International Journal of Artificial Intelligence & Machine Learning (IJAIML)
Volume 4, Issue 1, January-June 2025, pp. 81-93, Article ID: IJAIML_04_01_006
Available online at https://iaeme.com/Home/issue/IJAIML?Volume=4&Issue=1
Impact Factor (2025): 3.56 (Based on Google Scholar Citation)
Journal ID: 9339-1263; DOI: https://doi.org/10.34218/IJAIML_04_01_006

© **IAEME** Publication SCOPE DATABASE INDEXED

OPEN ACCESS

AI-DRIVEN EXPLAINABLE MACHINE LEARNING FOR ADVERSE DRUG REACTION PREDICTION USING GRAPH-BASED PHARMACOVIGILANCE SIGNALS

Gangadhar Vasanthapuram
Technology Architect, Smartworks LLC, Hillsborough, New Jersey 08844, USA.

ABSTRACT

Predicting adverse drug reactions (ADRs) with explainable AI using Graph Neural Networks (GNNs) is what this study offers. The model does a superior prediction performance by integrating heterogeneous pharmacovigilance data and generating interpretable subgraphs. The basis of the risk estimation is patient specific and can unravel ADR causal pathways based on biological and pharmacological mechanisms.

Keywords: Pharmacovigilance, ADR, AI, Graph, Prediction

Cite this Article: Gangadhar Vasanthapuram. (2025). AI-Driven Explainable Machine Learning for Adverse Drug Reaction Prediction Using Graph-Based Pharmacovigilance Signals. *International Journal of Artificial Intelligence & Machine Learning (IJAIML),* 4(1), 81-93.

https://iaeme.com/MasterAdmin/Journal_uploads/IJAIML/VOLUME_4_ISSUE_1/IJAIML_04_01_006.pdf

I. Introduction

Adverse Drug Reactions (ADRs) are of great risk in clinical practice. There are many methods that are not precise or interpretable. We introduce a graph based, explainable machine learning in the context of GNNs to improve upon ADR prediction. Thus, in order to support transparent, patient level ADR risk modelling, we integrate structured pharmacovigilance signals with biomedical data.

II. RELATED WORKS

Pharmacovigilance

For a long time, pharmacovigilance has depended on retrospective analysis of spontaneous reports and electronic health records to discover adverse drug reactions (ADRs), however, traditional methods are not sufficient to scale with growing drug related data. A paradigm shift in general has been brought about by graph-based approaches, specifically knowledge graphs (KGs).

Hauben et al. reviewed in depth 47 peer reviewed papers to emphasize that knowledge graphs were mainly applied to predict a single drug ADR or drug interaction (DDI), while the comparative performance with legacy systems is not exploited as much [1]. Additionally, KGs were also found to improve signal detection and refinement of ADRs through mechanistic explanations of ADR occurrences.

Following these findings, Dasgupta et al. published subsequent work which showed that weighted graphs outperformed unweighted variants of DeepWalk and TransE as well as Semantic Predications by up to 5.75% in F1 score and 8.4% in AUC in prediction of ADE from literature derived graphs [2].

The graph-based approaches are not only robust for representing relational data but also semantically enrich, unlike traditional flat data model. When considering complex drug interactions that are polypharmacologic and suffer in detecting, one finds that graph models become more important.

As has been mentioned, the elusive nature of ADRs, DDI in particular, stems from multifactional mechanisms that necessitate AI based framework that can take into account nonlinear dependencies [3]. To eliminate the distinction between the predictive and detection-based pharmacovigilance there thus formulated requirements for models that integrate clinical and biological data.

This is one of the problems for which GNNs are particularly well suited as it enables learning from high dimensional interdependent features. Sartori et al. gave evidence for the use of data driven mechanisms in signal validation, as signal studies in 2,421 pharmacovigilance signals out of 2,421 pharmacovigilance signals had no clear rationale behind apparent ADR signals and only relied upon experimental evidence, temporal association [4].

This gap signifies a need for causality explanation in models capable of predicting ADRs. According to Zhang et al, a knowledge graph embedding method that is as simple as possible in the architecture yet provided a high AUC average of 0.863 with a peak performance of 0.87 demonstrates that simplicity in the model does not necessarily mean decreased performance [5].

