed framework on the DrugBank Dataset. Though the final task for the DrugBank dataset is the prediction of the drug-drug interaction type, by comparing our results with related works, we achieve state-of-the-art results for our proposed framework on the drug-drug interaction type prediction task as well. We show that our proposed framework effectively models drug-drug interactions and despite the change in the prediction task, it can model the dynamics of a chemical interaction and lays the platform for further research on the Drug-Drug Interaction Domain and Bioinformatics.

B. PAPER ORGANIZATION

The following sections of the paper are divided as follows: Section II discusses the related literature concerning drug-drug interaction. Further, we also include literature concerning drug-target interaction using various deep learning techniques, the advent and popularity of Graph Neural Networks, and the applications of graph-based learning concerning organic and biochemistry. In section III we describe our proposed framework. Here we discuss the architecture and workflow of the training and testing process and the various graph operations we tested in our framework along with intrinsic details such as the hyperparameters and general layered architecture. In Section IV, we discuss the results obtained and compare various graph neural network models to show how our proposed framework outperforms the other experimented operations to a great extent. We show the need for the proposed methodology by validating the training process concerning stability. Further, we also discuss the importance of pre-training in our study by drawing comparisons with models trained with randomly initialized weights. In section V, we summarize the details of our entire methodology, and results, and lay the ground for further research related to this study. Finally in Section VI, we discuss the potential downfalls of our idea and implementation and how the mentioned limitations can be overcome with further research.

II. RELATED WORKS

The presence of Artificial Intelligence(AI) is apparent throughout all realms of science. Nevertheless, AI in chemistry has come up to be one of the most researched areas. Reference [14] successfully modeled QSAR which predicts the carcinogenic potency of aromatic amines and an FDA/OTR MultiCASE model predicting the carcinogenicity of pharmaceuticals. Similarly, its worth noting that [15] employed 2D similarity fingerprints for its efficiency and simplicity of computation for drug-drug interactions to prevent adverse effects.

The advent of Machine Learning(ML) and Neural Networks(NN) has raised interest in drug discovery. Reference [16] uses genomic features from the cell lines and chemical information from drugs and shows that it is possible to build in-silico-based multi-drug models to impute missing IC50 values with non-parametric machine learning algorithms. Reference [17] calculates drug-drug pair similarities using four features namely: drug phenotypic, therapeutic, chemical, and genomic properties with applied predictive models namely Naive Bayes, Decision Tree, k-nearest neighbor, logistic regression, and support vector machine(SVM). Reference [18] discusses a two stage hybrid approach model that identifies the positive instances using a feature based binary classifier, and then a Long Short Term Memory (LSTM) based classifier to classify positive instances while [19] uses Discriminative Vector Machines for accurate prediction of protein-protein interactions.

The Graph Neural Network(GNN) Model [20] approach is known to perform well in chemical graphs inclusive of but not limited to analyzing molecular structures, protein-protein interaction networks, and drug discovery. Graph Attention Network (GAT) [10] has been applied in various tasks and applications such as [21] and [22]. Reference [23] uses the Graph Convolution operation which allows the nodes to aggregate information from their immediate neighbors in the graph. Further, [24] makes use of spectral graph convolutions that leverage the eigenvalues and eigenvectors of the graph Laplacian matrix to define convolutional operations, enabling the propagation of information through the graph and [25] discussed the formulation of spectral graph convolution networks for directed graphs. GraphSAGE [11] was used in [26] and [27] and has proved to make a significant improvement in performance in transductive-based prediction tasks for large and elaborate graph networks.

III. METHODOLOGY

Figure 1 represents the overall framework of our proposed approach. The chemical SMILE string of each reactant is converted to a molecular graph with node feature vectors for each atom and bond. This is followed by self supervised pre-training to provide the model with a knowledge base followed by the classification model that predicts the adverse drug reactions.

A. INITIAL FEATURE EXTRACTION

1) GRAPH DEFINITION

The Chemical SMILE String of each reactant of the reaction is converted to a Molecular Graph in the Graph Space followed by feature vectors for each atom in the Euclidean Space. An atom is represented as a graph node while a bond is represented as a graph edge. Each node is represented by a 56-dimensional feature vector and each edge is represented by a 9-dimensional vector. Equation 1 depicts the graph definition.

$$\begin{aligned} \mathrm{G} &= (A, B) \\ \text{where, } \mathrm{G} &= \text{Molecular Graph} \\ \mathrm{A} &= [v_1, v_2, \ldots., v_{56}] \\ \mathrm{B} &= [e_1, e_2, \ldots .e_9] \end{aligned} \tag{1}$$

In eq 1, A refers to the 56 dimensional embedding vector for each atom of the graph, B refers to the 9 dimensional embedding vector for each bond.

2) NODE FEATURES

On the atomic level, the following features were used to featurise each vertex V of graph G:

- Formal Charge
- Degree of each Atom
- Hybridization
- Presence in Ring
- Aromaticity
- Atomic Mass
- Vander waal's Radius
- Covalent radius

Formal Charge refers to the charge carried by each atom in a molecule. Equation 2 describes the computation of formal charge.

$$\text{Formal Charge} = V - N + \frac{B}{2} \tag{2}$$

In (2), V stands for number of valence electrons, N stands for number of non bonding valence electrons and B stands for total number of electrons shared in bonds. The range for formal charge we have considered is from -3 to 3.

Degree refers to the number of neigboring atoms interlinked by bonds. The range of degree we have considered is from 0 to infinity.

Hybridization refers to the intermixing of atomic orbitals with same energy levels to produce a new type of atomic orbitals of the same number in accordance with the Valence Bond Theory(VBT). The type of hybridization gives the geometry of the compound and bond angle. It is indirectly calculated by computing the number of hybrid orbitals(described in (3)) of the central atom which is mapped to the hybridization type described in Table 2.

$$H = \frac{1}{2}(V + M - C + A) \tag{3}$$

where H stands for the number of atomic orbitals of central atom, V stands for Number for Valence Electrons of central atom, M stands for the number of surrounding monovalent atoms, C refers to charge on cation and A refers to the charge on anion.

Presence in Ring: A binary feature(1 or 0) whose value is 1 if the atom is part of a ring, otherwise 0.

Aromaticity of a compound refers to a property where cyclalkenes are conjugated. They enhance the stability of a molecule by delocalization of π-π electrons due to formation of more than 1 resonant structures.

Atomic Mass is the average relative mass of an atom as compared to an atom of carbon isotope 12 i.e C-12. We use the scaled form which is computed using (4)

$$\text{Atomic Mass Scaled} = \frac{\text{Mass of Atom} - 10.812}{116.092} \tag{4}$$

VanderWaal Radius of an atom refers to half the distance between the center of nuclei of two atoms held together by week vander waal's force. We use the scaled form which is computed using (5)

$$\text{VanderWaal's Radius Scaled} = \frac{\text{VDWR} - 1.5}{0.6} \tag{5}$$

where, VDWR refers to Vanderwaal's Radius.

Covalent Radius refers to the distance between the center of nuclei of two atoms when bound together by a single covalent bond. We use the scaled form which is computed in (6).

$$\text{Covalent Radius Scaled} = \frac{\text{CVR} - 0.64}{0.76} \tag{6}$$

where CVR refers to covalent radius.

All features are computed and aggregated using (7)

$$\begin{aligned} \text{Atom Vector} = \mathrm{FC} + \mathrm{D} + \mathrm{HYB} + \mathrm{Ring} \\ + \mathrm{ARO} + \mathrm{AM} + \mathrm{CV} + \mathrm{VDR} \end{aligned} \tag{7}$$

where FC refers to Formal Charge, D refers to Degree, HYB refers to the one hot vector for hybridization type, Ring is a binary variable if atom is a part of a ring, ARO refers to aromaticity, CV refers to covalent radius and VDR refers to Vander waal's radius.

TABLE 1. Literature survey.

