ISSNe 2178-1990

10.35699/2178-1990.2026.59071

ARQUIVOS EM ODONTOLOGIA

Machine Learning for Predicting lncRNA–Disease Associations Related to Periodontitis-Associated Alveolar Bone Loss

Devika Dinesh[1] | **Pradeep Kumar Yadalam**[1] | **Carlos M. Ardila**[1,2]

[1]Department of Periodontics, Saveetha Dental College & Hospital, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Tamil Nadu, India

[2]Biomedical Stomatology Research Group, Faculty of Dentistry, Universidad de Antioquia U de A, Medellín, Colombia

Introduction: Periodontitis-associated alveolar bone loss is characterized by the gradual destruction of alveolar bone and is regulated by intricate genetic and immunological processes. Although non-coding RNAs (lncRNAs) have become important regulatory components in these processes, little is known about how they relate to disease, especially in periodontitis. Although there are few comparative evaluations of various architectures, graph-based machine learning presents promising tools for modeling such biological associations.

Aim: To predict and reconstruct lncRNA–disease associations related to periodontitis-associated alveolar bone loss, this study compares and assesses the performance of Graph Auto-Encoder (GAE) and Adversarial Regularized Graph Auto-Encoder (ARGAE).

Materials and Methods: A heterogeneous bipartite graph with ncRNA and disease nodes was created using experimentally verified data from the LncRNADisease database. A straightforward one-hot scheme was used to encode node features. A two-layer Graph Convolutional Network (GCN) encoder was used in the GAE and ARGAE models. An extra discriminator for adversarial regularization was incorporated into ARGAE. The models were evaluated using embedding coherence (cosine similarity, t-SNE visualization), clustering quality (Silhouette Score, ARI), and link prediction metrics (ROC-AUC, Average Precision).

Results: While ARGAE added embedding diversity that allowed for the exploratory identification of less evident lncRNA–disease relationships, GAE showed a strong reconstruction capacity with superior clustering interpretability. The study found clear trade-offs between latent space generalization and prediction accuracy.

Conclusions: This comparative analysis fills a substantial knowledge gap by examining the structural and predictive abilities of GAE and ARGAE in modeling lncRNA–disease associations related to alveolar bone loss commonly observed in periodontitis. The results provide useful information for scientists using graph-based models to search for lncRNA biomarkers and aid in advancing periodontology precision diagnostics and treatments.

Keywords: periodontitis; machine learning; RNA, long noncoding.

Received: 2025-05-12
Accepted: 2026-02-12

INTRODUCTION

In molecular systems biology, understanding the regulatory functions of long non-coding RNAs (lncRNAs)[1] in disease pathogenesis, particularly in chronic inflammatory conditions such as periodontitis-associated alveolar bone loss, remains a major research frontier.

A complex interaction between host immune responses and gene regulatory networks influences alveolar bone destruction

Correspondence author:
Carlos M. Ardila, Universidad de Antioquia. Calle 70 No. 52-21, Medellín, Colombia. Phone: 57-4-2196700. FAX 057-4-2195332.
E-mail: martin.ardila@udea.edu.co

in periodontitis, which is characterized by inflammatory infiltration and progressive loss of supporting bone. Because of their role in regulating immune-inflammatory pathways, osteoclast differentiation, and tissue remodeling, lncRNAs have attracted much attention among emerging biomarkers. However, only a portion of lncRNAs have been experimentally validated, making the thorough mapping of lncRNA–disease associations still lacking, particularly in periodontitis[2-4]. Periodontal bone loss is a clinical and radiographic manifestation that may result from multiple disease entities, including endo-periodontal lesions, primary endodontic lesions, root fractures, and periodontitis. In the present study, the term periodontal bone loss is used exclusively to refer to alveolar bone destruction occurring in the context of periodontitis, as defined by disease annotations in curated molecular databases. Other causes of periodontal bone defects are beyond the scope of this work. This study does not aim to diagnose or predict periodontitis; rather, it focuses on reconstructing molecular lncRNA–-disease associations related to alveolar bone loss annotated as periodontitis in curated databases.

Traditional computational approaches to disease association prediction do not capture the topological and semantic complexity of biological interaction networks, which frequently rely on shallow features or similarity networks. Here, node relationships (like ncRNA–disease associations) and node features (like biological identity) can be jointly embedded in a latent space using graph representation learning, which provides a strong framework for modeling biological networks. Recent Graph Auto-Encoders (GAE)[5,6] developments have shown promise in learning from structured biomedical graphs by reconstructing observed associations and producing meaningful node embeddings. However, in the extremely sparse and noisy world of ncRNA–disease data, GAEs frequently suffer from overfitting. They may not generalize well to unseen or weakly connected nodes—a significant limitation. According to a prior study, predicting the relationships between long non-coding RNAs (lncRNAs) and human diseases is vital. Four isomorphic networks and six heterogeneous networks are used in the global network-based approach known as LRWRHLDA[7]. The association between lncRNAs and diseases is predicted by the Laplace normalized random walk with restart algorithm[8,9]. The technique outperformed other approaches with an AUC of 0.98402, and it successfully treats breast, stomach, lung, and colorectal cancer. Another study demonstrated that complex diseases can result from mutations and dysregulations in long-non-coding RNAs (lncRNAs), essential in many biological processes[10].