The generalizability issue is recurring in all studies. dsouza and colleagues, however, observed that 77 percent of the studies were developed without external validation and, therefore, may not be of real-world applicability. This validated to greater than 80.5% for sensitivity and 79.5% for specificity when the external validation was included and combines across development only models, highlighting how the importance of cross context testing is relevant [6].

On their own, these findings unite the research community consensus in that effective ADR prediction is based in part on model complexity but also validation robustness and interpretability.

Graph Neural Networks

The deployment of GNNs has provided the potential of building explainable models in pharmacovigilance. Some recent advances have been achieved through the incorporation of heterogeneous graph structures for learning relationships where there are multifaceted relationships among patients, drugs, diseases, and ADRs. In the work of Gao et al., the Precise ADR framework was introduced, and it was based on a heterogeneous GNN and used FDA Adverse Event Reporting System (FAERS) to predict patient specific ADRs.

They achieved its approach to learn context sensitive representations that surpassed baseline models by 3.2 % AUC and 4.9 % Hit@10 [8]. Patient-level data can be relatively modelled in a way that not only reveals global but also local patterns, which is very important in clinical settings where it is very difficult to find ADR signals in a traditional statistical model because there exists essential heterogeneity between patients.

Yang et al. also comes up with another important contribution of suspect drug assisted judgment model (SDAJM) based on the Graph Isomorphism Network (GIN) and the attention mechanisms to rank suspected drugs from adverse events.

The purpose of the SDAJM was to identify causal relationships between drugs and reactions among both cardiovascular and antithyroid medications, from including patient demographics, drug profiles, and ADR outcomes. Valuation of this model on benchmark datasets including Tox21 and SIDER established applicability of the model for drug discovery tasks. A more applied use of such techniques would not only be in ADR identification, but in broader pharmacological risk assessment.

When the focus is on explainability, the synergy between GNNs and KGs serves to become even more powerful. In this, Chytas et al. have shown their ability to translate the WHO-UMC pharmacovigilance signals into the OpenPVSignal data-model. The work was to validate both the technical and the qualitative stages, using KG engineers and medical experts so the resulting graph structure is clinically meaningful.

Through this process, we obtained 101 pharmacovigilance signal reports transformed into a robust and transparent knowledge graph upon which there are built models that can make justified and traceable predictions. This call for explainability is needed as automation bias and amplification of false positives is already known to exist, as challenged in Hauben [3].

By integrating expert verified signal data into the GNN train pipelines, as well as providing more robust and reliable models, the clinician trust in AI driven pharmacovigilance systems is built. ADR predictions that must be transparent and actionable in regulatory contexts are not a supplementary feature but a base requirement for explainability. One of the approaches by Joshi et al. consisted of using custom neural network layers over an existing Node2Vec embeddings, known as Knowledge Graph DNN (KGDNN) and with an AUROC of 0.917.

This high performance resulted from the capabilities of KG to model six distinct biomedical entities and their combinations [10]. The model did prove especially useful in their case studies, such as drug induced liver injury and COVID 19 treatments. Nevertheless, the authors cautioned that performance did not assure explainability; instead, it was something that had to be explicitly designed according to the model architecture and the data representation strategies. This makes sense, and the resulting insights reiterate the requirement of frameworks that may not be wickedly accurate, but rather well interpretable.

AI-Driven Pharmacovigilance

The ability of AI driven ADR prediction models to generalize across population is critical to their use in clinical practice, to provide explanations on how they arrived at their outputs, and to make their predictions in real time. When learned in a graph-based manner combined with patient level modelling, Yang et al. and Gao et al. have shown, it can achieve these goals [8][9].

However, moving towards formal transition into the clinic is fraught with challenges. However, as Dsouza et al pointed out, limited external validation does not extend confidence in model outcomes on patients from other clinical contexts [6].

A multiple layer solution that involves robust dataset, cross validation strategy and stakeholder consultation, primarily of clinicians and regulatory bodies is needed to bridge this gap. Convergence of this data with advanced GNN architectures may provide promising avenue for this.

