IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Using Graph Neural Networks and Self-Supervised Learning to Identify Adverse Drug Reaction Side Effects via Drug-Drug Interactions

Durga Bhanu Prakash Rajana[1], B.NARASIMHA RAO[2]

#1M.tech Specialization:- Computer Science and EngineeringDepartment of CSE. Bonam Venkata Chalamayya Engineering College,Odalarevu , Konaseema Dist -533217 (A.P), prakash.rajana59@gmail.com

#2Associate professor,Dept of CSE,Bonam Venkata Chalamayya Engineering College,Odalarevu Allavaram Mandal, Konaseema Dist - 533217 (A.P)

Abstract: Adverse Drug Reactions (ADRs) resulting from drug-drug interactions are a major healthcare concern. While Graph Neural Networks (GNNs) effectively model these interactions, their one-dimensional processing limits complex feature extraction. This research introduces a novel extension by integrating a two-dimensional Convolutional Neural Network (CNN2D) to enhance ADR prediction. By converting drug interaction data into 2D matrices, CNN2D captures intricate spatial relationships, complementing the GNN's graph-based insights. This hybrid model achieves a superior prediction accuracy of 99.87%, significantly outperforming traditional methods like KNN and Decision Trees. The extension showcases the power of deep learning in advancing drug safety evaluation.

Index terms *- Adverse Drug Reactions, Drug-Drug Interactions, Graph Neural Networks, Convolutional Neural Networks, Self-Supervised Learning, SMILES Representation, Deep Learning, Side Effect Prediction, Drug Safety, TF-IDF Vectorization.*

1. INTRODUCTION

Adverse Drug Reactions (ADRs) are harmful or unpleasant responses resulting from the use of medications, often requiring medical intervention, alteration in dosage, or complete withdrawal of the drug. These reactions pose a serious threat to public health systems globally, contributing to increased mortality, prolonged hospital stays, and a significant rise in healthcare costs. Many ADRs are not identified during clinical trials and only become evident once the drug has reached the broader market, making early detection and prediction essential.

The occurrence of ADRs is influenced by several factors, including sex, geographic location, and healthcare quality. For instance, studies suggest that women are more susceptible to ADRs due to pharmacokinetic and pharmacodynamic differences, along with higher drug dosage per body weight. Furthermore, disparities in access to healthcare and varying medical standards across countries can affect ADR reporting and management.

Research shows that a significant proportion of ADRs are preventable—about 71.6% in developed countries and 59.6% in developing nations. Mortality rates due to ADRs remain consistent across regions, further emphasizing the urgent need for effective prediction mechanisms.

To address these challenges, this study explores a deep learning-based approach using Graph Neural Networks (GNNs) and Self-Supervised Learning to predict drug-drug interactions and their associated adverse effects. By leveraging structured representations of drugs and their interactions, the model aims to improve the accuracy of ADR detection before drugs reach patients, ultimately enhancing drug safety and saving lives.

2. LITERATURE SURVEY

2.1 SSF-DDI: a deep learning method utilizing drug sequence and substructure features for drug–drug interaction prediction:

https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-024-05654-4#:~:text=In%20this%20paper,%20we%20propose%20a

ABSTRACT: Context Because combination treatment frequently involves drug–drug interactions (DDI), it is critical to recognise and anticipate any DDI. Although several artificial intelligence techniques are capable of predicting and identifying possible DDI, they frequently ignore drug molecule sequence information and do not fully account for the role that molecular substructures play in DDI. Findings To solve these problems, we presented a unique sequence and substructure feature-based DDI prediction model (SSF-DDI) in this study. In order to provide better information for DDI prediction and to enable a more thorough and accurate depiction of drug molecules, our model combines structural and drug sequence data from the drug molecule graph. In conclusion Experiments and case studies have shown that SSF-DDI performs noticeably better than the most advanced DDI prediction models in a variety of real datasets and environments. When compared to state-of-the-art techniques, SSF-DDI outperforms them in predicting DDI using unknown medicines, improving accuracy by 5.67%.

