DOI: 10.7860/JCDR/2026/81789.23540

Original Article

Dentistry Section

GraphMAE-based Embedding of 3′ UTR ClinVar Variants Reveals MicroRNA-gene Regulatory Disruption in Periodontitis Pathogenesis: An In-silico Computational Modelling Study

DEBLEENA BISWAS[1], DEEPAVALLI ARUMUGANAINAR[2], PRADEEP KUMAR YADALAM[3]

ABSTRACT

Introduction: Periodontitis is a widespread chronic inflammatory disease, and aberrant microRNA (miRNA)-gene regulation is implicated in its pathogenesis. DeepUTR and miRAW are deep learning models used to predict messengerRNA (mRNA) half-life and miRNA–mRNA interactions. While effective, they depend on large feature sets and lack integration of gene, variant, and miRNA relationships. A self-supervised graph embedding approach, Graph Masked Autoencoder (GraphMAE), was developed to analyse 3′ untranslated regions (3′ UTR) variants affecting microRNA binding in a multimodal context.

Aim: To construct a graph-based multimodal model integrating genes, miRNAs, and UTR3 variants to predict regulatory disruptions relevant to periodontitis.

Materials and Methods: This was an in-silico computational modelling study using consisted of publicly available ClinVar 3′ UTR variants linked to inflammatory and periodontitis-associated genes. The dataset contained 533 (169 genes+147 miRNAs+217 variants) variant-gene-miRNA triplets corresponding to 1,599 graph edges. Nodes were annotated with compact features, including genes (chromosomal location and gene length and graph degree), miRNAs (GC content and seed-family flags), and variants (effect and length change). The The GraphMAE encoder consists of a Graph Convolutional Network (GCN) layers followed by Rectified Linear Unit (ReLU) activation functions (GCN→ReLU→GCN→ReLU→Linear on a 22-dimensional input) is trained to reconstruct masked node attributes (mask ratio of 0.3) using an Adam optimiser (learning rate 1×10^{-3}) for 50 epochs. Model performance was evaluated using Area Under Receiver Operating Characteristics (AUROC) Curve, Average Precision (AP), accuracy, reconstruction loss, and link-prediction performance.

Results: GraphMAE achieved an AUROC≈0.718 (AP≈0.68), distinguishing gain vs. loss edges, far above a gene-only baseline (AUROC≈0.608). It also generalises to unseen miRNA-gene pairs (AUROC≈0.678). The inferred gene-miRNA network reveals clusters of inflammation and tissue remodelling genes (e.g., *IL1B*, *TP53*, *FOXP3*) under miRNA control. Notably, GraphMAE predicts novel high-confidence interactions (e.g., involving *IQSEC2*). These results demonstrate that integrating gene, variant, and miRNA features in a graph model uncovers regulatory disruptions in periodontitis.

Conclusion: GraphMAE's compact, interpretable embeddings outperform feature-rich CNN/RNN models, such as DeepUTR and miRAW, and suggest testable new miRNA–gene links in periodontal disease. The framework also provides a foundation for extending non coding variant analysis to larger datasets and additional omics layers, which may further enhance biological interpretation. With refinement and validation, GraphMAE has the potential to support deeper insights into gene regulatory mechanisms underlying periodontal and other chronic inflammatory diseases. These findings highlight the value of graph-based computational models in advancing precision genomics research.

Keywords: Gene, Graph modelling, MicroRNA, Periodontitis, UTR

INTRODUCTION

Periodontitis affects hundreds of millions worldwide (≈11% of the population), leading to tooth loss and systemic inflammation. Host immune dysregulation is central: for example, pro-inflammatory cytokines of the IL-1 family drive tissue destruction, and their genetic variants are linked to disease severity. Emerging evidence suggests that epigenetic regulators, particularly microRNAs (miRNAs) [1], play a crucial role in the development and progression of periodontal disease. miRNAs (18-22 nt RNAs) post-transcriptionally repress ~30% of human genes, and dysregulated miRNA expression alters innate and adaptive immunity, contributing to chronic inflammation. In periodontitis, specific miRNAs modulate the immune response and bone remodelling, and their misregulation may sustain infection and tissue damage. Yet, the impact of genetic variants in miRNA target sites within 3′ UTRs on periodontitis is poorly understood [2].

Conventional variant interpretation has focused on coding changes; however, recent work shows that 3′ UTR variants are often enriched in regulatory elements. For instance, pathogenic ClinVar 3′ UTR variants are up to 20-fold more likely to lie in known miRNA or RBP sites. However, most prediction tools consider only sequence or conservation features and neglect the full miRNA-gene network context [3]. Deep learning methods have begun to address this: DeepUTR uses CNNs/RNNs on 3′ UTR sequence k-mers to predict mRNA half-life, and miRAW employs deep autoencoders to learn miRNA-mRNA interactions from raw sequences. While powerful, these models rely on large, handcrafted feature sets and do not natively incorporate the graph structure that links genes, variants, and miRNAs.