Reference No.	Title	Techniques Used	Findings	Gap in literature
[28]	DeepH-DTA: Deep Learning for Predicting Drug-Target Interactions: A Case Study of COVID-19 Drug Repurposing	Employs a novel framework Heterogeneous Graph Attention (HGAT) model to efficiently model diverse topological representations of drugs. Additionally, two layers of bidirectional ConvLSTM capture spatio-sequential characteristics in drug sequences encoded in a SMILES format.	It addresses the simultaneous prediction challenges of chemical interactions between protein sequences and the homogeneity of drug candidates. Proposed a novel model, HGAT which helps in predicting how well a drug will bind to a target. A two layer bidirectional ConvLSTM has been added to improve the spatial and sequential characteristics of drug sequences which helps in recognizing the positions of components within drug molecules and understanding the order and dependencies for the same. Squeezed-Excited Dense Convolution Network has been implemented to understand the important features in the sequence of amino acids that make up the protein.	The framework lacks the use of transformer models, proven effective in various sequential data domains, for generating more informative sequential data representations. Moreover, the computational demands of Heterogeneous Graph Attention (HGAT) method are immensely high.
[29]	DIVERSE: Bayesian Data IntegratiVE Learning for Precise Drug ResponSE Prediction	Uses a Bayesian matrix factorization with importance weights, a tri- and bi-matrix factorization to model the relationships between cell lines, drugs, and gene interactions.	This method systematically integrates heterogeneous side information from diverse sources, by augmenting the precision of inferred associations. Hence, enabling a probabilistic modeling approach, providing a principled framework for capturing uncertainties in the inference process. The incorporation of importance weights emphasizes the significance of different sources of side information, allowing for a nuanced and weighted integration of diverse data.	The scope of improvement in predictive performance lies in incorporation of more detailed datasets of 3D features instead of the 2D features that have been used. Gibbs optimization can also be a good alternative in order to have a comprehensive understanding the uncertainity of the parameters. However it would take a large number of iterations of updating the parameters for the model to come to a stable state.
[30]	FingerDTA: A Fingerprint-Embedding Framework for Drug-Target Binding Affinity Prediction	Implements a Fingerprint-embedding framework for Drug-Target binding Affinity prediction (FingerDTA), leverages a Fingerprint-embedding strategy, where the framework employs Convolutional Neural Networks (CNN) to extract local patterns crucial for prediction accuracy. Simultaneously, fingerprints are utilized to encapsulate global information, ensuring a comprehensive representation of the molecular interactions involved in Drug-Target binding.	Incorporates essential information from the fingerprints of drugs and targets into a Convolutional Neural Network model,hence enhancing its efficacy in predicting drug-target binding affinity. Through this approach, fingerDTA demonstrates notable performance improvements. Additionally it proves valuable in the context of identifying potential drugs for deactivating COVID-19 through binding to the spike target.	The study faces a challenge in training the model due to a shortage of experimental affinity data, particularly when dealing with an extensive array of molecules during drug screening. The suggested future work entails exploring different network structures, such as graph neural networks, to refine compound representation.
[31]	NNDSVD-GRMF: A Graph Dual Regularization Matrix Factorization Method Using Non-Negative Initialization for Predicting Drug-Target Interactions	The proposed framework NNDSVD-GRMF implements a Non-Negative Double Singular Value Decomposition-based Graph Regularized Matrix Factorization	The NNDSVD-GRMF model performs Graph dual regularization which improves the model's ability to discern complicated data structures, thereby removing the risk of feature loss crucial for addressing the Drug-Target Interaction (DTI) problem. Furthermore, the integration of non-negative double singular value decomposition at the initiation phase of matrix factorization contributes to the refinement of convergence dynamics and elevates the quality of solutions yielded by the model.	While the proposed methods show better performance compared to state-of-the-art methods, their generalization to diverse datasets and scenarios needs further exploration. In addition to that, the computational complexity is not modelled to perform its best on large scale applications.
			The HOGCN when compared to the GCN and the SkipGNN for prediction of Drug Drug Interactions (DDI) on the DrugBank dataset proved that the HOGCN model consistently exhibits height-	

TABLE 1. *(Continued.)* Literature survey.

[32]	Predicting Biomedical Interactions With Higher-Order Graph Convolutional Networks	A novel framework, the higher-order graph convolutional network (HOGCN) has been proposed to gather features from neighbouring nodes of varying distances and employing a linear mixing approach to systematically acquire informative representations of biomedical entities.	ened confidence in predicting drug pair interactions that are more likely to be true positives, surpassing other models in this regard. Moreover, when dealing with false positive predictions in Drug-Drug Interactions (DDI), HOGCN effectively diminishes the predicted confidence, demonstrating its capacity to identify and remove false positive instances. This contrasts with GCN, which tends to exhibit overconfident predictions, highlighting HOGCN's success in averting false positives.	The model currently does not incorporate various physicochemical and biological properties of biomedical entities that could offer valuable supplementary information about interaction, making the prediction more accurate.
[33]	Drug-Target Interactions Prediction Based on Signed Heterogeneous Graph Neural Networks	The signed heterogeneous graph neural networks (SHGNNs) framework has been proposed which accomodates signed bipartite networks and also integrates auxiliary information from Drug-Drug Interactions (DDIs) and Protein-Protein Interactions (PPIs).	This framework tackles the task of message passing and aggregation within signed Drug-Target Interaction (DTI) networks, guaranteeing a streamlined and precise distribution of information throughout the network.The methodology has undergone rigorous testing across diverse configurations, encompassing variations in operational modes, embedding dimensions, and initial feature settings. The evaluation has ensured a thorough exploration of the method's performance under various conditions. Hence enhancing its robustness and applicability across a spectrum of scenarios.	Currently, there are no solutions to DTIs prediction with unknown drugs and targets.Furthermore, multi-modal node attributes have not been taken into account which would greatly improve the prediction performance.
[34]	Explainable Drug Repurposing Approach From Biased Random Walks	The proposed approach introduces a Markov process-based similarity methodology, leveraging a knowledge graph constructed from diverse data sources. These sources encompass experimental interactions between drugs and genes, associations between diseases and genes, and pharmacological indications of drugs in relation to diseases.	The datasets are represented as bipartite graphs in order to enable the natural construction of entropy-inspired similarity measures crucial for the knowledge graph. The graph structures are stored as sparse matrices which makes the performace more reliable and fast due to matrix operations. Additionally, the Markov process based drug recommendation system makes sure the model is stable in recommendation even with minute changes to the input data.	Since the model is dependent on a limited number of datasets, such as DrugBank and DisGeNet, there are chances of there being biases present in the model. As there are inconsistencies in naming conventions, the data fusion for the model cannot be done correctly.
[35]	A Novel Drug-Drug Indicator Dataset and Ensemble Stacking Model for Detection and Classification of Drug-Drug Interaction Indicators	The proposed ensemble methodology employs a combination of sophisticated classifiers, including the Gradient Boosting (GB) classifier, Adaboost, and Gaussian Naive Bayes (GNB), for the purpose of classifying drug indication types.	Combining the predictions of multiple classifiers, the ensemble benefits from the collective intelligence of each model. This irradicates individual weaknesses and uses the strengths of different classifiers, leading to a more robust and accurate predictive model. Essentially, GB excels in capturing complex relationships, AdaBoost emphasizes correcting misclassifications, and GNB is effective with continuous features. These features further help in adapting to more complicated data patterns.	Since ensemble is highly sensitive to noise, if the input dataset consists of noise, the model may not be able to perform well. Correspondingly, if the relationships between drugs evolve over time, the ensemble's adaptability might be limited to such dynamic changes.
[36]	A multimodal deep learning framework for predicting drug–drug interaction events	The DDIMDL method employs a multi-step pipeline integreted with multiple-layer neural networks	DDIMDL incorporates sub-models based on multiple-layer neural networks, providing a nuanced understanding of drug–drug interactions and performs much better than the conventional classification methods such as random forest (RF), k-nearest neighbor (KNN), and logistic regression (LR).	The model may not be as dynamically advanced as Graph Neural Networks (GNNs) in perceiving the complexity of the Drug Drug Interactions (DDIs) interdependence and complex architecture.

TABLE 1. ***(Continued.)*** **Literature survey.**

[37]	Drug-Drug Interaction Extraction Based on Transfer Weight Matrix and Memory Network	TM-RNN method is introduced as an enhancement to conventional multilayer bidirectional Long Short-Term Memory (LSTM) architectures. This improvement involves the incorporation of a transfer weight matrix to augment robustness. Additionally, a memory network is introduced to facilitate feature fusion within the model. The integration of these components aims to enhance the overall performance and capabilities of the LSTM-based approach.	The dynamic encoding capabilities to input sentences, capturing their long-term dependencies is because of the integration of a bidirectional Long Short-Term Memory (LSTM) architecture with a transfer weight matrix.This allows the model to generate networks with varying depths and structures, overcoming overfitting issues and enhancing robustness. For the DDI problem statement, it eliminates negative samples. Feature fusion based on the stored attentive weights avoids crosstalk of input signals.	The effectiveness of TM-RNN may be impacted when dealing with extremely long sentences or scenarios where two target drug entities are positioned far apart, introducing noise into the information. To address this challenge, a potential solution involves including a parser tree information which would capture structural closeness between drug entities, even if they are distant in the word sequence.
[38]	DEEPScreen: high performance drug–target interaction prediction with convolutional neural networks using 2-D structural compound representations	This framework utilizes 2-D structural compound representations as input features.	Since this framework is employing image-based representations instead of molecular fingerprints, they have a higher coverage of compound features making the DTI prediction performances more enhanced.	Since DEEPScreen only takes into account the features of compounds at the input level and treats the target proteins as labels, it only has predictions for 704 highly-studied proteins. Hence, without data augmentation techniques predictions for proteins without known interactions would not be possible.
[39]	Drug-Target Interaction Prediction Based on Multi-Similarity Fusion and Sparse Dual-Graph Regularized Matrix Factorization	The MSDGRMF framework is described as an ensembled computational approach. In addition to this, it provides a multi similarity fusion which aims to capture diverse and potentially useful information for predicting Drug Target Interactions (DTIs).	The proposed method can predict novel DTIs, providing insights for new drug design and drug repositioning which can lead to the model identifying potential interactions that were not previously known. The dual-graph regularized matrix factorization decomposes the drug-target interaction matrix into latent factor matrices for drugs and targets, capturing underlying patterns and relationships.	A possible limitation can be data dependence. Since the performance of the model heavily relies on the quality and completeness of the input data, there is a chance of noise or biased data affecting the generalization and prediction accuracy of the model.
[40]	Modular Multi–Source Prediction of Drug Side–Effects With DruGNN	Proposes a novel method DruGNN implementing Graph Neural Networks (GNNs) model to predict single drug side–effect.	Effectively addresses complex node-based classification problem with multiple classes hence handling Multi-Label Classification. In addition, to that implements supervised learning and semi supervised learning simultaneously for continual learning from splitting the training datasets.	There is a loss of information implied by using solely drug structure fingerprints instead of the full drug structure.There was also a lack of tissue specific information in the dataset that led to loss of recall.
[41]	Drug-Drug Interactions Prediction Based on Drug Embedding and Graph Auto-Encoder	Implements a model that is a combination of drug embedding from metapath2vec and VAE, and leveraging Graph Auto-Encoders for link prediction in the DDI network.	The Metapath2vec algorithm uses pre-defined sequences of node types to learn representations of nodes in the Drug Target Interaction (DTI) network. The VAE learns a probabilistic mapping of data as is employed to capture the chemical structure representations of drugs. Finally, the GAE is applied to predict additional edges.	The limitations could be the model becoming more biased towards predicting the more frequent types of DDI, if certain types of DDIs are significantly more prevalent than others in the dataset.
[42]	3DGT-DDI: 3D graph and text based neural network for drug–drug interaction prediction	The proposed novel 3DGT-DDI model integrates 3D molecular structure information, position embedding, and text features using SCIBERT.	The 3D graph structure model captures the spatial characteristics of drug entities, providing a more detailed representation. SCIBERT is used to understand the complex biomedical language and extract information from scientific literature. Finally, the position embedding, allows the model to consider the spatial arrangement of molecules and their relative positions.	One limitation of this model is the high compute that is required to render 3D graphs.