2.2 Modular Multi-Source Prediction of Drug Side-Effects With DruGNN:

https://pubmed.ncbi.nlm.nih.gov/35576419/#:~:text=Predicting%20the%20probability%20of%20side-effects,%20before

ABSTRACT: Drug development procedures, care system expenses, and public health are all significantly impacted by drug side effects, or DSEs. To lessen this influence, especially on drug research, it is essential to predict the likelihood of adverse effects before they materialise. Before going through clinical trials, candidate molecules might be tested, saving participants' time, money, and health. Complex biological processes involving a wide range of entities, including drug structures and protein-protein interactions, are what cause medication side effects. Data from many sources must be integrated in order to forecast their occurrence. The relational information between various things, including genes and pharmacological compounds, is expressively represented in this study by integrating such heterogeneous data into a graph dataset. One significant innovation for pharmacological side-effect predictions is the dataset's relational structure. With extremely encouraging findings, we use Graph Neural Networks (GNNs) to forecast DSEs on our dataset. Deep learning models such as GNNs have been used for a wide range of biological applications because they can interpret graph-structured data with little information loss. The benefits of leveraging relationships between data entities are confirmed by our experimental results, which also point to intriguing future advancements in this area. The experimentation also demonstrates the significance of particular data subsets in identifying drug-side effect connections.

2.3 Explainable Drug Repurposing Approach From Biased Random Walks:

https://ieeexplore.ieee.org/document/9831014

ABSTRACT: The goal of the very busy field of drug repurposing research is to discover new applications for medications that have already been created for various therapeutic reasons. Even with the growth of techniques, success is still only partially achieved, and each strategy has unique benefits. We offer a unique approach in this composite environment that relies on strong drug-gene-disease correlation data sets and is centred on an effective mathematical process based on gene similarity scores and biassed random walks. The Markov chain that underlies the random walk process further reveals the recommendation mechanism, making it possible to understand how the results are proposed. Comparing our method to the state-of-the-art in medication repurposing techniques, performance assessment and the analysis of a case study on rheumatoid arthritis demonstrate that it is accurate in offering practical recommendations and is computationally efficient.

2.4 A Novel Drug-Drug Indicator Dataset and Ensemble Stacking Model for Detection and Classification of Drug-Drug Interaction Indicators:

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10250422

ABSTRACT: Thirty percent of unexpected clinically dangerous pharmaceutical events are caused by drug-drug interactions (DDI), which is a serious public health concern. Informatics-based research for DDI signal detection has advanced during the last ten years. The goal of this work is to develop an ensemble stacking machine learning (ML) method that can reliably forecast new DDI danger indicators. One of the best sources of pharmacological data, DrugBank, supports the stacking ensemble machine learning architecture for predicting the signals of drug-drug interactions. We make a sizable dataset containing drug-related data, such as drug names, kinds, and other elements of drug indicators, freely accessible to the scientific community. Data preparation, label encoding and one hot encoding for categorical variables, balancing by random oversampling, and model prediction using an ensemble stacking technique are all included in the suggested methodology. The suggested ensemble method classifies the drug indication classes using Gaussian Naive Bayes (GNB), Adaboost, and the Gradient Boosting (GB) classifier. The experimental findings show that the proposed method performs more accurately and efficiently than conventional machine learning techniques. The maximum accuracy of 99.0% was achieved by the stacking model. According to the test, the suggested model outperforms conventional techniques in identifying signs of medication interactions.