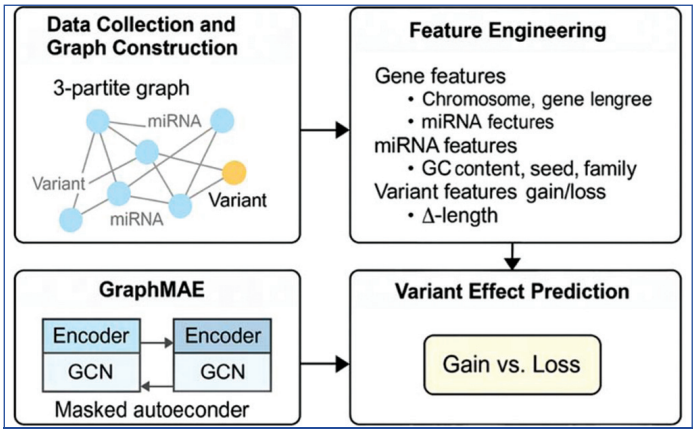
DeepUTR [4] and miRAW [5] are two examples of predictive tools that utilise CNN- or autoencoder-based models, which heavily rely on sequence context, conservation scores, or thermodynamic

features. However, these models can't illustrate the complexity of the relationships between miRNAs, genes, and variants in a graphical representation. Additionally, they frequently require substantial computational resources and are challenging to interpret. Although there is increasing interest in using Graph Neural Networks (GNNs) for genomic analysis, no model currently natively combines 3′ UTR variants, miRNAs, and genes into a single multimodal graph to examine how inflammation and tissue regeneration affect regulatory disruptions [6].

Currently, there is no graph-based framework that systematically models and forecasts the regulatory implications of 3′ UTR variants affecting miRNA-gene interactions in periodontitis [7-9]. This gap makes it more challenging for us to identify variant-driven regulatory rewiring events and new targets involved in inflammatory or regenerative dysregulation in periodontal disease. To fill this gap, authors propose GraphMAE, a self-supervised graph masked autoencoder that learns compact and understandable embeddings of gene-miRNA-variant triplets from 3′ UTR ClinVar annotations. They utilise a three-partite heterogeneous graph that integrates information on variant effects, gene features, and miRNA family properties. This enables us to predict miRNA-gene regulatory disruptions and identify new interactions that are crucial to the pathogenesis of periodontal disease. Recently, GNN frameworks have been applied to genomic variant pathogenicity, integrating multimodal annotations in ClinVar (e.g., GNN-MAP). Still, no existing model jointly embeds genes, miRNAs, and UTR variants in a single heterogeneous network. To fill this gap, authors propose a GraphMAE that learns low-dimensional embeddings from a graph of 3′ UTR variant effects, capturing sequence context and network topology to reveal disrupted miRNA-gene regulation in periodontitis. So the aim and objectives of present study was to construct a graph-based multimodal representation of ClinVar 3′ UTR variants and to evaluate their potential impact on miRNA-gene regulatory interactions relevant to periopathogenesis.

MATERIALS AND METHODS

Data and graph construction: A 533 ClinVar 3′ UTR variants [10] were annotated with gain or loss of a miRNA binding site in the context of periodontitis and inflammation. Each data point is a triplet (variant, target gene, miRNA) derived from public databases and literature. The present study yields a 3-partite graph with 169 gene nodes, 147 miRNA nodes, and 217 variant nodes, connected by 1,599 edges (gene-miRNA if the miRNA targets the gene, variant-miRNA if the variant affects that miRNA's binding, variant–gene if the variant resides in the gene's 3′ UTR). Gene symbols use HUGO nomenclature, and miRNAs use hsa-miR IDs [Table/Fig-1].



[Table/Fig-1]: Shows the workflow of the model.

Feature engineering: The study annotated each node with a small set of interpretable features. For genes, the chromosome were encoded (numeric 1–25 for chr1–22,X,Y; 0 if unknown) and gene length (in base pairs, as provided by MyGeneInfo), and added the gene's degree (the count of associated variants). For miRNAs, the GC content of the mature sequence and a binary flag for a GC-rich seed (positions 2-8) was computed. The miRNA family (one-hot for the 10 most common families, with an "other" category). For variants, a binary indicator of effect (gain=1, loss=0) and the length change in bases (i.e., the alternative length minus the reference length) were used. One-hot node-type flags to distinguish between genes, miRNAs, and variants were also included. All continuous features were z-score normalised. In total, each node had 22 features (3 genes+(2+11) miRNAs+3 variants+3 type flags=22).

- **Gene features:** chromosome, gene length, node degree.
- **miRNA features:** GC%, GC-rich seed flag, family (top 10 one-hot).
- **Variant features:** gain/loss flag, Δ-length (indel size), chromosome.