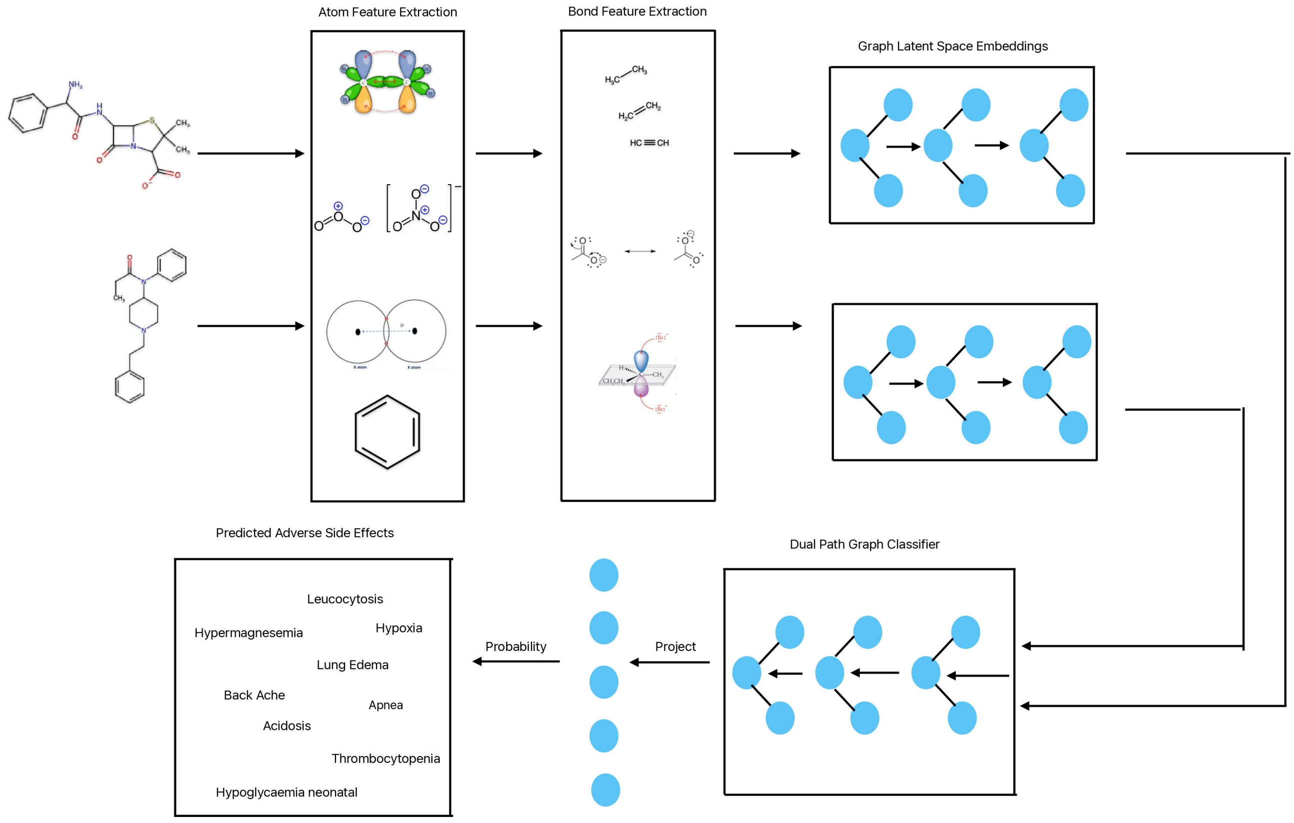

FIGURE 1. Overall framework of our proposed method.

TABLE 2. Hybridization type.

No. of Hybrid Oribitals	Hybridization
2	sp
3	sp^2
4	sp^3
5	sp^3d
6	sp^3d^2

3) EDGE FEATURES

The following bond features were used to featurize the edges E of graph G.

- Bond Type
- Conjugation
- Stereochemical Aspects

Bond Type refers to the type of bond between two atoms. It maybe a single, double or triple bond or aromatic.

Conjugation is represented by a binary variable(1 or 0). The value is 1 if the bond is a part of a conjugation i.e alternate single and double bond resulting in delocalization of π-π electrons forming different resonant structures, else its 0.

Stereochemical refers to the usage of E-Z stereochemical features around double bonds. Equation (8) describes the equation we used to featurize bonds.

$$\begin{aligned}\text{Bond Vector} = \text{Bond Type} + \text{Conjugation}\\ + \text{E-Z Features}\end{aligned} \tag{8}$$

B. SELF SUPERVISED PRE-TRAINING

Since Graph Neural Networks are prone to underfitting due to meaningless embedding vectors at certain times, we first pre-train the reactants before classification of their respective side-effects. We pre-train each reactant using a Variational Graph Autoencoder [43]. The goal behind this approach is to ensure the model learns the distribution of each reactant before the classification task. This provides a knowledge base to the model reducing the computation bounds to an extent thus increasing performance. Each reactant is pre-trained separately by two different models. At this stage, we do not perform any reaction-like modeling such as weight sharing or fusion to ensure the individual distributions of the reactants are learned first.

1) VARIATIONAL GRAPH AUTOENCODERS

We employ a Variational Graph Autoencoder(VAE) as our baseline model for self-supervised learning. The architecture of the Graph Autoencoder comprises an encoder and decoder. The encoder is usually the backbone which may comprise

TABLE 3. Terminology used in node and edge feature extraction phase.

Abbreviations	Explanation
G	Molecular Graph
A	Embedding Vector representing the Nodes(vertices) of the Graph
B	Embedding Vector representing the Edges of the Graph
V	Number of Valence Electrons of the Central Atom
N	Number of Non Bonding Valence Electrons
B	Total number of electrons shared in bonds
H	Number of Atomic Orbitals(Not to be confused with Hydrogen Atom when used in equation)
M	Number of surrounding monovalent atoms
C	Charge on Cation
A	Charge on Anion
VDR	VanderWaal's Radius
CV	Covalent Radius
AM	Atomic Mass

pre-trained weights or randomly initialized weights. Since state-of-the-art pre-trained graph models are quite sparse in the literature, we initialize the backbone with random weights by sampling out points using a normal distribution. The decoder model aims to reconstruct the input with minimal loss.

The encoder takes the reactant graph as input and encodes the input to its latent distribution. A point is randomly sampled from the latent distribution which is then reconstructed to form the original input. By providing a probabilistic view using Bayesian Inference, the model learns the distribution of the reactants. Equation 9 describes the workflow of the Variational Graph Autoencoder.

$$E(G) \rightarrow p(\frac{z}{G}) \rightarrow D(z) \tag{9}$$

In (9), E refers to the encoding function, which takes Graph G as input. z is a point in the latent space, and D refers to the decoding function. A point z is randomly sampled from the latent distribution $p(\frac{z}{G})$ which is then reconstructed to form the input as described above.

The loss function of this model is a combination of the reconstruction loss and a regularization term. The reconstruction loss ensures that the reconstructed and input node vectors are similar while the regularization term is added to the latent space. It ensures that the returned distribution is similar to a standard normal distribution. The regularization loss employed is the Kulback-Leibler Divergence Loss

$$L = ||x - d(z)||^2 + \mathrm{KL}[N(\mu_g, \sigma_g), N(0, I)] \tag{10}$$

where, x is the input node vectors, $d(z)$ is the reconstructed output of the decoder, μ_g, σ_g is the mean and variance of the latent space z respectively and I stands for the Identity term referring to the variance of a standard normal distribution.

2) ARCHITECTURE OF THE BACKBONE

For our final implementation, we employ two backbone variants:

- Backbone based on Graph Convolution
- Backbone based on Spectral Graph Convolution

Graph Convolution Based Backbone: The encoder comprises 3 layers of Graph Convolutions. The number of node features for each layer is tuned as $[128, 256, 512]$. Symmetric normalization is employed after each layer for faster convergence. The activation function for each layer is the Rectified Linear Unit(ReLU).