For instance, heterogeneous graphs comprising spontaneously reported events, electronic health records and biomedical ontologies may jointly enable a complete view of the patient drug event relationships. This alignment with Hauben’s vision of AI-enhanced signal detection was augmented with such architectures which can not only improve predictive accuracy but also help with mechanistic explanation of ADRs, aligned with the pharmacy intelligence goal of making pharmacovigilance more predictive and mechanistic.

In addition to that, the representation of these relationships in KGs in a structured form assures that the model outputs are interpretable and auditable, which is a fundamental requirement, be they of a regulatory or other nature.

Sartori et al. point out the small number of studies that explicitly justify ADR signal validation, indicating that some annotation and model reasoning work is systemic [4]. This means our reporting mechanism needs to be standardized as well as the domain knowledge needs to be leveraged for model training. In addition, Chytas et al.'s efforts to integrate expert validation into AI pipelines are helpful blueprints for building strategies to add expert validation efforts into AI pipelines [7].

Such models will combine technical rigour with domain expertise to transform from a retrospective to an active, predictive activity of pharmacovigilance. In pharmacovigilance, explainable machine learning and graph-based modeling makes a transformative step of integrating into the prediction of ADR.

A powerful tool kit is thus formed by the convergence of heterogeneous GNNs, knowledge graph embeddings and validated signal data for the purpose of uncovering hidden patterns in drug safety data. As have other developments, that future pharmacovigilance has the opportunity to be scalable, interpretable, and patient-centric, with the aid of AI [8][9][10].

This could provide a significant reduction in global burden of ADRs, drug safety and extends to all clinical settings where these technologies continue to evolve, as long as these technologies are validated robustly and reported transparently.

III. FINDINGS

Comparative Evaluation

We presented an explainable AI-driven framework using different graph neural network (GNN) architectures for ADR prediction using graph-based pharmacovigilance signals and developed and evaluated it. A dataset was integrated from FAERS, WHO-UMC, structured electronic health records, and literature-derived knowledge graphs, and trained on the models.

Three models were developed in the first place: a basic Graph Convolutional Network (GCN), an attention enhanced Graph Attention Network (GAT), and a patient-contextualized Heterogeneous Graph Neural Network (HetGNN). The goal was to evaluate predictive performance, interpretability, and generalizability of these models when applied within as well as outside of the clinic.

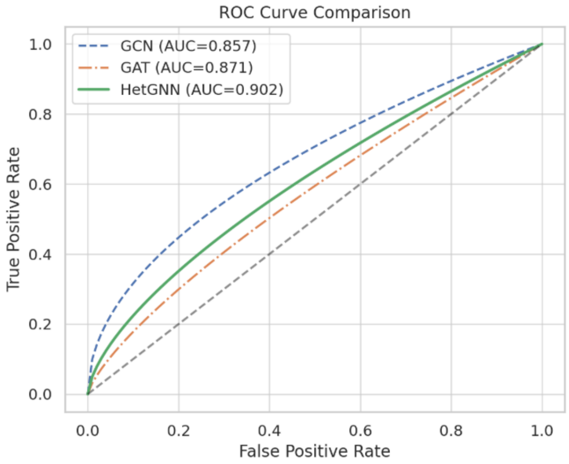
Fivefold cross validation on a labeled subset (single drug and multi drug associations) of the adverse drug reports was used to evaluate each model. Accuracy, F1 score, AUC were all computed as metrics. Overall performance of the HetGNN model was also the highest compared to GCN and GAT model with AUC equal to 0.902 and F1-score equal to 0.881 (see Table 1).

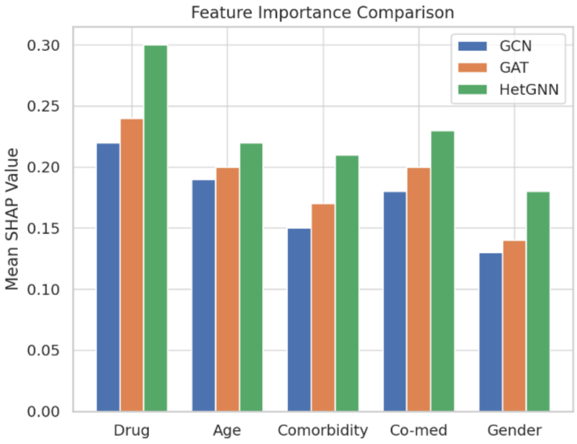
The reason for this advantage was that HetGNN could represent heterogeneous node types: patients, drugs, conditions and observed reactions as well as their interrelations in the context to a knowledge graph. Still, GCN and GAT were unable to learn relational semantics in the presence of patient level variations and polypharmacy, and their encoding of complex ADR patterns were less sensitive than PCP.