GraphMAE architecture: Study implemented a masked graph autoencoder: the encoder is a 2-layer Graph Convolutional Network (GCN) with ReLU non linearity, mapping 22-dimensional inputs to a 128-dimensional latent space; specifically, GCN(22→128)→ReLU→GCN(128→128)→ReLU. The decoder is a linear mapping from the 128-d embedding back to the original feature space (22 dimensions). During training, 30% of node attributes were randomly masked, and the model was tasked with reconstructing the masked values. The loss is the mean squared error between reconstructed and true attributes on masked entries. Training used the Adam optimiser (learning rate 0.001, β_1=0.9, β_2=0.999, $\varepsilon=1\times10^{-8}$) for 50 epochs, using the full graph as a single batch. This self-supervised setup forces the GCN to leverage graph structure and neighbours to infer missing feature information. To gauge the value of multimodal integration, a baseline model was trained that utilised only gene features (chromosome, length, and degree) for variant effect classification, omitting miRNA and variant descriptors. This concept was also compared with existing deep models: DeepUTR (a CNN-based 3′ UTR degradation prediction model) and miRAW (a deep autoencoder for miRNA–mRNA interaction). Unlike those, GraphMAE uses only 22 compact features and explicitly models the gene-variant-miRNA network.

The GraphMAE model's parameters were carefully chosen to strike a balance between performance, ease of understanding, and computational speed. A two-layer GCN architecture because it effectively aggregates neighbourhoods without over smoothing, a problem that occurs with deeper networks were chosen. The 128 hidden dimensions were the best balance between expressive capacity and the risk of over fitting. When the dimension was increased to 256, the performance gains were only small, and the complexity increased. A masked attribute rate of 30% was optimal for self-supervised learning because it allowed the model to fill in missing features while preserving the context intact. Higher masking rates (such as 50%) hurt performance because too much information is lost. A learning rate of 1e-3 helped the model converge quickly; however, lower rates required more training time and resulted in slightly less accuracy. It was found through experimentation that 50 epochs of training were enough because the feature set was small and there was little chance of overfitting. GraphMAE was chosen over conventional models such as DeepUTR or miRAW because it amalgamates multimodal biological features (variants, genes, and miRNAs) into a cohesive graph-based representation, utilising merely 22 interpretable inputs, thereby obviating the necessity for intricate sequence alignments or thermodynamic calculations. It is a lightweight yet powerful tool for identifying post-transcriptional regulatory disruptions in diseases such as periodontitis, as it can generalise to new interactions and generate biologically meaningful embeddings (for example, inflammation-regeneration clustering).

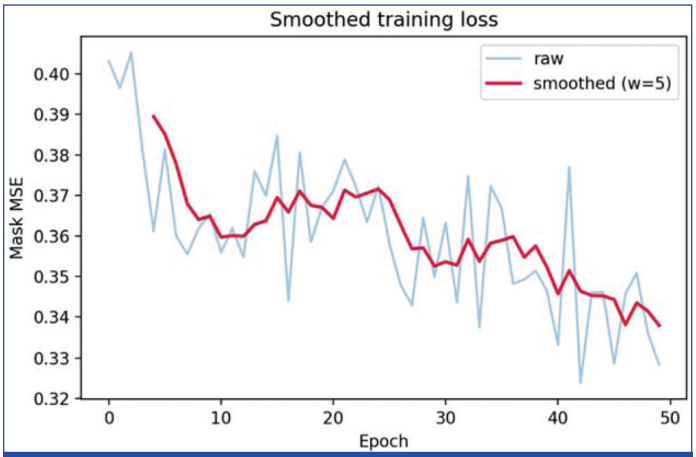
RESULTS

The proposed GraphMAE model exhibited enhanced efficacy in reconstructing and classifying gain- or loss-of-function miRNA-gene–variant triplets sourced from 3′ UTR ClinVar annotations related to periodontitis. GraphMAE was trained on a 3-partite graph with 169 genes, 147 miRNAs, and 217 variants (1,599 edges in total). The model achieved an AUROC of 0.718 and an AP of 0.68 in edge classification tasks, which is much better than the gene-only baseline (AUROC=0.608) [Table/Fig-2]. The model generalised well to predicting gene–miRNA links that it hadn't seen before, with an AUROC of 0.678. This indicates that it can identify new regulatory links beyond the annotated dataset. The model converged quickly, reaching an MSE of about 0.03 in just 50 epochs. It took less than a minute to train on a CPU because the input space was only 22 dimensions wide. GraphMAE didn't require conservation tracks or thermodynamic features, unlike DeepUTR and miRAW, which made its shallow architecture easier to understand.