Spectral Graph Convolution Based Backbone: In this case as well, the encoder comprises 3 layers. The number of node features for each layer is tuned as $[128, 256, 512]$. The size of the kernel filter is set to 3 for each layer. Similar to Graph Convolutions, symmetric normalization is applied after each layer followed by ReLU activation.

3) ARCHITECTURE OF THE DECODER

The architecture of the decoder follows a transpose pattern compared to the encoder i.e the latent-space encoding is reconstructed to the input distribution in the same manner in which the encoding operations took place. The decoder for both Graph convolution based and Spectral Graph Convolution based encoders contains 3 layers. The number of node features for each layer is tuned as $[512, 256, 128]$. Symmetric Normalization and ReLU activation is applied after each layer for faster convergence.

4) INITIALIZATION

The weights of all the layers of both the encoder and decoder are initialized concerning a Normal Distribution. The mean and variance used to sample weights are 0 and 1 respectively. All biases were initialized with zeros.

5) OPTIMIZER AND LEARNING RATE CYCLES

We use the Adam [44] optimizer for pre-training each reactant. The β_1 and β_2 parameters are 0.9 and 0.999 respectively. The model is trained up to 1000 epochs. The initial learning rate is set to 0.005. Further, we make use of the Cosine Annealing learning rate scheduler with a half-duty cycle of 10 epochs. Equation (11) describes the computation of learning rate [45].

$$\eta_t = \eta_{min} + \frac{1}{2}(\eta_{max} - \eta_{min})(1 + \cos\frac{T_{cur}}{T_i}\pi) \tag{11}$$

Figure 2 describes the variation of the learning rate in the pre-training stage.

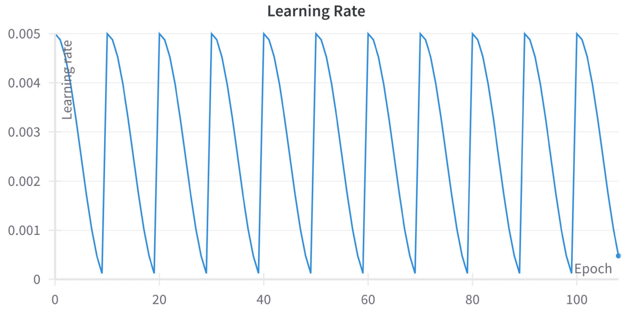

FIGURE 2. Cosine annealing learning rate variation.

C. TRAINING PHASE

The backbone weights post the self-supervised pre-training phase are transferred for the classification task. In this phase, we develop an architecture aimed at learning the underlying features of both reactants combined by a shared weight strategy through a dual path graph neural framework.

1) ARCHITECTURE OF DUAL PATH GRAPH NEURAL NETWORK

Figure 3 depicts the proposed architecture of our study. All model architectures take 2 graphs as input which represent the 2 reactants. Each reactant is forward propagated through the multi-input model as shown in Figure 3. The architecture contains 2 paths for forward propagation, analogous to a Dual Path Convolution Neural Network. Each sub-module or path for the reactants consists of 3 graph convolution or spectral graph convolution layers. The number of features learned for each layer increases in the order of 2^l where l represents the layer number ranging from 1 to 3. For our proposed framework, we tune the number of features for each layer as $[128, 256, 512]$ from layer 1 to layer 3. The kernel size for each layer is 3×3. The outputs of each layer of both the paths are activated by the ReLU activation function which is preceded by Symmetric Normalization. The results of the feed-forward computations of each path are then aggregated by computing the mean of the features across the set of nodes(Global Mean Pooling).

$$\text{Mean}(X) = \frac{1}{X}\Sigma_{x_i \in X} x_i \tag{12}$$

The aggregated features of each module are then fused using the Add operator(+). This is followed by a single linear layer with a nonlinear ReLU activation. The number of hidden units of the ReLU-activated linear layer is 1024. This is followed by another linear layer comprising 1317 neurons, symbolic of the 1317 side effects thus representing the classification layer, therefore it is activated by the sigmoid function to learn a probability distribution for the predicted side effects. Equations (13) and (14) describe the computation of the linear transformation and sigmoid classifier. The computation graph for gradient computation during back-propagation includes all layers and operations in the proposed framework i.e. all layers of the proposed architecture are trainable.

$$x' = xW^{\intercal} + b \tag{13}$$

$$\hat{y} = \frac{1}{1 + \exp(-x')} \tag{14}$$

where W represents the weights matrix and b indicates the bias.

In Section III-D, we describe in detail, the graph neural network operations that were experimented in our proposed training framework with respect to their individual tune-able parameters and propagation steps

2) LOSS FUNCTION

The negative log-likelihood function was considered as the objective function for training. The negative log-likelihood is a popular approach for classification problems since it satisfies Jensen's inequality for convexity, and is continuous and differentiable at all points, enabling deep learning models to reach an optimal solution quickly. Equation (15) describes the computation of the loss function.

$$\begin{aligned} l(x, y) &= \text{mean}([l_1, l_2, \ldots, l_n]) \\ l_n &= -y_n \log(p_n) - (1 - y_n)\log(1 - p_n) \end{aligned} \tag{15}$$

where, y_n is the actual label and p_n is the output of the final layer computed using (14).

3) OPTIMIZER AND HYPER-PARAMETERS

Adams Optimizer [44] was used as our principle optimization algorithm. The β_1 parameter was set to 0.9 and β_2 parameter was set to 0.999. The learning rate used for all experiments is 0.005. Further, we make use of a learning rate scheduler by reducing the learning rate by a factor of 10 in case of local optima and plateauing. We train all models for up to 750 epochs.

D. GRAPH NEURAL NETWORK OPERATIONS USED

In this section, we describe the architecture and hyper-parameters for each sub-model of our proposed training framework. Further we also describe the other operations such as Multi-layer perceptron(Linear), SAGE and Graph Attention which are not a part of our training framework but were used in experiments to validate our proposed framework.

1) SPECTRAL GRAPH CONVOLUTIONS(CHEBYSHEV CONVOLUTIONS)

We transfer the weights of the Spectral Graph Convolution backbone for each reactant as mentioned in Section III-B2 and fine-tune it for classification. The same filter size, number of layers and number of features for layer are maintained. Equations (16)-(19) describe the forward propagation computation employed for spectral graph convolutions [46].

$$X' = \Sigma_{k=1}^{K} Z^k . \Theta^k \tag{16}$$

where Z^k is a recursive variable computed as

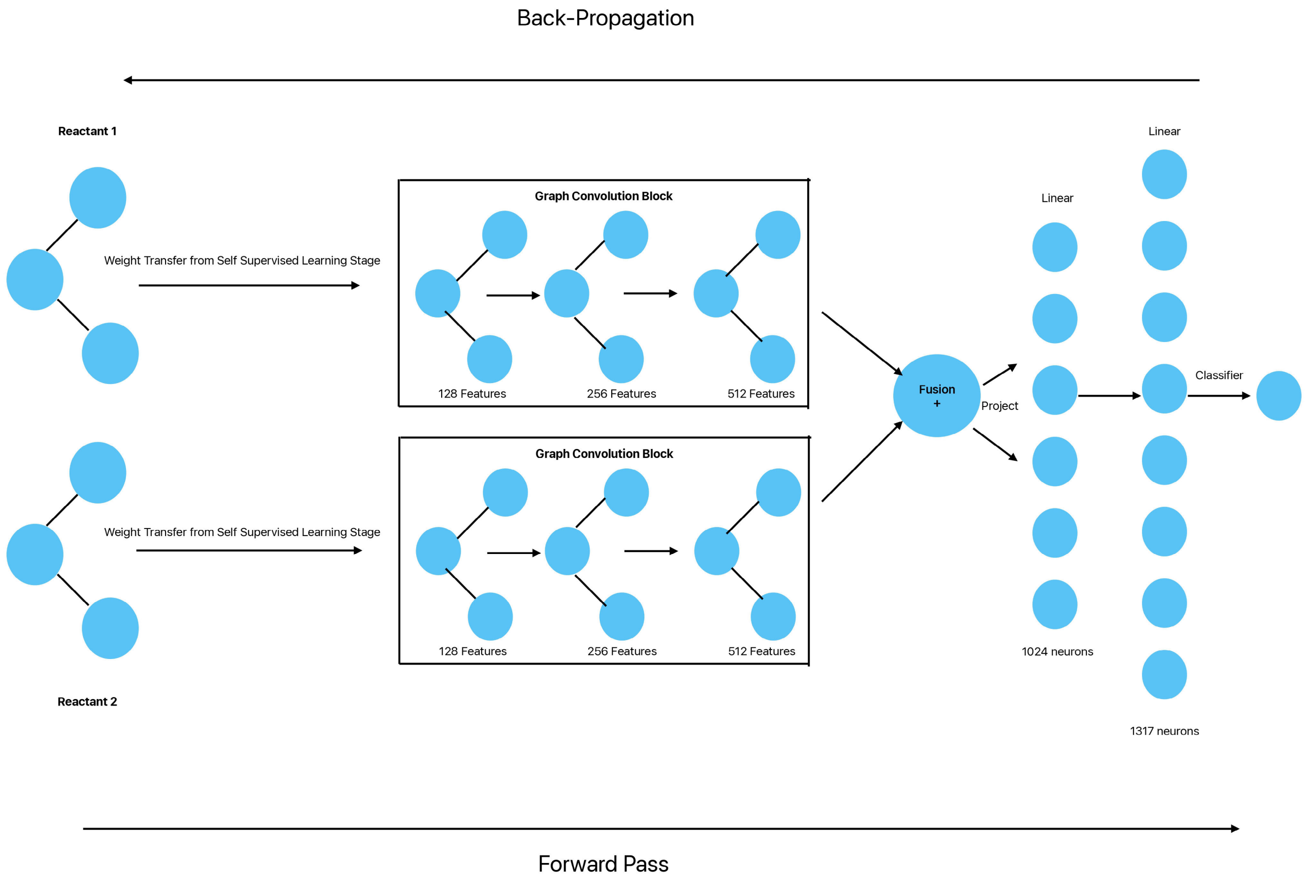

FIGURE 3. Dual Path GNN architecture which models a Drug-Drug Interaction with classifier layer to predict each side-effect that may occur.