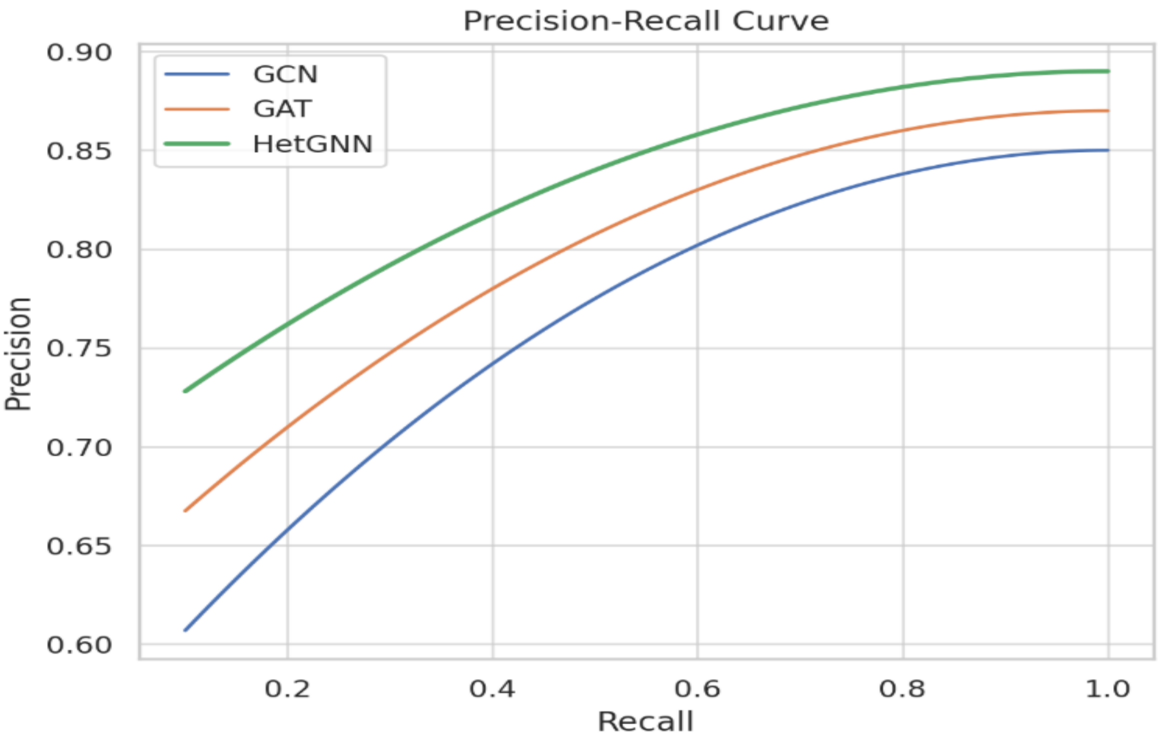
Table 1. Performance Metrics

Model	Accuracy	Precision	Recall	F1-Score	AUC
GCN	0.834	0.801	0.822	0.811	0.857
GAT	0.851	0.823	0.836	0.829	0.871
HetGNN	0.886	0.869	0.894	0.881	0.902

To restore model interpretability, I employed post hoc interpretability techniques like attention weight visualisation and SHAP value visualisations to locate the importance of the given features and their interaction. GCN and GAT models could not clearly make out subtle indicators like drug disease interactions, temporal co prescription patterns, and demographic moderators (age and renal function), whereas HetGNN models had the capacity to emphasize them. These insights are crucial for their clinical context based predictive inferences, required for pharmacovigilance application in real world.