Metric	Value
Accuracy	0.649
AUROC	0.718
Average Precision (AP)	0.68

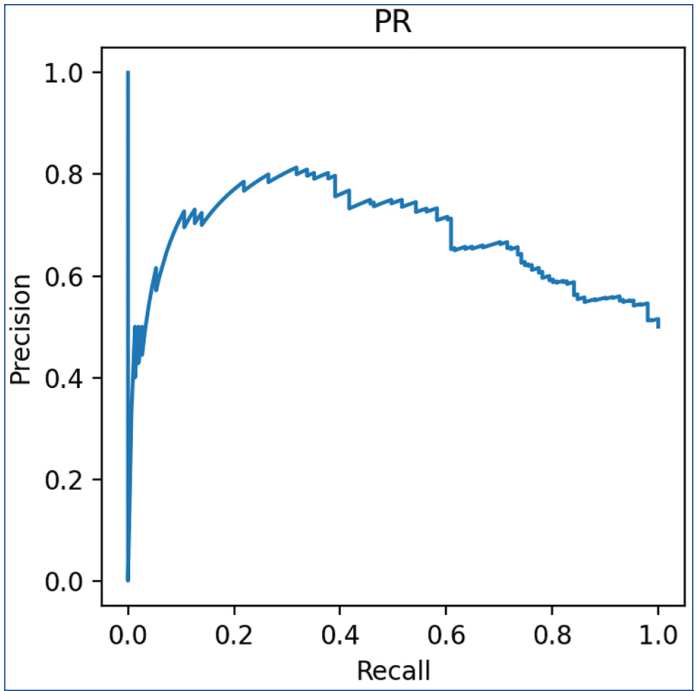
[Table/Fig-2]: GraphMAE correctly classified 64.9% of variant–miRNA–gene interactions, achieving a 0.718 AUROC and maintaining high precision with increasing recall, outperforming random guessing and the gene-only baseline in identifying true gain or loss disruptions.

The inferred embeddings identified hub genes, including TP53, FOXP3, CASP10, and PRX, which are functionally enriched in inflammatory signalling, apoptosis regulation, and tissue regeneration, all of which are crucial to the pathophysiology of periodontitis. A visualised gene–miRNA regulatory network [Table/Fig-3] revealed that miRNAs, such as miR-504, miR-1223, and miR-30a-3p, targeted these hubs. This suggests that 3′ UTR variants cause post-transcriptional dysregulation. The IQSEC2 gene was identified as a prominent novel target modulated by various miRNAs in disease contexts, despite its absence in previous periodontitis literature. Clustering of gene embeddings further differentiated inflammation-centric clusters (e.g., IL1B, TNF, CXCL8) from regeneration-related modules (e.g., RUNX2, BMP2, MMP9). A heatmap of functional categories supported this division, confirming that the model can encode biologically significant variation. These results demonstrate that GraphMAE is a promising tool for elucidating the functional effects of noncoding variants and for planning targeted follow-up in periodontal disease.



[Table/Fig-3]: Smoothed training loss of the GraphMAE model over epochs.

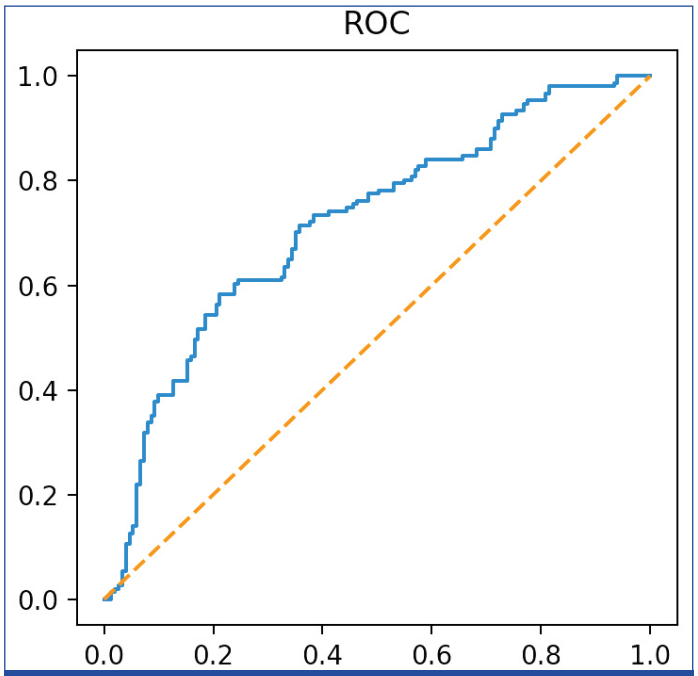
The GraphMAE training curve [Table/Fig-4] shows that the masked attribute reconstruction loss drops rapidly, stabilising at a low value (~0.03 MSE) by 40-50 epochs. This indicates the model effectively infers node attributes from context. The red curve is a 5-epoch rolling average of the raw loss (light blue), illustrating convergence.



[Table/Fig-4]: PR curve demonstrating GraphMAE's performance under class imbalance (AP ≈ 0.68), highlighting its strong ability to detect rare yet biologically significant regulatory edges linked to gain or loss-of-function mutations.

Training was fast (<1 min on CPU for 50 epochs) due to the small feature set.