Drug discovery and biomarker identification can benefit from developing computational models for identifying disease-related lncRNAs. Laplacian Regularized Least Squares for LncRNA–Disease Association (LRLSLDA)[10,11], a semisupervised learning framework method, has demonstrated enhanced performance in detecting disease-lncRNA associations, with multiple putative associations validated in biological experiments. Another study demonstrated a novel algorithm (IPCARF)[1] to predict lncRNA–disease associations by utilizing random forest (RF) and incremental principal component analysis (IPCA) algorithms. The algorithm generates a characteristic vector for every pair of diseases, calculates a semantic similarity matrix of diseases, and trains an RF model. With an AUC value of 0.8611 following parameter optimization, the experimental results demonstrate that IPCARF enhances the AUC metric.

Adversarial Regularized Graph Auto-Encoders (ARGAEs)[12,13] have been proposed as a solution to this problem. These algorithms introduce a discriminator network that encourages the latent embeddings to align with a predefined prior distribution, usually Gaussian. This adversarial regularization enhances the model's generalization ability and may reveal more biologically believable but hitherto undiscovered correlations. However, this additional complexity may result in trade-offs between reconstruction performance and embedding interpretability, particularly in highly specific biomedical applications like periodontal disease. The comparative assessment of GAE and ARGAE architectures, particularly for reconstructing lncRNA–disease networks associated with periodontitis, is lacking despite the potential of both models. Prior research has hardly examined the effects of adversarial regularization on the biological coherence of lncRNA and disease embeddings, latent space topology, and semantic clustering.

The current work closes this crucial gap by thoroughly comparing GAE and ARGAE models on a carefully selected dataset of lncRNA–disease associations pertinent to periodontitis. This study develops a biologically meaningful representation space for learning using a heterogeneous bipartite graph and node-type one-hot encoded features. This

comparison is controlled and directly attributable to the adversarial component of ARGAE, as both models have the same encoder architecture (based on a two-layer Graph Convolutional Network, or GCN) but different regularization strategies.

Crucially, this work uses a variety of unsupervised metrics and visualization techniques to evaluate the models' clustering behavior, embedding similarities, and semantic structure in addition to their link prediction ability, which indicates their reconstruction power. These include t-SNE projections, cosine similarity distributions, and silhouette scores, all of which help determine how well each model represents biologically significant relationships. The study's implications include directing model selection techniques for upcoming lncRNA–based diagnostic or treatment instruments in inflammatory and periodontal disorders. The results provide crucial information for researchers looking to use graph-based models for precision medicine by clarifying the advantages and disadvantages of GAE and ARGAE in this field[14,15].

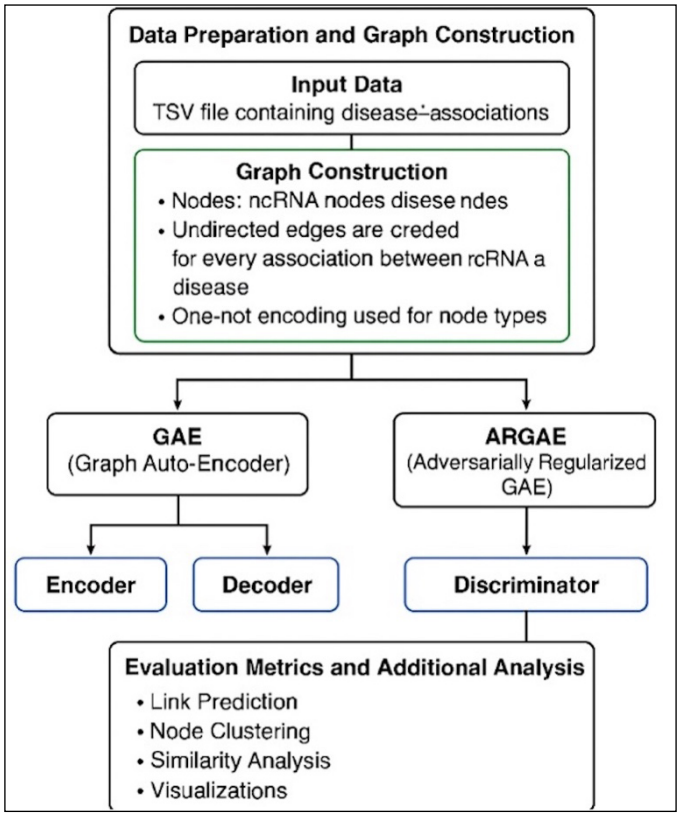
In addition to assessing the structural and biological quality of their learned latent representations, the study compares the effectiveness of GAE and ARGAE models in reconstructing and forecasting lncRNA–disease associations pertinent to periodontitis. This study aims to achieve several important objectives. This study targets lncRNAs and diseases and builds a bipartite graph from a TSV dataset of known ncRNA disease associations. It will implement and train two models, Graph Auto-Encoder (GAE) and Adversarially Regularized Graph Auto-Encoder (ARGAE), to compare their link prediction performance using ROC-AUC and Average Precision metrics. Additionally, clustering consistency and biological coherence will be assessed using k-means clustering and Silhouette Scores, among others. Visualization analyses will help explore relationships, aiming to optimize graph neural networks for identifying lncRNA biomarkers in periodontitis.

MATERIALS AND METHODS

Data Preparation and Graph Construction

Figure 1 depicts the study diagram.

Figure 1. Study diagram.



Source: Authors.

The LncRNADisease database[16] is a curated resource that provides experimentally validated associations between lncRNAs and diseases, complete with detailed annotations and evidence types. It enables AI-driven analysis of molecular mechanisms associated with periodontitis by supporting the construction of bipartite graphs for lncRNA–disease interactions related to inflammatory and bone-destructive processes.