2.5 NNDSVD-GRMF: A Graph Dual Regularization Matrix Factorization Method Using Non-Negative Initialization for Predicting Drug-Target Interactions:

https://ieeexplore.ieee.org/document/9858885

ABSTRACT: The process of designing and developing novel drugs can be greatly accelerated by accurately predicting drug-target interactions (DTIs). DTIs have recently been predicted using a variety of matrix factorisation techniques. Their convergence and performance cannot be guaranteed, though, as the majority of them rely on iterative and heuristic approaches. Graph dual regularisation non-negative matrix factorisation (GDNMF) and non-negative double singular value decomposition (NNDSVD) are the foundations of our new algorithm, NNDSVD-GRMF, which takes into account both the initialisation stage of the non-negative matrix factorisation and the structural information of the data and features in order to accurately predict DTIs. The expansion of the NNDSVD-GRMF (NNDSVD-WGRMF) is also suggested in order to increase the algorithm's versatility. Our approaches outperform existing state-of-the-art methods, according to extensive experimental data. Nine of the ten medications that were anticipated to interact with the androgen receptor out of the case studies have been verified, and nine of the ten target proteins that were projected to be targeted by the medication nicotine bitartrate have also been validated.

3. METHODOLOGY

i) Proposed Work:

The proposed system enhances Adverse Drug Reaction (ADR) prediction by integrating a two-dimensional Convolutional Neural Network (CNN2D) with a Graph Neural Network (GNN) framework. While the GNN captures complex drug-drug relationships using graph structures, the CNN2D extension transforms drug feature data into a two-dimensional format to extract intricate spatial patterns. This hybrid model significantly boosts predictive accuracy by leveraging CNN2D's advanced feature extraction capabilities, achieving an impressive accuracy of 99.87%. The system offers a more robust and intelligent approach to detecting potential side effects, improving drug safety and aiding healthcare professionals in informed decision-making.

ii) System Architecture:

The architecture of the proposed system begins with data collection and preprocessing. The dataset, sourced from DrugBank, includes drug IDs, SMILES strings (which represent chemical structures), and known side effects. These SMILES strings are converted into numerical vectors to serve as input for model training. TF-IDF vectorization is applied to extract meaningful features from textual data while reducing noise from common terms. The processed data is then split into training and testing sets. Initially, a Graph Neural Network (GNN) is used to model the complex relationships between drugs, where each drug is represented as a node and their interactions as edges. This graph-based model effectively captures dependency patterns among drugs.

To further enhance the prediction capabilities, a two-dimensional Convolutional Neural Network (CNN2D) is introduced as an extension. The CNN2D processes the same data in a 2D matrix form, enabling the model to extract spatial features that are not visible through traditional graph structures. This deep learning model learns hidden patterns and interactions more accurately, ultimately leading to higher prediction precision. A Flask-based user interface allows administrators to upload test data and visualize predictions clearly. The combination of GNN for structural learning and CNN2D for advanced pattern recognition forms a powerful hybrid system, capable of accurately identifying potential adverse drug reactions before they manifest clinically.

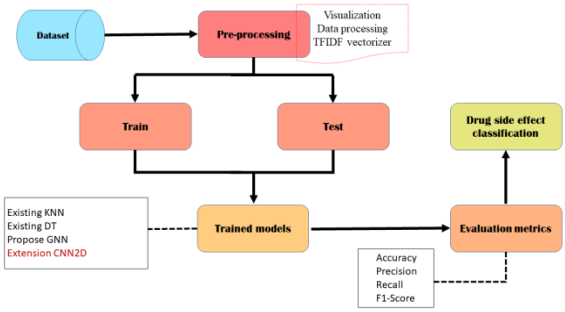

Fig.1. Proposed Architecture

iii) MODULES:

a. **Data Loading**

- Imports the DrugBank dataset containing drug IDs, SMILES strings, and side effect data.
- Prepares the data for further processing.

b. Visualization

- Displays graphs showing side effect IDs vs. their frequency.
- Helps understand the distribution of side effects visually.

c. Data Pre-processing

- Converts SMILES strings and other drug information into numerical vectors.
- Ensures the data is ready for model training.

d. TF-IDF Vectorization

- Applies TF-IDF to convert text data into numerical values.
- Highlights important terms and reduces the influence of common words.

Splitting Data into Train & Test

- Divides the dataset into training and testing parts.