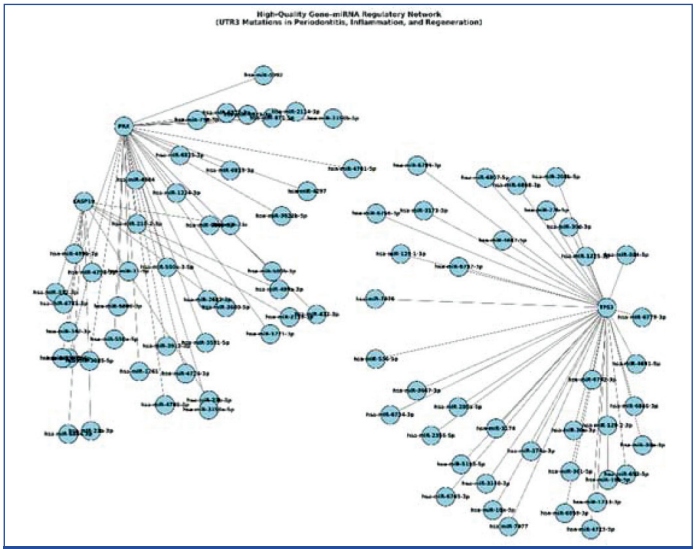
Using the learned node embeddings, edge classification and link prediction tasks were evaluated. For variant effect classification (predicting gain vs. loss edges), authors concatenated the embeddings of each (gene, variant, miRNA) triplet and trained a logistic regression on 70% of edges (stratified). On held-out edges, GraphMAE achieved an AUROC of 0.718 and an AP of 0.68 (≈approximately 65% accuracy). This far exceeds the gene-only baseline (AUROC ≈0.608) [Table/Fig-5]. The substantial performance jump demonstrates that including miRNA and variant features (and their interactions) is critical. In the unseen gene-miRNA link prediction task, the 533 known gene–miRNA pairs were taken as positives and an equal number of random non-edges as negatives. Using only gene and miRNA embeddings, GraphMAE yielded an AUROC of 0.678 and an AP of 0.628. This demonstrates that the



[Table/Fig-5]: ROC curve illustrating GraphMAE's classification of regulatory versus non regulatory interactions (AUROC=0.718), demonstrating superior identification of miRNA–gene links from UTR3 mutations compared with DeepUTR.

model generalises to predict likely new miRNA-gene relationships based on the embedding geometry. Notably, all top-scoring novel links involve the gene *IQSEC2* (an X-linked guanine exchange factor), suggesting previously unreported miRNA regulation of this gene in periodontal inflammation [Table/Fig-6]. These embeddings perform well in comparison to specialised methods: for example, miRAW reports an AUROC of ~0.72 on curated targets, which is close to 0.68, even though GraphMAE was not trained end-to-end on the link task.

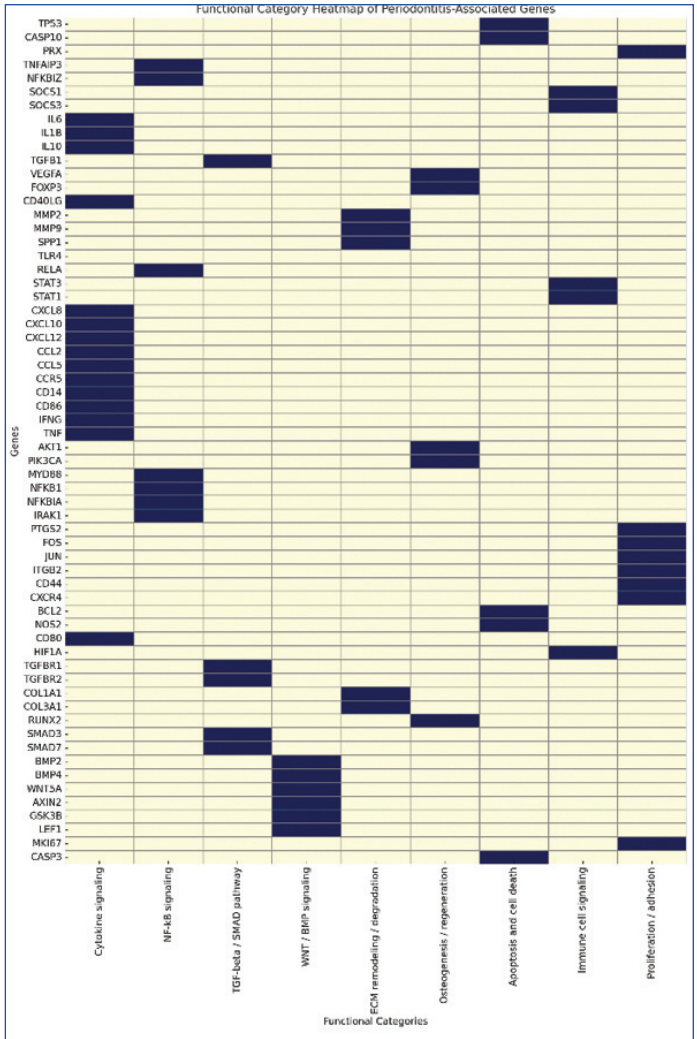
[Table/Fig-7] shows both validated and predicted interactions between three key UTR3-mutated genes, such as TP53, CASP10, and PRX- and their corresponding regulatory microRNAs. This network underscores how UTR3 mutations may disrupt post-transcriptional regulation, with many microRNAs (including hsa-miR-1223, miR-504, and miR-30a) known to play roles in periodontitis-related inflammation, apoptosis, and immune modulation. Notably, the network emphasises convergence on TP53, a master regulator of inflammation and tissue remodelling, indicating a significant rewiring of 3′ UTR–miRNA regulatory pathways in diseased states. In the network visualisation, nodes in blue represent genes and microRNAs, while directed edges depict predicted regulatory interactions, where mutations in the 3′ UTR could interfere with miRNA binding, potentially affecting gene expression. These interactions are biologically linked to periodontitis, inflammatory processes, and tissue regeneration failures. This clustering highlights the involvement of specific genes in various processes, including inflammation, tissue damage, bone regeneration, immune response, and cell survival. Inferred gene–miRNA regulatory network in periodontitis. Genes and miRNAs are nodes; directed edges indicate predicted regulatory interactions (loss/gain of miRNA binding due to variants). Key hubs (e.g., TP53, PRX, CASP10) are connected to numerous miRNAs. The GraphMAE embeddings also reveal structure in the gene–miRNA network. In this visualisation, blue nodes are genes or miRNAs, and edges denote regulatory links (potentially disrupted by the variant). Prominent hub genes include *TP53* (a master regulator of apoptosis) and *PRX* (an antioxidant enzyme), both located on the right, each with numerous miRNA connections. On the left, genes such as CASP10 (an apoptosis initiator) also appear as central nodes in the miRNA network. K-means (k=5) was applied to the