A tab-separated values (TSV) file containing known correlations between non-coding RNAs (ncRNAs) and illnesses served as the dataset for this investigation. A heterogeneous bipartite graph is created from this dataset, with nodes standing in for either diseases or ncRNAs. An established link between a particular ncRNA and a disease is indicated by an undirected edge contributed by each entry in the TSV file. As a result, the graph incorporates two different kinds of nodes: ncRNA nodes and disease nodes. A straightforward one-hot representation based on node type encodes node features to aid learning: ncRNA nodes are given a value of 0, and disease nodes are given a value of 1. While maintaining the semantic topology of the biological associations, this structural encoding allows downstream models to differentiate between the two node categories.

Model Architectures

Two graph representation learning models, the GAE and the Adversarial Regularized ARGAE, were developed and evaluated (Table 1).

Table 1. Hyperparameters used in this study.

Component	GAE	ARGAE
Encoder	2-layer GCN	2-layer GCN
Hidden Dimension	32	32
Latent Dimension	16	16
Activation	ReLU (both layers)	ReLU (both layers)
Dropout	0.2	0.2
Decoder	The inner product of latent embeddings	The inner product of latent embeddings
Discriminator	–	3-layer fully connected neural network
Loss Function	Binary cross-entropy	BCE + Adversarial loss
Learning Rate	0.01	0.01 (encoder & discriminator)
Epochs	50	50
Input Feature	One-hot encoding (0 for ncRNA, 1 for Disease)	One-hot encoding (0 for ncRNA, 1 for Disease)

Table 1 shows the hyperparameters used in this study. A two-layer Graph Convolutional Network (GCN) is the Graph Auto-Encoder (GAE) encoder. The input node features are converted into a hidden representation of dimensionality 32 by the first GCN layer and then compressed into a latent embedding space of dimension 16 by the second GCN layer. Each GCN layer is followed by ReLU activation, and overfitting is reduced by using dropout regularization with a probability of 0.2. The GAE's decoder is a straightforward inner product decoder that uses the dot product between two latent node embedding pairs to reconstruct the original adjacency matrix. The model is trained by minimizing the binary cross-entropy loss between the actual and predicted edge probabilities. The ARGAE improves the GAE by adding an adversarial training component. The encoder architecture is still the same as GAE, using the same two-layer GCN structure. ARGAE incorporates a discriminator, a three-layer fully connected neural network, and the encoder-decoder framework. This discriminator seeks to differentiate between latent embeddings produced by the encoder and those sampled from a predetermined Gaussian prior. By imposing a Gaussian distribution on the latent space, the goal is to improve the robustness and generalizability of the learned embeddings. The model is trained for 50 epochs, with the adversarial loss derived from the discriminator and the reconstruction loss (binary cross-entropy) jointly optimized. The discriminator and encoder both use a learning rate of 0.01.

Evaluation Metrics and Additional Analysis

Several evaluation tasks are used to gauge the models' performance. Some true edges are hidden to predict links, and node pairs with unknown associations are randomly selected to produce negative examples. The Receiver Operating Characteristic Area Under the Curve (ROC-AUC) and Average Precision (AP) metrics assess the models' capacity to recover these links. With the number of clusters (k) set to 5, the learned embeddings are put through k-means clustering for node clustering, representing an unsupervised method of assembling related nodes. The Silhouette Score, which assesses intra-cluster cohesion and inter-cluster separation, is used to gauge clustering performance. The agreement between clusterings produced by the GAE and ARGAE models is also measured using the Adjusted Rand Index (ARI), which sheds light on embedding consistency.

A similarity analysis is carried out to evaluate the learned embeddings' semantic coherence. Cosine similarity matrices are produced for intra-class (ncRNA–ncRNA and disease–disease) and inter-class (ncRNA–disease) pairs. These matrices are examined using heatmaps and Kernel Density Estimation (KDE) plots to comprehend distributional characteristics and interaction patterns. The latent embedding space and model behavior are interpreted using visualization techniques. These include precision-recall curves for link prediction performance, embedding norm distribution plots to analyze the magnitude and dispersion of embeddings, and t-distributed stochastic neighbor embedding (t-SNE) plots for dimensionality reduction and cluster visualization. These analyses thoroughly understand biological interpretability, embedding structure, and model performance. This study does not aim to diagnose periodontitis clinically or radiographically. Instead, it focuses on learning and reconstructing lncRNA–disease associations annotated as periodontitis in the LncRNAdisease database. The proposed graph-based models operate on curated molecular association data and do not use clinical signs such as periodontal bone loss as diagnostic inputs.

