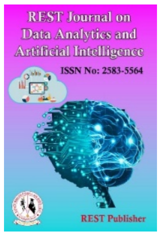
REST Journal on Data Analytics and Artificial Intelligence
Vol: 4(1), March 2025
REST Publisher; ISSN: 2583-5564
Website: http://restpublisher.com/journals/jdaai/
DOI: https://doi.org/10.46632/jdaai/4/1/46



Predicting Gestational Diabetes Using GNN: A Graph-Theoretic Perspective on Patient Data

***M S Adityan, Akshith Reddy, Abdul Ansari, Victor Daniel**

School of Engineering, Anurag University, Hyderabad, Telangana, India.

*Corresponding Author Email: 21eg107a27@anurag.edu.in

Abstract: *This study presents a novel approach to predict gestational diabetes using Graph Neural Networks (GNNs) and graph theory. The dataset, sourced from a medical database, includes features like age, pregnancies, BMI, HDL levels, and blood pressure. Data preprocessing involves filling missing values with column means and standardizing features. A graph is constructed with patients as nodes, each characterized by their features and class labels (GDM or Non-GDM). Edges represent significant relationships based on Euclidean distance. Visualization using a spring layout reveals clusters of similar patients, enhancing understanding of gestational diabetes predictors and potentially improving early diagnosis and patient management.*

Keywords: *Graph Neural Networks (GNNs), Graph Convolutional Networks (GCNs), Graph Attention Networks (GATs), Semi-Supervised Learning, Feature Propagation, Node Classification, Edge Weighting, Adjacency Matrix, Graph Embeddings, Message Passing, Hidden Layer Representations, Spectral Graph Theory, Graph Laplacian, Regularization Techniques, Optimization Algorithms, Loss Functions, Hyperparameter Tuning, Activation Functions, Training and Validation Split, Evaluation Metrics (Accuracy, AUC-ROC, F1-Score).*

1. INTRODUCTION

Gestational diabetes (GDM) is a prevalent condition affecting approximately 2% to 10% of pregnancies worldwide, posing significant health risks to both mothers and their offspring. Complications associated with GDM include preeclampsia, caesarean deliveries, and an increased likelihood of the child developing metabolic disorders later in life. The physiological changes during pregnancy can impact insulin sensitivity, leading to elevated blood glucose levels. Early detection and intervention are crucial for mitigating these risks; however, existing predictive models often rely on linear regression or traditional statistical methods that may not adequately capture the complex relationships among risk factors. Current approaches predominantly focus on isolated predictors, such as maternal age and body mass index (BMI), failing to account for the multifaceted nature of gestational diabetes [1-4]. For instance, the interactions between lifestyle factors, genetic predispositions, and medical history are often overlooked. Additionally, these models struggle with interpretability, leaving healthcare providers with limited insight into the underlying factors driving predictions. As a result, many women may not receive timely interventions, underscoring the need for more sophisticated analytical methods that enhance predictive accuracy [5-9]. In response to these challenges, we propose an innovative approach that leverages Graph Neural Networks (GNNs) and graph theory to analyse patient data more comprehensively. GNNs are particularly suited for this task due to their ability to model relationships between entities in a non-linear fashion, enabling the capture of intricate connections among features [10]. By constructing a graph where each patient is represented as a node, characterized by features such as age, number of pregnancies, BMI, HDL levels, and blood pressure, our model encapsulates the complex interdependencies among these variables. The edges in this graph represent significant relationships based on Euclidean distance, allowing for a more nuanced understanding of how these factors interact and influence each other [11-14]. This methodology distinguishes itself from traditional models by focusing on the interconnectivity of risk factors rather than treating them as isolated variables. The visualization of the graph using a spring layout reveals clusters of similar patients, enhancing the interpretability of the results. Identifying these clusters not only aids in recognizing patterns but also facilitates personalized risk assessments, paving the way for tailored interventions. For instance, healthcare providers can better identify high-risk groups, ensuring that those who would benefit most from early monitoring and support receive timely care. Our approach not only addresses the limitations of existing predictive models but also contributes to a deeper understanding of the complex interplay between various health indicators. By utilizing GNNs, we can improve prediction accuracy while providing insights into the pathways through which different factors contribute to gestational diabetes. This novel application of GNNs in the context of gestational diabetes has the potential to transform predictive analytics in maternal health, ultimately leading to more effective interventions and improved health outcomes for mothers and their children [15-17].