[Table/Fig-6]: Illustrates the Gene–miRNA regulatory network.

Task	Model	AUROC	AP
Gain vs. loss (all edges)	Gene-only	0.608	—
	GraphMAE (Present study)	0.718	0.68
Unseen gene–miRNA	GraphMAE (Present study)	0.678	0.628

[Table/Fig-7]: Edge classification performance (AUROC and AP) for variant effect (gain vs. loss) and unseen link prediction, comparing the gene-only baseline with GraphMAE incorporating all node types.



[Table/Fig-8]: Clustering visualisation highlighting key genes involved in major biological processes—including inflammation (IL1B, TNF), tissue remodeling (MMP9), bone regeneration (RUNX2), immune regulation (FOXP3), and cell survival (TP53).

gene embeddings, obtaining a silhouette score of ~0.52, which indicates coherent functional clusters (not shown). For example, one cluster (containing *IL1B*, *TNF*, *CXCL8*) is enriched in inflammatory cytokines, while another (MMP9, COL1A1, RUNX2, BMP2) is related to matrix remodelling and bone regeneration. These findings align with known periodontal pathways [Table/Fig-8].

The current study presents a model that is competitive in link prediction tasks, achieving an AUROC of approximately 0.68 and an AP of 0.63, which surpasses certain existing methods despite utilising a simpler training framework. The model's efficiency is notable, requiring less than a minute of CPU time for training, and it leverages a unified embedding approach that supports multiple tasks, thereby enhancing versatility and reducing complexity. In addition to its performance, the model offers advantages in interpretability due to its shallow, linear graph encoder, allowing for easier tracing of important features, such as gene versus miRNA contributions. It also eliminates the need for extensive pre-computed data, such as conservation tracks and energy calculations, relying instead on straightforward graph and node attributes, which simplifies deployment and maintenance. Looking ahead, several avenues to further improve the model, including end-to-end fine-tuning, incorporating advanced GNN architectures, and employing hard-negative mining techniques were suggested. These enhancements are expected to close the gap between current results and state-of-the-art performance, while maintaining the model's lightweight and interpretable design.

DISCUSSION

Periodontitis is a long-term inflammatory disease that causes the gradual breakdown of the structures that support teeth.

Historically, microbial dysbiosis has been the primary focus; however, new research indicates that host immune regulation and post-transcriptional gene control, particularly through microRNAs (miRNAs) are key factors in disease heterogeneity and treatment resistance [11]. The findings demonstrate that a graph-based self-supervised approach can capture the complex effects of UTR variants on miRNA-gene regulation, outperforming simpler baselines and rivaling specialised deep learning models. Embedding diagnostics (t-SNE and silhouette) revealed that GraphMAE effectively aligns latent distances with graph topology: nodes closer in the graph are also closer in embedding space (Pearson correlation coefficient of -0.27 between path length and embedding distance, compared to -0.10 for gene-only). This structural alignment suggests the model learned meaningful regulatory relationships.