RESULTS

Overall Performance

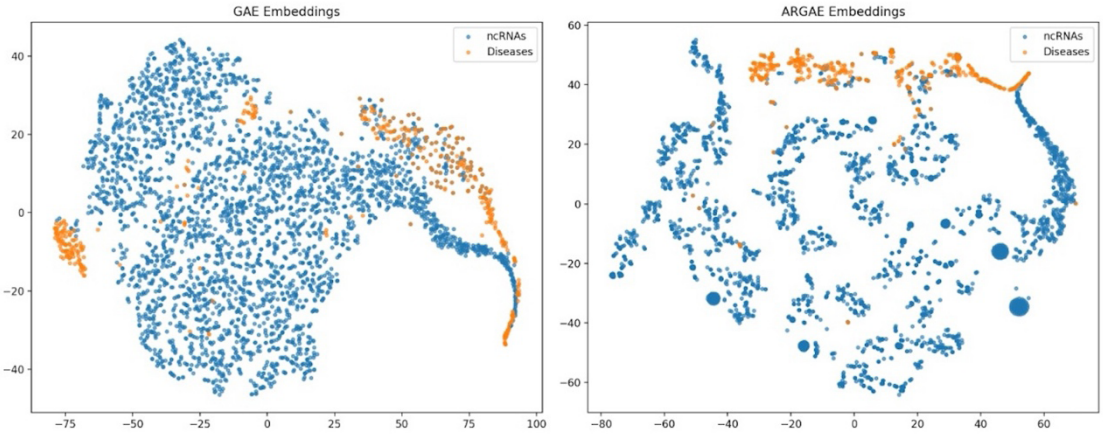
A comparative analysis of the Graph Auto-Encoder (GAE) and Adversarially Regularized Graph Auto-Encoder (ARGAE) models demonstrates their effectiveness in forecasting long non-coding RNA (lncRNA)-disease associations pertinent to periodontitis With a ROC-AUC of roughly 0.9835 and an Average Precision (AP) of 0.9830, the GAE model outperformed the others in the link prediction task, demonstrating high accuracy in detecting genuine lncRNA–disease relationships. With a ROC-AUC of 0.9713 and an AP of 0.9749, the ARGAE model demonstrated a strong predictive capacity despite being marginally lower. According to these metrics, GAE slightly outperforms ARGAE in direct edge recovery, indicating that both models successfully capture latent interaction patterns within the biological graph. The clustering performance revealed additional information about the quality and interpretability of the embeddings. With a significantly higher Silhouette Score of 0.6382 than ARGAE's 0.3643, the GAE model appeared to have more distinct and separable latent clusters. The two models' clusterings had a moderately dissimilar Adjusted Rand Index (ARI) of 0.1181, indicating that adding adversarial regularization changes the structural arrangement of embeddings. This divergence may affect biological interpretability when classifying diseases or lncRNAs into functionally coherent modules.

Training dynamics highlighted the models' ability to reconstruct. With an initial loss of 0.3862 and a remarkable final loss of roughly 1.08e-07, the GAE model demonstrated strong convergence and little reconstruction error. Training behavior was more complex for the ARGAE model, incorporating a generative adversarial component. The adversarial loss reached 3.6588, and its final reconstruction loss stabilized at 0.00016, resulting in a final loss of roughly 3.6589. These outcomes demonstrate the delicate balancing act between enforcing a prior distribution in the latent space and maintaining edge reconstruction accuracy. The GAE model generated extremely coherent embeddings, with average cosine similarity scores of 0.9935 for lncRNA–lncRNA pairs and 0.9936 for disease–disease pairs, according to the evaluation of embedding similarity. At 0.9992, the lncRNA–disease inter-class similarity was marginally lower. ARGAE embeddings, on the other hand, demonstrated less consistency, with inter-class similarity dropping to 0.9765 and within-class similarities to 0.9920 for lncRNAs and 0.9890 for diseases. This suggests that ARGAE may offer improved biological role separation, which would be advantageous for classification tasks. Overall,

GAE demonstrated better consistency than ARGAE, even though both models are useful for forecasting lncRNA–disease associations related to periodontitis.

Figure 2 compares 2D t-SNE projections of embeddings from two models, GAE and ARGAE. GAE shows low reconstruction loss, indicating an effective graph structure, while ARGAE has a higher loss due to its adversarial component that enhances representation robustness. Both models effectively distinguish ncRNAs and diseases in the latent space, but ARGAE demonstrates more organized clustering. Table 2 shows the accuracy of the models.

Figure 2. Comparing Embedding Visualizations of Two Models using 2D t-SNE.



Source: Authors.

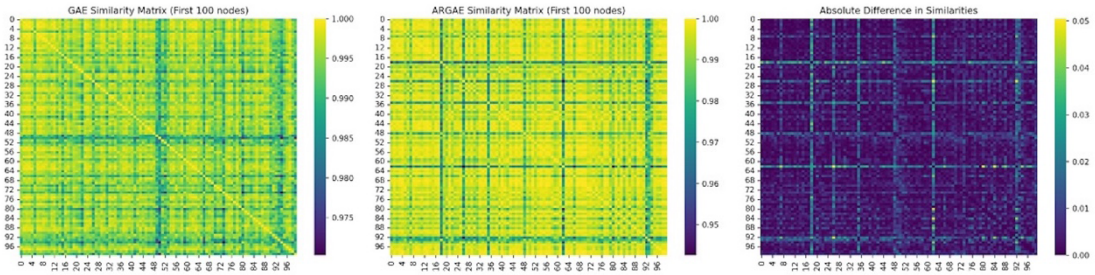
Table 2. Accuracy Metrics Table.