$$Z^1 = X \tag{17}$$

$$Z^2 = \hat{L}.X \tag{18}$$

$$Z^3 = 2.\hat{L}.Z^{k-1} - Z^{k-2} \tag{19}$$

Here, $\hat{L}$ stands for the normalized Laplacian Matrix computed as per (20)

$$\hat{L} = \frac{2L}{\lambda_{max}} - I \tag{20}$$

The Laplacian Matrix L is computed as per (21)

$$L = D - A \tag{21}$$

where, D is the Diagonal Matrix of degrees of each node and A is the Adjacency Matrix. λ_{max} refers to the highest Eigen value of the graph Laplacian.

2) GRAPH CONVOLUTIONS

We transfer the weights of the Graph Convolution Backbone for each reactant as described in Section III-B2 and fine tune the model for classification. The same number of layers and features for each layer are maintained. Equations (22)-(24) describe the forward propagation steps for Graph Convolution layers.

$$X' = \hat{D}^{\frac{-1}{2}}\hat{A}\hat{D}^{\frac{-1}{2}}X\Theta \tag{22}$$

$$\hat{A} = A + I \tag{23}$$

$$\hat{D}_{ii} = \Sigma_{j=0}\hat{A}_{ij} \tag{24}$$

where, $\hat{A}$ is the Adjacency Matrix with inserted self loops and $\hat{D}$ is a diagonal matrix with degrees of each node.

3) GRAPH ATTENTION NETWORK

Graph Attention Networks (GAT) are implemented in a graph when some nodes are more important than the others. We make use of the enhanced Graph Attention Operator [47] for our experiments. Equations (25)-(27) describe the forward propagation for Graph Attention layers.

$$x'_i = \alpha_{i,i}\Theta_s x_i + \Sigma_j \alpha_{i,i}\Theta_t x_j, \tag{25}$$

$$\alpha_{i,j} = \text{softmax}(e_{ij}) \tag{26}$$

where,

$$e_{ij} = a(Wh_i, Wh_j) \tag{27}$$

where, α refers to the attention coefficients, e_{ij} is the importance of node j's features to node i. W is the weight matrix and h is the set of node features $h = [h_1, h_2, \ldots, h_N]$. a refers to a shared-attention mechanism mapping represented by $a : R^{F'}XR^{F'} \rightarrow R$. This function

TABLE 4. Terminology used in equation of graph neural network operations implemented.

Abbreviation	Explanation
L	Laplacian Matrix
D	Diagonal Matrix of Degrees of Each Node
A	Adjacency Matrix
$\hat{L}$	Normalized Laplacian Matrix
$\hat{A}$	Adjacency Matrix with inserted self loops

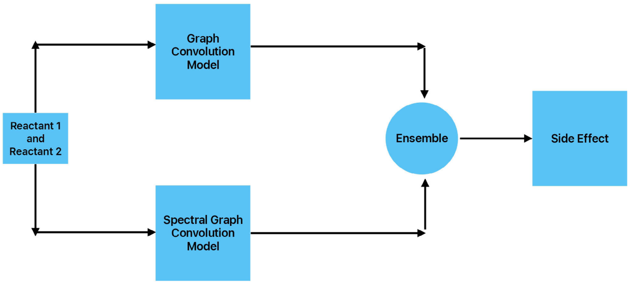

FIGURE 4. Ensemble framework.

computes the attention coefficients. F refers to the number of features in each node and F' refers to the number of node features after projecting the node features to a different cardinality.

We vary the number of output features as $[128, 256, 512]$ for each layer and number of attention heads as $[8, 4, 1]$. A dropout with noise of 0.40 is added to each layer to prevent overfitting [48].

4) MULTI LAYERED PERCEPTRON

The architecture of the Multi Layered Perceptron(MLP) incorporates an input layer and an output layer in addition to 6 hidden layers. Each layer consists of Batch normalization to reduce internal covariance shift [49], an activation layer using ReLU and regularisation using dropout to prevent over-fitting in the 1D space. Equation (28) describe the forward propagation step for each linear layer.

$$x' = \text{ReLU}(xW^{\intercal} + b) \tag{28}$$

5) GRAPHSAGE

The GraphSAGE is structured with three layers, each having $[128, 256, 512]$ node features respectively. The features of each layer is aggregated by computing the arithmetic mean across the nodes. Equations (29) describe the forward propagation steps of SAGE [11].

$$x_i' = W_1 x_i + W_2.\text{mean}_{j \in N(i)}(x_j) \tag{29}$$

E. ENSEMBLE

We combine the predictions of the fine-tuned Graph Convolution Model and Spectral Graph Convolution based dual path Graph Neural Network Model using a weighted average technique of the two models symbolizing the importance of the parent models.We consider the best of both models and architectures to come to a final conclusion thus we experiment on ensemble architectures for spectral GCN and vanilla GCN frameworks. Further, owing to the transductive nature of both these models, ensemble is aimed to combine the underlying features of both models to come to an ideal fit that can model the mapping from Drug pairs to side effects. Figure 4 describes the methodology used to ensemble the two models. Equation (30) and (31) describes the ensemble process and the final prediction step in our proposed framework respectively.

$$p = \text{Sigmoid}(w_1 z_1 + w_2 z_2) \tag{30}$$

$$y = \begin{cases} 1 & \text{if } p \geq 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{31}$$

where, $w_1 < 1$ and $w_2 = 1 - w_1$ are the weights of the Two Input Graph Convolution and Two Input Spectral graph convolution model respectively while z_1, z_2 are the predictions of the Two Input Graph Convolution and Two Input Spectral graph convolution model respectively.

IV. RESULTS AND DISCUSSION

A. HARDWARE AND SOFTWARE REQUIREMENTS

1) PRE-PROCESSING

All graphs were pre-processed on a 2022 M2 Macbook Pro. The binaries of each graph were divided into 8 folds to set up multiple processes. Upto 8 concurrent processes were created making use of all the 8 cores of the hardware to convert all SMILE Strings to PyTorch Graph Tensors. The total time taken to pre-process and create all the graph tensors was 5 hours.

2) MODEL TRAINING AND TESTING

All models were trained on a 2022 M2 Macbook Pro and Acer Predator Helios Neo 16 13th Gen Intel Core i7. The configurations of the M2 Macbook Pro are inclusive of an 8GB unified memory with a 256GB Hard disk storage. The number of cores of the CPU and GPU are 8 and 10 respectively. The time taken on average for model training was about 2 to 3 hours. It is to be noted that the GPU wasn't used for training or validation due to constraints in the Torch Geometric framework concerning Apple Silicon.

The configurations of Acer Predator Helios Neo 16 13th Gen Intel Core i7 are inclusive of 16 GB RAM and 956 GB Hard disk storage. The number of cores of CPU and logical processors are 16 and 24 respectively. The time taken on average for model training was about 1.5 to 2 hours. The NVIDIA GeForce RTX 4050 GPU has been used for training as well as validation.

3) SOFTWARE

All the source code for our study was written in Python 3.11. We used PyTorch as our primary framework for data preparation, pre-processing, model construction, training, and testing. For processing the chemical SMILE strings and converting them to molecular graphs, we use the RdKit library. For the computation of metrics such as precision and accuracy, we use Torchvision metrics for multi-label classification problems.

B. DATASET DESCRIPTION

The TWOSIDES PolyPharmacy Side Effects Dataset was employed in our study and was used for all experiments. It is to be noted that this is the only available benchmark dataset on side effects prediction through drug-drug interaction. The DrugBank Dataset uses drug-drug interactions as well, however, for the task of prediction of interaction type, which we later use to experimentally validate our proposed framework. The dataset consists of several drug-drug interaction pairs that are associated with multiple side effects. Each drug pair is associated with one too many side effects such as dizziness, chest pains, etc. The dataset comprises the following features:

- "ID1" refers to the ID of the first drug taking part in the interaction.
- "ID2" refers to the ID of the second drug taking part in the interaction.
- "X1" refers to the first reactant of the chemical reaction. It is a SMILES string that represents the molecular structure of the drug.
- "X2" refers to the second reactant of the chemical reaction. It is a SMILES string that represents the molecular structure.
- "Y" refers to the side effect label as a result of the drug-drug interaction which is mapped to a specific side effect.