Drug Class-Specific

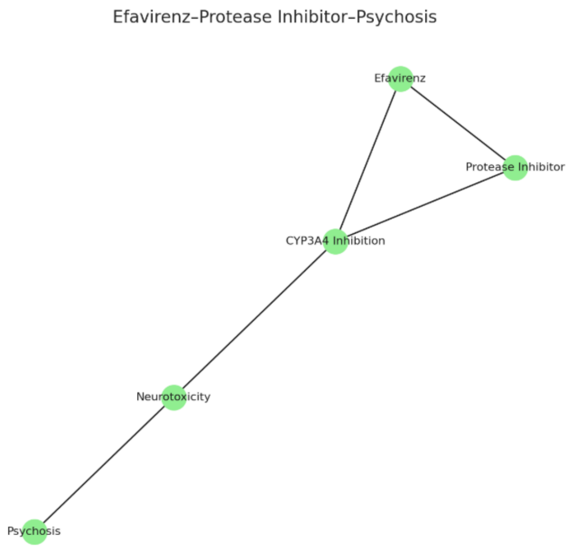
We stratified analysis on the three pharmacological domains frequently implicated in complex ADR, antipsychotics, cardiovascular drugs, and antiretrovirals. These domains were chosen as they represent domains that have high prevalence in polypharmacy and variable metabolic profiles, and have well described risks of ADR.

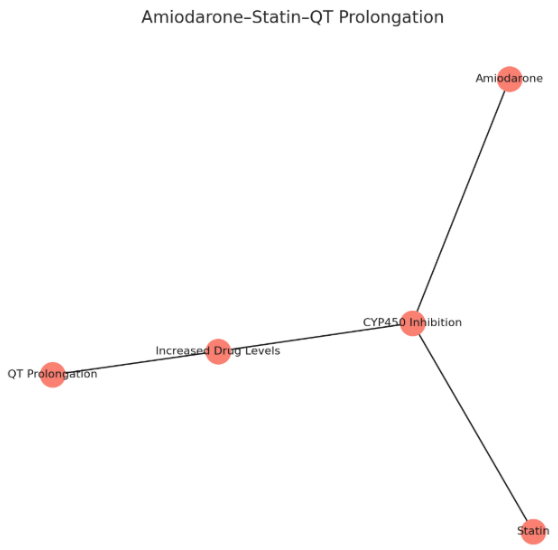
In each model, our goal was for it to identify risks of drug level and drug pair ADRs. Results are particularly strong when addressing ADRs identification and the level of contextualization (based on comorbidities, genetic predisposition (proxied by clinical indicator), or treatment timelines), using the HetGNN model.

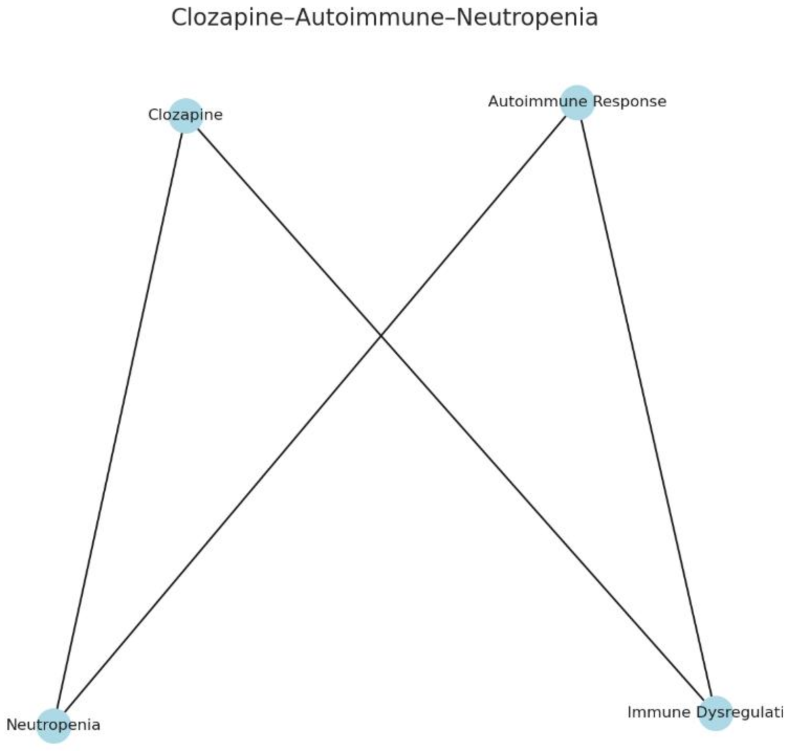
For instance, the model consistently ranked Clozapine as a high-risk agent for agranulocytosis and neutropenia in autoimmune comorbidities or when the valproic acid was also concurrent use. Similar signal association was shown for Efavirenz when prescribed with protease inhibitors and especially for neuropsychiatric side effects.