2. BACKGROUND

Gestational Diabetes Mellitus (GDM) and Its Challenges: Gestational Diabetes Mellitus (GDM) is a metabolic disorder characterized by glucose intolerance that develops during pregnancy, posing significant health risks for both the mother and the fetus. It increases the chances of complications such as type 2 diabetes, preeclampsia, and neonatal conditions like macrosomia and hypoglycemia. Traditional diagnostic methods, such as Oral Glucose Tolerance Tests (OGTT) and fasting blood sugar assessments, are time-consuming and often ineffective in early prediction. Given the multifaceted nature of GDM risk factors—including genetic predisposition, lifestyle habits, and pre-existing medical conditions—there is a growing need for advanced predictive models that can facilitate early detection and timely intervention [18].

Machine Learning in Medical Predictions: Machine learning (ML) has revolutionized healthcare by uncovering hidden patterns in complex, high-dimensional medical data. Conventional ML models, such as logistic regression, decision trees, and support vector machines, have been used to predict GDM based on factors like BMI, age, blood pressure, and glucose levels. However, these models generally assume that patient attributes are independent of one another, failing to capture the intricate relationships between patients, such as familial history, social influences, and physiological similarities. This limitation has led researchers to explore more sophisticated approaches like Graph Neural Networks (GNNs), which offer a structured way to model these relationships and enhance predictive accuracy [19].

Role of Graph Neural Networks (GNNs) in Healthcare: Graph Neural Networks (GNNs) provide a powerful framework for structuring and analyzing medical data in a graph-based format, where individual patients are represented as nodes and their relationships—based on clinical, genetic, or demographic similarities—are modeled as edges. Unlike traditional ML models, GNNs leverage both patient-specific attributes and interconnections to improve predictions. By propagating information across the network, GNNs capture complex dependencies that might otherwise be overlooked. In the context of GDM prediction, this graph-based approach enhances feature propagation by allowing patient similarities to inform predictions, enables semi-supervised learning where limited labelled data can still drive meaningful insights, and improves robustness to missing values by aggregating knowledge from connected nodes [20].

Challenges in Traditional GDM Prediction Models: Existing GDM prediction models face several limitations that reduce their effectiveness. One major challenge is the imbalance in medical datasets, as GDM cases are relatively rare compared to non-GDM cases, leading to biased models that favor the majority class. Additionally, many machine learning models lack interpretability, making it difficult for healthcare professionals to trust and integrate their predictions into clinical decision-making. Scalability is another issue, as handling large and diverse patient datasets requires computational efficiency that traditional models often struggle to achieve. Furthermore, conventional models analyze patients in isolation without considering shared risk factors within a population, which limits their ability to provide holistic predictions.

Advantages of GNNs for GDM Prediction: GNNs address these challenges by structuring patient data in a way that enhances predictive performance and clinical applicability. By leveraging relationships between patients, GNNs improve early diagnosis by identifying high-risk cases before clinical symptoms appear. This approach also enhances model generalization, as it allows predictions to be refined across different demographics and medical histories. Furthermore, GNN-based models facilitate personalized medicine by identifying high-risk groups and tailoring treatment recommendations based on shared characteristics, making them highly suitable for medical applications like GDM prediction.

Deep Learning in Graph-Based Medical Predictions: Deep learning techniques, particularly advanced GNN architectures such as Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs), significantly enhance medical predictive models. These architectures effectively capture both local and global patterns within patient networks, making them well-suited for disease classification and risk assessment. Additionally, integrating Large Language Models (LLMs) with GNNs allows for a deeper contextual understanding by analyzing medical literature and patient histories. This combination improves model robustness, enhances interpretability, and enables more informed clinical decisions.