There are 4,649,441 drug-drug interaction pairs with 645 drugs, each associated with a side effect. Since several of the drug-drug interaction pairs represented the same reactants and were associated with a single label only, we combine all the similar drug-drug interactions by their reactant IDs along with their side effect labels and formulate a multi-label classification problem. There are a total of 1317 side effects that may be possible for each drug-drug interaction. On performing this step, the number of drug-drug interaction pairs was condensed down to 63473. Further, we also mapped the label of the side effect to the side effect name. The condensed dataset comprises the following features:

- "ID1" refers to the ID of the first drug taking part in the interaction.
- "ID2" refers to the ID of the second drug taking part in the interaction.
- "X1" refers to the first reactant of the chemical reaction. It is a SMILES string that represents the molecular structure of the drug.
- "X2" refers to the second reactant of the chemical reaction. It is a SMILES string that represents the molecular structure.
- "Side Effect Name" is a list of side effects caused by the associated drug-drug interaction.

Following which we perform a train-validation-test split on the drug pairs. Table 5 contains the number of samples in each split. The drug-drug pairs were randomly split as mentioned in the dataset split policy of TwoSides Polypharmacy Side Effects.

The number of samples for each train, test and validation are mentioned in Table 5

TABLE 5. Number of samples for each split.

Train	Validation	Test
45000	14500	500

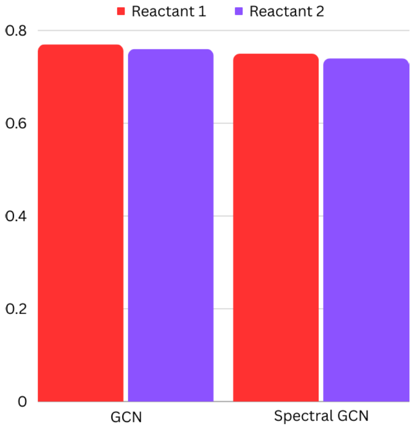

FIGURE 5. Comparison of reconstructed mean square error between GCN and spectral GCN.

C. PERFORMANCE OF GRAPH VARIATIONAL AUTOENCODER

As discussed in III-B, we employ a Variational Autoencoder with Graph-based data as Input as our basis for self-supervised pretraining. The purpose of such an implementation is to provide an enhanced knowledge base to the model by learning the distribution of each of the reactants. We experiment with 2 backbones, one based on Graph Convolutions and the other on Spectral Graph Convolutions, and achieve excellent results in both cases. Figure 5 shows the mean square errors obtained on the reconstruction of reactant graphs for both graph convolution and spectral graph convolution-based backbones.

The mean square errors obtained validate the use of a Variational Autoencoder before classification. The overall mean square loss obtained for Graph Convolution and Spectral Graph Convolution is only 0.76 and 0.75 respectively with spectral graph convolution slightly outperforming the graph convolution network. Since the reconstruction loss for both the reactants is similar under the same hyperparameters and architecture, we can infer that the latent space distribution is a good representation of the initial feature vectors of the nodes of the molecular graphs, thus validating the self-supervised training phase. The Variational Autoencoder has effectively captured the hidden and underlying patterns of the reactants enabling us to transfer this knowledge base to our classification framework. These results provide the basis

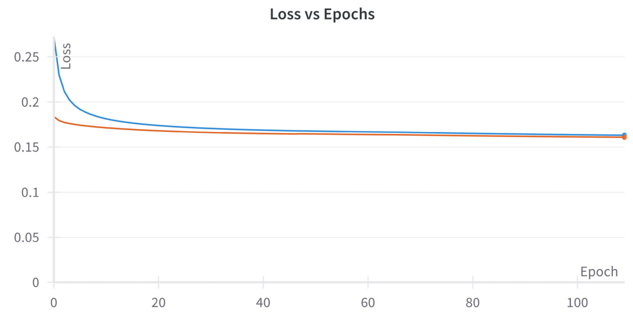

FIGURE 6. Train and validation loss of 2-Input spectral graph convolution network.

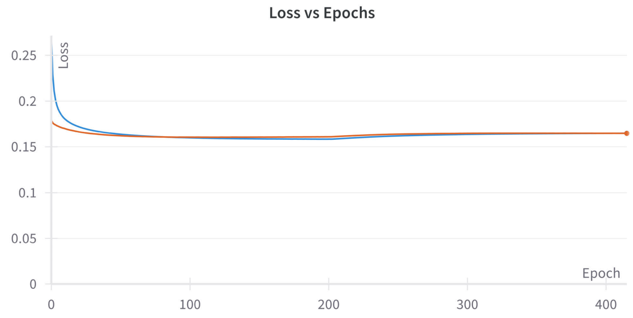

FIGURE 7. Train and validation loss of 2-Input graph convolution network.

for the classification model enabling us to achieve state-of-the-art results on the same.

D. STABILITY OF THE DUAL PATH GRAPH TRAINING FRAMEWORK

As discussed in III we use a 2-input dual path graph neural network. The motivation behind our proposed framework is the need to effectively model drug-drug interactions. Training multi-path graph neural networks can be unstable creating complicated computation graphs, thus resulting in poor performance and deeming the model unreliable. However, we managed to achieve stable training curves for both Graph Convolution and Spectral Graph Convolution-based models. Figure 6 and Figure 7 show the train and test loss progression with time for spectral graph convolution and graph convolution network respectively.

This curve was logged in real-time during the alternate training and validation process. We can see that with time, the performance of the model gradually improved eventually converging to an optimal solution after 80 to 100 epochs. The similarity of convergence in both train and validation losses indicates that there is no overfitting thus making the proposed framework reliable. Further, as noted in the performance of the autoencoders, architectures based on spectral convolution operators perform slightly better than the graph convolution operator as seen in the training curves. On average for both models, we obtain a classification loss of only 0.175 on the train and 0.177 on the test, hence validating the architecture of our proposed framework and deeming it as stable.

TABLE 6. Overall results of all experiments displayed in percentage.

Operation	Self Supervised Pretraining	Precision	Accuracy
Multi-Layer Perceptron(Linear)	-	20	52
Graph Convolution	No	38.97	75
Spectral Graph Convolution	No	45.48	79.72
Graph Attention	No	2	10
Graph Attention	Yes	5	17
SAGE	No	52.11	80.04
Graph Convolution	Yes	67.95	82
Spectral Graph Convolution	Yes	71	84
Ensemble of Graph Convolution and Spectral Graph Convolution	**Yes**	75	**90**

E. EXPERIMENTS CONDUCTED

To validate our proposed methodology and since our study lays the foundation for research in side effect prediction through drug-drug interaction, we provide results for several experiments that we conducted for various graph neural network operations and ensemble methods. The following are the experiments we conducted:

- Multi-Layer Perceptron with ReLU activated hidden layers for both paths of the proposed architecture.
- Graph Convolution with and without self-supervised pre-training for both paths of the proposed architecture.
- Spectral Graph Convolution with and without self-supervised pre-training for both paths of the proposed architecture
- Graph Attention Operation with and without self-supervised pre-training
- Graph SAGE Operation without self-supervised pre-training.
- Ensemble of Graph Convolution based and Spectral Graph Convolution based Dual Path Graph Neural Network.

F. RESULTS

Table 6 displays the precision and accuracy for all experiments that we conducted. As seen in Table 6, we achieve excellent results with a precision of 75% and 90% with our proposed framework, thus validating our hypothesis of using a dual path network to represent the underlying features of both the reactants effectively. To validate the requirement of our proposed framework and indicate its state-of-the-art performance, we conduct experiments with different graph-based operators namely Graph Attention and SAGE in addition to linear operators used in Multi-Layer Perceptron-based models. Further, we also show the importance and requirement of the self-supervised pretraining stage by performing experiments with models by randomly initializing their weights.

We perform experiments with no pre-training stage by randomly initializing the weights of the models of the

TABLE 7. Comparison of results on Drug-Drug interactions on DrugBank interaction type task.

Reference	Accuracy	Precision	F1
[36]	0.88	0.85	0.76
[41]	-	0.94	0.76
[50]	0.99	0.99	0.99
[51]	-	0.94	0.92
[52]	0.82	0.79	0.95
Ours	**0.992**	**0.992**	**0.992**

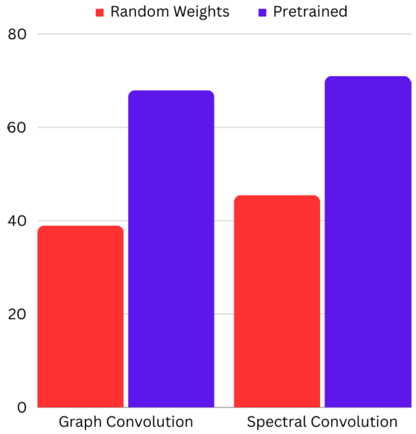

FIGURE 8. Comparison of pretrained and randomly initialized models.

framework using a standard normal distribution i.e. mean as 0 and standard deviation as 1. Figure 8 displays the difference in the performance of our proposed framework with fine-tuning of pre-trained weights and randomly initializing the weights before the side effect prediction task.

As indicated in figure 8, the difference in precision between the fine-tuned framework and randomly initialized is around 25% for the Graph Convolution based model and 28% for the Spectral Graph Convolution based model indicating a huge difference in performance. Such a disparity between the two frameworks validates the purpose of the pretraining phase, which contributes to a 25% rise in performance.

We also perform experiments with Graph Attention and SAGE operators. On SAGE, we achieve a precision of 52% and an accuracy of 80% which is 23% and 10% lesser than the performance of our proposed framework on precision and accuracy respectively. We do not perform self-supervised pre-training for the SAGE experiments owing to their ability to generate node embeddings using node features for previously unseen data [11]. SAGE is inherently inductive which means it can generalize on unseen nodes as well, which is a necessary requirement in drug-drug interactions. Therefore, no pre-training stage was required for SAGE. On Graph Attention, the precision obtained is 2% while the accuracy is 10%. Further, the results are marked as unstable since we do not achieve stable training when the Graph Attention operator is used in our proposed framework.