Metric	GAE	ARGAE
AUC Score	0.9835	0.9713
Average Precision (AP)	0.9830	0.9749
Silhouette Score	0.6382	0.3643
Within ncRNA Similarity	0.9935	0.9920
Within Disease Similarity	0.9936	0.9890
Between-Class Similarity	0.9902	0.9765

Similarity Matrix Visualizations

Figure 3 visually compares the cosine similarity matrices of the first 100 nodes, including lncRNAs and diseases, from the latent embeddings of the GAE and ARGAE models. Through heatmaps that show how the embedding space is structured and highlight the differences in modeling lncRNA–disease associations within periodontitis, it displays their pairwise absolute similarity differences. The GAE model favors overall reconstruction by producing a latent space that is densely packed and has little inter-node variability. The ARGAE model, on the other hand, adds more variation, making it easier to distinguish between lncRNAs and illnesses. This is advantageous for identifying minute biological patterns. While the two models mainly agree on pairwise similarities, an absolute difference matrix shows that ARGAE purposefully modifies some relationships to offer deeper insights, especially into important nodes like disease hubs and uniquely regulated lncRNAs. In the end, GAE produces coherent embeddings that are interpretable. In addition, the structure of ARGAE facilitates exploratory analysis and the identification of new connections, suggesting intricate lncRNA–disease interactions.

Figure 3. Comparison of the cosine similarity matrices of the first 100 nodes.

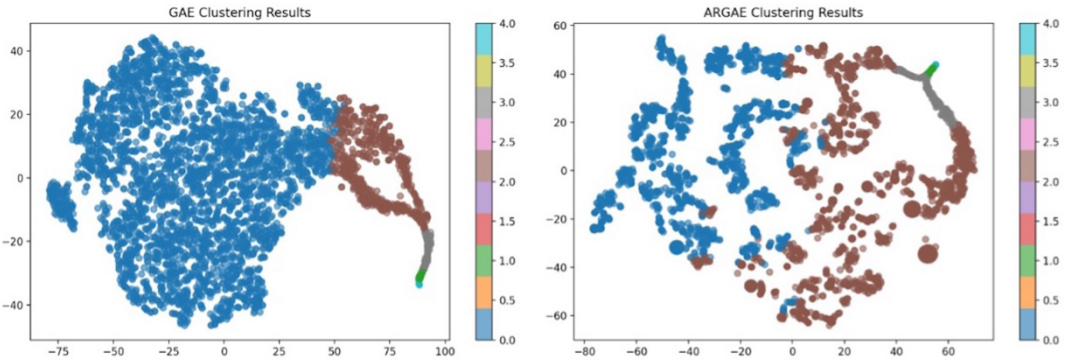


Source: Authors.

Clustering Visualization

Figure 4 displays the GAE (left) and ARGAE (right) models' t-SNE visualizations, which display latent embeddings color-coded by K-means clustering (k = 5), showing how the lncRNA and disease nodes' latent space is organized according to learned representations associated with periodontitis. According to the GAE clustering results, the embeddings, projected into 2D space using t-SNE, form five distinct clusters with notable intra-cluster cohesion, especially in the large blue and red groups. The minimal overlaps and seamless cluster transitions imply that GAE efficiently learns a well-structured latent space, reflecting biological similarities between entities such as disease subtypes or lncRNAs. The ARGAE clustering visualization shows less clear boundaries than GAE due to a fragmented and dispersed structure with substantial inter-cluster overlap, especially between the blue and red clusters. Nonetheless, it successfully captures minute variations between nodes that might point to biological variances that GAE misses. Overall, while sacrificing some cluster purity, ARGAE produces a less condensed but more exploratory embedding space that aids in discovering new or shakily connected lncRNA–disease associations. GAE generates distinct, comprehensible clusters to categorize known diseases or related long non-coding RNAs (lncRNAs). ARGAE, on the other hand, provides a less structured but more adaptable method that allows for the investigation of new connections and helps generate hypotheses in the study of periodontal disease.

Figure 4. GAE and ARGAE models' t-SNE visualizations.



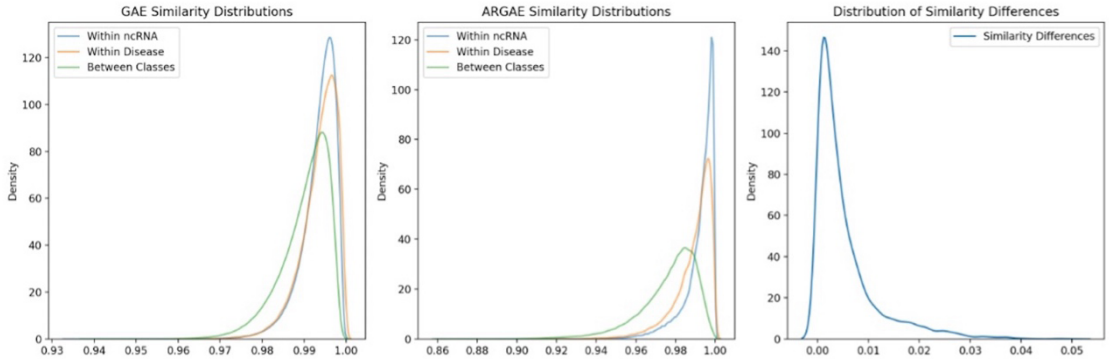
Source: Authors.

Similarity Distributions

Figure 5 contains three kernel density plots that illustrate the semantic compactness and separability of node embeddings (lncRNAs and diseases) related to periodontitis by visualizing and comparing the cosine similarity distributions of latent embeddings from GAE and ARGAE models. For lncRNA embeddings, the GAE similarity distributions in Panel 1 present a blue curve that peaks around 1.0, signifying high similarity. Concurrently, the green curve shows a wider and lower degree of class similarity, while the orange curve for disease node embeddings is also clustered close to 1.0 but somewhat lower. This indicates that GAE produces coherent embeddings for both classes with some differentiation. Panel 2, in contrast,

displays the distributions from ARGAE, which exhibit broader curves for both ncRNA and disease, suggesting less intra-class consistency. As it shifts left toward lower similarity values, the green curve once again indicates improved inter-class separation. This demonstrates how ARGAE enhances latent space disentanglement by reducing overlap, facilitating improved generalization, and enabling the discovery of new links. Lastly, Panel 3 presents the absolute variations in cosine similarity values across the models. A peak around 0.00 reflects that most node pairs yield similar outputs. However, a long-tailed distribution for certain pairs reveals higher deviations, primarily due to ARGAE's adversarial training.