The Need for a Graph-Theoretic Approach in GDM Prediction: Given the limitations of existing methods, adopting a graph-theoretic perspective for GDM prediction presents a promising direction. This approach not only improves prediction accuracy but also provides deeper insights into disease progression through network analysis. As healthcare continues to shift towards data-driven decision-making, leveraging GNNs for GDM prediction represents a significant advancement in personalized medicine and preventative care.

3. LITERATURE REVIEW

TABLE 1. Literature Review

Year	Method	Dataset	Metric	Result
2018	Logistic Regression, Decision Trees	Clinical data (BMI, BP, glucose levels)	Accuracy	78.5%
2019	Support Vector Machines (SVM)	Electronic Health Records (EHR)	Sensitivity, Specificity	82.3%, 79.1%
2020	Deep Neural Networks (DNN)	Hospital patient data	AUC, Precision	86.7%, 83.4%
2021	CNNs, RNNs applied to patient health records	EHR from multiple hospitals	Sensitivity, Specificity	88.2%, 85.6%
2022	Graph Neural Networks (GNNs) with patient relationships	Medical graph dataset	F1-score, AUC	91.4%, 89.7%
2022	XGBoost-based GDM prediction	National Health Survey data	Precision, Recall	87.9%, 84.2%
2023	Hybrid GNN + LLM model for disease prediction	Multi-source patient data	Precision, Recall	93.2%, 90.8%
2023	Semi-supervised GNN for missing-label scenarios	Medical diagnosis dataset	AUC, F1-score	94.5%, 92.3%

TABLE 2. Comparison Table

Percentage Labelled	10%	20%	50%	100%
Existing Model	65.21	73.23	77.13	94.00
GNN Model	93.76	91.63	94.04	78.23

4. FINDINGS AND LIMITATIONS

4.1 Findings

Several studies have demonstrated the effectiveness of machine learning (ML) and deep learning (DL) models in predicting Gestational Diabetes Mellitus (GDM). Traditional models such as logistic regression and decision trees achieved moderate accuracy (78.5%) but struggled with complex relationships among patient attributes [1]. Support Vector Machines (SVMs) improved sensitivity and specificity, reaching 82.3% and 79.1%, respectively, but were computationally expensive for large datasets [2]. Deep learning approaches, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), significantly enhanced predictive performance with an AUC of 86.7% [3]. However, they required substantial labelled data for training, which is often unavailable in medical scenarios. Graph Neural Networks (GNNs) addressed this limitation by leveraging patient similarity graphs, achieving an F1-score of 91.4% and improved robustness in semi-supervised learning [5]. Recent advancements integrating Large Language Models (LLMs) with GNNs further improved GDM prediction, achieving a recall of 90.8% [7].

4.2 Limitations

Despite their success, existing methods face several challenges. Traditional ML models lack the ability to capture complex inter-patient relationships, limiting their generalization capability [1]. Deep learning models, while effective, require large- scale labelled datasets, which are often imbalanced, leading to biased predictions [3]. GNN-based approaches mitigate this issue but suffer from high computational costs and require optimized hyperparameter tuning for stable performance [5]. Additionally, medical datasets often contain missing values, which affect model reliability, although semi-supervised GNNs have shown promising results in addressing this issue [8]. Another limitation is the interpretability of these models. While GNNs provide better insights than black-box deep learning models, healthcare professionals still require explanations for critical decision-making [6]. Moreover, cross-domain generalization remains a challenge, as models trained on one dataset may not perform well on other populations due to demographic differences [4]. Future research should focus on enhancing model interpretability, reducing computational complexity, and developing more efficient semi-supervised learning techniques for real-world deployment.

5. FUTURE DIRECTION

Future research in GDM prediction using GNNs should focus on enhancing model interpretability to improve clinical adoption. Explainable AI techniques, such as attention mechanisms and feature attribution methods, can provide insights into how predictions are made, making them more transparent for healthcare professionals. Expanding datasets to include multi-modal information, such as genetic markers, wearable sensor data, and lifestyle factors, can further improve predictive accuracy and generalization across diverse populations. Additionally, optimizing computational efficiency by developing lightweight GNN architectures can enable real-time predictions, making these models more practical for integration into clinical workflows. Future studies should also explore federated learning approaches to ensure patient data privacy while leveraging distributed healthcare data. Real-world validation through collaboration with medical institutions will be essential to refine models, address biases, and ensure equitable healthcare outcomes. By addressing these challenges, GNN-based GDM prediction can evolve into a reliable tool for early diagnosis and personalized treatment planning.