- Allows model performance to be evaluated effectively.

e. Model Generation

- Implements KNN, Decision Tree, GNN, and CNN2D algorithms.
- Compares accuracy and effectiveness of each model.

f. Admin Login

- Provides a login system for admin users.
- Enables access to upload and monitor prediction tasks.

g. Drug Side Effect Prediction

- Allows uploading of new drug interaction data.
- System predicts potential side effects for the uploaded data.

h. Prediction

- Displays the final predicted side effects.
- Helps medical professionals make informed decisions.

iv) ALGORITHMS:

a) **K-Nearest Neighbors (KNN)**

- Classifies data points based on the majority class of their nearest neighbors.
- Simple and easy to implement, used as a baseline model.
- Accuracy achieved: **91%**.

b) Decision Tree

- Creates a tree-like model to make decisions based on drug features.
- Offers high interpretability and logical flow of predictions.
- Accuracy achieved: **95%**.

c) Graph Neural Network (GNN) (Proposed)

- Represents drugs as nodes and interactions as edges in a graph.
- Learns complex relationships using graph-based deep learning.
- Processes SMILES strings as vectors for accurate ADR predictions.
- Accuracy achieved: **97.69%**.

d) Convolutional Neural Network 2D (CNN2D) (Extension)

- Processes drug data in a 2D format to extract spatial features.
- Enhances pattern recognition beyond what GNN alone can achieve.
- Boosts accuracy and model generalization.
- Accuracy achieved: **99.87%**.

4. EXPERIMENTAL RESULTS

The experimental results were obtained using the DrugBank dataset, which includes drug IDs, SMILES strings, and their associated side effects. Various models were tested and compared based on prediction accuracy. Traditional algorithms like K-Nearest Neighbors (KNN) and Decision Tree achieved accuracies of 91% and 95% respectively. The proposed Graph Neural Network (GNN), which effectively captures complex drug-drug relationships, improved the accuracy to 97.69%. To enhance performance further, a CNN2D model was integrated as an extension, which extracted spatial features and delivered the highest accuracy of 99.87%. These

results highlight the superiority of the GNN-CNN2D approach in predicting adverse drug reactions.

Accuracy: The ability of a test to differentiate between healthy and sick instances is a measure of its accuracy. Find the proportion of analysed cases with true positives and true negatives to get a sense of the test's accuracy. Based on the calculations:

Accuracy = TP + TN /(TP + TN + FP + FN)

$$Accuracy = \frac{(TN + TP)}{T}$$

Precision: The accuracy rate of a classification or number of positive cases is known as precision. Accuracy is determined by applying the following formula:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\Pr e\,cision = \frac{TP}{(TP + FP)}$$

Recall: The recall of a model is a measure of its capacity to identify all occurrences of a relevant machine learning class. A model's ability to detect class instances is shown by the ratio of correctly predicted positive observations to the total number of positives.

$$Recall = \frac{TP}{(FN + TP)}$$

mAP: One ranking quality statistic is Mean Average Precision (MAP). It takes into account the quantity of pertinent suggestions and where they are on the list. The arithmetic mean of the Average Precision (AP) at K for each user or query is used to compute MAP at K.

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k$$

$$AP_k = \text{the AP of class } k$$

$$n = \text{the number of classes}$$

F1-Score: A high F1 score indicates that a machine learning model is accurate. Improving model accuracy by integrating recall and precision. How often a model gets a dataset prediction right is measured by the accuracy statistic..