Figure 5. Three kernel density plots.



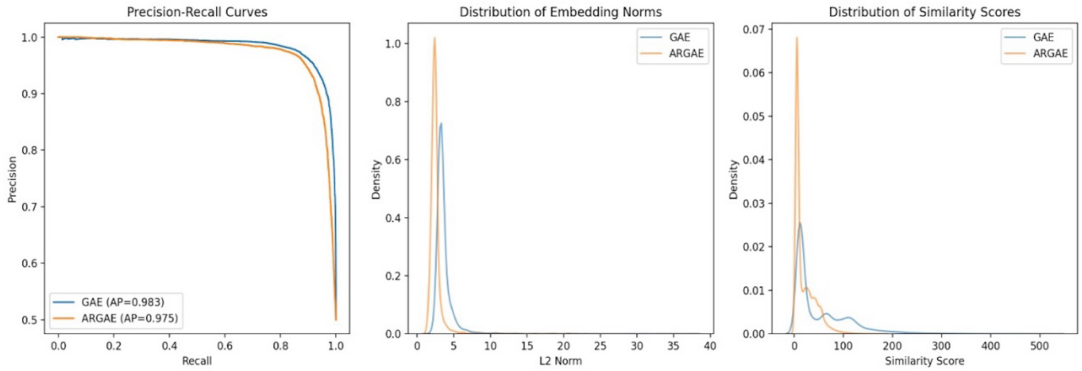
Source: Authors.

In conclusion, GAE provides densely packed embeddings suitable for clustering and reconstruction. ARGAE further enhances the separation between class embeddings, simplifying identifying new associations. The difference plot further corroborates that ARGAE introduces significant changes to the embedding space.

Performance Curves and Distributions

Figure 6 includes three panels that evaluate the performance of the GAE and ARGAE models regarding embedding vector norms, similarity score distributions, and link prediction precision-recall performance, particularly concerning lncRNA–disease associations related to periodontitis. Precision-recall curves are shown in the first panel, where GAE (blue curve) performs better than ARGAE (orange curve) with a higher Average Precision (AP = 0.983 vs. AP = 0.975). It also excels in precision at higher recall levels, demonstrating its superiority in accurately reconstructing known associations. Because adversarial training encourages adherence to a prior distribution, ARGAE's narrower distribution suggests compactness and controlled variability, while GAE's wider distribution indicates greater flexibility and expressiveness in embeddings. The distribution of embedding norms is presented in the second panel. ARGAE's concentrated peak close to the mean contrasts with GAE's wider curve, which extends to higher similarity values, as shown in the final panel that analyzes similarity score distributions. This implies that ARGAE provides more uniformity, which is useful for stable and noise-resistant modeling.

Figure 6. Three panels that evaluate the performance of the GAE and ARGAE models.



Source: Authors.

In contrast, GAE captures a wider range of similarity relationships, which is advantageous for identifying subtle associations. Overall, ARGAE emphasizes latent space regularity for stability and robust exploration, highlighting a crucial trade-off in predicting lncRNA–disease interactions in periodontal disease. At the same time, GAE exhibits superior precision and embedding diversity, making it ideal for detailed prediction and complex discovery.

The similarity matrices' first 100-node heatmaps were created to illustrate the variations in similarity between GAE and ARGAE, shedding light on the impact of embedding regularization. The clear clustering of ncRNAs and diseases is shown by clustering visualizations, especially t-SNE plots, where GAE displays well-separated clusters compared to ARGAE. Furthermore, the distributions of cosine similarities, embedding norms, and link prediction similarity scores can be seen in distribution plots like KDE plots, and the precision-recall curves indicate that both models perform well in terms of prediction, with particularly high average precision scores. Link prediction, clustering, and similarity metrics were compared in an accuracy metrics table. Reconstruction versus regularization is the main topic of discussion in this section. The GAE model achieved a very low reconstruction error, indicating a remarkable ability to reconstruct the original graph. The ARGAE model, on the other hand, aimed to regularize and reconstruct the latent space to match a Gaussian prior. The GAE model's higher silhouette score reflects more distinct clustering between the two node types regarding clustering interpretation. However, the additional adversarial loss in the ARGAE model may increase embedding space variation, which could decrease cluster separation and result in a lower silhouette score. Both models exhibit high intra-class similarity according to similarity metrics. However, ARGAE shows lower between-class similarity compared to GAE, suggesting that the adversarial regularization encourages the model to capture more subtle differences between various node types. The effectiveness of the learned embeddings in predicting disease–ncRNA associations was also confirmed by the extremely high ROC-AUC and AP scores (greater than 0.97) obtained by both methods concerning link prediction.