To further improve performance, we experiment with model ensembling. The final stage of our proposed framework includes an ensemble of Graph Convoluton and Spectral Graph Convolution. On performing experiments, we achieve excellent results outperforming all the other models. The precision obtained is 75% and the accuracy is 90%. Though the side effects are not mentioned by doctors or clinical experts, our deep learning framework is able to capture the hidden patterns and offers excellent results by modeling the dynamics of a drug-drug interaction. This shows that our proposed framework is successfully able to develop a relationship between drug-drug interaction and its potential side effects.

G. EXPERIMENTAL VALIDATION USING DRUGBANK

Due to the lack of widespread datasets and previous benchmarks on the side-effect prediction task, we aim to validate our model on the DrugBank [52] Dataset. The DrugBank dataset is used for predicting the interaction type. It contains 191,707 drug pairs with 1706 drugs. Each drug pair is mapped to one of 86 different interaction types thus formulating a multi-class classification problem. The results we obtained show that despite the difference in the task, our proposed framework is effectively able to model drug-drug interactions consistently when validated over several metrics. No major changes were made to the architecture described in section III except the final classifier head which was changed to 86 neurons with Softmax Activation to model a multi-class problem. Table 7 contains the results obtained by us along with a comparison of related works.

As observed in table 7, we achieve state-of-the-art results over 3 metrics on the interaction type task. We compare our proposed framework with 5 other related works that use deep learning techniques to predict the drug-drug interaction type. We achieve an accuracy, precision, and F1 score of 99.2% on this task thus further validating our model's ability to capture the underlying features with respect to chemical interactions. We have included the F1 score as well giving equal weight to both true positives and true negatives. This shows that our approach precisely detects both true positives and true negatives.

Through this experimental validation, we show that our model is able to adapt to diverse tasks related to drug-drug interactions. This enables clinical experts to further develop and trust our model considering its ability to provide state-of-the-art results on other drug-drug interaction tasks as well. The self-supervised pretraining stage has been effective in producing node embeddings for each atom considering both chemical and spatial properties. This along with the dual-path mode not only provides state-of-the-art results on metrics but also empirically models the dynamics of a chemical interaction.

V. CONCLUSION

Through this study, we explore a wide range of methods to effectively model drug-drug interactions by performing an extensive literature survey. We attempted a deep learning approach using Graph Neural Networks to capture the hidden patterns of each drug by using their spatial and physical characteristics from their molecular structure. Through this approach, we model drug-drug interactions to predict the side effects that it may cause. We divide the proposed framework into several phases and achieve excellent results on the same. The direction taken in drug-drug interactions by leveraging the chemical properties of drugs and their representations as molecular graphs is unique and opens doors for further research and improvements on the same.

The usage of the Graph Variational Autoencoder to pre-train the distribution of each reactant played a major role in the final performance. The overall reconstruction means a square error of only 0.75 for each reactant showing that the autoencoder was able to effectively map the reactants to a highly featured latent space distribution. Further, the similarity in the performance for each reactant showed that the same model configurations could be used, providing simplicity to our proposed approach. The transfer of weights from the reactant distribution task to the downstream classification had a major role in the overall performance. The average performance gain in precision was a massive 25% which showed the need for a trained knowledge base for complex graph-level tasks involving molecular representations.

To validate the performance and requirement of our proposed framework, we conduct several experiments using different Graph Neural Network operators and by randomly initializing the weights of the classification model. The precision and accuracy obtained on the test set for these experiments were significantly less than that of our proposed framework. Through our novel approach, we obtain a precision of 75% and an accuracy of 90% on the test dataset. This makes our architecture extremely reliable for side effect prediction considering its likelihood of accurately predicting 3 symptoms with high confidence for every 4 positive instances on unseen data.

Further, due to the lack of previous benchmarks on the side-effect prediction task, we show an experimental validation that our proposed framework effectively models drug-drug interactions using the DrugBank [52] dataset. We compare our results on the interaction type prediction task with related benchmarks and achieve a state of the art results. We obtain 99% on precision, recall, and F1, thereby validating our proposed framework. Through the interaction type prediction task, we show that our proposed approach is able to effectively model the dynamics and intrinsic patterns of drug-drug interactions.

Since the results obtained are positive and convincing, this opens a path for further research on the avenue of drug-drug interactions and using Graph Neural Networks to model the same. The results we obtained indicate that adverse drug reactions and their potentially fatal side effects can be predicted early without the need for patient reports, and having to depend on clinical trials performed on a sample size to report such adverse reactions. Since this is a recent research area, further research into such frameworks and modeling drug-drug interactions to predict side effects is required to ensure that the overall system based on deep learning in this field can be trusted by patients, doctors, and researchers.

VI. LIMITATIONS

A. COMPLEXITY VS INTERPRETABALITY

Since the framework comprises of several stages before the classification and considering the complexity of graph neural networks, especially in modeling chemical interactions, the overall architecture of the model is highly complex making it less interpretable [53]. Despite efforts to model chemical interactions by using the weights fusion method and the properties of drugs to initialize the nodes and edges of the molecular graphs, further steps need to be taken in developing more interpretable models in future research considering the framework is aimed to be used by patients, doctors and all clinical staff.

B. LACK OF WIDESPREAD DATASETS

The TwoSIDES Polypharmacy Side Effects Dataset, which was used for our study is the only available dataset that maps Drug-Drug Interaction using SMILE strings of chemical compounds to their respective side effects. The presence of only a single dataset may result in an overall biased model since it is trained and tested only on a single kind of distribution. Training and testing our proposed framework on more datasets, when available, will increase the overall reliability of the model and the bounds of its ability to generalize.

REFERENCES

[1] J. J. Coleman and S. K. Pontefract, "Adverse drug reactions," *Clin. Med.*, vol. 16, no. 5, p. 481, 2016.

[2] H. Jiang, Y. Lin, W. Ren, Z. Fang, Y. Liu, X. Tan, X. Lv, and N. Zhang, "Adverse drug reactions and correlations with drug–drug interactions: A retrospective study of reports from 2011 to 2020," *Frontiers Pharmacol.*, vol. 13, Aug. 2022, Art. no. 923939.

[3] D. Galeano, S. Li, M. Gerstein, and A. Paccanaro, "Predicting the frequencies of drug side effects," *Nature Commun.*, vol. 11, no. 1, p. 4575, Sep. 2020.

[4] S. Watson, O. Caster, P. A. Rochon, and H. den Ruijter, "Reported adverse drug reactions in women and men: Aggregated evidence from globally collected individual case reports during half a century," *EClinicalMedicine*, vol. 17, Dec. 2019, Art. no. 100188.

[5] M. T. Angamo, L. Chalmers, C. M. Curtain, and L. R. E. Bereznicki, "Adverse-Drug-Reaction-Related hospitalisations in developed and developing countries: A review of prevalence and contributing factors," *Drug Saf.*, vol. 39, no. 9, pp. 847–857, Sep. 2016.

[6] C. Kim and N. Tatonetti, "Prediction of adverse drug reactions associated with drug-drug interactions using hierarchical classification," *bioRxiv*, Feb. 2021.

[7] C. Palleria, A. Di Paolo, C. Giofrè, C. Caglioti, G. Leuzzi, A. Siniscalchi, G. De Sarro, and L. Gallelli, "Pharmacokinetic drug-drug interaction and their implication in clinical management," *J. Res. Med. Sciences*, vol. 18, no. 7, p. 601, 2013.

[8] F. Del Pup and M. Atzori, "Applications of self-supervised learning to biomedical signals: A survey," *IEEE Access*, vol. 11, pp. 144180–144203, 2023.

[9] I. D. Mienye and Y. Sun, "A survey of ensemble learning: Concepts, algorithms, applications, and prospects," *IEEE Access*, vol. 10, pp. 99129–99149, 2022.

[10] P. Velič ković, G. Cucurull, A. Casanova, A. Romero, P. Lió, and Y. Bengio, "Graph attention networks," 2017, *arXiv:1710.10903*.

[11] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1–19.

[12] R. Jovanovic, M. Palk, S. Bayhan, and S. Voss, "Applying graph neural networks to the decision version of graph combinatorial optimization problems," *IEEE Access*, vol. 11, pp. 38534–38547, 2023.

[13] T. Bilot, N. E. Madhoun, K. A. Agha, and A. Zouaoui, "Graph neural networks for intrusion detection: A survey," *IEEE Access*, vol. 11, pp. 49114–49139, 2023.

[14] A. M. Richard and R. Benigni, "AI and SAR approaches for predicting chemical carcinogenicity: Survey and status report," *SAR QSAR Environ. Res.*, vol. 13, no. 1, pp. 1–19, Jan. 2002.

[15] S. Vilar, R. Harpaz, E. Uriarte, L. Santana, R. Rabadan, and C. Friedman, "Drug—Drug interaction through molecular structure similarity analysis," *J. Amer. Med. Inform. Assoc.*, vol. 19, no. 6, pp. 1066–1074, Nov. 2012.