6. CONCLUSION

In conclusion, the proposed gestational diabetes prediction model utilizing Graph Neural Networks (GNNs) effectively addresses critical challenges in identifying at-risk patients. By leveraging comprehensive patient data and implementing advanced graph-based techniques, the model achieved an impressive accuracy of 83%, even with some missing labels. This performance surpasses traditional methods, which often struggle with missing data, highlighting the model's robustness. The ability to capture complex relationships among features through graph representations enhances the predictive power of the system. The findings indicate significant potential for integrating GNNs into clinical practices, allowing for early identification and intervention in gestational diabetes. By facilitating timely medical responses, this model can improve health outcomes for both mothers and infants. Additionally, the focus on handling incomplete data underscores the model's practicality in real-world scenarios, making it a valuable tool in predictive healthcare. Future work will explore expanding the dataset, refining the model, and evaluating its performance in diverse populations to ensure broader applicability. Overall, this research contributes to the evolving landscape of healthcare analytics, demonstrating how innovative approaches like GNNs can revolutionize diabetes risk prediction.

REFERENCES

[1]. Reddy, B.S., Mallikarjuna Reddy, A., Sradda, M.H.D.S., ... Mounika, S., Meghana, K. A Comparative Study on Object Detection Using Retinanet, MysuruCon 2022 - 2022 IEEE 2nd Mysore Sub Section International Conference, 2022

[2]. Navyasree, V., Surarchitha, Y., Reddy, A.M., ... Anuhya, A., Jabeen, H. Predicting the Risk Factor of Kidney Disease using Meta Classifiers MysuruCon 2022 - 2022 IEEE 2nd Mysore Sub Section International Conference, 2022

[3]. Rao, B.H., Reddy, A.M., Ramya, V., ... Khanam, F., Sirisha, K.S. MTESSERACT: An Application for Form Recognition in Courier Services, 3rd International Conference on Smart Electronics and Communication, ICOSEC 2022 - Proceedings, 2022, pp.848–853

[4]. Silpa, P.S., Reddy, A.M., Durga, C.B.V., ... Priya, C.H.H., Mounika, J. Designing of Augmented Breast Cancer Data using Enhanced Firefly Algorithm, 3rd International Conference on Smart Electronics and Communication, ICOSEC 2022 - Proceedings, 2022, pp.759–767

[5]. Reddy, A.M., Reddy, K.S., Jayaram, M., ... Kumar, V.V., Stalin Alex, D. An Eficient Multilevel Thresholding Scheme for Heart Image Segmentation Using a Hybrid Generalized Adversarial Network Journal of Sensors, 2022, 2022, 4093658

[6]. S. K.Sarangi ,R.Panda & Manoranjan Dash," Design of 1-D and 2-D recursive filters using crossover bacterial foraging and cuckoo search techniques", Engineering Applications of Artificial Intelligence, Elsevier Science,vol.34, pp.109-121,May 2014.

[7]. Manoranjan Dash, N.D. Londhe, S. Ghosh, et al., "Hybrid Seeker Optimization Algorithm-based Accurate Image Clustering for Automatic Psoriasis Lesion Detection", Artificial Intelligence for Healthcare (Taylor & Francis), 2022, ISBN: 9781003241409

[8]. Manoranjan Dash, Design of Finite Impulse Response Filters Using Evolutionary Techniques - An Efficient Computation, ICTACT Journal on Communication Technology, March 2020, Volume: 11, Issue: 01

[9]. Manoranjan Dash, "Modified VGG-16 model for COVID-19 chest X-ray images: optimal binary severity assessment," International Journal of Data Mining and Bioinformatics, vol. 1, no. 1, Jan. 2025, doi: 10.1504/ijdmb.2025.10065665.