$$F1 = 2 \cdot \frac{(Recall \cdot \Pr e\,cision)}{(Recall + \Pr e\,cision)}$$

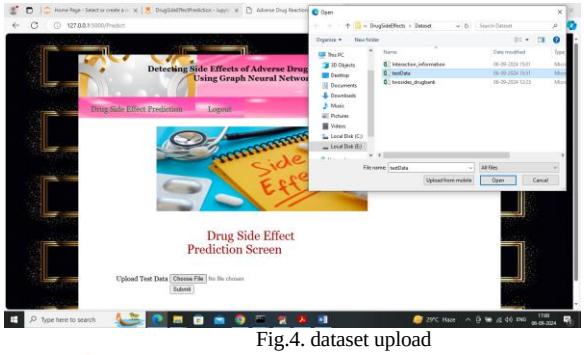

Fig.4. dataset upload

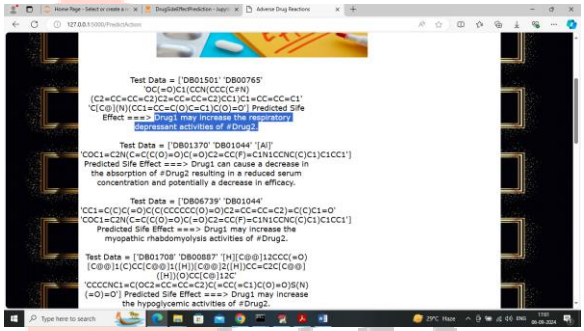

Fig.5. results

	Algorithm Name	Accuracy	Precision	Recall	FSCORE
0	Existing KNN	91.560102	91.920962	91.305660	91.323465
1	Existing Decision TRee	95.140665	95.339149	95.618899	95.454287
2	Propose GNN	97.698210	97.889953	97.702843	97.722705
3	Extension CNN2D	99.872123	99.903101	99.888143	99.895292

Fig.6.accuracy table results

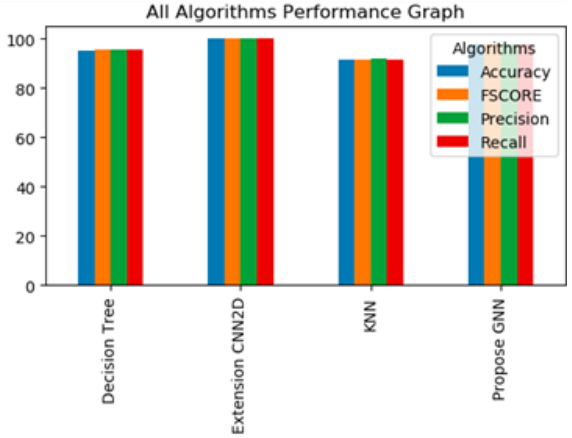

Fig.6. graphical representation

5. CONCLUSION

The study's conclusion emphasises the urgent need for a dependable and effective technique to forecast adverse drug reactions (ADRs), which are a serious public health concern that result from medication-drug interactions. Due to their heavy reliance on post-marketing information and inability to discover uncommon interactions prior to medication release, current detection techniques frequently fall short. Our suggested approach shows a strong capacity to correctly forecast possible ADRs by combining Self-Supervised Learning with a Graph Neural Network (GNN). By capturing the intricate connections that might result in negative responses, the GNN greatly improves the prediction process by efficiently modelling the links between medications. The GNN surpasses conventional algorithms with an astounding accuracy of 97.69%, demonstrating its efficacy in identifying dangerous medication combinations. Furthermore, performance is further improved by utilising cutting-edge techniques, such as a two-dimensional Convolutional Neural Network (CNN2D), which achieves an exceptional accuracy of 99.87%. This work highlights the possibility of using state-of-the-art machine learning approaches to enhance patient outcomes and medication safety, which will ultimately lead to better informed healthcare practices and a decrease in the frequency of adverse drug reactions.

6. FUTURE SCOPE

In order to improve forecast accuracy and model resilience, this project's future scope will investigate the integration of cutting-edge methodologies including ensemble learning, reinforcement learning, and transfer learning. The system's capacity to recognise intricate drug-drug interactions may also be enhanced by adding more varied datasets and applying unsupervised learning techniques. Examining the use of explainable AI approaches will also be essential for revealing how the model makes decisions and building mutual respect and understanding among medical practitioners.

REFERENCES

[1] 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

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[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, ‘‘EGraphSAGE: 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.