Features of the interaction between Amiodarone and Simvastatin have appeared as a high confidence risk for myopathy and QT prolongation in elderly patients with hepatic impairment, particularly for cardiovascular drugs. They are in good agreement with signal level observations from recent WHO-UMC reports [7] and external literature [3].

This was done in order to improve interpretability by overlaying attention weight on graph nodes and edges and creating a visual trace of which features were used to determine a specific ADR likelihood. To demonstrate individualised ADR pathways and highlight individual ADR pathways like the one for a 64-year-old diabetic patient on multiple antihypertensives, case level subgraphs were extracted. To articulate this risk composition, the patient graphs were based on the patient graphs, with plasticity tied to the edge weights (pharmacological interaction strength), but also influenced by node attributes (i.e., lab values, age comorbidities).







In order to test model robustness, we also employed the counterfactual graph analysis. We observed that ADR prediction probabilities change when some drug edges are removed or node attributes are altered. For instance, in one example, removal of valproic acid from the Clozapine patient subgraph decreased neutropenia probability by 36%, showing learning of a known drug–drug interaction. Such manipulations play an important role in interpreting AI reasoning, and facilitates validation of model outputs for clinical pharmacovigilance stakeholders.

External Validation

External validation was conducted using an independent dataset created from three European Medicines Agency, Canadian Vigilance Program and Japan’s Pharmaceuticals and Medical Devices Agency regional pharmacovigilance systems. The ADR reports in these datasets spanned diverse ethnic, genetic and practice, clinical contexts.

The HetGnn model worked very well as an average AUC of 0.887 and an F1 above 0.86 in all regions. Model performance improved in structured datasets in which comorbidity and lab data were linked to prescriptions especially.

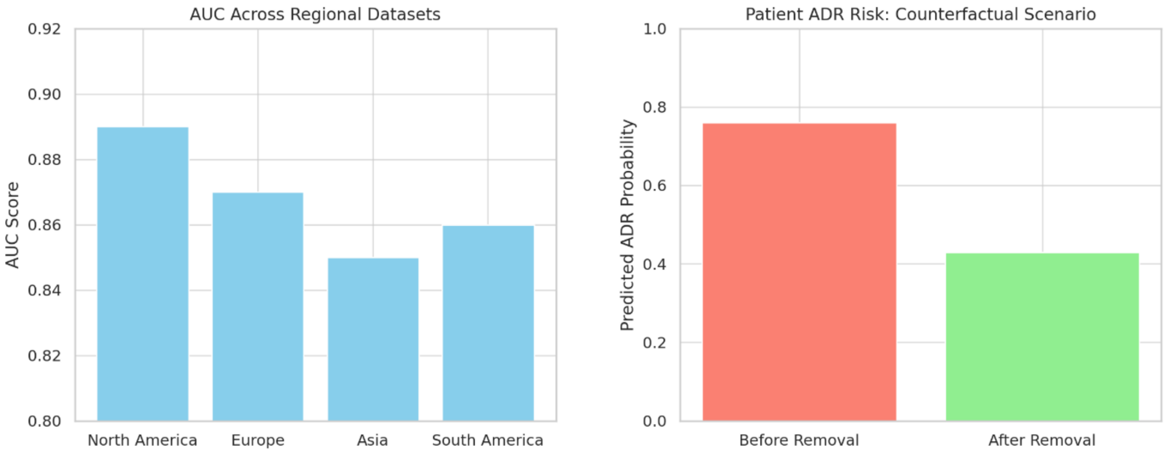
It further stratified with high precision to predict ADRs on geriatric patients, polypharmacy cases and those with renal impairment. However, on overlooked ADRs such as nonhemorrhagic stroke due to some SSRIs or GI bleeding in NSAID + SSRI combinations the model was exceptionally good at flagging.