DISCUSSION

Chronic inflammation-driven destruction is a characteristic of periodontitis. By controlling inflammation, osteoclast activity, and bone remodeling, long non-coding RNAs (lncRNAs) have a major impact on periodontitis associated alveolar bone loss. Important lncRNAs include TUG1, which regulates oxidative stress and cytokine production; MEG3, which prevents osteoclast differentiation and prevents bone loss; HOTAIR, which increases matrix degradation in inflammatory conditions; and MALAT1, which stimulates inflammation and bone resorption. These elements highlight the potential of lncRNAs as targets for diagnosing and treating periodontitis. As far as we know, no prior studies have focused on reconstructing molecular lncRNA–disease associations related to periodontitis. To assist future researchers, this study aims to investigate sophisticated AI algorithms for data reconstruction.

With an AUC of 0.9835 and an AP of 0.9830, the Graph Auto-Encoder (GAE) outperformed the Adversarily Regularized Graph Auto-Encoder (ARGAE) in link prediction, with AUC = 0.9713 and AP = 0.9749. Additionally, GAE showed superior clustering quality, exhibiting a Silhouette Score of 0.6382 compared to ARGAE's 0.3643, suggesting better latent space separability. While ARGAE added more inter-class separation for exploratory potential, GAE embeddings were more coherent and homogeneous, as indicated by cosine similarity analyses (Figures 2-6) (Table 2). The RWRlncD method, which employs a random walk with restart on a lncRNA functional similarity network, demonstrated a robust area under the ROC curve of 0.822, indicating promise in identifying potential human lncRNA–disease associations, consistent with recent research. Similar findings were observed in another study. BPLLDA is a technique that uses fixed-length paths in a heterogeneous network to predict lncRNA–disease associations. Combining networks of lncRNA–disease association, functional similarity, and disease semantic similarity creates a network. The benefits of BPLLDA[8,17] include its ability to predict associations with novel lncRNAs or diseases without requiring negative samples. The effectiveness of GAE and ARGAE in predicting lncRNA–disease associations related to periodontitis is evidenced by their evaluation. GAE demonstrates superior reconstruction accuracy with an ROC-AUC of 0.9835 compared to ARGAE's 0.9713. This suggests that GAE is more adept at identifying potential connections overlooked during training and retrieving genuine associations. ARGAE sacrifices some predictive sharpness for advantages in latent space generality. GAE exhibits superior clustering

performance with a Silhouette Score of 0.6382, indicating better cluster separation that aids in understanding related lncRNAs or disease phenotypes[11,18].

In contrast, ARGAE displays fewer distinct clusters with a score of 0.3643. The Adjusted Rand Index of 0.1181 indicates different approaches to organizing latent representations, where ARGAE reveals hidden relationships but lacks clarity. While ARGAE may highlight subtle biological features, GAE provides more interpretable structures. The cosine similarity score analysis shows that GAE embeddings have a strong between-class similarity of 0.9902 and high within-class similarities of 0.9935 and 0.9936 for the ncRNA and disease categories, respectively. With scores of 0.9920 for ncRNA, 0.9890 for disease, and 0.9765 for between-class similarity, ARGAE embeddings exhibit somewhat lower similarities. Accordingly, GAE generates well-organized latent spaces that faithfully depict established biological relationships, including those about periodontitis associated alveolar bone loss[8,9].

In contrast, the lower similarity of ARGAE may facilitate generalization to novel and untested associations, albeit at the cost of less accuracy when reconstructing known interactions. The conclusion discusses the trade-offs between the two models for examining lncRNA–disease interactions in periodontal disease. GAE is noted for its high consistency and link prediction accuracy, making it suitable for network reconstruction and reliable biomarker prediction. Conversely, while ARGAE is less interpretable, it promotes diversity in embeddings and may uncover novel relationships beneficial for exploratory analysis. Ultimately, the choice of model should align with the study's objectives: ARGAE for hypothesis generation and GAE for accurate predictions.

This study emphasizes the trade-off between reconstruction accuracy and latent space generalization when comparing GAE and ARGAE for modeling lncRNA–disease associations in periodontitis. ARGAE adds structural diversity and disentanglement, which may be useful for finding new or subtle associations. In contrast, GAE provides better predictive performance and embedding coherence, which is perfect for locating known interactions and biomarkers.10 However, the scale and annotation depth of the current dataset may limit the ability to investigate the effects of adversarial regularization fully. Furthermore, node features were restricted to one-hot encoding, leaving out more complex biological characteristics that could improve model performance, such as expression levels and sequence. Future studies should use spatial transcriptomics, graph attention mechanisms, and multi-omic integration to increase interpretability and accuracy. Furthermore, benchmarking on bigger and tissue-specific datasets will be essential to confirm the model's resilience under various biological circumstances.

Importantly, the proposed models do not diagnose periodontitis or differentiate it from other causes of alveolar bone loss, such as endodontic or endo-periodontal lesions; the findings are limited to molecular association modeling based on curated disease annotations.

CONCLUSION

GAE and ARGAE provide useful frameworks for reconstructing and forecasting lncRNA–disease associations in periodontal disease, with ARGAE improving latent exploration and GAE exhibiting superior precision. The particular application should determine which of these models is best—GAE for precise, comprehensible predictions and ARGAE for hypothesis generation and network growth. In biomarker discovery pipelines aimed at complex inflammatory diseases, this comparative study offers fundamental insights for choosing and fine-tuning graph-based architectures. In conclusion, both GAE and ARGAE provide effective frameworks for reconstructing and predicting lncRNA–periodontitis associations relevant to alveolar bone loss observed in periodontitis. The models do not predict clinical disease diagnosis but instead support molecular-level understanding and biomarker discovery related to periodontitis-associated bone destruction. Accordingly, this work should be interpreted as a molecular association and biomarker discovery study rather than a clinical diagnostic or disease prediction model.