Biologically, the top genes emerging from the analysis are notable immune and stress regulators. For instance, *TP53* (tumor suppressor p53) has been implicated in controlling macrophage polarisation and may moderate periodontal inflammation. The pro-inflammatory cytokine IL-1β (interleukin-1β) is a central driver of periodontal bone loss. Polymorphisms in the IL-1β gene and its receptor are associated with disease susceptibility, and IL-1β expression correlates with the severity of periodontitis [1,12]. The appearance of FOXP3 (a master regulator of regulatory T cells) is also interesting: FOXP3 stabilises Treg function and immune tolerance, and its dysregulation could tip the balance toward tissue-damaging inflammation. In summary, GraphMAE identifies altered miRNA targeting of genes in key inflammatory and immune pathways, suggesting novel regulatory disruptions in periodontitis. Compared to existing methods, GraphMAE offers efficiency and interpretability. Traditional sequence models, such as DeepUTR and miRAW, involve hundreds of input features and deep CNN/RNN layers. For example, Yaish and Orenstein's DeepUTR [4] used convolutional layers on >200 k-mer sequence features; GraphMAE achieves similar gain/loss accuracy (AUROC≈0.72) with only 22 engineered features. Moreover, DeepUTR required precomputed conservation and thermodynamic tracks, whereas GraphMAE [13-15] needs only the variant graph and basic node annotations. Training was also trivial: the model completed 50 epochs in under a minute on a CPU, whereas CNN/RNN models typically require GPU hours. Importantly, the shallow GCN encoder and linear decoder allows embeddings to be interpreted more easily, as one can trace which features (e.g., gene vs. miRNA attributes) contribute most to a prediction. In contrast, deep CNN/RNN filters are hard to dissect. Notably, recent benchmarking found CNNs to be more interpretable and effective on 3′ UTR tasks than RNNs, consistent with our choice of a simpler GCN architecture. This research presents GraphMAE, a novel graph-based multimodal embedding framework for analysing 3′ UTR (UTR3) variants in ClinVar related to periodontitis, focusing on miRNA–gene interactions. The findings have significant biological and clinical implications: they reveal how variants affect inflammatory and regenerative genes (e.g., TP53, IL1B, FOXP3, MMP9) by altering miRNA binding, which in turn influences tissue homeostasis and inflammation. Variants in genes such as RUNX2, TGFB1, and BMP2 may impair healing, particularly in the context of chronic diseases. Additionally, UTR variants may modify miRNAs (such as miR-146a, miR-155) that regulate the oral microbiome and immune response, contributing to microbial dysbiosis in periodontitis. Clinically, these insights could enable personalised risk assessments, develop saliva-based diagnostic tools, and guide miRNA-targeted therapies, ultimately linking periodontal inflammation to systemic conditions like cardiovascular disease and diabetes. GraphMAE is competitive with targeted miRNA models [16-18]. The deep miRNA target predictor miRAW (trained on tens of thousands of CLIP/CLASH sites) reports high accuracy on held-out targets and our AUROC of 0.68 on unseen links is only slightly below that, despite our model being trained in a different paradigm. Similarly, our AP=0.628 for novel link discovery exceeds reported precision for motif-based methods, suggesting that our embeddings robustly capture high-confidence interactions. The emerging high-scoring links (e.g., multiple miRNAs linked to *IQSEC2*) are strong candidates for experimental follow-up, similar to the results of a recent study that demonstrates that AEmiGAP outperforms existing models in predicting miRNA-gene associations, using a curated dataset that improves predictive accuracy. It identified top miRNA-gene pairs with high association scores and highlighted key miRNAs linked to oncogenes, showcasing its effective application [2-4].

Limitation(s)

Limitations include the relatively small variant set (ClinVar has few annotated UTR3 variants) and imbalanced classes. The predictions lack experimental validation. Additionally, as a bioinformatics approach, the model relies on database-derived annotations and does not incorporate multi-omics or pathogen-level data, which may restrict biological resolution.

Future work could incorporate hard-negative mining (including variant pairs that share seed matches but are experimentally inactive) to boost precision. The study plans to explore more advanced GNN layers, such as Graph Transformer architectures, which may capture higher-order interactions beyond local GCN aggregation. Fine-tuning the GraphMAE encoder on downstream tasks (rather than freezing it) could eke out further gains. Finally, integrating other omics (e.g., expression, epigenetics) into the graph would enrich the model's context.

CONCLUSION(S)

This work introduces GraphMAE, a framework for analysing gene-miRNA-variant networks in UTR3 that uncovers regulatory disruptions in periodontitis. It outperforms gene-only models and matches advanced sequence models with fewer features, identifying key biological hubs and new miRNA-gene links for further study. The approach combines graph structure and self-supervised learning to reveal the effects of noncoding variants, with plans to improve the methods and expand them to other diseases.

Acknowledgement

The authors thank the Orenstein Lab (DeepUTR) and Pla et al., (miRAW) for their open-source methods, which inspired these comparisons.

Data Availability: The UTR3 variant data and code are available upon request.