[16] M. P. Menden, F. Iorio, M. Garnett, U. McDermott, C. H. Benes, P. J. Ballester, and J. Saez-Rodriguez, "Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties," *PLoS ONE*, vol. 8, no. 4, Apr. 2013, Art. no. e61318.

[17] A. Kastrin, P. Ferk, and B. Leskošek, "Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning," *PLoS ONE*, vol. 13, no. 5, May 2018, Art. no. e0196865.

[18] D. Huang, Z. Jiang, L. Zou, and L. Li, "Drug–drug interaction extraction from biomedical literature using support vector machine and long short term memory networks," *Inf. Sci.*, vols. 415–416, pp. 100–109, Nov. 2017.

[19] Z.-W. Li, Z.-H. You, X. Chen, L.-P. Li, D.-S. Huang, G.-Y. Yan, R. Nie, and Y.-A. Huang, "Accurate prediction of protein–protein interactions by integrating potential evolutionary information embedded in PSSM profile and discriminative vector machine classifier," *Oncotarget*, vol. 8, no. 14, pp. 23638–23649, Apr. 2017.

[20] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 61–80, Jan. 2009, doi: 10.1109/TNN.2008.2005605.

[21] W. Song, Z. Xiao, Y. Wang, L. Charlin, M. Zhang, and J. Tang, "Session-based social recommendation via dynamic graph attention networks," in *Proc. 12th ACM Int. Conf. Web Search Data Mining*, Jan. 2019, pp. 555–563.

[22] H. Tak, J.-w. Jung, J. Patino, M. Todisco, and N. Evans, "Graph attention networks for anti-spoofing," 2021, *arXiv:2104.03654*.

[23] G. Qian, A. Abualshour, G. Li, A. Thabet, and B. Ghanem, "PU-GCN: Point cloud upsampling using graph convolutional networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 11678–11687.

[24] S. Parisot, S. I. Ktena, E. Ferrante, M. Lee, R. G. Moreno, B. Glocker, and D. Rueckert, "Spectral graph convolutions for population-based disease prediction," in *Proc. 20th Int. Conf.*, 2017, pp. 177–185.

[25] Y. Ma, J. Hao, Y. Yang, H. Li, J. Jin, and G. Chen, "Spectral-based graph convolutional network for directed graphs," 2019, *arXiv:1907.08990*.

[26] W. W. Lo, S. Layeghy, M. Sarhan, M. Gallagher, and M. Portmann, "E-GraphSAGE: A graph neural network based intrusion detection system for IoT," in *Proc. NOMS - IEEE/IFIP Netw. Operations Manage. Symp.*, Apr. 2022, pp. 1–9.

[27] Y. Cui, C. Shao, L. Luo, L. Wang, S. Gao, and L. Chen, "Center weighted convolution and GraphSAGE cooperative network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5508216.

[28] M. Abdel-Basset, H. Hawash, M. Elhoseny, R. K. Chakrabortty, and M. Ryan, "DeepH-DTA: Deep learning for predicting drug-target interactions: A case study of COVID-19 drug repurposing," *IEEE Access*, vol. 8, pp. 170433–170451, 2020.

[29] B. G. Paltun, S. Kaski, and H. Mamitsuka, "DIVERSE: Bayesian data IntegratiVE learning for precise drug ResponSE prediction," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 19, no. 4, pp. 2197–2207, Jul. 2022.

[30] X. Zhu, J. Liu, J. Zhang, Z. Yang, F. Yang, and X. Zhang, "FingerDTA: A fingerprint-embedding framework for drug-target binding affinity prediction," *Big Data Mining Analytics*, vol. 6, no. 1, pp. 1–10, Mar. 2023.

[31] J. Zhang and M. Xie, "NNDSVD-GRMF: A graph dual regularization matrix factorization method using non-negative initialization for predicting drug-target interactions," *IEEE Access*, vol. 10, pp. 91235–91244, 2022.

[32] K. KC, R. Li, F. Cui, and A. R. Haake, "Predicting biomedical interactions with higher-order graph convolutional networks," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 19, no. 2, pp. 676–687, Mar. 2022.

[33] M. Chen, Y. Jiang, X. Lei, Y. Pan, C. Ji, and W. Jiang, "Drug-target interactions prediction based on signed heterogeneous graph neural networks," *Chin. J. Electron.*, vol. 33, no. 1, pp. 231–244, Jan. 2024.

[34] F. Castiglione, C. Nardini, E. Onofri, M. Pedicini, and P. Tieri, "Explainable drug repurposing approach from biased random walks," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 20, no. 2, pp. 1009–1019, Mar. 2023.

[35] S. Abbas, G. Avelino Sampedro, M. Abisado, A. S. Almadhor, T.-H. Kim, and M. Mohamed Zaidi, "A novel drug-drug indicator dataset and ensemble stacking model for detection and classification of drug-drug interaction indicators," *IEEE Access*, vol. 11, pp. 101525–101536, 2023.

[36] Y. Deng, X. Xu, Y. Qiu, J. Xia, W. Zhang, and S. Liu, "A multimodal deep learning framework for predicting drug–drug interaction events," *Bioinformatics*, vol. 36, no. 15, pp. 4316–4322, Aug. 2020.

[37] J. Liu, Z. Huang, F. Ren, and L. Hua, "Drug-drug interaction extraction based on transfer weight matrix and memory network," *IEEE Access*, vol. 7, pp. 101260–101268, 2019.

[38] A. S. Rifaioglu, E. Nalbat, V. Atalay, M. J. Martin, R. Cetin-Atalay, and T. Doğan, "DEEPScreen: High performance drug–target interaction prediction with convolutional neural networks using 2-D structural compound representations," *Chem. Sci.*, vol. 11, no. 9, pp. 2531–2557, 2020.

[39] M. Lian, W. Du, X. Wang, and Q. Yao, "Drug-target interaction prediction based on multi-similarity fusion and sparse dual-graph regularized matrix factorization," *IEEE Access*, vol. 9, pp. 99718–99730, 2021.

[40] P. Bongini, F. Scarselli, M. Bianchini, G. M. Dimitri, N. Pancino, and P. Lió, "Modular multi–source prediction of drug side–effects with DruGNN," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 20, no. 2, pp. 1211–1220, Mar. 2023.

[41] S. Purkayastha, I. Mondal, S. Sarkar, P. Goyal, and J. K. Pillai, "Drug-drug interactions prediction based on drug embedding and graph auto-encoder," in *Proc. IEEE 19th Int. Conf. Bioinf. Bioengineering (BIBE)*, Oct. 2019, pp. 547–552.

[42] H. He, G. Chen, and C. Yu-Chian Chen, "3DGT-DDI: 3D graph and text based neural network for drug–drug interaction prediction," *Briefings Bioinf.*, vol. 23, no. 3, May 2022, Art. no. bbac134.

[43] T. N. Kipf and M. Welling, "Variational graph auto-encoders," 2016, *arXiv:1611.07308*.

[44] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[45] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," 2016, *arXiv:1608.03983*.

[46] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 1–11.

[47] S. Brody, U. Alon, and E. Yahav, "How attentive are graph attention networks?" 2021, *arXiv:2105.14491*.

[48] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[49] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[50] M. R. Karim, M. Cochez, J. B. Jares, M. Uddin, O. Beyan, and S. Decker, "Drug-drug interaction prediction based on knowledge graph embeddings and convolutional-LSTM network," in *Proc. 10th ACM Int. Conf. Bioinf., Comput. Biol. Health Informat.*, Sep. 2019, pp. 113–123.

[51] J. Zhu, C. Che, H. Jiang, J. Xu, J. Yin, and Z. Zhong, "SSF-DDI: A deep learning method utilizing drug sequence and substructure features for drug–drug interaction prediction," *BMC Bioinf.*, vol. 25, no. 1, p. 39, Jan. 2024.

[52] D. S. Wishart et al., "DrugBank 5.0: A major update to the DrugBank database for 2018," *Nucleic Acids Res.*, vol. 46, no. 1, pp. D1074–D1082, Jan. 2018.

[53] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.

OMKUMAR CHANDRAUMAKANTHAM received the B.Tech. and M.Tech. degrees (Hons.) in CSE from JNTU-Anantapur, in 2010 and 2013, respectively, and the Ph.D. degree in cyber security from Anna University, Chennai, in 2020. He is currently an Assistant Professor (Sr. G) with the School of Computer Science Engineering, Vellore Institute of Technology, Chennai Campus, has a distinguished academic and professional journey. He embarked on his teaching career with the SRM-Easwari Engineering College, Chennai, from 2013 to 2016, where he discovered his passion for teaching and honed his pedagogical skills. His research interests include the IoT and deep learning, leading to the filing of an international patent and the publication of 25 research articles in reputable journals, establishing an H-index value of six.

SRINITISH SRINIVASAN was born in Chennai, India, in 2002. He is currently pursuing the bachelor's degree in computer science and engineering with specialization in artificial intelligence and machine learning with Vellore Institute of Technology, Chennai Campus. He has been involved in research work in agriculture, pattern recognition, and bio-informatics using deep learning. His research interests include optimization of systems, bio-informatics, and generative AI.

VARENYA PATHAK is currently pursuing the bachelor's degree in computer science and engineering with specialization in artificial intelligence and robotics with Vellore Institute of Technology, Chennai Campus. She has been involved in research work in bio-informatics, path planning for autonomous underwater vehicles, and generative AI. Her research interests include optimization systems for real world scenarios, computer vision, and robotics.