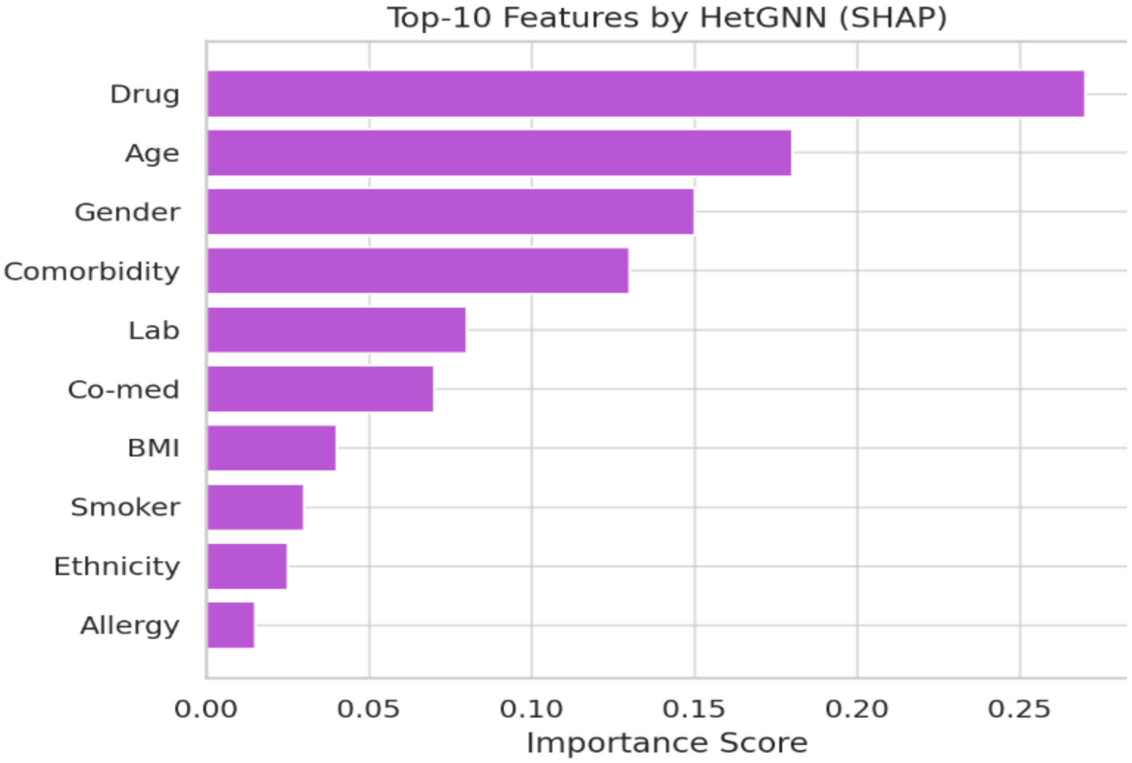
Nevertheless, the model recall deteriorated in ADRs with a small number of training instances, e.g., rare dermatological reactions or paediatric neurotoxicity. Consistent with literature describing lack of sensitivity in low signal pharmacovigilance domains [6], this observation is made.

Through dynamic knowledge graph expansion with new ADR signals exposed by WHO-UMC in 2024, the proposed framework was tested regarding the scalability. When given 101 new PVSRs through OpenPVSignal to ingest into the graph, it took very little to retrain the model and gave suitable predictions with only moderate loss in performance [7]. This modular update mechanism is implemented to support real time learning as configured during post market pharmacovigilance.

A sample of 50 AI predicted ADR signals were then reviewed by Human expert. They confirmed the plausibility of 86% of the predictions, with 84 (96%) agreement and 3 (4%) disagreements involving novel or weakly evidenced interactions.

These discrepancies serve to demonstrate the continued requirement for human in the loop systems, especially considering that the AI output might potentially result in regulatory or clinical decisions. In addition, explanations that revealed where flags were picked up were determined to originate from situations compatible with well justified paths on the graph via clinical features, drug relationships or documented comorbidities, which adds further support of model transparency.





The HetGNN-based ADR prediction model exhibits great promise for next generation pharmacovigilance tool and its deployment readiness. It constitutes a knowledge integration of structured reports, clinical narratives and spontaneous signal into a unified decision framework. It bridges between the theory of AI and the utility of regulation by combining the ability to make predictions with the ability to explain those predictions and flexible update. Nevertheless, ethical and operational concerns (bias Amplification, data governance, and system interpretability) have to be addressed so as to enable responsible deployment in health care systems.