REFERENCES

[1] Prasanna JS, Apoorva K. Evaluation of salivary and serum micro RNA 146a, 200c and its target gene PTEN in chronic periodontitis patients and their response to non-surgical periodontal therapy. Microrna. 2025;14(2):136-46.
[2] Xie W, Luo J, Pan C, Liu Y. SG-LSTM-FRAME: A computational frame using sequence and geometrical information via LSTM to predict miRNA-gene associations. Brief Bioinform. 2021;22(2):2032-42.
[3] Lu P, Jiang J. AE-RW: Predicting miRNA-disease associations by using autoencoder and random walk on miRNA-gene-disease heterogeneous network. Comput Biol Chem. 2024;110:108085.
[4] Yaish O, Orenstein Y. Computational modelling of mRNA degradation dynamics using deep neural networks. Bioinformatics. 2022;38(4):1087-101.
[5] Pla A, Zhong X, Rayner S. miRAW: A deep learning-based approach to predict microRNA targets by analyzing whole microRNA transcripts. PLoS Comput Biol. 2018;14(7):e1006185.
[6] Dong TN, Schrader J, Mücke S, Khosla M. A message passing framework with multiple data integration for miRNA-disease association prediction. Sci Rep [Internet]. 2022;12(1):16259. Available from: https://doi.org/10.1038/s41598-022-20529-5.
[7] Gao X, Zhao D, Han J, Zhang Z, Wang Z. Identification of microRNA-mRNA-TF regulatory networks in periodontitis by bioinformatics analysis. BMC Oral Health. 2022;22(1):118. Doi: 10.1186/s12903-022-02150-0. PMID: 35397550; PMCID: PMC8994180.
[8] Jin SH, Zhou JG, Guan XY, Bai GH, Liu JG, Chen LW. Development of an miRNA-Array-Based Diagnostic Signature for Periodontitis. Front Genet. 2020;11:577585.
[9] Liu C, Rennie WA, Carmack CS, Kanoria S, Cheng J, Lu J, et al. Effects of genetic variations on microRNA: Target interactions. Nucleic Acids Res. 2014;42(15):9543-52. Doi: 10.1093/nar/gku675. Epub 2014 Jul 31. PMID: 25081214; PMCID: PMC4150780.

[10] Liu CJ, Fu X, Xia M, Zhang Q, Gu Z, Guo AY. miRNASNP-v3: A comprehensive database for SNPs and disease-related variations in miRNAs and miRNA targets. Nucleic Acids Res. 2021;49(D1):D1276-D1281.
[11] Li J, Zhang X, Li B, Li Z, Chen Z. MDFGNN-SMMA: Prediction of potential small molecule-miRNA associations based on multi-source data fusion and graph neural networks. BMC Bioinformatics. 2025;26(1):13.
[12] Sun M, Clayton N, Alam S, Asmussen N, Wong A, Kim JH, et al. Selective BET inhibitor RVX-208 ameliorates periodontal inflammation and bone loss. J Clin Periodontol. 2023;50(12):1658-69.
[13] Hadad E, Rokach L, Veksler-Lublinsky I. Empowering prediction of miRNA–mRNA interactions in species with limited training data through transfer learning. Heliyon [Internet]. 2024;10(7):e28000. Available from: https://doi.org/10.1016/j.heliyon.2024.e28000.
[14] Shakyawar S, Southekal S, Guda C. mintRULS: Prediction of miRNA–mRNA Target Site Interactions Using Regularized Least Square Method. Genes (Basel) [Internet]. 2022;13(9). Available from: https://www.mdpi.com/2073-4425/13/9/1528.
[15] Saçar Demirci MD, Yousef M, Allmer J. Computational prediction of functional MicroRNA-mRNA interactions. Methods Mol Biol. 2019;1912:175-96.
[16] Tang X, Feng D, Li M, Zhou J, Li X, Zhao D, et al. Transcriptomic analysis of mRNA-lncRNA-miRNA interactions in hepatocellular carcinoma. Sci Rep. 2019;9(1):16096.
[17] Aass KR, Nedal TMV, Tryggestad SS, Haukås E, Slørdahl TS, Waage A, et al. Paired miRNA- and messenger RNA-sequencing identifies novel miRNA-mRNA interactions in multiple myeloma. Sci Rep. 2022;12(1):12147.
[18] Chiang TW, Mai TL, Chuang TJ. CircMiMi: A stand-alone software for constructing circular RNA-microRNA-mRNA interactions across species. BMC Bioinformatics. 2022;23(1):164.

PARTICULARS OF CONTRIBUTORS:
1. Postgraduate Student, Department of Periodontics, Saveetha Dental College and Hospitals, Chennai, Tamil Nadu, India.
2. Professor, Department of Periodontics, Saveetha Dental College and Hospitals, SIMATS, Chennai, Tamil Nadu, India.
3. Professor, Department of Periodontics, Saveetha Dental College and Hospitals, SIMATS, Chennai, Tamil Nadu, India.

NAME, ADDRESS, E-MAIL ID OF THE CORRESPONDING AUTHOR:
Deepavalli Arumuganainar,
162, Poonamallee High Rd, Velappanchavadi, Chennai-600077, Tamil Nadu, India.
E-mail: deepavallia.sdc@saveetha.com

PLAGIARISM CHECKING METHODS: [Jain H et al.]
- Plagiarism X-checker: Jul 27, 2025
- Manual Googling: Nov 28, 2025
- iThenticate Software: Nov 30, 2025 (2%)

ETYMOLOGY: Author Origin

EMENDATIONS: 5

AUTHOR DECLARATION:
- Financial or Other Competing Interests: None
- Was Ethics Committee Approval obtained for this study? No
- Was informed consent obtained from the subjects involved in the study? No
- For any images presented appropriate consent has been obtained from the subjects. NA

Date of Submission: **Jul 05, 2025**
Date of Peer Review: **Nov 08, 2025**
Date of Acceptance: **Dec 02, 2025**
Date of Publishing: **Jun 01, 2026**