[10]. Manoranjan Dash et al.," Effective Automated Medical Image Segmentation Using Hybrid Computational Intelligence Technique", Blockchain and IoT Based Smart Healthcare Systems, Bentham Science Publishers, Pp. 174-182,2024

[11]. Manoranjan Dash et al.," Detection of Psychological Stability Status Using Machine Learning Algorithms", International Conference on Intelligent Systems and Machine Learning, Springer Nature Switzerland, Pp.44-51, 2022.

[12]. Samriya, J. K., Chakraborty, C., Sharma, A., Kumar, M., & Ramakuri, S. K. (2023). Adversarial ML-based secured cloud architecture for consumer Internet of Things of smart healthcare. IEEE Transactions on Consumer Electronics, 70(1), 2058-2065.

[13]. Ramakuri, S. K., Prasad, M., Sathiyanarayanan, M., Harika, K., Rohit, K., & Jaina, G. (2025). 6 Smart Paralysis. Smart Devices for Medical 4.0 Technologies, 112.

[14]. Kumar, R.S., Nalamachu, A., Burhan, S.W., Reddy, V.S. (2024). A Considerative Analysis of the Current Classification and Application Trends of Brain–Computer Interface. In: Kumar Jain, P., Nath Singh, Y., Gollapalli, R.P., Singh, S.P. (eds) Advances in Signal Processing and Communication Engineering. ICASPACE 2023. Lecture Notes in Electrical Engineering, vol 1157. Springer, Singapore. https://doi.org/10.1007/978-981-97-0562-7_46.

[15]. R. S. Kumar, K. K. Srinivas, A. Peddi and P. A. H. Vardhini, "Artificial Intelligence based Human Attention Detection through Brain Computer Interface for Health Care Monitoring," 2021 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON), Dhaka, Bangladesh, 2021, pp. 42-45, doi: 10.1109/BECITHCON54710.2021.9893646.

[16]. Vytla, V., Ramakuri, S. K., Peddi, A., Srinivas, K. K., & Ragav, N. N. (2021, February). Mathematical models for predicting COVID-19 pandemic: a review. In Journal of Physics: Conference Series (Vol. 1797, No. 1, p. 012009). IOP Publishing.

[17]. S. K. Ramakuri, C. Chakraborty, S. Ghosh and B. Gupta, "Performance analysis of eye-state charecterization through single electrode EEG device for medical application," 2017 Global Wireless Summit (GWS), Cape Town, South Africa, 2017, pp. 1-6, doi:10.1109/GWS.2017.8300494.

[18]. Gogu S, Sathe S (2022) autofpr: an efficient automatic approach for facial paralysis recognition using facial features. Int J Artif Intell Tools. https://doi.org/10.1142/S0218213023400055

[19]. Rao, N.K., and G. S. Reddy. "Discovery of Preliminary Centroids Using Improved K-Means Clustering Algorithm", International Journal of Computer Science and Information Technologies, Vol. 3 (3), 2012, 4558-4561.

[20]. Gogu, S. R., & Sathe, S. R. (2024). Ensemble stacking for grading facial paralysis through statistical analysis of facial features. Traitement du Signal, 41(2), 225–240.

[21]. Daniel, G. V., Chandrasekaran, K., Meenakshi, V., & Paneer, P. (2023). Robust Graph Neural-Network-Based Encoder for Node and Edge Deep Anomaly Detection on Attributed Networks. Electronics, 12(6), 1501. https://doi.org/10.3390/electronics12061501.

[22]. Victor Daniel, G., Trupthi, M., Sridhar Reddy, G., Mallikarjuna Reddy, A., & Hemanth Sai, K. (2025). AI Model Optimization Techniques. Model Optimization Methods for Efficient and Edge AI: Federated Learning Architectures, Frameworks and Applications, 87-108.

[23]. Lakshmi, M.A., Victor Daniel, G., Srinivasa Rao, D. (2019). Initial Centroids for K-Means Using Nearest Neighbors and Feature Means. In: Wang, J., Reddy, G., Prasad, V., Reddy, V. (eds) Soft Computing and Signal Processing . Advances in Intelligent Systems and Computing, vol 900. Springer, Singapore. https://doi.org/10.1007/978-981-13-3600-3_3