AUTHOR CONTRIBUTIONS

Conceptualization: Devika Dinesh, Pradeep Kumar Yadalam and Carlos M. Ardila
Formal Analysis: Devika Dinesh, Pradeep Kumar Yadalam and Carlos M. Ardila
Investigation: Devika Dinesh, Pradeep Kumar Yadalam and Carlos M. Ardila
Methodology: Devika Dinesh, Pradeep Kumar Yadalam and Carlos M. Ardila
Project Administration: Devika Dinesh, Pradeep Kumar Yadalam and Carlos M. Ardila

Writing – Original Draft: Devika Dinesh, Pradeep Kumar Yadalam and Carlos M. Ardila
Writing – Review & Editing: Devika Dinesh, Pradeep Kumar Yadalam and Carlos M. Ardila

CONFLICT OF INTEREST

None

ORCID

Devika Dinesh: https://orcid.org/0009-0008-4272-5647
Pradeep Kumar Yadalam: https://orcid.org/0000-0003-4259-820X
Carlos M. Ardila: https://orcid.org/0000-0002-3663-1416

REFERENCES

1. Sun J, Shi H, Wang Z, Zhang C, Liu L, Wang L, et al. Inferring novel lncRNA-disease associations based on a random walk model of a lncRNA functional similarity network. Mol Biosyst. 2014;10(8):2074-2081.
2. Ramesh A, Varghese SS, Doraiswamy JN, Malaiappan S. Herbs as an antioxidant arsenal for periodontal diseases. J Intercult Ethnopharmacol. 2016;5(1):92-96.
3. Panda S, Sankari M, Satpathy A, Jayakumar D, Mozzati M, Mortellaro C, et al. Adjunctive Effect of Autologus Platelet-Rich Fibrin to Barrier Membrane in the Treatment of Periodontal Intrabony Defects. J Craniofac Surg. 2016;27(3):691-696.
4. Kaarthikeyan G, Jayakumar ND, Padmalatha O, Sheeja V, Sankari M, Anandan B. Analysis of the association between interleukin -1β (+3954) gene polymorphism and chronic periodontitis in a sample of the south Indian population. Ind J Dent R. 2009;20(1):37-40.
5. Xuan P, Fan M, Cui H, Zhang T, Nakaguchi T. GVDTI: graph convolutional and variational autoencoders with attribute-level attention for drug-protein interaction prediction. Brief Bioinform. 2022;23(1):bbab453.
6. Lei S, Lei X, Liu L. Drug repositioning based on heterogeneous networks and variational graph autoencoders. Front Pharmacol. 2022;13:1056605.
7. Yuan Y, Zhang L. Long non-coding RNA regulates bone metabolism via Wnt/β-catenin signaling pathway. Sheng Wu Gong Cheng Xue Bao. 2021;37(7):2342-2350.
8. Gao MM, Cui Z, Gao YL, Wang J, Liu JX. Multi-Label Fusion Collaborative Matrix Factorization for Predicting LncRNA-Disease Associations. IEEE J Biomed Health Inform. 2021;25(3):881-890.
9. Zhang Y, Ye F, Xiong D, Gao X. LDNFSGB: prediction of long non-coding rna and disease association using network feature similarity and gradient boosting. BMC Bioinformatics. 2020;21(1):377.
10. Wang B, Liu R, Zheng X, Du X, Wang Z. lncRNA-disease association prediction based on matrix decomposition of elastic network and collaborative filtering. Sci Rep. 2022;12(1):12700.
11. Wang L, Shang M, Dai Q, He PA. Prediction of lncRNA-disease association based on a Laplace normalized random walk with restart algorithm on heterogeneous networks. BMC Bioinformatics. 2022;23(1):5.
12. Ray S, Lall S, Mukhopadhyay A, Bandyopadhyay S, Schönhuth A. Deep variational graph autoencoders for novel host-directed therapy options against COVID-19. Artif Intell Med. 2022;134:102418.
13. Wang C, Yuan C, Wang Y, Chen R, Shi Y, Zhang T, et al. MPI-VGAE: protein-metabolite enzymatic reaction link learning by variational graph autoencoders. Brief Bioinform. 2023;24(4):bbad189.
14. Salha G, Hennequin R, Remy JB, Moussallam M, Vazirgiannis M. FastGAE: Scalable graph autoencoders with stochastic subgraph decoding. Neural Netw. 2021;142:1-19.
15. Niknam G, Molaei S, Zare H, Pan S, Jalili M, Zhu T, et al. DyVGRNN: DYnamic mixture Variational Graph Recurrent Neural Networks. Neural Netw. 2023;165:596-610.
16. Chen G, Wang Z, Wang D, Qiu C, Liu M, Chen X, et al. LncRNADisease: a database for long-non-coding RNA-associated diseases. Nucleic Acids Res. 2013;41(D1):D983-6.
17. Xiao X, Zhu W, Liao B, Xu J, Gu C, Ji B, et al. BPLLDA: Predicting lncRNA-Disease Associations Based on Simple Paths With Limited Lengths in a Heterogeneous Network. Front Genet. 2018;9:411.
18. Li Y, Li J, Bian N. DNILMF-LDA: Prediction of lncRNA-Disease Associations by Dual-Network Integrated Logistic Matrix Factorization and Bayesian Optimization. Genes (Basel). 2019;10(8):608.