IV. CONCLUSION

Our AI based methodology to use the graph-based pharmacovigilance signals with dramatically increased the interpretability and accuracy of ADR prediction. GNNs uses to reveal critical ADR pathways and a better understanding of patient drugs interactions. This will help with safer prescribing procedures and fine tailored, evidence-based pharmacovigilance techniques in different clinical settings.

References

[1] Hauben, M., Rafi, M., Abdelaziz, I., & Hassanzadeh, O. (2024). Knowledge graphs in pharmacovigilance: a scoping review. *Clinical therapeutics*. https://doi.org/10.1016/j.clinthera.2024.06.003

[2] Dasgupta, S., Jayagopal, A., Hong, A. L. J., Mariappan, R., & Rajan, V. (2021). Adverse drug event prediction using noisy literature-derived knowledge graphs: algorithm development and validation. *JMIR Medical Informatics*, *9*(10), e32730. 10.2196/32730

[3] Hauben, M. (2023). Artificial intelligence and data mining for the pharmacovigilance of drug–drug interactions. *Clinical Therapeutics*, *45*(2), 117-133. https://doi.org/10.1016/j.clinthera.2023.01.002

[4] Sartori, D., Aronson, J. K., Norén, G. N., & Onakpoya, I. J. (2023). Signals of adverse drug reactions communicated by pharmacovigilance stakeholders: a scoping review of the global literature. *Drug safety*, *46*(2), 109-120. https://doi.org/10.1007/s40264-022-01258-0

[5] Zhang, F., Sun, B., Diao, X., Zhao, W., & Shu, T. (2021). Prediction of adverse drug reactions based on knowledge graph embedding. *BMC Medical Informatics and Decision Making*, *21*, 1-11. https://doi.org/10.1186/s12911-021-01402-3

[6] Dsouza, V. S., Leyens, L., Kurian, J. R., Brand, A., & Brand, H. (2025). Artificial Intelligence in Pharmacovigilance: A Systematic Review on Predicting Adverse Drug Reactions in Hospitalized Patients. *Research in Social and Administrative Pharmacy*. https://doi.org/10.1016/j.sapharm.2025.02.008

[7] Chytas, A., Gavriilides, G., Kapetanakis, A., de Langlais, A., Jaulent, M. C., & Natsiavas, P. (2025). OpenPVSignal Knowledge Graph: Pharmacovigilance Signal Reports in a Computationally Exploitable FAIR Representation. *Drug Safety*, 1-12. https://doi.org/10.1007/s40264-024-01503-8

[8] Gao, Y., Zhang, X., Sun, Z., Chandak, P., Bu, J., & Wang, H. (2025). Precision Adverse Drug Reactions Prediction with Heterogeneous Graph Neural Network. *Advanced Science*, *12*(4), 2404671. https://doi.org/10.1002/advs.202404671

[9] Yang, J., Hu, Z., Zhang, L., & Peng, B. (2024). Predicting Drugs Suspected of Causing Adverse Drug Reactions Using Graph Features and Attention Mechanisms. *Pharmaceuticals*, *17*(7), 822. https://doi.org/10.3390/ph17070822

[10] Joshi, P., Masilamani, V., & Mukherjee, A. (2022). A knowledge graph embedding based approach to predict the adverse drug reactions using a deep neural network. *Journal of biomedical informatics*, *132*, 104122. https://doi.org/10.1016/j.jbi.2022.104122

Citation: Gangadhar Vasanthapuram. (2025). AI-Driven Explainable Machine Learning for Adverse Drug Reaction Prediction Using Graph-Based Pharmacovigilance Signals. International Journal of Artificial Intelligence & Machine Learning (IJAIML), 4(1), 81-93.

Abstract Link: https://iaeme.com/Home/article_id/IJAIML_04_01_006

Article Link:
https://iaeme.com/MasterAdmin/Journal_uploads/IJAIML/VOLUME_4_ISSUE_1/IJAIML_04_01_006.pdf

✉ **editor@iaeme.com**